Recovering structured signals in high dimensions via non-smooth convex optimization: Precise performance analysis

Thesis by
Christos Thrampoulidis

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
Defended May 17, 2016

Christos Thrampoulidis
ORCID: [0000-0001-9053-9365]

ACKNOWLEDGEMENTS

"*As you set out for Ithaka hope the voyage is a long one, full of adventure, full of discovery. (...) Wise as you will have become, so full of experience, you will have understood by then what these Ithakas mean."*

Ithaka, C.P. Cavafy.

Admittedly, the journey to this thesis has not been a paved way; it has gone through difficult and sometimes lonely paths, and it has been full of ups and downs. However, during these five years I have been extremely lucky to have been surrounded by wonderful people -mentors, colleagues, friends, family- that have helped me the most in all aspects. It is thanks to them that "this voyage has been full of adventure, full of discovery and full of experience", and that I have grown both as a person and as a researcher. This is an opportunity to recognize my appreciation to them.

First, I would like to express my gratitude to my advisor Prof. Babak Hassibi for his kindness, generosity, and patience; for the unstressful work environment and the academic freedom that he has provided me with, and for the guidance throughout this journey. His brightness, his enthusiasm, his clarity of thought, and his ability to communicate complex concepts -from a very diverse set of fields- in the most transparent ways have been an enormous inspiration. His provocative questions and careful criticism have taught me to see the bigger picture, not only when presenting my work, but also as a guidance to new directions. His constant encouragement has helped me grow confidence as a researcher. I am also grateful for several thought-provoking and always intriguing non-technical discussions, and for the opportunities he has provided me to travel and attend a number of conferences held all around the world.

Next, gratitude is extended to the rest of the members of my thesis committee: Professors P. P. Vaidyanathan, Joel Tropp, Adam Wierman, and Venkat Chandrasekaran, for their time, advice, and critical evaluation of my thesis. I would like to express special thanks to Professors Vaidyanathan and Tropp for their encouragement, support, and advice that also extended beyond research to career planning. I am thankful to Prof. Vaidyanathan for his kindness and friendliness during our everyday encounters and short discussions in the hallways of the Moore Laboratory. I thank Joel Tropp for some of the most useful, enriching and stimulating courses that I have taken at Caltech; I will always admire and be inspired by his remarkable pro-

foundness, clarity and simplicity in teaching and in presenting his work.

I feel indebted to all my collaborators: Samet Oymak, Subhmonesh Bose, Ehsan Abbasi, Ashkan Panahi, Weiyu Xu, Linqi (Daniel) Guo, Kishore Jaganathan, and Navid Azizan, for their patience, for all the things that I learnt from them, and for all the fun and long hours that we spent thinking and learning together! Some of the major contributions and ideas of this thesis developed during long whiteboard discussions with Samet, Ehsan, and Ashkan. Thanks to Samet for introducing me to many of the problems addressed in this thesis. His unique combination of fast problem-solving skills, of hard work and of constant enthusiasm always inspired me. I also have some very good memories with Samet during our trip in Hawaii for ISIT. Thanks to Ehsan for all the technical work that we accomplished together. But even more so, I wish to thank him for his support and encouragement, for his endless patience, for being a great listener and a good friend. Thank you Ehsan, for the always pleasant and fun overnight shifts in the lab, for the handy Persian expressions that you taught me, for our amazing trip to Chicago, and for all the stories we shared together. Thanks to Ashkan and Daniel for our (rather) short but very productive collaboration. Thanks to Weiyu for helpful discussions and encouraging feedback. The works that we did together with Bose, Kishore, and Navid are not part of this thesis, but are by no means less important. Bose had the difficult task to be my first collaborator and coauthor. I am grateful to him for his patience, for teaching me the importance of communicating research results in a transparent way, and for his advice during the multiple times that I felt lost. Thanks to Κισώρος (cf., Kishore) for our joint work that led us win the Qualcomm Innovation Fellowship. I always admire his patience and calmness, and I confess that I sometimes laugh at his “jokes"! My collaboration with Navid has been only very recent, but one of the most pleasant ones, and I wish it continues.

Many thanks go to all my labmates over these five years for the things that they have taught me, and for our fascinating conversations! Thanks to the “older generation": Amin, Teja, Wei, Samet, Bose and Ahn. Special thanks to Amin Khajehnejad for introducing me to the culture of the lab when I first arrived at Caltech, and for teaching me many useful things (including, but by no means limited to, access to free food). Thank you to Matt, Kishore, Wael, Ramya, Ehsan, Navid, James, Anatoly, and Fariborz for creating an especially friendly and pleasant environment in the lab during the last two years. Thank you for your support, for putting up with my strange habits (yes, you may now take the clock and put it back on the wall),

and for being great listeners to my everyday complaints. I would especially like to express my gratitude to my good friend Wael, whom I have known since my very first days at Caltech.

Also, thank you to all other students in Caltech's Electrical Engineering and Applied Mathematics Departments for friendly interactions, advice and intriguing discussions. Special thanks go to Roarke, Mark, Richard, and Yong Sheng. I would also like to thank the secretarial staff for putting up with me and answering all my questions: Shirley, Tanya, Katie, Anne, Terecita, and Lucinda. My thanks also extend to the friendly, welcoming, and always smiling administrative staff at Caltech's International Student Programs Office and at Chandler Café.

Outside Caltech, I am grateful to Professors G. Moustakides, D. Toumpakaris, A. Birbas, A. Tzes, G. Bitsoris, T. Stouraitis, and S. Fassois from the University of Patras for their motivation and support when applying for graduate studies in the US. I am especially thankful to Prof. Moustakides for being a wonderful advisor during my undergraduate years, but also for the constant encouragement and guidance that I have been receiving from him since then. My gratitude extends to Prof. Toumpakaris for selflessly putting on long hours to help me complete my application files. I would also like to thank the "Andreas Mentzelopoulos Scholarships for the University of Patras" for supporting my studies during my first year at Caltech. A very special thank you belongs to my teacher Mr. Dimitris Aretakis and to my friend Stefanos Aretakis for teaching me and for making me love math.

Thank you to my very first international friends at Caltech: Aleksander, Christophe, Wael, Krishna, Mark, and Roarke, who have helped me the most during the intense first-year of courses, and with whom I have some awesome stories to remember. Thanks to the Greeks at Caltech and at JPL for all the fun and relaxing moments: for our loud lunches at Chandler, for the unforgettable "Vassilopita" nights, and for our everyday loud conversations and jokes at Caltech's gym. Special thanks to Costas S., Theodore, Christos, Costas A., Marilena, and Vassilis for being the first ones to welcome me at Caltech. Thanks to my friend Dimitris for being the best host during my visits in Chicago! Thanks to my conference-travel buddy Nicoló for some unforgettable nights, whether at Hong-Kong or at Urbana-Champaign. I would also like to thank my friends back in Greece for all the fun and relaxing moments during my summer visits over there. Thank you to my childhood friends in Kastellokampos. Special thanks to Evangelia, Costas, and Nikos for their encouragement.

Among the many wonderful people that I have met during this journey, I would like to especially acknowledge Panagiotis Vergados, Juan Andrés Muniz, Wael Halbawi and Georgia Papadakis, who have been great friends and have kept me sane and happy over the years. Thank you for your honesty, for your kindness, and for caring for me not only when the sun is shining, but most importantly during the storms. (Yes, there were storms despite the California weather!) Thank you for all the memories that we have shared; thank you for every single day. Panagiotis, additional thanks go to you for the long hours that you selflessly put on proofreading parts of this thesis and providing invaluable feedback on my thesis presentation!

I am also thankful to my extended family in Veroia, Ptolemaida, Thessaloniki and Patras for always being there for me, for their love, and for their heartfelt support during all these years. Thank you for always cutting a cake for my birthday even if I had to blow out the candles and eat my piece of it over Skype!

Finally, my deepest gratitude belongs to my family: to my father Kleanthis, whose dignity, diligence and ambition are a constant source of inspiration to me; to my mother Vassiliki, whose positive energy and kindness are a constant reminder to always smile; and to Manolis, the best and most caring brother in the world. I have no words to thank you for your unconditional love, for all my sweet memories, for all the values that you have taught me, for all that I am and that I have ever accomplished, but to dedicate this dissertation to you.

To my parents Vassiliki and Kleanthis,
and to my brother Manolis.

Στους γονείς μου Βασιλική και Κλεάνθη,
και στον αδερφό μου Μανώλη.

ABSTRACT

The typical scenario that arises in modern large-scale inference problems is one where the ambient dimension of the unknown signal is very large (e.g., high-resolution images, recommendation systems), yet its desired properties lie in some low-dimensional *structure* such as, sparsity or low-rankness. In the past couple of decades, *non-smooth convex optimization* methods have emerged as a powerful tool to extract those structures, since they are often computationally efficient, and also they offer enough flexibility while simultaneously being amenable to performance analysis. Especially, since the advent of Compressed Sensing (CS) there has been significant progress towards this direction. One of the key ideas is that *random* linear measurements offer an efficient way to acquire structured signals. When the measurement matrix has entries iid from a wide class of distributions (including Gaussians), a series of recent papers have established a complete and transparent theory that *precisely* captures the performance in the *noiseless* setting. In the more practical scenario of *noisy* measurements the performance analysis task becomes significantly more challenging and corresponding *precise* and *unifying* results have hitherto remained scarce. The available class of optimization methods, often referred to as *regularized M-estimators*, is now richer; additional factors (e.g., the noise distribution, the loss function, and the regularizer parameter) and several different measures of performance (e.g., squared-error, probability of support recovery) need to be taken into account.

This thesis develops a novel analytical framework that overcomes these challenges, and establishes precise asymptotic performance guarantees for regularized M-estimators under Gaussian measurement matrices. In particular, the framework allows for a unifying analysis among different instances (such as the Generalized LASSO, and the LAD, to name a few) and accounts for a wide class of performance measures. Among others, we show results on the mean-squared-error of the Generalized-LASSO method and make insightful connections to the classical theory of ordinary least squares and to noiseless CS. Empirical evidence is presented that suggests the Gaussian assumption is not necessary. Beyond iid measurement matrices, motivated by practical considerations, we study certain classes of random matrices with orthogonal rows and establish their superior performance when compared to Gaussians.

A prominent application of this generic theory is on the analysis of the bit-error

rate (BER) of the popular *convex-relaxation* of the Maximum Likelihood decoder for recovering BPSK signals in a massive Multiple Input Multiple Output setting. Our precise BER analysis allows comparison of these schemes to the unattainable Matched-filter bound, and further suggests means to provably boost their performance.

The last challenge is to evaluate the performance under *non-linear* measurements. For the Generalized LASSO, it is shown that this is (asymptotically) equivalent to the one under noisy linear measurements with appropriately scaled variance. This encompasses state-of-the art theoretical results of *one-bit CS*, and is also used to prove that the optimal quantizer of the measurements that minimizes the estimation error of the Generalized LASSO is the celebrated Lloyd-Max quantizer.

The framework is based on Gaussian process methods; in particular, on a new strong and tight version of a classical comparison inequality (due to Gordon, 1988) in the presence of additional convexity assumptions. We call this the *Convex Gaussian Min-max Theorem* (CGMT).

CONTENTS

Acknowledgements . . . iii
Abstract . . . viii
Contents . . . x
List of Figures . . . xii
List of Tables . . . xvii
Chapter I: Introduction . . . 1
Chapter II: Background, Literature Survey and Summary of Contributions . . . 7
2.1 Compressed Sensing . . . 7
2.2 Noiseless Case . . . 10
2.3 Noisy Case: The Challenge . . . 22
2.4 Thesis Contributions & Organization . . . 23
Chapter III: The Convex Gaussian Min-max Theorem . . . 27
3.1 Gaussian Comparison Inequalities . . . 27
3.2 Gaussian Min-max Theorem . . . 30
3.3 Convex Gaussian Min-max Theorem (CGMT) . . . 33
3.4 Proof of the CGMT . . . 37
Chapter IV: The Squared-error of Regularized M-estimators . . . 40
4.1 Introduction . . . 40
4.2 General Result . . . 44
4.3 Separable M-estimators . . . 51
4.4 Survey of Relevant Literature . . . 56
Chapter V: Analysis Framework . . . 60
5.1 How it Works . . . 60
5.2 An Example . . . 62
Chapter VI: Specific Examples . . . 68
6.1 M-estimators without Regularization . . . 68
6.2 Ridge Regularzation . . . 70
6.3 Cone-constrained M-estimators . . . 72
6.4 Generalized-LASSO . . . 78
6.5 Square-root LASSO . . . 79
6.6 Sparse Recovery via the LASSO . . . 80
6.7 Group-Sparse Recovery via the Group-LASSO . . . 82
6.8 Low-rank Matrix Recovery via the Trace-LASSO . . . 83
6.9 Robust Estimators . . . 83
6.10 Numerical Simulations . . . 84
Chapter VII: Noise Sensitivity of the Generalized-LASSO . . . 88
7.1 Introduction . . . 89
7.2 Revisiting Least Squares . . . 92
7.3 Least-squares Meets Compressed Sensing . . . 93

7.4 The NSE of Generalized LASSO in Gaussian Noise 97
7.5 Constrained LASSO 103
7.6 ℓ_2-LASSO 105
7.7 ℓ_2^2-LASSO 109
7.8 The NSE of Generalized LASSO with Arbitrary Fixed Noise 114
7.9 The Worst-Case NSE of Generalized LASSO 121
Chapter VIII: Beyond iid Ensembles: Isotropically Random Orthogonal Matrices 123
8.1 Introduction 124
8.2 Results 126
8.3 Proof Outline 130
Chapter IX: Beyond Squared-error: General Performance Metrics 133
9.1 Introduction 133
9.2 Review: ℓ_2-reconstruction Error 135
9.3 Lipschitz Performance Metrics 136
9.4 Support Recovery 138
9.5 Proofs 140
Chapter X: Application: The Bit-Error Rate of the Box-Relaxation Optimization 143
10.1 BPSK Signal Recovery 144
10.2 Implications 145
10.3 Extensions 148
Chapter XI: Non-linear Measurements 151
11.1 Motivation & Contribution 152
11.2 Results 157
11.3 Application: Optimal q-bit Quantization 160
Chapter XII: Conclusions and Future work 164
Appendix A: Proofs for Chapter 3 184
Appendix B: Proofs for Chapter 4 189
Appendix C: Proofs for Chapter 6 227
Appendix D: Calculating the Summary Parameters 236
Appendix E: Proofs for Chapter 8 240
Appendix F: Proofs for Chapter 10 250
Appendix G: Proofs for Chapter 11 255
Appendix H: A Note on Simple Denoising 266

LIST OF FIGURES

Number *Page*

2.1 Illustration of the Null-space condition (Proposition 2.2.1). 11

2.2 Illustration of the distance of a vector to the scaled subdifferential $\lambda\partial f(\mathbf{x}_0)$ and to the cone of subdifferential cone$(\partial f(\mathbf{x}_0))$. 16

5.1 Schematic representation of the CGMT framework. The first step involves equivalently expressing the regularized M-estimator as a min-max Primary Optimization (PO) (cf. (3.11a)). (These problems are hard to directly analyze and are thus shown in red.) The CGMT Theorem 3.3.1 associates with the (PO) an Auxiliary Optimization (AO) problem that is simpler to analyze (hence, depicted in green). The second step of the framework involves simplifying the (AO) into an optimization problem that only involves scalar variables. This makes possible the convergence analysis that follows as a third step and leads to a deterministic Scalar Performance Optimization (SPO). The last step involves using the (SPO) to conclude about the original regularized M-estimator. 61

6.1 Using the predictions of Theorem 4.2.1 to analytically compare the performance of different instances of M-estimators. Here, we compare a least-absolute deviations (LAD) to a least-squares (LASSO) loss function in (6.16) for sparse signal estimation under sparse noise. The normalized squared error is plotted as a function of the sparsity-level $\bar{s}$ of the noise at the high-SNR regime. The noise is sparse with sparsity level $\bar{s}$ and nonzero entries are i.i.d $\mathcal{N}(0, \sigma^2)$ and $\sigma^2 \to 0$. Also, the sparsity level of the unknown signal is fixed to be 0.1 and the normalized number of measurements is $\delta = 3/5$. 77

6.2 Squared error of the l_1-Regularized LAD with Gaussian ($\circ$) and Bernoulli ($\square$) measurements as a function of the regularizer parameter λ for two different values of the normalized number of measurements, namely $\delta = 0.7$ and $\delta = 1.2$. Also, $\mathbf{x}_{0,i} \overset{\text{iid}}{\sim} p_x(x) = 0.9\delta_0(x) + 0.1\phi(x)/\sqrt{0.1}$ and $\mathbf{z}_j \overset{\text{iid}}{\sim} p_z(z) = 0.7\delta_0(z) + 0.3\phi(z)$ for $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. For the simulations, we used $n = 768$ and the data were averaged over five independent realizations. 85

6.3 Comparing the squared error of the ℓ_1-Regularized LAD with the corresponding error of the LASSO. Both are plotted as functions of the regularizer parameter λ, for two different values of the normalized measurements, namely $\delta = 0.7$ and $\delta = 1.2$. The noise and signal are iid sparse-Gaussian as follows: $\mathbf{x}_{0,i} \overset{\text{iid}}{\sim} p_x(x) = 0.9\delta_0(x) + 0.1\phi(x)/\sqrt{0.1}$ and $\mathbf{z}_j \sim p_z(z) = 0.9\delta_0(z) + 0.1\phi(z)$ with $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. For the simulations, we used $n = 768$ and the data were averaged over five independent realizations. 86

6.4 Squared error of the ℓ_1-Regularized M-Estimator with Huber-loss as a function of the regularizer parameter λ. Here, $\delta = 0.7$, $\mathbf{x}_0 \overset{\text{iid}}{\sim} p_x(x) = 0.9\delta_0(x) + 0.1\phi(x)/\sqrt{0.1}$ and $p_z(z) = 0.9\delta(z) + 0.1\eta(z)$ with $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $\eta(z) = \frac{1}{\pi(1+z^2)}$. For the simulations, we used $n = 1024$ and the data are averaged over 5 independent realizations. . 87

7.1 NSE heatmap for ℓ_1 minimization based on Theorem 7.5.1. The x and y axes are the sparsity and measurements normalized by the ambient dimension. To obtain the figure, we plotted the heatmap of the function $-\log \frac{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}{m-\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}$ (clipped to ensure the values are between $[-10, 5]$). 104

7.2 Regions of operation of the ℓ_2-LASSO. 107

7.3 We consider the ℓ_1-penalized ℓ_2-LASSO problem for a k sparse signal in $\mathbb{R}^n$. For $\frac{k}{n} = 0.1$ and $\frac{m}{n} = 0.5$, we have $\lambda_{\text{crit}} \approx 0.76$, $\lambda_{\text{best}} \approx 1.14$, $\lambda_{\text{max}} \approx 1.97$. 108

7.4 Illustration of the region $\mathcal{R}_{\text{ON}}$ and of the map function (Defn. 7.7.2) for $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^n$ a k-sparse vector. map^{-1} maps the value of the regularizer λ in (7.3) to a value in $\mathcal{R}_{\text{ON}}$. $\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$ and $\overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau)$ are computed as in (7.40). 112

7.5 Numerical validation of Theorem 7.7.1 for $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^n$ a k-sparse vector. Measured values of the NSE(σ) are averages over 50 realizations of $\mathbf{A}, \mathbf{v}$. The theorem accurately predicts NSE(σ) as $\sigma \to 0$. The results support our claim that aNSE = wNSE. λ_{best} is the value of the optimal regularizer as predicted by Lemma 2.2. . . . 113

7.6 NSE of the C-LASSO with ℓ_1-regularization. The unknown signal $\mathbf{x}_0 \in \mathbb{R}^{500}$ is 5-sparse. The number of measurements m varies from 0 to 360. We plot the empirical NSE assuming $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$ for several values of σ. The solid black line corresponds to the bound of Theorem 7.8.1. The dashed line corresponds to the phase transition line of noiseless CS, namely $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ (cf. Theorem 2.2.1). . . . 116

7.7 Figure 7.7 illustrates the bound of Theorem 7.8.2, which is given in red for $n = 340$, $m = 140$, $k = 10$ and for $\mathbf{A}$ having $\mathcal{N}(0, \frac{1}{m})$ entries. The upper bound of Theorem 7.6.1, which is asymptotic in m and only applies to i.i.d. Gaussian $\mathbf{z}$, is given in black. In our simulations, we assume $\mathbf{x}_0$ is a random unit norm vector over its support and consider both i.i.d. $\mathcal{N}(0, \sigma^2)$ as well as non-Gaussian noise vectors $\mathbf{z}$. We have plotted the realizations of the normalized error for different values of λ and σ. As noted, the bound of Theorem 7.6.1 is occasionally violated since it requires very large m, as well as, i.i.d. Gaussian noise. On the other hand, the bound of Theorem 7.8.2 always holds. 119

8.1 Illustration of Theorem 8.2.1 for $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^{256}$ a 10-sparse vector. Simulation results support the claim that aNSE = wNSE. Furthermore, randomly sampled Discrete Cosine Transform (DCT) and Hadamard (HDM) matrices appear to have same NSE performance as IRO matrices. Measured values of the NSE are averages over 25 realizations. 128

8.2 Illustration of Theorem 8.2.2. The bound exceeds the corresponding bound for Gaussian matrices. We have chosen $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^n$, a k-sparse vector. 129

9.1 Performance of the Square-root Lasso with respect to $\Psi(\mathbf{x}) = \frac{1}{\sqrt{n}}\|\mathbf{x}\|_2$ (Red Line) and $\Psi(\mathbf{x}) = \frac{1}{n}\|\mathbf{x}\|_1$ (Blue line) as a function of λ. The theoretical prediction follows from Theorem 9.3.1. For the simulations, we used $n = 256$, $\delta = 0.8$, $\rho = 0.1$, SNR=0.5 and the data are averaged over five independent realizations. 137

9.2 Probability of successful recovery of the on-support and of the off-support entries as a function of λ for two different values of the normalized measurements, namely $\delta = 0.8$ (solid) and $\delta = 1.2$ (dashed). The theoretical prediction (shown in solid/dashed lines) follows from Theorem 9.4.1. For the simulation points (shown with squares and circles), we used $n = 256$, SNR= 0.5, $\epsilon = 10^{-3}$, $\rho = 0.1$ and the data are averaged over five independent realizations of the problem. . 139

10.1 Bit error rate performance of the Boxed Relaxation: BER as a function of SNR for different values of the ratio $\delta = \lceil m/n \rceil$ of receive to transmit antennas. The theoretical prediction follows from Theorem 10.1.1. For the simulations, we used $n = 512$. The data are averages over 20 independent realizations of the channel matrix and of the noise vector for each value of the SNR. 146

10.2 Bit error rate of the Box Relaxation Optimization (BRO) in (10.1) in comparison to the Matched Filter Bound (MFB) for $\delta = 0.7$ (dashed lines) and $\delta = 1$ (solid lines). The red curves follow the formula of Thm. 10.1.1, the green ones correspond to (10.4), and, BER^{MFB} of (10.5) is in blue. 148

10.3 Bit error rate of the Box Relaxation Optimization (BRO) in (10.6) as a function of the SNR for BPSK, 4-PAM and 8-PAM signals. The theoretical prediction follows from Theorem 10.3.1. For the simulations, we used n = 512 and $\delta = 1.2$. The data are averages over 20 independent realizations of the channel matrix and of the noise vector for each value of the SNR. 149

11.1 Squared error of the ℓ_1-regularized LASSO with non-linear measurements ($\square$) and with corresponding linear ones ($\star$) as a function of the regularizer parameter λ; both compared to the asymptotic prediction. Here, $g_i(x) = \mathrm{sign}(x + 0.3z_i)$ with $z_i \sim \mathcal{N}(0, 1)$. The unknown signal $\mathbf{x}_0$ is of dimension $n = 768$ and has $\lceil 0.15n \rceil$ non-zero entries. The different curves correspond to $\lceil 0.75n \rceil$ and $\lceil 1.2n \rceil$ number of measurements, respectively. Simulation points are averages over 20 problem realizations. 155

11.2 Squared error of the ℓ_1-regularized LASSO as a function of the regularizer parameter for noisy 1-bit measurements $g_i(x) = \text{sign}(x + 0.3z_i)$. Here, $\mathbf{x}_0$ is sparse with $p_{X_0}(+1) = p_{X_0}(0) = 0.05$, $p_{X_0}(+1) = 0.9$. The theoretical prediction is obtained by Corollary 11.2.1. Finally, $\delta = 0.75$, $n = 512$, and the simulation points represent averages over 20 realizations. 160

11.3 Squared error of the group-sparse LASSO as a function of the regularizer parameter compared to the asymptotic predictions for noisy 1-bit measurements $g_i(x) = \text{sign}(x+0.3z_i)$. Here, $\mathbf{x}_0$ is group-sparse: it is composed of $t = 512$ blocks of block size $b = 3$, and each block is zero with probability 0.95, otherwise its entries are iid $\mathcal{N}(0, 1)$. The theoretical prediction is obtained by Corollary 11.2.1. Finally, $\delta = 0.75$, and the simulation points represent averages over 20 realizations. 161

11.4 Illustration of the equivalence result of Theorem 11.2.1 applied to quantized measurements. 162

B.1 Graphs of the Moreau envelope functions of a quadratic (left) and of the absolute value (right), for different values of the parameter τ. Moreau envelopes are always *smooth* under-estimators of the original function. 223

LIST OF TABLES

Number *Page*

D.1 Closed form upper bounds for $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$. . . . 236

Chapter 1

INTRODUCTION

"Big data" and accompanying terms such as "data-analytics" and "data-science" have grown to become some of the hottest and most overused buzzwords over the past few years [Mac16; Pel15; Flo16]. This terminology has, by now, proliferated across the world of academia and has invaded a lot of academic conversations, presentations, and research agendas. Certainly, there is a lot of hype over big data, but no one can argue against the following fact: today's world is awash with data originating from numerous different disciplines (e.g. image processing, wireless communications, sensor networks, machine learning, financial data, genomics signal processing, and DNA microarrays) that are gathered by all means at an increasingly fast pace, and major efforts are underway to extract valuable information from them. A common theme among such instances of massive automatic data collection is *data outputs, which are comprised of many observations/measurements, but even more so, of a larger and larger number of variables of interest*. This is very different from the traditional assumption behind classical tools in estimation theory, under which only a few well-chosen variables are of interest.

To make ideas concrete, consider the fundamental statistical inference task of recovering an unknown signal from noisy linear measurements:

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}, \tag{1.1}$$

Henceforth, $\mathbf{y}$ denotes the vector of (say) m measurements, $\mathbf{x}_0$ is the unknown signal comprised of n variables, $\mathbf{A}$ is the measurement matrix and $\mathbf{z}$ is the noise vector. In the "classical world", the pervasive modeling assumption is that the number of variables n is fixed and small while m grows large. Here, there is a complete set of tools that can derive good estimates $\hat{\mathbf{x}}$ of $\mathbf{x}_0$. Importantly, an intellectually clean theoretical framework accompanies this set of tools with *performance guarantees* under all sorts of different settings (e.g., in the presence of outliers, deviations in the model, and so on). *Unfortunately, the classical tools and theory are mostly inadequate to cope with the dimensionality explosion that we experience in today's applications.*

In modern inference problems, unknown signals live in high-dimensional spaces. Hence, *the number of variables n is no longer small* (e.g., think of $\mathbf{x}_0$ representing

a large-scale image produced from magnetic resonance imaging (MRI), the human genome, or a vector of transmitted symbols in a massive MIMO scenario). Moreover, it is common that *the number of measurements m is less than the number of variables n*. For example, there are many genes, but only few patients with a given genetic disease, and, in applications like MRI there is not enough time to collect many observations [Don+00]. On the face of it, this makes the problem ill-posed (as the system of Equation in (1.1) becomes underdetermined). Fortunately, the signal of interest is often constrained structurally so that it only has a few degrees of freedom relative to its ambient dimension. For instance, MRI images often admit sparse representations in appropriate transform domains [CRT06], transmitted symbols belonging to finite constellations (e.g. m-PAM, m-QAM) only take values belonging to a finite alphabet, and covariance matrices are often well approximated by low-rank matrices. In summary, modern inference procedures and accompanying theory are developed in view of the following distinguishing features of *high-dimensional inference* problems:

(i) large number of variables to be estimated (large n),

(ii) (often) compressed measurements ($m < n$),

(iii) signals typically possess low dimensional *structure* (e.g., sparsity, low-rankness).

In this context, a new set of high-dimensional signal processing and statistics tools is required, ones that have the following favorable properties: the ability to reveal those structures and operate under a compressed number of measurements; computational efficiency; robustness to outliers, to model misspecification, and to missing data; and also, optimality guarantees. Among different approaches, convex-optimization based ones are often preferred since they offer enough flexibility and at the same time are usually amenable to analysis, simultaneously, with regard to computation, asymptotic theory and intuitive interpretation.

Over the past couple of decades, *non-smooth convex optimization* has emerged as a powerful structure-extracting tool for high-dimensional inference. These procedures obtain estimates $\hat{\mathbf{x}}$ of the unknown signal $\mathbf{x}_0$ by solving convex programs of the form:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \ \mathcal{L}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \lambda f(\mathbf{x}). \tag{1.2}$$

Henceforth, $\mathcal{L}$ represents a convex loss function that penalizes the residual, f is a convex (typically non-smooth) regularizer, and, $\lambda > 0$ is a regularizer parameter. Often, such estimators are referred to as *regularized M-estimators*, which includes for example, ℓ_1-penalized least-squares (aka LASSO), penalized least-absolute deviations (aka LAD), and, regularized Maximum-likelihood estimators. Regularized M-estimators have been around for at least twenty-years and have enjoyed great success in practice. In fact, the non-regularized versions (i.e., $f = 0$) of (1.2) correspond to the "plain vanilla" regression M-estimators, which were proposed and have been analyzed under the classical statistical setting since at least the 70s [Hub11]. The idea of adding a non-smooth regularizer to exploit the underlying structure of the unknown signal is also relatively old [CM73; SS86], but it appears to have gained significant popularity and attention starting in the mid 90's [Tib96; CDS98] and even more so about a decade ago in the context of *Compressed Sensing* [CRT06; Don06a].

The convex nature of (1.2) can in principle lead to corresponding tractable numerical algorithms. In particular, many of these programs (e.g. ℓ_1 and nuclear-norm penalized least-squares) are instances of convex conic programs, and so they can be solved in polynomial time using (say) interior point methods [BV09]. However, such standard solvers for convex programming, are often prohibitively computationally intense for modern large-scale data sets. This has led to an increasing interest in deriving and analyzing the convergence properties of simpler first-order methods (e.g., projected gradient-descent) that aim to make (1.2) scalable in high-dimensions (e.g., see [TA16; ZL15; ORS15; Bru+14] and references therein). Related *algorithmic* efforts involve designing solvers that can solve (1.2) in a distributed manner among different machines (e.g., [Rec+11]).

Rather than algorithmic issues, this thesis studies the fundamental *analytical* questions related to the inference performance of (1.2):

How good an estimator of the true unknown signal $\mathbf{x}_0$
is the solution $\hat{\mathbf{x}}$ *of the regularized M-estimator in* (1.2)*?* (Q.1)

A solution of the optimization in (1.2) consists of the estimate $\hat{\mathbf{x}}$ and the corresponding optimal cost, i.e., the minimum value of the objective function. Observe that the objective function in (1.2) is only a surrogate and so its optimal value is not by itself informative about the quality of estimation. A useful procedure, which is often employed in practice, in order to assess the estimation performance is through

cross-validation. The idea here is to split the data set (in our context, the pairs $(\mathbf{y}, \mathbf{A})$) into a training set and a validation set: an estimate $\hat{\mathbf{x}}$ is obtained by solving (1.2) based on the training set and its quality is evaluated on the rest of the data.

In this thesis we follow an *analytical approach*: we assume that the measurement matrix $\mathbf{A}$ is realized from the ensemble of $m \times n$ matrices with entries iid standard normal and derive an *exact* asymptotic characterization of the estimation quality of regularized M-estimators. The assumption on the random nature of the measurement matrix is by now a benchmark in the field of Compressed Sensing and high-dimensional signal-processing[1]: randomly generated matrices can *provably* yield good estimates of $\hat{\mathbf{x}}_0$ from compressed high-dimensional measurements [FR13; EK12; Boc+15]. Matrices sampled from the Gaussian ensemble have been traditionally useful in analytical works in random matrix theory and Compressed Sensing has been no exception to that rule[2]. In fact, one of the finest and most elegant (analytical) successes of the field corresponds to an exact characterization of the absolute minimum number of measurements required as a function of the structural complexity of the unknown signal, in order for convex optimization algorithms of the form in (1.2) to perfectly recover the signal in the *absence of noise*. These are known as *phase-transition* results in the literature of noiseless Compressed Sensing (see Section 2.2 for a survey of references).

One of the main contributions of this thesis is an extension of these results to the noisy case. When compared to the noiseless setting, the analysis under the presence of noise is not only more practical but is also inherently more challenging since: (a) one has to predict the precise value of the estimation error, rather than just discriminating between perfect recovery or not; (b) the performance depends not only on the number of measurements but also on the noise and signal statistics; (c) the optimization itself involves additional parameters that contribute to the final prediction.

Extensions of the theory to matrices (a) with entries drawn iid from other measurement ensembles, (b) with random orthogonal rows, and (c) ones that are elliptically

[1] For example random matrices have been known to be useful for dimensionality-reduction purposes since at least the mid 80's [JL84]

[2] Admittedly, this is a very special case of possible distributions of $\mathbf{A}$; in a large extent this is driven by the fact that it allows us to rely on some remarkable properties that govern the Gaussian ensemble. However, it should be noted that many relevant results obtained in random matrix theory for the Gaussian ensemble enjoy a *universality* property, i.e. they actually hold for a wider class of probability distributions. We will see later in Section 2.2 that this has recently proved to be the case for the Compressed Sensing problem as well [OT15].

distributed are also discussed. For instance, we derive explicit formulae characterizing the estimation performance under a certain class of orthogonal matrices (the isotropically random ones), and establish their *superior* performance when compared to Gaussians. Notably, we empirically observe that the same formulae continue to hold true for random Discrete Cosine Transform (DCT) and Hadamard matrices, which are often preferred in practice since they allow for fast multiplication and reduced storage complexity.

An important feature of the *exact* nature of the estimation predictions derived in this thesis is that *they can be used to compare performance between different instances of regularized M-estimators*. This lays the groundwork towards *developing a complete theory of regularized M-estimators in the high-dimensional regime* that involves providing rigorous answers to optimality questions regarding the choice of the involved parameters:

- *What is the optimal loss function and regularizer, under different settings, e.g., in the presence of outliers, particular structure of* $\mathbf{x}_0$*?*
- *What is the minimum achievable squared error in each one of those scenarios? Under what conditions can* $\mathbf{x}_0$ *be recovered with zero error?*
- *How may the regularizer parameter* λ *be optimally tuned?*
- *How does the sampling ratio* $\delta = m/n$ *affect the error?*
- *How robust is the estimation to deviations from the linear model in* (1.1)*?*

In the course of this thesis, we provide answers to some of the questions above[3]. For instance, we answer the last question by evaluating the performance of (say) regularized least-squares (aka Generalized LASSO) under measurements of the form $\mathbf{y} = g(\mathbf{A}\mathbf{x}_0)$, where g is a possibly unknown, random and nonlinear link function that aims to capture potential model miss-spesifications in (1.1).

Nonlinear measurements of this form might also arise *by design* (e.g. quantized measurements), in which case, our theoretical results lead to new opportunities in the *optimal design* of the nonlinear link function (e.g. by choice of the thresholds

[3]It is worth repeating that the high-dimensional regime of interest differs from the classical statistical regime. As such, the answers to these questions are expected to be (and in fact, they are) in general different than predicted by the classical theory of M-estimation that dates back to at least the 70's [Hub73; Hub11].

and levels of quantization). For an illustration, we prove that the optimal quantizer of the measurements that minimizes the estimation error of the Generalized LASSO is the celebrated Lloyd-Max quantizer.

Another prominent application of the developed theory is on the study of convex relaxation type decoders used in wireless communication settings with massive numbers of transmitting and receiving antennas. Owing to their tractability, such schemes are very well established in practical systems. Yet, the questions remain: *What is their bit-error rate performance? How do they compare to Maximum-Likelihood decoders?* We address these questions, the answers of which we further exploit by suggesting algorithmic improvements to boost their performance.

As we will see, the analysis is based on Gaussian process methods. In particular, at the heart of it lies a tight and extended version of a classical comparison inequality, proved by Gordon in 1988, in the presence of additional convexity assumptions. We call this the *Convex Gaussian Min-max Theorem* (CGMT). The CGMT might be of independent interest and may have applications that go beyond the scope of this dissertation.

Chapter 2

BACKGROUND, LITERATURE SURVEY AND SUMMARY OF CONTRIBUTIONS

The chapter begins with a survey on the theory of phase-transitions of convex optimization in *noiseless* linear inverse problems, which has been developed in a series of recent papers [DT09a; Sto09b; Cha+12; BLM+15; Ame+13; Sto13b], and which is an essential precursor to the material of this thesis. The rest of the chapter discusses in detail the scope and contributions of this dissertation.

Section 2.1 introduces some key ideas that have emerged from the existing theory of Compressed Sensing (CS). Section 2.2 reviews the theory of phase transitions in noiseless CS in some detail and surveys the relevant literature. In Section 2.3 we see that the presence of noise imposes additional challenges in the analysis. This leads us to Section 2.4, where we set the main objectives of this thesis and survey its contributions on a chapter by chapter basis.

2.1 Compressed Sensing

In broad terms, the field of Compressed Sensing (CS) studies the essential problem of recovering signals with low-dimensional structures from high-dimensional underdetermined measurements, which arises in many modern applications (tomography, accelerated MRI, radio interferometry, to name a few). The prototypical example is that of sparse recovery (or approximation), in which case the unknown signal is sparse (or approximately sparse), from linear (noisy) underdetermined measurements [CT06]. Another celebrated instance of Compressed Sensing is the problem of low-rank matrix completion, which arises in applications like predicting customer ratings or customer purchases for a recommendation system, and in system identification in control [RFP10]. Sparse approximation and related problems have been of interest since at least the early 90s, while some ideas can be traced even earlier in the literature. A series of works in the early 2000s studied the analytical performance of classical sparse approximation algorithms such as Basis Pursuit (BP) [CDS98] and Orthogonal Matching Pursuit (OMP) [TG07]. The celebrated papers [CT06] of Candes, Tao and Romberg and [Don06a] of Donoho initiated tremendous research activity over the last decade under the name of *Compressed Sensing* (CS). Although high-level, three of the most fruitful and successful ideas

that were developed during these years are as follows:

(i) *Exploit the underlying low-dimensional structure of the unknown signal*. Sparsity is only one such example of structure. Other often encountered examples include signals that are block-sparse, low-rank, slow-varying, take values over a finite alphabet, and so on. It has been recently recognized that recovery and analysis techniques that were initially developed for the problem of signal recovery extend naturally to other kinds of structures [Cha+12; Ame+13; FM14; OTH13b; TAH16]. Of course, such a *unifying* viewpoint has, among others, the clear advantage of enlarging the scope and applicability of the developed theory.

(ii) *Use of random measurement matrices*. The value of randomness in the measurement matrix model was recognized in the early work of Candes, Romberg and Tao [CT06] and has remained crucial in most subsequent literature. Randomness can be expressed in various forms (e.g. entrees sampled iid from various distributions, randomly subsampled Fourier matrices, etc.) and often guarantees the required incoherence property between the sampling matrix and the unknown vector, which makes the recovery problem well-posed. Moreover, the randomness turns out to be crucial in establishing analytical results. From a practical perspective, the randomness assumption is most relevant in applications where one has the freedom of designing the measurement matrix. Yet, valuable intuitions can be gained in instances where this is not the case.

(iii) *Use of recovery methods that are based on convex programming techniques*. This idea can be traced back very early in the literature [CM73; SS86]. In the context of sparse approximation the idea that gives rise to BP is to replace the original ℓ_0-minimization formulation of the problem by its convex relaxation, the ℓ_1-minimization. Of course, the same idea goes beyond sparsity and extends to more general notions of structure. As already mentioned, the advantage of convex methods is that they often lead to tractable numerical algorithms as well as to insightful statistical performance analyses.

The analysis and results in this thesis are also governed by the same ideas.

Linear Inverse Problems and Convex Optimization

The classical setting of CS is that of linear inverse problems, which assume noisy linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$ of an unknown, but structured, signal $\mathbf{x}_0 \in \mathbb{R}^n$. To keep things general, we do not specify the particular structure of $\mathbf{x}_0$, although is assumed known to us; it could be sparsity, group-sparsity, low-rankness, and so on. We are particularly interested in the scenario of compressed measurements, i.e. $m < n$, and the goal is that of estimating $\mathbf{x}_0$.

Towards this goal, *non-smooth convex optimization* techniques have emerged as a powerful technique. As already mentioned in Chapter 1, these methods produce an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ by solving (1.2). The loss function $\mathcal{L}$ aims to fit the final estimate to the observations based on the linear measurement model. On the other hand, the regularizer function f aims to exploit the particular structure of the unknown signal $\mathbf{x}_0$. For instance, it is by now well-understood in the CS literature that ℓ_1-regularization promotes sparsity, $\ell_{1,2}$-regularization is appropriate for group sparsity, and nuclear-norm-regularization promotes low-rank solutions. In fact, there are principled ways to construct such convex regularizer functions based on the idea of representing the low-dimensional structure of $\mathbf{x}_0$ as a decomposition into a *few* well-selected *atoms* [Cha+12]. The atomic-decomposition framework has roots in non-linear approximation [Jon92; Bar93] and was formally introduced in the context of noisy linear inverse problems under compressed measurements by Chandrasekaran et. al. in [Cha+12]. The framework explains in a principled and insightful way why ℓ_1-minimization and nuclear-norm are natural candidates for sparse and low-rank recovery, respectively, and generalizes the construction to several other types of low dimensional structures [Cha+12]. It should be mentioned that other recipes for associating convex regularizers to corresponding low-dimensional structures have been considered in the literature (e.g., [Bac10; BCW10]). A detailed review of all these goes beyond the scope of this thesis.

For the purposes of our discussion it is important to note that f in (1.2) aims to promote the structure of $\mathbf{x}_0$ and that typically good choices correspond to non-smooth functions (e.g. ℓ_1-norm, nuclear-norm). The fact that f is typically non-smooth imposes additionally challenges in the assessment of the estimation performance of (1.2), since the solution $\hat{\mathbf{x}}$ does not admit a closed-form expression (for example, this would be the situation in the case of a quadratic $\mathcal{L}$ function with a quadratic regularization, also known as ridge-regression).

As discussed, the main contribution of this thesis is providing exact answers to

Question (Q.1). We start by discussing the noiseless case (i.e. $\mathbf{z} = \mathbf{0}$ in (1.1) in Section 2.2, where recent studies provide an exact answer through a mathematically clean, elegant and general theory. Answering Question (Q.1) in the presence of noise is more challenging, and this is what this thesis focuses on addressing. Corresponding exact results in the literature are scarce and limited to specific instances of (1.2).

2.2 Noiseless Case

In the absence of noise, the measurements satisfy $\mathbf{y} = \mathbf{A}\mathbf{x}_0$. Naturally then (1.2) reduces to the following constrained convex minimization problem[1]:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{y}=\mathbf{A}\mathbf{x}} f(\mathbf{x}). \tag{2.1}$$

Since measurements are noiseless, we hope that the unknown signal $\mathbf{x}_0$ can be recovered exactly, i.e. $\hat{\mathbf{x}} = \mathbf{x}_0$. Consequently, the fundamental question (Q.1) essentially reduces to the following:

Under what conditions is the solution $\hat{\mathbf{x}}$ *of* (2.1) unique and equal to $\mathbf{x}_0$? (Q.2)

Null-space Condition

When $\hat{\mathbf{x}} = \mathbf{x}_0$ is the unique solution of (2.1), we say that the program succeeds, otherwise it fails. A necessary and sufficient condition for success of (2.1) is known as the "null-space condition" and is given in the lemma below. Let $\mathcal{N}(\mathbf{A})$ denote the null-space of the measurement matrix $\mathbf{A}$ and $\mathcal{T}_f(\mathbf{x}_0)$ the tangent cone of f at $\mathbf{x}_0$, as defined below:

Definition 2.2.1 (Tangent Cone)**.** The tangent cone $\mathcal{T}_f(\mathbf{x}_0)$ of f at $\mathbf{x}_0$ is defined as the closure of the conic hull of the set of descent directions $\mathcal{D}_f(\mathbf{x}_0)$ of f at $\mathbf{x}_0$:

$$\mathcal{D}_f(\mathbf{x}_0) := \{\mathbf{v} \mid f(\mathbf{x}_0 + \mathbf{v}) \leq f(\mathbf{x}_0)\}.$$

Proposition 2.2.1 (Null-space Condition)**.** $\mathbf{x}_0$ *is the unique minimizer of* (2.1) *iff* $\mathcal{N}(\mathbf{A}) \cap \mathcal{D}_f(\mathbf{x}_0) = \{\mathbf{0}\}$*, or equivalently,*

$$\mathcal{N}(\mathbf{A}) \cap \mathcal{T}_f(\mathbf{x}_0) = \{\mathbf{0}\}. \tag{2.2}$$

The proof of the proposition is almost straightforward but is included for completeness. See Figure 5.1 for a simple schematic representation of the null-space condition for the case of sparse recovery using ℓ_1-minimization.

[1] Besides convex relaxation based schemes, other signal recovery methods such as greedy pursuits and combinatorial algorithms have also been proposed and analyzed in the relevant literature. See for example [NT09] and references therein.

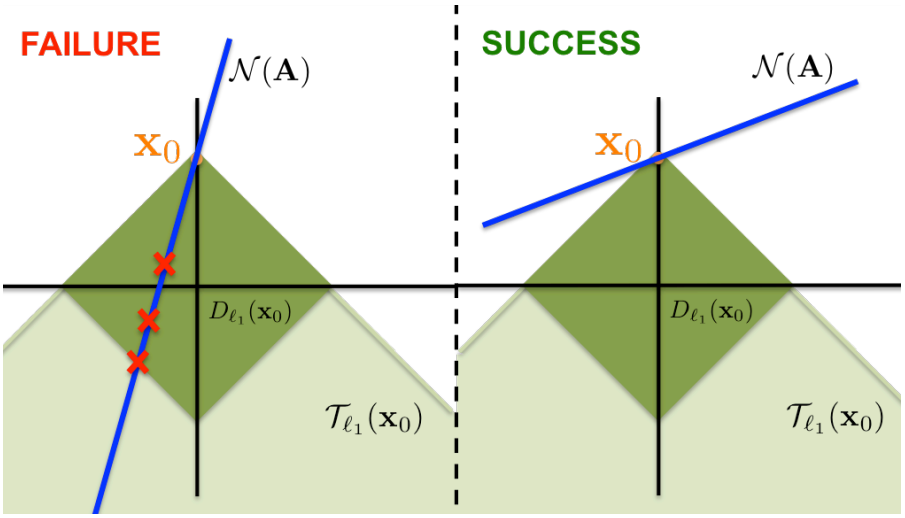

Figure 2.1: Illustration of the Null-space condition (Proposition 2.2.1).

Proof. (*of Proposition 2.2.1*). It is convenient to change the variable in the optimization in (2.1) to the *error vector* $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$. This gives

$$\min_{\mathbf{A}\mathbf{w}=0} f(\mathbf{x}_0 + \mathbf{w}). \tag{2.3}$$

We show that $\hat{\mathbf{w}} = 0$ is the unique solution to this minimization iff $\mathcal{N}(\mathbf{A}) \cap \mathcal{T}_f(\mathbf{x}_0) = \{\mathbf{0}\}$. Let $\mathbf{v} \in \mathcal{N}(\mathbf{A}) \cap \mathcal{D}_f(\mathbf{x}_0)$. Clearly, $\mathbf{A}\mathbf{v} = 0$ and $\mathbf{v}$ is feasible in (2.3). Moreover, $f(\mathbf{x}_0 + \mathbf{v}) \leq f(\mathbf{x}_0)$ by definition of the set of descent directions. Combined, $\mathbf{v}$ is a minimizer of (2.3), which completes the proof. The equivalence of the two conditions in the statement of the proposition follows by Definition 2.2.1 and the fact that $\mathcal{N}(\mathbf{A})$ is a linear subspace. □

Condition 2.2.1 is geometric in nature: "When does the null-space of the measurement matrix not intersect (other than at $\mathbf{0}$, of course) the tangent cone?". Checking this for deterministic matrices is hard. However, it turns out to be tractable when $\mathbf{A}$ possesses specific randomness properties. When $\mathbf{A}$ is realized from some probability ensemble, then it is desirable to satisfy Condition (2.2) *with high probability* (whp) over the matrix realization.

Gaussian Matrices: Escape through a mesh & Gaussian width

Suppose that the entries of $\mathbf{A}$ are sampled iid from a standard normal distribution. It is well-known that the null-space of an iid Gaussian matrix is *isotropically random*. Then, the question becomes: "When does a random subspace (cf. the null-space of $\mathbf{A}$) miss a fixed cone (cf. the tangent cone $\mathcal{T}_f(\mathbf{x}_0)$) with high probability?". The

answer to this question was given by Gordon in 1988 [Gor88] and is known as the "escape through a mesh lemma". Gordon proved and used the lemma in a different context; Rudelson and Vershynin first noticed its relevance to the CS problem in 2006 [RV06], in the context of sparse signal recovery.

Before stating the lemma, we will introduce two very useful concepts, namely the "*minimum conic singular value*" (mCSV) and the "*conic Gaussian width*".

Definition 2.2.2 (Minimum conic singular value)**.** Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. The minimum conic singular value of $\mathbf{A}$ with respect to a cone $\mathcal{K} \subset \mathbb{R}^n$ is defined as,

$$\sigma_{\min}(\mathbf{A};\mathcal{K}) = \inf_{\mathbf{w}\in\mathcal{K}\cap\mathcal{S}^{n-1}} \|\mathbf{A}\mathbf{w}\|_2.$$

Henceforth, $\mathcal{S}^{n-1}$ denotes the unit sphere in $\mathbb{R}^n$. To see the relevance of Definition 2.2.2 to our discussion, observe that

$$\sigma_{\min}(\mathbf{A};\mathcal{T}_f(\mathbf{x}_0)) > 0 \quad \Rightarrow \quad (2.2) \text{ holds.} \tag{2.4}$$

Also, note that $\sigma_{\min}(\mathbf{A};\mathbb{R}^n)$ is the minimum singular value of $\mathbf{A}$.

Definition 2.2.3 (conic Gaussian width)**.** Let $\mathbf{h} \in \mathbb{R}^n$ have entries iid standard normal. The Gaussian width of a cone (not necessarily convex) $\mathcal{K} \subset \mathbb{R}^n$ is denoted by $\omega(\mathcal{K})$ and is defined as:

$$\omega(\mathcal{K}) := \mathbb{E}[\sup_{\mathbf{w}\in\mathcal{K}\cap\mathcal{S}^{n-1}} \mathbf{h}^T\mathbf{w}],$$

where the expectation is over the randomness of $\mathbf{h}$.

The Gaussian width is a geometric measure of the size of the cone and plays a central role in asymptotic convex geometry [AAGM15; LT91].

Proposition 2.2.2 (Escape through a mesh)**.** *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ have entries iid standard normal and $\mathcal{K}$ be a cone in $\mathbb{R}^n$. Then, for any $t > 0$, it holds with probability at least $1 - e^{-t^2/2}$ that*

$$\sigma_{\min}(\mathbf{A};\mathcal{K}) \geq \sqrt{m-1} - \omega(\mathcal{K}) - t.$$

In essence Proposition 2.2.2 is contained in [Gor88]. The result as presented above is drawn from [Tro15]. Its proof is based on *Gaussian process methods* and specifically on *Gordon's Gaussian Min-max Theorem* (GMT) [Gor88]. GMT plays a central role in this thesis, but we defer this discussion along with a proof of Proposition 2.2.2 to Chapter 3.2.

Combining Proposition 2.2.2 with (2.4) and Proposition 2.2.1, shows that, when the measurement matrix is Gaussian, the convex program (2.1) succeeds with exponentially high probability, as long as the number of measurements satisfies

$$m \geq (\omega(\mathcal{T}_f(\mathbf{x}_0)) - t)^2 + 1. \tag{2.5}$$

In this sufficient condition for successful recovery, the role of the regularizer function f and the particular structure of $\mathbf{x}_0$ are summarized by the Gaussian width of the tangent cone $\omega(\mathcal{T}_f(\mathbf{x}_0))$. Certainly, (2.5) is alone a remarkable result. Yet, it would be of limited practical use unless $\omega(\mathcal{T}_f(\mathbf{x}_0))$ can be computed for interesting regularizers and for corresponding structures. Thankfully, it will be soon shown that this indeed the case! Towards this direction, note from (2.5) that any upper bound on the Gaussian width translates to a sufficient lower bound on the required number of measurements for successful recovery. Importantly, it turns out that for many examples of structured signals that are encountered in practice, there exist good choices of the regularizer function such that

$$\omega^2(\mathcal{T}_f(\mathbf{x}_0)) \ll n. \tag{2.6}$$

Therefore, under iid Gaussian design matrices the convex optimization (2.1) successfully recovers $\mathbf{x}_0$ whp (over $\mathbf{A}$) with number of measurements that is (much) less than the ambient dimension n of the signal.

Rudelson and Vershynin [RV06] were the first to derive an upper bound on the Gaussian width in the case of k-sparse recovery with ℓ_1-regularization and concluded that $\approx 8k \log(n/k)$ number of measurements are sufficient. In 2009, Stojnic performed a more careful analysis using a convex optimization duality argument and obtained a sharper upper bound [Sto09b]. Through simulations he observed this bound to be *tight* (asymptotically with respect to the problem dimensions), i.e. a greater number of measurements than it leads to success whp, while fewer of them leads to failure whp. This sharp transition between success and failure is known as "*phase transition*" in CS, as we discuss next. It was soon realized that Stojnic's upper bounding technique could be extended to other related problems. Oymak & Hassibi used it to study the low-rank recovery problem with nuclear-norm minimization [OH10]. Chandrasekaran et al. realized the necessary abstractions behind Stojnic's technique and phrased it in terms of convex-geometric notions such as the "tangent cone", "cone of subdifferential", etc. [Cha+12]. Together with subsequent works by Amenlunxen et. al. [Ame+13] and Foygel & Lester [FM14], this led to

a clean recipe that derives upper bounds on the Gaussian width for general convex regularizers. The derived bounds appeared to be tight via simulations, and this favorable property was proved in [Ame+13, Thm. 4.3] (see also [FM14, Prop. 1]).

The recipe for controlling the conic Gaussian width is based on polarity. We will review the basic idea next. This will also allow us to introduce the relevant geometric concepts of "*Gaussian-distance squared*" and "*statistical dimension*", which will turn out to play key role in the results of this thesis, as well.

Calculating the Gaussian width: Gaussian distance squared

The contents of this section largely follow the treatment in [Tro15]. The technique was developed in a series of works [Sto09b; Cha+12; Ame+13; FM14].

Recalling the definition of the Gaussian width, it follows that

$$(\omega(\mathcal{T}_f(\mathbf{x}_0)))^2 \leq (\mathbb{E}[\sup_{\mathbf{w}\in\mathcal{T}_f(\mathbf{x}_0)\cap\mathcal{B}^{n-1}} \mathbf{h}^T\mathbf{w}])^2 \leq \mathbb{E}([\sup_{\mathbf{w}\in\mathcal{T}_f(\mathbf{x}_0)\cap\mathcal{B}^{n-1}} \mathbf{h}^T\mathbf{w}])^2 =: \delta(\mathcal{T}_f(\mathbf{x}_0)), \tag{2.7}$$

where: (i) for the first inequality we have enlarged the constraint set in the maximization to be over the intersection of the cone with the unit ball $\mathcal{B}^{n-1}$, rather than with the unit sphere, (ii) for the second inequality we have used Jensen's inequality. The quantity on the right-hand side (RHS) of (2.7) is known in the literature as the "*statistical dimension*" of the tangent cone and is denoted by $\delta(\mathcal{T}_f(\mathbf{x}_0))$ [Ame+13]. Amelunxen et. al. showed that, compared to the Gaussian width, the statistical dimension delivers a better summary parameter of the size of a cone since it canonically extends the dimension of a subspace to the class of convex cones, and it satisfies many elegant identities [Ame+13, Prop. 3.1]. However, the two notions are very closely related; in fact, it can be shown [Ame+13, Prop. 2] that

$$(\omega(\mathcal{T}_f(\mathbf{x}_0)))^2 \leq \delta(\mathcal{T}_f(\mathbf{x}_0)) \leq (\omega(\mathcal{T}_f(\mathbf{x}_0)))^2 + 1, \tag{2.8}$$

where of course the lower bound is a restatement of (2.7).

As we show next, there is a principled way to derive upper bounds on the statistical dimension of the tangent cone. In fact, it is shown that (in most interesting cases) these bounds are sharp, asymptotically, in the problem dimensions. Thus, in view of (2.8), they translate to sharp numerical estimates of the Gaussian width.

Upper bounding the statistical dimension is based on a polarity argument. First, observe that

$$\delta(\mathcal{T}_f(\mathbf{x}_0)) := \mathbb{E}[\text{dist}^2(\mathbf{h}, (\mathcal{T}_f(\mathbf{x}_0))^\circ)], \tag{2.9}$$

where we have used the convexity of f (see for example [**troppBowling**]) and the dist function is used to denote the distance of a vector to a set. Formally, for a nonempty, convex, closed set C,

$$\text{dist}(\mathbf{x}, C) = \inf_{\mathbf{v}\in C} \|\mathbf{x} - \mathbf{v}\|_2.$$

Convexity and closeness assures that the infimum is attained at a unique point lying in the set C. Also, $(\cdot)^\circ$ is used to denote the polar of a cone[2]. A classical result in convex analysis characterizes the polar of the tangent cone in terms of the subdifferential of the function [Roc97, Thm. 23.7]. This polarity correspondence is key.

Recall here that the subdifferential of f at $\mathbf{x}_0$ is the set of vectors:

$$\partial f(\mathbf{x}_0) = \left\{\mathbf{s} \in \mathbb{R}^n \mid f(\mathbf{x}_0 + \mathbf{v}) \geq f(\mathbf{x}_0) + \mathbf{s}^T\mathbf{v}, \forall \mathbf{v} \in \mathbb{R}^n\right\},$$

and is always a compact and convex set [Roc97]. Also, if $\mathbf{x}_0$ is not a minimizer of f, then $\partial f(\mathbf{x}_0)$ does not contain the origin. For any nonnegative number $\tau \geq 0$, we denote the, scaled (by τ), subdifferential set as

$$\tau \cdot \partial f(\mathbf{x}_0) = \{\tau\mathbf{s} \mid \mathbf{s} \in \partial f(\mathbf{x}_0)\},$$

and, for the conic hull of the subdifferential we write

$$\text{cone}(\partial f(\mathbf{x}_0)) = \{\mathbf{s} \mid \mathbf{s} \in \tau \cdot \partial f(\mathbf{x}_0), \text{ for some } \tau \geq 0\}.$$

Proposition 2.2.3 (Polarity, [Roc97])**.** *Let f be proper convex and such that $\mathbf{x}_0$ is not a minimizer of f. Then,*

$$(\mathcal{T}_f(\mathbf{x}_0))^\circ = \text{cone}(\partial f(\mathbf{x}_0)).$$

Clearly then,

$$\delta(\mathcal{T}_f(\mathbf{x}_0)) = \mathbb{E}[\text{dist}^2(\mathbf{h}, \text{cone}(\partial f(\mathbf{x}_0)))] =: \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))). \tag{2.10}$$

[2] As a reminder, the polar $\mathcal{K}^\circ$ of a cone $\mathcal{K}$ is the *closed convex cone* defined as $\mathcal{K}^\circ := \{\mathbf{v} \mid \mathbf{v}^T\mathbf{x} \leq 0 \text{ for all } \mathbf{x} \in \mathcal{K}\}$.

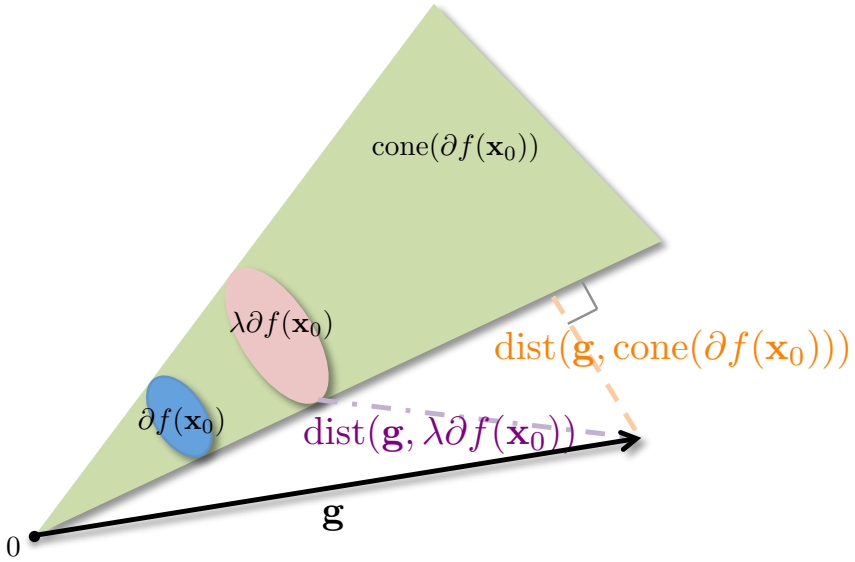

Figure 2.2: Illustration of the distance of a vector to the scaled subdifferential $\lambda\partial f(\mathbf{x}_0)$ and to the cone of subdifferential $\mathrm{cone}(\partial f(\mathbf{x}_0))$.

Above, we have introduced another notation for the statistical dimension, which is indicative of the fact that it corresponds to the "*Gaussian distance squared to the cone of subdifferential*". For this, it is not hard to see that

$$\delta(\mathcal{T}_f(\mathbf{x}_0)) = \mathbf{D}(\mathrm{cone}(\partial f(\mathbf{x}_0))) = \mathbb{E}\big[\inf_{\tau\geq 0} \mathrm{dist}^2(\mathbf{h}, \tau\cdot\partial f(\mathbf{x}_0))\big] \leq \inf_{\tau\geq 0} \mathbf{D}(\tau\partial f(\mathbf{x}_0)), \tag{2.11}$$

where the last equality follows since the distance to a union of sets equals the minimum distance to any of its members, and, we have defined the "*Gaussian distance squared to the scaled subdifferential*":

$$\mathbf{D}(\tau\partial f(\mathbf{x}_0)) := \mathbb{E}[\mathrm{dist}^2(\mathbf{h}, \tau\cdot\partial f(\mathbf{x}_0))]. \tag{2.12}$$

For many commonly encountered examples of regularizer functions f and associated structures of $\mathbf{x}_0$, $\mathbf{D}(\tau\partial f(\mathbf{x}_0))$ can be computed for $\tau \geq 0$. Then, the minimum value over all such parameters τ provides a (numerical) upper bound to the statistical dimension (correspondingly to the Gaussian width). This elegant recipe was developed in [Cha+12; Ame+13]. Moreover, it is shown in [Ame+13, Thm. 4.3] and [FM14, Prop. 1] that these bounds are asymptotically sharp, as the problem dimensions grow large. We refer the reader to the original references for the exact statements of this result; for our purposes, it suffices to remember that (for most

cases of interest):

$$\mathbf{D}(\mathrm{cone}(\partial f(\mathbf{x}_0))) \approx \inf_{\tau \geq 0} \mathbf{D}(\tau \partial f(\mathbf{x}_0)). \tag{2.13}$$

As a mere illustration, when f is the ℓ_1-norm and $\mathbf{x}_0$ is k-sparse, $\partial f(\mathbf{x}_0)$ has a well-known simple characterization and $\mathbf{D}(\tau \partial f(\mathbf{x}_0))$ admits simple closed-form expressions in terms of the tail distribution $Q(\tau)$ of a standard Gaussian (e.g., Appendix D):

$$\mathbf{D}(\tau \partial f(\mathbf{x}_0)) = k(1+\tau)^2 + (n-k)(2(1+\tau^2)Q(\tau) - \sqrt{2/\pi}\tau e^{-\frac{\tau^2}{2}}). \tag{2.14}$$

The minimum of this expression over $\tau \geq 0$ is equal to the statistical dimension and is easy to numerically evaluate. Alternatively, one can obtain a closed form upper bound by evaluating $\mathbf{D}(\tau \partial f(\mathbf{x}_0))$ at $\tau = \sqrt{2\log(n/k)}$, which yields a simple closed-form expression:

$$\mathbf{D}(\mathrm{cone}(\partial f(\mathbf{x}_0))) \leq 2k(\log(n/k) + 3/4). \tag{2.15}$$

Following the same recipe, it can be shown that for f, the nuclear norm, and $\mathbf{x}_0 = \mathrm{vec}(\mathbf{X}_0)$ of a rank-r matrix $\mathbf{X}_0 \in \mathbb{R}^{n \times n}$,

$$\mathbf{D}(\mathrm{cone}(\partial f(\mathbf{x}_0))) \leq 6nr. \tag{2.16}$$

We refer the reader to Appendix D for some details on these calculations and for more examples. A useful observation amounts to the fact that the above upper bounds do not depend on the specific values of $\mathbf{x}_0$; rather, they only depend on the degree of structure they possess, i.e. on the sparsity level and the rank, respectively.

The take-away message here is that one can compute asymptotically sharp estimates of the statistical dimension of the tangent cone via the Gaussian distance squared to the scaled subdiferential. These estimates translate (in view of (2.5) and of (2.8)) to explicit expressions on the minimum required number of measurements for successful recovery.

Sharp phase-transitions

The nullspace condition of Proposition 2.2.1 provides a sufficient and necessary condition for the success of (2.1). Gordon's escape through a mesh Lemma was used to show that $\approx \omega(\mathcal{T}_f(\mathbf{x}_0))^2$ number of measurements are sufficient for the nullspace condition to hold (cf. (2.5)). Remarkably, this much number of measurements

is also necessary. This fact, which had earlier been observed empirically in [Sto09b; OH10; Cha+12] was proved by Amelunxen et. al. [Ame+13] in 2014. (The same result was also proved by Stojnic in an independent effort for the case of sparse recovery with ℓ_1-minimization [Sto13b]. See the next section for a detailed literature survey.)

Theorem 2.2.1 (Phase transitions in noiseless Linear inverse problems, [Ame+13])**.** *Let $\mathbf{x}_0 \in \mathbb{R}^n$ be a fixed vector and f be a proper, convex function. Suppose $\mathbf{A}$ has entries iid standard normal, noiseless linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ and consider the minimization in (2.1). For all $p \in (0, 1)$,*

$$\begin{aligned} m \leq \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) - \alpha_p\sqrt{n} \;\Rightarrow\; &\text{(2.1) succeeds with probability} \leq p, \\ m \geq \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) + \alpha_p\sqrt{n} \;\Rightarrow\; &\text{(2.1) succeeds with probability} \geq 1-p, \end{aligned} \tag{2.17}$$

where $\alpha_p := \sqrt{8\log(4/p)}$.

Precise and general results

When the measurement matrix has entries iid Gaussian, Theorem 2.2.1 provides a *precise* and *unifying* answer to question Q.2. This is in contrast to early results in the field which instead were *order-wise* and/or problem-specific. Order-wise results correspond to bounds on the required number of measurements to succeed that involve unknown (or loose) constants.

Apart from the mathematical challenge per se and the resulting elegant and transparent theory, there are several further benefits that come along with precise and general results.

(i) They permit comparing the performance of different instances of (2.1), those resulting from different choices of the regularizer function f. This in turn leads to principled recipes to optimally choose the regularizer function (e.g. [Cha+12]).

(ii) They can be used to study the convergence rates of fast iterative solvers of (2.1). Please see [ORS15].

(iii) They answer questions regarding time-data tradeoffs that occur in modern data analysis [CJ13; Bru+14]. Such tradeoffs refer to the ability to reduce the computational complexity of an inference procedure when one has access to increasingly large datasets [CJ13].

Naturally, owing to their precise nature, the results of this thesis inherit these benefits, as well.

Universality

In view of Theorem 2.2.1 the noiseless compressed sensing problem as posed in question Q.2 has been completely solved in the case where the random measurement matrix $\mathbf{A}$ has entries iid Gaussian.

The Gaussian assumption is appealing mainly for two reasons: (i) It opens the door to a very rich set of probabilistic tools available in the literature for the Gaussian ensemble. Notable examples that we saw being critical in the establishment of Theorem 2.2.1 include Gaussian process inequalities (such as Proposition 2.2.2), and the Gaussian concentration of Lipschitz functions (see Proposition 3.1.1). (ii) Results that hold under this assumption enjoy a remarkable *universality* property in that they continue to hold for a fairly broad family of other ensembles.

The universality property of the Gaussian distribution is by now well established in random matrix theory; important results, such as the semi-circle law, were first shown to hold for Gaussian matrices and were subsequently proved to hold for much broader classes of random matrices [Tao12; Joh06]. But does this apply to the noiseless compressed sensing problem? Is the phase-transition result of Theorem 2.2.1 universal?

Extensive empirical investigations had been reported in the literature suggesting that this is indeed the case [DT09b]. Bayati et. al. [BLM+15] were the first to rigorously demonstrate universality of the phase-transition of ℓ_1-minimization over a class of random ensembles beyond Gaussians. Only very recently, Oymak & Tropp [OT15] have extended this result to a broader class of measurement models. Even more importantly, they succeed in establishing the universality property under the general setting of Theorem 2.2.1, thus significantly broadening its scope and its implications to measurement matrices that have entries iid following a broad class of probability distributions.

But, what happens beyond iid measurement ensembles? Does the universality property of Theorem 2.2.1 extend to such cases? Certainly, there are important examples of random measurement models that fall outside the class of iid matrices for which answering these questions becomes important. A prime example includes random matrices with orthogonal rows. For instance, the use of matrices formed by randomly subsampled rows of Fourier, discrete-cosine and Hadamard matri-

ces is appealing in practice since such matrices allow for fast multiplication and reduced storage complexity. For the specific class of Isotropically Random Orthogonal (IRO) matrices, i.e. matrices that are sampled uniformly from the manifold of row-orthogonal matrices satisfying $\mathbf{A}\mathbf{A}^T = \mathbf{I}_m$, the answer to the second question above is affirmative and easy to prove. It is a well-known fact that the nullspace of an IRO matrix, which is what matters for the performance of (2.1) thanks to the nullspace condition Proposition 2.2.1, is an isotropically random orthogonal subspace in $\mathbb{R}^n$ of dimension $n - m$. In particular, this means that it follows the same distribution as the nullspace of an iid Gaussian random matrix, which in turn leads to the phase-transitions being the same.

As part of this thesis, we will establish that this is no more the case in the noisy setting: the performance of convex-type methods under IRO matrices is superior to that of Gaussians.

Literature Survey

As mentioned, the work on phase transitions of non-smooth convex optimization used to recover structured signals from noiseless linear measurements is an essential precursor to the material of this thesis. Hence, we have discussed it above in detail. Here, we put together together a narrative description of the relevant contributions starting from the seminal works of Candes & Tao and of Donoho all the way to the papers that establish Theorem 2.2.1. As discussed, this line of work attempts to characterize the minimum number of measurements, say m_*, as a function of the structural complexity of $\mathbf{x}_0$ and of the choice of f, such that $\mathbf{x}_0$ is the unique solution of (2.1) with probability approaching 1 if and only if $m > m_*$.

The early works in the field studied this question in the context of sparse signal recovery and ℓ_1-minimization; they showed that it can recover a sparse signal $\mathbf{x}_0$ from fewer observations than the ambient dimension n [CT06; Don06b; DT09a]. On the one hand, Candes & Tao assumed the measurement matrix $\mathbf{A}$ satisfies certain restricted isometry properties and provided an “order-optimal" (with very loose constants) upper bound on m_*. On the other hand, when $\mathbf{A}$ has entries iid Gaussian, Donoho and Tanner obtained an asymptotically precise upper bound on m_*, via polytope angle calculations and related ideas from combinatorial geometry. The results of Donoho and Tanner were latter extended to weighted ℓ_1-minimization and were supplemented with robustness guarantees in [XH11]. However, the combinatorial geometry approach has proved hard to extend to regularizers whose set of

sub-gradients is non-polyhedral (the most representative such example is nuclear-norm minimization for the low-rank recovery problem, see for example [RXH11] for some early loose performance bounds using this approach).

In early 2005, Rudelson & Vershynin [RV06] proposed a different approach to studying ℓ_1-minimization that uses Gordon's Gaussian Min-max Theorem (GMT) (specifically, a corollary of it known as the "escape through a mesh" lemma [Gor88]). Stojnic refined this approach and obtained an empirically sharp upper bound on m_* both for sparse and group-sparse vectors [Sto09b; Sto09a]. This approach is simpler than that of Donoho & Tanner and extends to very general settings. Oymak & Hassibi [OH10] used it to study the low-rank recovery problem, and later, Chandrasekaran et al. [Cha+12] developed a geometric framework and were able to analyze general structures and convex regularizers f, while clarifying the key role played in the analysis by the geometric concept of "Gaussian width" [Gor88]. See also [MT14; FM14] for extensions to other signal recovery problems.

The works discussed thus far only derive upper bounds on m_*. Matching lower bounds that prove the asymptotic tightness of the former (known as *phase-transition*) are even more recent. Bayati et. al [BLM+15] rigorously demonstrates the phase transition phenomenon for ℓ_1-minimization. The analysis is based on a state evolution framework for an iterative Approximate Message Passing (AMP) algorithm inspired by statistical physics, which was earlier introduced by Donoho et. al [DMM09; BM11]. Amelunxen et. al. [Ame+13] took a different route; using tools from conic integral geometry they established for the first time that previous results of [Cha+12] were tight. In particular, they showed that: (a) a phase transition almost always exists for general convex regularizers f; (b) that it can be located exactly by computing the "statistical dimension" (which is very related to the "Gaussian width", but has some extra favorable properties); and (c) that it is possible to give accurate upper and lower bounds for the statistical dimension. Subsequently, Stojnic [Sto13b] combined his earlier approach, whith was based on Gordon's GMT, with a convex duality argument and used this to prove that his earlier bounds on ℓ_1 and $\ell_{1,2}$ were asymptotically tight. (A similar observation was also reported in [Ame+13, Rem. 2.9].) Stojnic's approach deserves special credit under the prism of our work, since it essentially motivated and inspired most of our contributions on the study of the precise reconstruction error under noisy measurements using Gaussian process methods.

2.3 Noisy Case: The Challenge

The noisy setting is significantly more challenging than the noiseless one. To begin with, the addition of noise, which can potentially follow many different distributions, leads to a much richer class of recovery optimization problems. In particular, compared to (2.1), the minimization in (1.2) offers the additional flexibility of choosing different loss functions. On a same note, (1.2) poses additional questions regarding the choice of the regularizer parameter λ and how it affects the recovery performance. Moreover, in the presence of noise, it is in general too optimistic to expect exact recovery of the true unknown signal (as did in the noiseless case). Instead, a more reasonable goal is that of obtaining a good estimate of it, but there can be a plethora of different ways to quantify this. Perhaps the most popular and widely-used measure of performance is the *squared-error* $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2$, which measures the deviation of the estimate $\hat{\mathbf{x}}$ from the true signal $\mathbf{x}_0$ in ℓ_2-norm. However, depending on the specific application other measures might be more appropriate. For instance, in sparse recovery it is often of interest to guarantee that $\hat{\mathbf{x}}$ reveals the correct support (i.e., location of non-zero entries) of $\mathbf{x}_0$. Hence, the challenge becomes that of providing guarantees for a variety of performance measures.

In short, in the presence of noise, a general and precise theory that would resemble that of noiseless Compressed Sensing as presented in Section 2.2 should be such that it addresses the following rich set of questions.

(Q.a) Can we obtain precise and general characterization of the recovery performance of (1.2) as a function of all the involved parameters (e.g., loss function, regularizer, regularizer parameter, noise-distribution)?

(Q.b) Can we do so in the context of a mathematically clean and transparent analysis framework?

(Q.c) How are the results related to those of noiseless compressed sensing? Is it possible to obtain those as special cases?

(Q.d) To what extent do the error formulae obtained for iid Gaussian matrices continue to hold true for random matrices from other ensembles?

(Q.e) Is it possible to obtain guarantees for various measures of performance (e.g., squared-error, probability of support recovery)?

(Q.f) What if the measurements are non-linear? Is it still meaningful to use (1.2) for the recovery? Equivalently, how robust is the performance of (1.2) to model miss-specifications?

2.4 Thesis Contributions & Organization

This dissertation extends the theory and the results of *noiseless* Compressed Sensing to the more challenging and practically important case of *noisy measurements*. In a fashion similar to the former results discussed in Section 2.2, we consider a random Gaussian model for the measurement matrix. We obtain results that are *precise* and *general*; hence, they enjoy the favorable properties of corresponding results in Section 2.2.

In particular, we *develop a novel analytical framework, which provides accurate answers to all the questions raised in Section 2.3*. Interestingly, the framework is based on Gaussian process inequalities; more specifically, it relies on a novel strengthened version of Gordon's Gaussian Min-max Theorem (GMT) in the presence of convexity, which we call the *Convex Gaussian Min-max Theorem (CGMT)*. Note that the original GMT is the basis of the "escape through a mesh" Proposition 2.2.2, which in turn is key in the analysis of noiseless CS. Overall, this *creates a coherent and elegant story that makes our understanding of the behavior of convex signal recovery methods with Gaussian measurements very clear.*

For ease of reference, we detail the contributions on a chapter by chapter basis below. Browsing through the opening paragraphs of each chapter should also serve as an overview of its scope.

Chapter 3

Chapter 3 establishes the Convex Gaussian Min-Max Theorem (CGMT), which is key to developing the analysis framework. The chapter begins with an introduction of the popular Slepian's Lemma and classical uses of it. This leads us to Gordon's comparison theorem that is a non-trivial extension of Slepian's result proved in 1988. The CGMT is a tight and strengthened version of Gordon's original result when combined with additional convexity assumptions, and might be of independent interest with applications that go beyond the scope of this dissertation. The proof of the theorem is also included in this chapter.

Some technical material is deferred to Appendix A.

Chapter 4

Chapter 4 studies the squared-error performance of regularized M-estimators (cf. (1.2)). More specifically, it establishes in a single theorem an asymptotically *precise* expression for the squared error $\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2$ of (1.2). The scope of the theorem is very *general* since it is valid under only very mild regularity assumptions on the loss functions, on the regularizer functions and on the noise distribution. Essentially, this chapter provides an answer to Question (Q.a) when performance is measured via the squared-error. The study reveals a new summary parameter, termed the *expected Moreau envelope*, that plays a central role in the error characterization, and is in fact a generalization of the Guassian squared distance that appeared earlier in Section 2.2. The chapter concludes with a detailed survey of the relevant literature on precise performance guarantees for regularized M-estimators.

Appendix B includes the proof of the theorem and of related useful results, such as properties of the expected Moreau envelope.

Chapter 5

Chapter 5 describes the general framework to analyze the recovery performance of (1.2). The framework is based on the CGMT and consists of four major steps, which are all explained here. It is the backbone for the proofs of the vast majority of the results that appear in the thesis. To better illustrate the steps involved, we outline how the framework is used to prove the theorem of Chapter 4. Apart from technical details, the basic mechanics are easy to explain, thus making the analysis transparent and providing an affirmative answer to Question (Q.b).

Chapter 6

In Chapter 6, we present results after applying the general theorem of Chapter 4 to specific popular instances of (1.2) and obtain instance-specific error expressions. Some of the instances considered include M-estimators without regularization, Regularized Least-squares (aka Generalized LASSO), Regularized Least Absolute Deviation (LAD) and more. We then analyze these error expressions to answer a number of interesting questions such as "what is the minimum number of measurements required for stable recovery? How does this number depend on regularization?", "Are there problem instances for which specific choices of loss and regularizer functions achieve the MMSE performance?". By the end of the chapter, we present simulation results that illustrate the validity of the theoretical predictions.

All proofs are deferred to the corresponding Appendix C.

Chapter 7

Chapter 7 further specializes the general results of Chapter 4 to the squared-error performance of regularized LASSO (aka generalized LASSO) in the regime of *high-SNR*. Specifically, it considers noise distribution of finite variance σ^2 and studies the normalized squared error (NSE) : $\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2/\sigma^2$. At high-SNR (i.e. small σ), the NSE obtains its worst-case value and the main result of the chapter explicitly characterizes that (this is known as *noise-sensitivity* study). The derived formulae are in closed-form and admit insightful interpretations: (i) They reveal clear connections to the results of Section 2.2 on noiseless compressed sensing, thus answering Question (Q.c). (ii) They are interpreted as natural extensions of classically known error expressions for ordinary least-squares. Moreover, their simple nature is used to derive recipes for the optimal tuning of the regularizer parameter. An important differentiating feature of many results in this chapter is that they are *non-asymptotic*.

Chapter 8

The content of Chapter 8 is motivated by Question (Q.d): "To what extent are the error expressions derived in previous chapters for Gaussian matrices *universal* over other random ensembles"? For matrices that are Isotropically Random Orthogonal (IRO), we precisely characterize the squared error performance of regularized least-squares and prove that it is *superior* to the error performance of Gaussians. In particular, this is in contrast to the corresponding result in the noiseless case, where we saw in Section 2.2 that the phase-transitions of the two ensembles match. Interestingly, we empirically observe the following universality property of IRO matrices: the derived error formulae for IRO matrices hold true for random DCT and Hadamard matrices.

The main idea of the proof is also given in this section, while some technical details are deferred to Appendix E.

Chapter 9

All the results from the previous chapters consider the squared-error reconstruction performance of (1.2). Chapter 8 takes a step forward by answering Question (Qe). In particular, it extends the applicability of the CGMT framework and the precise results that it yields to more general Lipschitz performance metrics. For concreteness, the focus is primarily on regularized least-squares. For an illustration, we characterize the probability of correct support recovery of the LASSO.

Chapter 10

This chapter presents an important application of the generic results of the previous chapters. The standard relaxation of the ML decoder for Binary Phase-Shift Keying signal transmission in a Massive Multiple Input Multiple Output setting, often called the Box relaxation optimization (BRO), is an instance of (1.2). The BRO is very popular in practice, but its bit error rate (BER) performance has hitherto remained unknown. Using results from Chapter 9, we precisely characterize the BER of the BRO. This let us compare performance to the unattainable matched-filter bound: we show a 3dB divergence in the square case of an equal number of transmitting to receiving antennas. We then discuss extensions to other signal constellations and potential (provable) improvements of the BRO when combined with local methods.

The proofs are deferred to Appendix F.

Chapter 11

Chapter 11 answers Question (Q.f). In particular, it studies the squared-error performance of the Generalized LASSO under a non-linear measurement model of the form $\mathbf{y} = g(\mathbf{A}\mathbf{x}_0)$ for some (potentially) non-linear, random and/or unknown link function g (e.g. quantized measurements). The main result of the chapter establishes an interesting equivalence of the LASSO performance under non-linearities to the already known results on the LASSO performance under linear measurements. This result has several implications worth exploring. For instance, it encompasses state-of-the art theoretical results of *one-bit Compressed Sensing* and generalizations to higher levels of quantization. Also, it is used at the end of the chapter to design optimal quantizers. Interestingly, we prove that the optimal quantizer of the measurements that minimizes the estimation error of the Generalized LASSO is the celebrated Lloyd-Max quantizer.

As usual, all proofs are deferred to Appendix G.

Chapter 12

The final chapter concludes with some brief remarks on various directions for future research that are suggested by the analysis methods and results presented in this thesis.

Chapter 3

THE CONVEX GAUSSIAN MIN-MAX THEOREM

This chapter establishes the Convex Gaussian Min-max Theorem (CGMT), which is a key result of this thesis. To arrive at the CGMT, we first present the classical Slepian's and Gordon's comparison Theorems in Section 3.1. A popular corollary of Gordon's result, called the Gaussian Min-max Theorem (GMT), is derived next in Section 3.2. We also demonstrate how the GMT leads to the "escape through a mesh" Proposition 2.2.2, which was shown earlier in Section 2.2 to play a central role in the study of phase-transitions in noiseless Compressed Sensing. The CGMT is stated in Section 3.3 and is interpreted as an extended and tight version of the GMT. Its proof is included in the last Section 3.4.

3.1 Gaussian Comparison Inequalities

Gaussian comparison theorems are powerful tools in probability theory. They establish probabilistic inequalities between functions of Gaussian processes (e.g. their maximum values) based on known relations on their first and second order moments, and they have various applications (see for example [LT91, Ch. 3.3] for an introduction).

Perhaps the most celebrated of those results is Slepian's Lemma, which dates back to 1962 [Sle62]. We state the lemma below and discuss a popular application of it.

Slepian's Lemma

Lemma 3.1.1 (Slepian's Lemma)**.** *Let* $\{X_i\}_{i=1}^N$, $\{Y_i\}_{i=1}^N$ *be two Gaussian processes with the same mean* μ_i *and the same variance* σ_i^2 *such that* $\forall i, i'$*:*

$$\mathbb{E}X_i X_{i'} \geq \mathbb{E}Y_i Y_{i'}.$$

Then, for any $c \in \mathbb{R}$*,*

$$\mathbb{P}(\max_i X_i \geq c) \leq \mathbb{P}(\max_i Y_i \geq c).$$

In words, Slepian's lemma says that for a Gaussian process Y_i that is more uncorrelated than another Gaussian process X_i, it holds:

if c is an *upper* bound on the $\max_i Y_i$, then so it is for $\max_i X_i$.

The intuition behind this seemingly simple but deep result is clear. The process Y_i is more uncorrected, hence it is more probable that it takes larger values than the process X_i. Also, for the Gaussian ensemble, the first two moments alone capture its characteristics. We refer the reader to [LT91, Ch. 3.1] for a proof.

Slepian's lemma has proved to be useful in several contexts. The textbook application of it is that of computing a high-probability upper bound on the maximum singular value of an iid Gaussian matrix. We demonstrate this, aiming to familiarize the reader with the uses of Gaussian process inequalities and to introduce some ideas that will be key to our subsequent discussion.

Maximum singular value of Guassian matrices

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ have entries iid Gaussian and consider its maximum singular value:

$$\sigma_{\max}(\mathbf{A}) = \|\mathbf{A}\|_2 = \max_{\substack{\|\mathbf{u}\|_2=1\\ \|\mathbf{w}\|_2=1}} \mathbf{u}^T\mathbf{A}\mathbf{w}. \tag{3.1}$$

Towards using Slepian's lemma to compute a high-probability *upper* bound on $\sigma_{\max}(\mathbf{A})$, let $\gamma \in \mathbb{R}$, $\mathbf{g} \in \mathbb{R}^m$ and $\mathbf{h} \in \mathbb{R}^n$ have iid $\mathcal{N}(0,1)$ entries and define the following two Gaussian processes each indexed by $\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right]$:

$$X_{\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right]} = \mathbf{u}^T\mathbf{A}\mathbf{w} + \gamma\|\mathbf{u}\|_2\|\mathbf{w}\|_2, \tag{3.2a}$$

$$Y_{\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right]} = \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w}. \tag{3.2b}$$

Clearly, the two processes have mean 0. Also, a simple calculation yields

$$\begin{aligned}\mathbb{E}[X_{\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right]}X_{\left[\begin{smallmatrix}\mathbf{u}'\\ \mathbf{w}'\end{smallmatrix}\right]}] - \mathbb{E}[Y_{\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right]}Y_{\left[\begin{smallmatrix}\mathbf{u}'\\ \mathbf{w}'\end{smallmatrix}\right]}] &= = (\mathbf{u}^T\mathbf{u}')(\mathbf{w}^T\mathbf{w}') + \|\mathbf{u}\|_2\|\|\mathbf{w}\|_2\|\mathbf{u}'\|_2\|\mathbf{w}'\|_2\\ &\quad - \|\mathbf{w}\|_2\|\mathbf{w}'\|_2(\mathbf{u}^T\mathbf{u}') - \|\mathbf{u}\|_2\|\mathbf{u}'\|_2(\mathbf{w}^T\mathbf{w}')\\ &= (\|\mathbf{u}\|_2\|\mathbf{u}'\|_2 - \mathbf{u}^T\mathbf{u}')(\|\mathbf{w}\|_2\|\mathbf{w}'\|_2 - \mathbf{w}^T\mathbf{w}') \geq 0,\end{aligned}$$

with equality if $\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right] = \left[\begin{smallmatrix}\mathbf{u}'\\ \mathbf{w}'\end{smallmatrix}\right]$ (thus, both processes have the same variance). Consequently, the processes defined in (3.2) satisfy the conditions of Slepian's lemma, from which it follows that for all $c \in \mathbb{R}$ [1],

$$\mathbb{P}(\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} X_{\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right]} \geq c) \leq \mathbb{P}(\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} Y_{\left[\begin{smallmatrix}\mathbf{u}\\ \mathbf{w}\end{smallmatrix}\right]} \geq c). \tag{3.3}$$

[1] Formally, note that Slepian's lemma is stated for processes indexed on *discrete sets*. A simple compactness argument leads to (3.3) (see for example [RXH11, Prop. 1].

Now, observe that

$$\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} X_{\begin{bmatrix}\mathbf{u}\\ \mathbf{w}\end{bmatrix}} = \max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} \mathbf{u}^T\mathbf{A}\mathbf{w} + \gamma.$$

This is already very similar to the quantity of interest $\sigma_{\max}(\mathbf{A})$ in (3.1). In fact, with a simple symmetrization trick, we can get rid of the "disturbing term" γ. The simple idea is to condition on the sign of γ, which is positive or negative with equal probability $1/2$. With this and using the observation

$$\mathbb{P}(\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} X_{\begin{bmatrix}\mathbf{u}\\ \mathbf{w}\end{bmatrix}} \geq c \mid \gamma \geq 0) \geq \mathbb{P}(\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} \mathbf{u}^T\mathbf{A}\mathbf{w} \geq c),$$

we find that

$$\mathbb{P}(\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} X_{\begin{bmatrix}\mathbf{u}\\ \mathbf{w}\end{bmatrix}} \geq c) \geq \frac{1}{2}\mathbb{P}(\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} \mathbf{u}^T\mathbf{A}\mathbf{w} \geq c). \tag{3.4}$$

Next, we evaluate the RHS in (3.3): performing the maximization over $\mathbf{u}$ and $\mathbf{w}$ for Y is straightforward and gives

$$\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} Y_{\begin{bmatrix}\mathbf{u}\\ \mathbf{w}\end{bmatrix}} = \|\mathbf{g}\|_2 + \|\mathbf{h}\|_2.$$

This, when combined with (3.4) and (3.3) yields

$$\mathbb{P}(\max_{\|\mathbf{u}\|_2=1,\|\mathbf{w}\|_2=1} \mathbf{u}^T\mathbf{A}\mathbf{w} \geq c) \leq 2 \cdot \mathbb{P}(\|\mathbf{g}\|_2 + \|\mathbf{h}\|_2 \geq c). \tag{3.5}$$

By Gaussian concentration of Lipschitz functions, $\|\mathbf{g}\|_2 + \|\mathbf{h}\|_2$ concentrates around $\sqrt{m} + \sqrt{n}$ which in view of (3.5) implies that the probability that $\sigma_{\max}(\mathbf{A})$ (significantly) exceeds $\sqrt{m} + \sqrt{n}$ is very small. Formally, set $c = \sqrt{m} + \sqrt{n} + t$ in (3.5) and use Proposition 3.1.1 below, from which each one of the events $\{\|\mathbf{g}\|_2 \geq \sqrt{m} + t/2\}$ and $\{\|\mathbf{h}\|_2 \geq \sqrt{n} + t/2\}$ occurs with probability at most $\exp(-t^2/8)$, to conclude with the following high-probability upper bound on $\sigma_{\max}(\mathbf{A})$:

$$\mathbb{P}(\sigma_{\max}(\mathbf{A}) \geq \sqrt{m} + \sqrt{n} + t) \leq 4e^{-t^2/8}.$$

Proposition 3.1.1 (Gaussian Lipschitz concentration)**.** *(e.g.,[BLM13, Theorem 5.6]) Let $\mathbf{w} \in \mathbb{R}^n$ have entries i.i.d. $\mathcal{N}(0,1)$ and $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz[2] . Then, $Var[f(\mathbf{w})] \leq L^2$. Furthermore, for all $t > 0$, each one of the events $\{f(\mathbf{w}) > \mathbb{E}f(\mathbf{w}) + t\}$ and $\{f(\mathbf{w}) < \mathbb{E}f(\mathbf{w}) - t\}$ occurs with probability no greater than $\exp\left(-t^2/(2L^2)\right)$.*

[2] We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz with constant L or is L-Lipschitz if $|f(\mathbf{w}) - f(\mathbf{u})| \leq L\|\mathbf{w} - \mathbf{u}\|$ for all $\mathbf{w}, \mathbf{u} \in \mathbb{R}^n$.

Gordon's comparison Theorem

In this section, we present Gordon's comparison theorem, which is a non-trivial extension of Slepian's lemma that was proved by Gordon in 1986 [Gor85]. The theorem establishes a probabilistic comparison between the min-max of two *doubly-indexed* Gaussian processes $\{X_{ij}\}$ and $\{Y_{ij}\}$ based on conditions on their corresponding covariance structures. See for example [LT91, Ch. 3.1] for a proof.

Theorem 3.1.1 (Gordon's comparison theorem)**.** *Let* $\left\{X_{ij}\right\}$ *and* $\left\{Y_{ij}\right\}$, $1 \le i \le I$, $1 \le j \le J$, *be centered Gaussian processes such that*

$$\begin{cases} \mathbb{E}X_{ij}^2 = \mathbb{E}Y_{ij}^2, & \text{for all } i, j, \\ \mathbb{E}X_{ij}X_{ik} \ge \mathbb{E}Y_{ij}Y_{ik}, & \text{for all } i, j, k, \\ \mathbb{E}X_{ij}X_{\ell k} \le \mathbb{E}Y_{ij}Y_{\ell k}, & \text{for all } i \ne \ell \text{ and } j, k. \end{cases}$$

Then, for all $c \in \mathbb{R}$,

$$\mathbb{P}(\min_i \max_j X_{ij} \le c) \le \mathbb{P}(\min_i \max_j Y_{ij} \le c).$$

Theorem 3.1.1 is powerful since it applies to any pair of processes that satisfy the imposed conditions. In the next section, we present a corollary of it, which follows by application of the theorem to the two specific Gaussian processes that were earlier introduced in (3.2) (only this time viewed as doubly-indexed on $\mathbf{u}, \mathbf{w}$).

3.2 Gaussian Min-max Theorem

Theorem 3.2.1 ((GMT))**.** *Let* $\mathbf{A} \in \mathbf{R}^{m\times n}$, $\gamma \in \mathbb{R}$, $\mathbf{g} \in \mathbb{R}^m$ *and* $\mathbf{h} \in \mathbb{R}^n$ *have entries iid standard normal. Let* $S_{\mathbf{w}}, S_{\mathbf{u}}$ *compact sets, and* $\psi(\mathbf{w}, \mathbf{u})$ *a continuous function. Define,*

$$\overline{\Phi}(\mathbf{A}, \gamma) = \min_{\mathbf{w}\in S_{\mathbf{w}}} \max_{\mathbf{u}\in S_{\mathbf{u}}} \mathbf{u}^T\mathbf{A}\mathbf{w} + \gamma\|\mathbf{u}\|_2\|\mathbf{w}\|_2 + \psi(\mathbf{w}, \mathbf{u}), \tag{3.6a}$$

$$\phi(\mathbf{g}, \mathbf{h}) = \min_{\mathbf{w}\in S_{\mathbf{w}}} \max_{\mathbf{u}\in S_{\mathbf{u}}} \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w}, \mathbf{u}). \tag{3.6b}$$

Then, for all $c \in \mathbb{R}$,

$$\mathbb{P}(\overline{\Phi}(\mathbf{A}, \gamma) \le c) \le \mathbb{P}(\phi(\mathbf{g}, \mathbf{h}) \le c).$$

The result was essentially proved by Gordon in 1988 [Gor88] as a corollary of 3.1.1. The version presented here requires an additional compactness argument when compared to the original result [Gor88, Lem. 3.1]; see Appendix A for details and a proof.

Theorem 3.3.1 asserts that the lower tail probability of $\overline{\Phi}(\mathbf{A}, \gamma)$ is upper bounded by that of $\phi(\mathbf{g}, \mathbf{h})$, or equivalently,

if c is a high probability *lower* bound on $\phi(\mathbf{g}, \mathbf{h})$, so it is for $\overline{\Phi}(\mathbf{A}, \gamma)$.

Escape through a mesh

Next, we apply the GMT Theorem 3.2.1 to prove the "escape through a mesh" Proposition 2.2.2. The proof is instructive.

Consider the setup of Proposition 2.2.2, i.e. $\mathbf{A} \in \mathbb{R}^{m \times n}$ has entries iid Gaussian and $\mathcal{K} \subset \mathbb{R}^n$ is a cone. The mCSVof $\mathbf{A}$ can be written as:

$$\sigma_{\min}(\mathbf{A}; \mathcal{K}) = \min_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} \|\mathbf{A}\mathbf{w}\|_2 = \min_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} \max_{\|\mathbf{u}\|_2 = 1} \mathbf{u}^T \mathbf{A} \mathbf{w}. \tag{3.7}$$

We apply the GMT for $\mathcal{S}_{\mathbf{w}} = \mathcal{K} \cap \mathcal{S}^{n-1}$, $\mathcal{S}_{\mathbf{u}} = \mathcal{S}^{m-1}$ and $\psi(\mathbf{w}, \mathbf{u}) = 0$. The min-max optimization in (3.6b) is easy to evaluate:

$$\begin{aligned} \min_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} \max_{\|\mathbf{u}\|_2 = 1} \|\mathbf{w}\|_2 \mathbf{g}^T \mathbf{u} + \|\mathbf{u}\|_2 \mathbf{h}^T \mathbf{w} &= \min_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} \|\mathbf{w}\|_2 \|\mathbf{g}\|_2 + \mathbf{h}^T \mathbf{w} \\ &= \|\mathbf{g}\|_2 - \max_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} (-\mathbf{h})^T \mathbf{w}. \end{aligned} \tag{3.8}$$

Moreover, the min-max optimization in (3.6a) is almost in the desired form that appears in (3.7), except from the disturbing term $\gamma \|\mathbf{w}\|_2 \|\mathbf{u}\|_2$. We can get rid of this term by a simple symmetrization trick that is very similar to the one that led to (3.4) earlier: the idea is again to condition on the sign of γ (see the proof of Theorem 3.3.1(i) for details). Omitting the details here, it can be shown that

$$\mathbb{P}\Big(\min_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} \max_{\|\mathbf{u}\|_2 = 1} \mathbf{u}^T \mathbf{A} \mathbf{w} \le c \Big) \le 2 \cdot \mathbb{P}\Big(\min_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} \max_{\|\mathbf{u}\|_2 = 1} \mathbf{u}^T \mathbf{A} \mathbf{w} + \gamma \|\mathbf{u}\|_2 \|\mathbf{w}\|_2 \le c \Big).$$

Combining this with (3.8) and applying the GMT, yields

$$\mathbb{P}(\sigma_{\min}(\mathbf{A}; \mathcal{K}) \le c) \le 2 \cdot \mathbb{P}(\|\mathbf{g}\|_2 - \max_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} (-\mathbf{h})^T \mathbf{w} \le c). \tag{3.9}$$

Thus, *the GMT translates the problem of lower-bounding the mCSVof* $\mathbf{A}$ *to the much simpler task of lower bounding the auxiliary min-max optimization in* (3.8). It can be easily shown that $\|\mathbf{g}\|_2$ and $\max_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} (-\mathbf{h})^T \mathbf{w}$ are both 1-Lipschitz functions of $\mathbf{g}$ and $\mathbf{h}$ respectively. Therefore, by Proposition 3.1.1, each one of the following events occurs with probability at most $\exp(-t^2/8)$ [3]:

$$\{\|\mathbf{g}\|_2 \le \sqrt{m-1} - t/2\} \quad \text{and} \quad \{ \max_{\mathbf{w} \in \mathcal{K} \cap \mathcal{S}^{n-1}} (-\mathbf{h})^T \mathbf{w} \le \omega(\mathcal{K}) - t/2 \}.$$

[3] For a detailed derivation, see for example [OTH13b, App. B].

Here, we have recognized that $\mathbb{E}[\max_{\mathbf{w}\in\mathcal{K}\cap\mathcal{S}^{n-1}}(-\mathbf{h})^T\mathbf{w}] = \omega(\mathcal{K})$ as in Definition 2.2.3. Putting these together and setting $c = \sqrt{m-1} - \omega(\mathcal{K}) - t$ proves that

$$\mathbb{P}\left(\sigma_{\min}(\mathbf{A};\mathcal{K}) \leq \sqrt{m-1} - \omega(\mathcal{K}) - t\right) \leq 4e^{-t^2/8}. \tag{3.10}$$

Remark 3.2.0.1. The escape through a mesh result was first proved by Gordon in [Gor88]. The version that appears in Proposition 2.2.2 is due to Chandrasekaran et. al. [Cha+12]. The proof presented above is slightly modified. In particular, the conditioning trick that gets rid of the disturbing "γ-term" appears to be new in this setting and will soon prove to be critical in establishing a stronger version of the GMT in Section 3.3. However, compared to the probability bound of Proposition 2.2.2 the constants in (3.10) are slightly looser.

A tight version?

A natural question that arises concerns the *tightness* of the bounds obtained via the GMT. To become explicit, suppose that $\phi(\mathbf{g},\mathbf{h})$ concentrates around some constant μ, in the sense that for all $t > 0$, the events

$$\{\phi(\mathbf{g},\mathbf{h}) \leq \mu - t\} \quad \text{and} \quad \{\phi(\mathbf{g},\mathbf{h}) \geq \mu + t\},$$

each occur with *low* probability. Of course, $\mu - t$ is then a high-probability lower bound to $\phi(\mathbf{g},\mathbf{h})$, but also this bound is *tight* since it is accompanied by a corresponding high-probability upper bound, namely $\mu + t$, whose value can be made arbitrarily close to the former. The GMT implies that $\mu - t$ is also a high-probability lower bound on $\Phi(\mathbf{A})$. But, it gives *no* information on how much $\Phi(\mathbf{A})$ is allowed to deviate from this.

In the coming section, we show that under additional *convexity* assumptions, the GMT is *tight* in the sense discussed above. We call this the *Convex Gaussian Min-max Theorem* (CGMT) and it constitutes one of the main theoretical contributions of this thesis. There are two critical observations that lead to this conclusion. First, if we can get rid of the term $\gamma\|\mathbf{u}\|_2\|\mathbf{w}\|_2$ in (3.6a), then the remaining objective function consists of a bilinear term in $\mathbf{w}$ and $\mathbf{u}$ and the function $\psi(\mathbf{w},\mathbf{u})$. Therefore, it is convex-concave[4] in its two arguments as long as ψ is convex-concave. This leads to the second observation: from Sion's min-max principle, we can flip the order of a min-max optimization in which the objective is convex-concave and the

[4] A function $f(\mathbf{x},\mathbf{y})$ is convex-concave if it is convex in its first argument $\mathbf{x}$ and concave in the second $\mathbf{y}$.

constraint sets are convex. As the proof shows, applying the GMT to the (now flipped) max-min problem, translates to the desired upper bound on the original optimization.

In fact, the CGMT moves even further. While Gordon's result only relates the optimal costs of the two involved min-max optimizations, the CGMT establishes a tight relation between the optimal solutions. This result is crucial to the analysis in the rest of the chapters of this thesis.

3.3 Convex Gaussian Min-max Theorem (CGMT)

The CGMT is a tight version of Gordon's Theorem 3.2.1 in the presence of additional convexity assumptions. The setup of the theorem is similar to that of the GMT.

In particular, let $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{g} \in \mathbb{R}^m$, $\mathbf{h} \in \mathbb{R}^n$, $\mathcal{S}_{\mathbf{w}} \subset \mathbb{R}^n$, $\mathcal{S}_{\mathbf{u}} \subset \mathbb{R}^m$ and $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. With these, consider the following two min-max optimization problems and their corresponding optimal costs.

$$\Phi(\mathbf{A}) := \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \mathbf{u}^T\mathbf{A}\mathbf{w} + \psi(\mathbf{w},\mathbf{u}), \tag{3.11a}$$

$$\phi(\mathbf{g},\mathbf{h}) := \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w}\mathbf{u}). \tag{3.11b}$$

Further denote $\mathbf{w}_\Phi := \mathbf{w}_\Phi(\mathbf{A})$ and $\mathbf{w}_\phi := \mathbf{w}_\phi(\mathbf{g},\mathbf{h})$ any optimal minimizers in (3.11a) and (3.11b), respectively.

Observe that the optimization in (3.11b) is the same as the one in (3.6b). On the other hand, the optimization in (3.11a) is missing the term "$\gamma\|\mathbf{w}\|_2\|\mathbf{u}\|_2$" when compared to (3.6a).

Henceforth, we refer to the two optimization problems in (3.11a) and (3.11b) as the *Primary Optimization* (PO) and *Auxiliary Optimization* (AO), respectively.

We are now ready to state the Convex Gaussian Min-max Theorem.

Theorem 3.3.1 (CGMT)**.** *In* (3.11)*, let* $\mathcal{S}_{\mathbf{w}}, \mathcal{S}_{\mathbf{u}}$ *be compact sets,* $\psi(\cdot,\cdot)$ *be continuous on* $\mathcal{S}_{\mathbf{w}} \times \mathcal{S}_{\mathbf{u}}$*, and* $\mathbf{A}, \mathbf{g}$ *and* $\mathbf{h}$ *all have entries iid standard normal. The following statements are true:*

(i) For all $c \in \mathbb{R}$*,*

$$\mathbb{P}(\ \Phi(\mathbf{A}) < c\ ) \leq 2\mathbb{P}(\ \phi(\mathbf{g},\mathbf{h}) \leq c\ ).$$

(ii) *Further assume that* $\mathcal{S}_{\mathbf{w}}, \mathcal{S}_{\mathbf{u}}$ *are convex sets and* ψ *is convex-concave on* $\mathcal{S}_{\mathbf{w}} \times \mathcal{S}_{\mathbf{u}}$. *Then, for all* $c \in \mathbb{R}$,

$$\mathbb{P}(\ \Phi(\mathbf{A}) > c\ ) \leq 2\mathbb{P}(\ \phi(\mathbf{g}, \mathbf{h}) \geq c\ ). \tag{3.12}$$

In particular, for all $\mu \in \mathbb{R}, t > 0$,

$$\mathbb{P}(\ |\Phi(\mathbf{A}) - \mu| > t\ ) \leq 2\mathbb{P}(\ |\phi(\mathbf{g}, \mathbf{h}) - \mu| \geq t\ ). \tag{3.13}$$

(iii) *Let* $\mathcal{S}$ *be an arbitrary open subset of* $\mathcal{S}_{\mathbf{w}}$ *and* $\mathcal{S}^c = \mathcal{S}_{\mathbf{w}}/\mathcal{S}$. *Denote* $\Phi_{\mathcal{S}^c}(\mathbf{A})$ *and* $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h})$ *the optimal costs of the optimizations in* (3.11a) *and* (3.11b), *respectively, when the minimization over* $\mathbf{w}$ *is now constrained over* $\mathbf{w} \in \mathcal{S}^c$. *If there exist constants* $\overline{\phi}$, $\overline{\phi}_{\mathcal{S}^c}$ *and* $\eta > 0$ *such that*

(a) $\overline{\phi}_{\mathcal{S}^c} \geq \overline{\phi} + 3\eta$,

(b) $\phi(\mathbf{g}, \mathbf{h}) < \overline{\phi} + \eta$ *with probability at least* $1 - p$,

(c) $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h}) > \overline{\phi}_{\mathcal{S}^c} - \eta$ *with probability at least* $1 - p$,

then,

$$\mathbb{P}(\mathbf{w}_{\Phi}(\mathbf{A}) \in \mathcal{S}) \geq 1 - 4p.$$

The probabilities are taken with respect to the randomness of $\mathbf{A}$, $\mathbf{g}$ and $\mathbf{h}$. A few remarks are in place.

Remarks

Remark 3.3.0.2 (Statement (i)). The inequality in the first statement of the CGMT is essentially no different than what Theorem 3.2.1 states:

if c is a high probability *lower* bound for the optimal cost $\phi(\mathbf{g}, \mathbf{h})$ of the (AO), so it is for the optimal cost $\Phi(\mathbf{A})$ of the (PO).

As already mentioned, in contrast to the GMT, the minimax optimization in (3.11a) does not include the term "$\gamma\|\mathbf{w}\|_2\|\mathbf{u}\|_2$". The "price" paid for this, is the multiplicative factor of two. Note however that this factor does not affect the essence of the result since the scenarios of interest are those for which $\mathbb{P}(\phi(\mathbf{g}, \mathbf{h}) \leq c)$ is close to zero. What is more, in most of the applications where the GMT is useful, the optimization problem involved is in the form of (3.11a) rather than that of (3.6a).

(For instance, recall the proof of the escape through a mesh theorem.) One reason behind this, is that under convexity assumptions on $\mathcal{S}_{\mathbf{w}}$, $\mathcal{S}_{\mathbf{u}}$ and ψ the (PO) is a *convex* program, which is generally more likely to be encountered in applications compared to the always non-convex program in (3.6a)[5]. Convexity, is also critical for establishing the second statement of the theorem.

Remark 3.3.0.3 (Statement (ii)). This is the main contribution of the CGMT and it holds only after imposing appropriate convexity assumptions providing a counterpart to statement (i):

if c is a high probability *upper* bound for the optimal cost $\phi(\mathbf{g}, \mathbf{h})$ of the (AO), so it is for the optimal cost $\Phi(\mathbf{A})$ of the (PO).

Combining the two statements, yields the concentration result in (3.13):

if $\phi(\mathbf{g}, \mathbf{h})$ concentrates around μ, so does $\Phi(\mathbf{A})$.

Remark 3.3.0.4 (Lipschitzness and normal concentration). Inequality (3.13) becomes interesting when μ is chosen so that $\phi(\mathbf{g}, \mathbf{h})$ concentrates around it. In this case, the probability in the right-hand side of (3.12) is vanishing, indicating that $\Phi(\mathbf{A})$ concentrates around the same value. In particular, we can apply (3.13) for $\mu = \mathbb{E}\phi(\mathbf{g}, \mathbf{h})$. It is shown in Lemma A.2.0.2 in Appendix A that $\phi(\mathbf{g}, \mathbf{h})$ is Lipschitz in $(\mathbf{g}, \mathbf{h})$. It then follows from the Gaussian concentration property of Lipschitz functions (see Proposition 3.1.1) that $\phi(\mathbf{g}, \mathbf{h})$ is normally concentrated around its mean $\mathbb{E}\phi(\mathbf{g}, \mathbf{h})$. Thus, we obtain Corollary 3.3.1 below.

Corollary 3.3.1. *Consider the same setup as in Theorem 3.3.1 and let the assumptions of statement (ii) therein hold. Further, define* $R_{\mathbf{w}} := \max_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \|\mathbf{w}\|_2$ *and* $R_{\mathbf{u}} := \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \|\mathbf{u}\|_2$. *Then, for all* $t > 0$,

$$\mathbb{P}\left(|\Phi(\mathbf{A}) - \mathbb{E}\phi(\mathbf{g}, \mathbf{h})| > t \right) \leq 4 \exp\left(-t^2/(4R_{\mathbf{w}}^2 R_{\mathbf{u}}^2)\right).$$

Remark 3.3.0.5 (On the convexity assumptions). The proof of the second statement shows that the critical step is being able to flip the order of the min-max operation in the (PO) problem without changing its optimal cost. The convexity conditions as specified in the second statement of the theorem guarantee that this is possible. Note however that these conditions are only sufficient. In principle, it might be possible

[5] the component $\gamma\|\mathbf{w}\|_2\|\mathbf{u}\|_2$ causes the min-max optimization in (3.6a) to be non-convex even when $\mathcal{S}_{\mathbf{w}}$,$\mathcal{S}_{\mathbf{u}}$ are convex and ψ convex-concave.

to flip the order of min-max under milder conditions in which case statement (ii) would continue to hold. For instance, flipping the order of min-max remains valid even under the weaker assumption of a quasi-convex-concave function, [Sio+58, Thm. 3.4].

Remark 3.3.0.6 (Statement (iii)). In the presence of convexity, the optimal cost of the (PO) concentrates to the same value to which the (AO) does. Statement (iii) uses this fact to conclude with an even stronger statement, only this time regarding the *minimizers* of two optimizations. Simply combining conditions (a)–(c) of the statement, implies that $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h}) > \phi(\mathbf{g}, \mathbf{h})$ whp. Of course, the left-hand side is always no smaller than the right-hand side. If a strict inequality holds (as implied by the conditions), then it is easy to conclude that $\mathbf{w}_\phi \in \mathcal{S}$ holds whp. The power of the theorem is that the same conclusion holds not only for the minimizer of the (AO), but also, for the minimizer $\mathbf{w}_\Phi$ of the (PO).

Remark 3.3.0.7 (An asymptotic version of Statement (iii)). All three statements of the CGMT hold non-asymptotically in the problem dimensions m and n. The following corollary is a version of the theorem that holds when the problem dimensions grow to infinity.

Corollary 3.3.2 (Asymptotic CGMT)**.** *Using the same notation as in Theorem 3.3.1 and under the convexity conditions of statement (ii), suppose there exists constants* $\overline{\phi} < \overline{\phi}_{\mathcal{S}^c}$ *such that* $\phi(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \overline{\phi}$ *and* $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \overline{\phi}_{\mathcal{S}^c}$ [6] *. Then,*

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{w}_\Phi(\mathbf{A}) \in \mathcal{S}) = 1.$$

Remark 3.3.0.8 (Applications). The CGMT plays a central role in this thesis and is one of its main theoretical contributions. In the subsequent chapters, it is used as a key ingredient to analyze the estimation performance of non-smooth convex optimization methods. The analysis is tight owing to the tight nature of the theorem itself. At this point, it is worth mentioning that the CGMT, while applicable to the problem of precise performance analysis, cannot be used to show tightness of the lower bound of Proposition 2.2.2 on the mCSV[7]. The theorem does not apply since the constraint sets $\mathcal{S}_\mathbf{w}$, $\mathcal{S}_\mathbf{u}$ involved in (3.7) are not convex. The CGMT might be of

[6] For a sequence of random variables $\{\mathcal{X}^{(n)}\}_{n\in\mathbb{N}}$ and a constant $c \in \mathbb{R}$ (independent of n), we write $\{\mathcal{X}^{(n)}\}_{n\in\mathbb{N}} \xrightarrow{P} c$, to denote convergence *in probability*, i.e. $\forall \epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left(|\mathcal{X}^{(n)} - c| > \epsilon\right) = 0$.

[7] We remark, however, that the corresponding lower bound $\sqrt{m} - \sqrt{n}$ for $\sigma_{\min}(\mathbf{A})$ is indeed *asymptotically* tight in the regime $n/m \to (0, 1), n \to \infty$ by the Bai-Yin's law [BY93].

individual interest and may have applications that go beyond this application. With this in mind, we have chosen to present it above in its most general version.

3.4 Proof of the CGMT

Proof of statement (i)

As discussed, this is an almost direct consequence of the GMT (Theorem 3.2.1). Yet we need to get rid of the term "$\gamma\|\mathbf{w}\|_2\|\mathbf{u}\|_2$" in (3.6a) in the GMT. The argument is rather simple but critical for the rest of the statements of the theorem. We will show that

$$\mathbb{P}\left(\Phi(\mathbf{A}) \leq c\right) \leq 2\mathbb{P}\left(\overline{\Phi}(\mathbf{A},\gamma) \geq c\right). \tag{3.14}$$

Once this is established, the claim follows directly by the GMT. To prove (3.14), fix $\mathbf{A}$ and $g < 0$ and denote

$$f_1(\mathbf{w},\mathbf{u}) = \mathbf{u}^T\mathbf{A}\mathbf{w} + \psi(\mathbf{w},\mathbf{u}) \quad \text{and} \quad f_2(\mathbf{w},\mathbf{u}) = \mathbf{u}^T\mathbf{A}\mathbf{w} + \gamma\|\mathbf{w}\|_2\|\mathbf{u}\|_2 + \psi(\mathbf{w},\mathbf{u}).$$

Clearly, $f_1(\mathbf{w},\mathbf{u}) \geq f_2(\mathbf{w},\mathbf{u})$ for all $(\mathbf{w},\mathbf{u}) \in \mathcal{S}_\mathbf{w} \times \mathcal{S}_\mathbf{u}$. We may then write,

$$\begin{aligned}\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} f_1(\mathbf{w},\mathbf{u}) = f_1(\mathbf{w}_1,\mathbf{u}_1) &\geq f_1(\mathbf{w}_1,\mathbf{u}) \text{ for all } \mathbf{u}\in\mathcal{S}_\mathbf{u} \\ &\geq \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} f_2(\mathbf{w}_1,\mathbf{u}) \geq \min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} f_2(\mathbf{w},\mathbf{u}).\end{aligned}$$

This proves $\Phi(\mathbf{A}) \geq \overline{\Phi}(\mathbf{A},\gamma)$, when $g < 0$. From this and from the independence of γ and $\mathbf{A}$, for all $c \in \mathbb{R}$:

$$\mathbb{P}\left(\overline{\Phi}(\mathbf{A},\gamma) \leq c \mid \gamma < 0\right) \geq \mathbb{P}\left(\Phi(\mathbf{A}) \leq c \mid \gamma < 0\right) = \mathbb{P}(\Phi(\mathbf{A}) \leq c).$$

When combined with $\gamma \sim \mathcal{N}(0,1)$, the above yields the desired inequality (3.14):

$$\mathbb{P}\left(\overline{\Phi}(\mathbf{A},\gamma) \leq c\right) = \frac{1}{2}\mathbb{P}\left(\overline{\Phi}(\mathbf{A},\gamma) \leq c \mid \gamma > 0\right) + \frac{1}{2}\mathbb{P}\left(\overline{\Phi}(\mathbf{A},\gamma) \leq c \mid \gamma < 0\right) \geq \frac{1}{2}\mathbb{P}(\Phi(\mathbf{A}) \leq c).$$

Proof of statement (ii)

The additional convexity assumptions imposed in statement (ii) of the theorem are critical here. By assumption, the sets $\mathcal{S}_\mathbf{w}, \mathcal{S}_\mathbf{u}$ are non-empty, compact and convex. Furthermore, the function $\mathbf{u}^T\mathbf{A}\mathbf{w}+\psi(\mathbf{w},\mathbf{u})$ is continuous, finite[8] and convex-concave

[8] A continuous function on a compact set is bounded from the Weierstrass extreme value theorem.

on $\mathcal{S}_{\mathbf{w}} \times \mathcal{S}_{\mathbf{u}}$. Thus, we can apply the minimax result in [Roc97, Corollary 37.3.2] to exchange "min-max" with a "max-min" in (3.11a)[9]:

$$\Phi(\mathbf{A}) = \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \mathbf{u}^T\mathbf{A}\mathbf{w} + \psi(\mathbf{w},\mathbf{u}).$$

It is convenient to rewrite the above as

$$-\Phi(\mathbf{A}) = \min_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \max_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} -\mathbf{u}^T\mathbf{A}\mathbf{w} - \psi(\mathbf{w},\mathbf{u}).$$

Then, using the symmetry of $\mathbf{A}$, we have that for any $c \in \mathbb{R}$:

$$\mathbb{P}\left(-\Phi(\mathbf{A}) \le c\right) = \mathbb{P}\left(\min_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \max_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \left\{\mathbf{u}^T\mathbf{A}\mathbf{w} - \psi(\mathbf{w},\mathbf{u})\right\} \le c\right).$$

Thus, we may apply[10] statement (i) of Theorem 3.3.1 (with the roles of $\mathbf{w}$ and $\mathbf{u}$ flipped):

$$\begin{aligned}\mathbb{P}\left(-\Phi(\mathbf{A}) < c\right) &\le 2\mathbb{P}\left(\min_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \max_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \left\{\|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} - \psi(\mathbf{w},\mathbf{u})\right\} \le c\right) \\ &= 2\mathbb{P}\left(\min_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \max_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \left\{-\|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} - \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} - \psi(\mathbf{w},\mathbf{u})\right\} \le c\right), \quad (3.15)\end{aligned}$$

where the last equation follows because of the symmetry of $\mathbf{g}$ and $\mathbf{h}$. To continue, note that

$$\begin{aligned}\min_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \max_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \left\{-\|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} - \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} - \psi(\mathbf{w},\mathbf{u})\right\} = \\ - \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \left\{\|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} + \psi(\mathbf{w},\mathbf{u})\right\},\end{aligned}$$

and further apply the minimax inequality [Roc97, Lemma 36.1] which requires that for all $\mathbf{g}, \mathbf{h}$,

$$\begin{aligned}\max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \left\{\|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w},\mathbf{u})\right\} \\ \le \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}}} \left\{\|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w},\mathbf{u})\right\} := \phi(\mathbf{g},\mathbf{h}).\end{aligned}$$

These, when combined with (3.15), give $\mathbb{P}\left(-\Phi(\mathbf{A}) < c\right) \le 2\mathbb{P}\left(-\phi(\mathbf{g},\mathbf{h}) \le c\right)$. Apply this for $c = -(\mu + t)$ and combine with statement (i) of the theorem for $c = \mu - t$, to conclude with (3.13) as desired.

[9] Flipping the order of min-max remains valid even under the weaker assumption of a *quasi*-convex-concave function ψ, [Sio+58, Thm. 3.4]. Hence, (3.12) holds in this case too by the same argument.

[10] Observe that the signs of $\mathbf{u}^T\mathbf{A}\mathbf{w}$, $\mathbf{g}^T\mathbf{u}$ and $\mathbf{h}^T\mathbf{w}$ do not matter because of the assumed symmetry in the distributions of $\mathbf{A}$, $\mathbf{g}$ and $\mathbf{h}$.

Proof of statement (iii)

Consider the following event

$$\mathcal{E} = \{\Phi_{\mathcal{S}^c}(\mathbf{A}) \geq \overline{\phi}_{\mathcal{S}^c} - \eta \ , \ \Phi(\mathbf{A}) \leq \overline{\phi} + \eta\}.$$

In this event, it is not hard to check using assumption (a) that $\Phi_{\mathcal{S}^c} > \Phi$, or equivalently $\mathbf{w}_\Phi \in \mathcal{S}$. Thus, it suffices to show that $\mathcal{E}$ occurs with probability at least $1 - 4p$.

Indeed, from statement (i) of the theorem and assumption (c),

$$\mathbb{P}(\Phi_{\mathcal{S}^c}(\mathbf{A}) < \overline{\phi}_{\mathcal{S}^c} - \eta) \leq 2\mathbb{P}(\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h}) \leq \overline{\phi}_{\mathcal{S}^c} - \eta) \leq 2p.$$

Also, from statement (ii) of the theorem and assumption (b),

$$\mathbb{P}(\Phi(\mathbf{A}) > \overline{\phi} + \eta) \leq 2\mathbb{P}(\phi(\mathbf{g}, \mathbf{h}) \geq \overline{\phi} + \eta) \leq 2p.$$

Combining the above displays the claim follows from a union bound.

Proof of Corollary 3.3.2

Call $\eta := (\overline{\phi}_{\mathcal{S}^c} - \overline{\phi})/3 > 0$. By assumption, for any $p > 0$ there exists $N := N(\eta, p)$ such that the events $\{\phi < \overline{\phi} + \eta\}$ and $\{\phi_{\mathcal{S}^c} > \overline{\phi}_{\mathcal{S}^c} - \eta\}$ occur with probability at least $1 - p$ each, for all $n > N$. Then, for all $n > N$, we can apply Theorem 3.3.1(iii) to conclude that $\mathbf{w}_\Phi(\mathbf{A}) \in \mathcal{S}$ with probably at least $1 - 4p$. Since this holds for all $p > 0$, the proof is complete.

Chapter 4

THE SQUARED-ERROR OF REGULARIZED M-ESTIMATORS

In this chapter, we consider a very general setup of structured signal estimation under the noisy linear measurement model in (1.1). Our focus is on the regime of *high-dimensions* where both the dimensions of the ambient space n and the number of measurements m are large. We assume the noise vector $\mathbf{z}$ is generated from some distribution density in $\mathbb{R}^m$, say $p_{\mathbf{z}}$, and we model prior structural information on the unknown signal $\mathbf{x}_0$ by assuming that it is sampled from an n-dimensional probability density $p_{\mathbf{x}_0}$. For the recovery of the signal, we use convex regularized M-estimators as in (1.2). We derive an asymptotically precise characterization of the (mean) squared-error performance of this general class of convex optimization estimators when the measurement matrix is iid Gaussian.

We introduce some notation and formally set up the problem in Section 4.1. The main theorem (Theorem 4.2.1) is presented next in Section 4.2, where its features and implications are also discussed. Theorem 4.2.1 is specialized to instances of M-estimators with separable loss and regularizer functions in Section 4.3. Later, in Chapter 6, we include numerous specific examples and numerical simulations. Also, in Chapter 5, we introduce the mechanics that lead to the proof of Theorem 4.2.1

4.1 Introduction

Regularized M-estimators

Regularized M-estimators obtain an estimate $\hat{\mathbf{x}}$ of the unknown $\mathbf{x}_0$ from the vector of observations $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$ is via solving the *convex* program

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \ \mathcal{L}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \lambda f(\mathbf{x}). \tag{4.1}$$

The *loss function* $\mathcal{L} : \mathbb{R}^m \to \mathbb{R}$ measures the deviation of $\mathbf{A}\hat{\mathbf{x}}$ from the observations $\mathbf{y}$, the *regularizer* $f : \mathbb{R}^n \to \mathbb{R}$ aims to promote the particular structure of $\mathbf{x}_0$, and, the regularizer parameter $\lambda > 0$ balances between the two. Henceforth, both $\mathcal{L}$ and f are assumed to be *convex*. Different choices of the loss function and of the regularizer give rise to a number of popular and widely-uesd estimators, including the following:

- Ordinary Least-Squares (LS) ($\mathcal{L}(\cdot) = (1/2)\|\cdot\|_2^2, f(\cdot) = 0$).
- Ridge regression ($\mathcal{L}(\cdot) = (1/2)\|\cdot\|_2^2, f(\cdot) = \|\cdot\|_2^2$).
- LASSO ($\mathcal{L}(\cdot) = (1/2)\|\cdot\|_2^2, f(\cdot) = \|\cdot\|_1$). Popular sparse recovery algorithm. The acronym was introduced in [Tib96]. To distinguish from the ℓ_2-LASSO defined below, we often refer to this version as the ℓ_2^2-LASSO. The "least-squares" nature of the loss function corresponds to a maximum likelihood estimator for the case when $\mathbf{z}$ is Gaussian.
- ℓ_2- (or, Square-root) LASSO, ($\mathcal{L}(\cdot) = \|\cdot\|_2$). A sparse-recovery algorithm similar in nature to the LASSO but there exists differences between them, e.g. tuning of the regularizer parameter of the ℓ_2-LASSO does not require knowledge of the standard deviation of the noise [BCW11; OTH13b].
- Generalized-LASSO, ($\mathcal{L}(\cdot) = (1/2)\|\cdot\|_2^2$ or $\mathcal{L}(\cdot) = \|\cdot\|_2$). A natural generalization of the LASSO to arbitrary convex (and, typically non-smooth) regularizers f, e.g. nuclear norm, $\ell_{1,2}$ norm (Group-LASSO, [YL06a]) and discrete total variation.
- Regularized LAD ($\mathcal{L}(\cdot) = \|\cdot\|_1$). Least Absolute Deviation algorithms are known to have robust properties in linear regression models (e.g. [RT95]). Also, they perform particularly well in the presence of heavy-tailed errors [Wan13] and *sparse* noise [WM10; FM14; TH14].
- Huber-loss ($\mathcal{L}(\cdot) = \sum_{j=1}^m h_\rho(\cdot)$). The Huber-loss function with parameter $\rho > 0$ is defined as

$$h_\rho(v) = \begin{cases} \frac{v^2}{2} & , |v| \le \rho, \\ \rho|v| - \frac{\rho^2}{2} & , \text{otherwise}, \end{cases} \tag{4.2}$$

  i.e. it is quadratic in small values of v but grows linearly for large values of v. It describes a popular *robust* estimator that is well-analyzed in the classical statistics setting [Hub11].
- Support Vector Machines regression, ($\mathcal{L}(\cdot) = \|\cdot\|_\epsilon, f(\cdot) = \|\cdot\|_2^2$). Here, $\|\mathbf{x}\|_\epsilon = \sum_i |\mathbf{x}_i|_\epsilon$, where $|x|_\epsilon = |x| - \epsilon$ if $|x| \ge \epsilon$ and 0, otherwise, is the Vapniks epsilon-insensitive norm; ϵ can be thought of as the resolution at which we want to look at the data [EPP00].

The list above is of course not exhaustive but illustrates the richness of the family of estimators represented by (4.1).

Notation

We gather here the basic notation that is used throughout the work.

Convex Analysis: For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, we let $\partial f(\mathbf{x})$ denote the subdifferential of f at $\mathbf{x}$ and $f^*(\mathbf{y}) = \sup_{\mathbf{x}} \mathbf{y}^T\mathbf{x} - f(\mathbf{x})$ its Fenchel conjugate. The Moreau envelope function of f at $\mathbf{x}$ with parameter τ is defined by

$$\mathrm{e}_f(\mathbf{x};\tau) := \min_{\mathbf{v}} \frac{1}{2\tau}\|\mathbf{x}-\mathbf{v}\|_2^2 + f(\mathbf{v}).$$

The optimal value in the minimization above is denoted by $\mathrm{prox}_f(\mathbf{x};\tau)$. When writing $\mathbf{x}_* = \arg\min_{\mathbf{x}} f(\mathbf{x})$, we let the operator arg min return any one of the possible minimizers of f.

Limits and Derivatives: For a real-valued (not necessarily differentiable) convex function f on $\mathbb{R}$ denote

$$f'_+(v) := \sup_{s\in\partial f(v)} |s|.$$

Also, write $\lim_{x\to c^+} f(x)$ for the one-sided limit of f at c, as x approaches from above. For a function $g(x,\tau)$ that is continuously differentiable on $\mathbb{R}^2$ we write $g'(x,\tau)$ or $g_1(x,\tau)$ for the derivative with respect to the first variable, and, $g_2(x,\tau)$ for the derivative with respect to the second variable.

Probability: The symbols $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ denote the probability of an event and the expectation of a random variable, respectively. For a sequence of random variables $\{\mathcal{X}^{(n)}\}_{n\in\mathbb{N}}$ and a constant $c \in \mathbb{R}$ (independent of n), we write $\{\mathcal{X}^{(n)}\}_{n\in\mathbb{N}} \xrightarrow{P} c$, to denote convergence *in probability*, i.e. $\forall\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left(|\mathcal{X}^{(n)} - c| > \epsilon\right) = 0$. We write $\mathcal{X} \sim p_X$ to denote that the random variable $\mathcal{X}$ has a density p_X. If $\mathcal{X}$ is a vector random variable with entries iid, then we use $\overset{\text{iid}}{\sim}$. Also, $\mathcal{X} \sim \mathcal{N}(\mu,\sigma^2)$ denotes a Gaussian random variable with mean μ and variance σ^2.

We reserve the letters $\mathbf{g}$ and $\mathbf{h}$ to denote standard Gaussian vectors (with iid entries $\mathcal{N}(0,1)$) of dimensions m and n, respectively. Similarly, G and H are reserved to denote (scalar) standard normal random variables.

Setup

Linear Asymptotic Regime: Our study falls into the linear asymptotic regime in which the problem dimensions m and n grow proportionally to infinity with

$$m/n \to \delta \in (0,\infty).$$

Measurement matrix: The entries of $\mathbf{A} \in \mathbb{R}^{m \times n}$ are i.i.d. $\mathcal{N}(0, \frac{1}{n})$. The normalization of the variance ensures that the rows of $\mathbf{A}$ are approximately unit-norm; this is necessary in order to properly define a signal-to-noise ratio.

Unknown (structured) signal: Let $\mathbf{x}_0 \in \mathbb{R}^n$ represent the unknown signal vector that is sampled from a probability density $p_{\mathbf{x}_0} \in \mathbb{R}^n$ with one dimensional marginals that are independent of n. Note, that we do *not* necessarily require that the entries of $\mathbf{x}_0$ be iid. The signal $\mathbf{x}_0$ is assumed independent of $\mathbf{A}$.

Information about the structure of $\mathbf{x}_0$ is encoded in $p_{\overline{\mathbf{x}}_0}$. For instance, to study an $\mathbf{x}_0$ which is sparse, it is typical to assume that its entries are i.i.d. $\mathbf{x}_{0,i} \sim (1-\rho)\delta_0 + \rho q_{\mathbf{X}_0}$, where $\rho \in (0, 1)$ becomes the normalized sparsity level, $q_{\mathbf{X}_0}$ is a scalar p.d.f. and δ_0 is the Dirac delta function[1].

Regularizer: We consider regularizers $f : \mathbb{R}^n \to \mathbb{R}$ that are proper continuous *convex* functions.

Loss function: The loss function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$ is proper continuous and convex. Without loss of generality, we assume for simplicity that $\min_{\mathbf{v}} \mathcal{L}(\mathbf{v}) = 0$. Finally, we impose a natural normalization condition as follows: for all $n \in \mathbb{N}$ and all constants $c > 0$ there exists constant $C > 0$, such that $\|\mathbf{v}\|_2 \leq c\sqrt{n} \implies \sup_{\mathbf{s} \in \partial \mathcal{L}(\mathbf{v})} \|\mathbf{s}\|_2 \leq C\sqrt{n}$.

Noise vector: The noise vector $\mathbf{z} \in \mathbb{R}^m$ follows a probability distribution $p_{\mathbf{z}} \in \mathbb{R}^m$ with one dimensional marginals that are independent of n. Also, it is independent of the measurement matrix $\mathbf{A}$.

Sequence of problem instances: Formally, our result applies on a sequence of problem instances $\{\mathbf{x}_0, \mathbf{A}, \mathbf{z}, \mathcal{L}, f, m\}_{n \in \mathbb{N}}$ indexed by n such that the properties listed above hold for all members of the sequence and for all $n \in \mathbb{N}$. (We do not write out the subscripts n for arguments of the sequence in order to not overload notation). Every such sequence generates a sequence $\{\mathbf{y}, \hat{\mathbf{x}}\}_{n \in \mathbb{N}}$ where $\mathbf{y} := \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, and,

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \ \mathcal{L}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \lambda f(\mathbf{x}). \tag{4.3}$$

Here, $\lambda > 0$ is a fixed regularizer parameter.

[1] Such models in place for studying structured signals have been widely used in the relevant literature, e.g. [DJ94; DMM11; DJM13]. In fact, the results here continue to hold as long as the marginal distribution of $\mathbf{x}_0$ converges to a given distribution (as in [BM12]).

Estimation error: Solving (4.3) aims to recover $\mathbf{x}_0$. We assess the quality of the estimator $\hat{\mathbf{x}}$ with the "empirical squared error" (or simply, "squared-error") defined as: $\frac{1}{n}\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2$. Note, that this is a random quantity owing to the randomness of $\mathbf{A}$, $\mathbf{z}$ and $\mathbf{x}_0$. Our main theorem precisely evaluates its high probability limit as $n \to \infty$.

4.2 General Result

Key Assumption

As already hinted in the introduction the functions $\mathcal{L}$, f and the distributions $p_{\mathbf{z}}$ and $p_{\mathbf{x}_0}$ determine the error performance indirectly through "summary functionals" related to the Moreau-envelope approximations. The assumption below is an in-probability convergence requirement on the sequence of Moreau-envelopes, and defines those summary functionals. It also involves a rather natural growth restriction on the loss function in the presence of noise to handle instances where the noise may have unbounded moments.

Assumption 4.2.1 (Summary functionals L and F)**.** *We say that Assumption 4.2.1 holds if:*

(a) For all $c \in \mathbb{R}$ and $\tau > 0$, there exist continuous functions $L : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$ and $F : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$ such that[2]

$$m^{-1}\left\{\mathrm{e}_{\mathcal{L}}\left(c\mathbf{g}+\mathbf{z};\tau\right)-\mathcal{L}(\mathbf{z})\right\} \xrightarrow{P} L\left(c,\tau\right) \quad \text{and}$$
$$n^{-1}\left\{\mathrm{e}_{f}\left(c\mathbf{h}+\mathbf{x}_0;\tau\right)-f(\mathbf{x}_0)\right\} \xrightarrow{P} F\left(c,\tau\right),$$

(b) At least one of the following holds. There exists constant $C > 0$ such that $\frac{\|\mathbf{z}\|_2}{\sqrt{m}} \leq C$ with probability approaching 1 (w.p.a.1), or, $\sup_{\mathbf{v}\in\mathbb{R}^m} \sup_{\mathbf{s}\in\partial\mathcal{L}(\mathbf{v})} \|\mathbf{s}\|_2 < \infty$ for all $m \in \mathbb{N}$.

Assumption 4.2.1 is rather mild: as discussed later in Section 4.2, it holds naturally under very generic settings. Yet, it is of key importance since it defines the functionals L and F, which are necessary ingredients involved in the error prediction of (4.3). The main theorem in its most general form will require some extra (continuity and growth) properties on the functionals L and F. Those will most often be naturally inherited from corresponding easy-to-verify, and in cases well-studied, properties of the Moreau envelope functions.

[2]The convergence above is in probability over $\mathbf{z} \sim p_{\mathbf{z}}, \mathbf{x}_0 \sim p_{\mathbf{x}_0}, \mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_m)$ and $\mathbf{h} \sim \mathcal{N}(0, \mathbf{I}_n)$.

Theorem

Assumption 4.2.1 provides us with the basic terminology needed for the statement of the main theorem. Technically, a few additional mild constraint qualifications are required. We present those immediately after the statement of the main result (see Assumption 4.2.2). The proof of the theorem is deferred to Appendix B. An outline is given earlier in Section 5.2.

Theorem 4.2.1 (Master Theorem)**.** *Let* $\hat{\mathbf{x}}$ *be a minimizer of the Generalized M-estimator in* (4.3) *for fixed* $\lambda > 0$*. Further let Assumptions 4.2.1 and 4.2.2 hold. If the following convex-concave minimax scalar optimization*

$$\inf_{\substack{\alpha\geq 0\\ \tau_g>0}} \sup_{\substack{\beta\geq 0\\ \tau_h>0}} \mathcal{D}(\alpha,\tau_g,\beta,\tau_h) := \frac{\beta\tau_g}{2} + \delta\cdot L\left(\alpha,\frac{\tau_g}{\beta}\right) - \frac{\alpha\tau_h}{2} - \frac{\alpha\beta^2}{2\tau_h} + \lambda\cdot F\left(\frac{\alpha\beta}{\tau_h},\frac{\alpha\lambda}{\tau_h}\right) \tag{4.4}$$

has a unique minimizer α_**, then, it holds in probability that*

$$\lim_{n\to\infty}\frac{1}{n}\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2 = \alpha_*^2.$$

We will often refer to the optimization problem in (4.4) as the *Scalar Performance Optimization* (SPO) problem.

A few important remarks are in place here (a detailed discussion follows in Section 4.2): (i) The convergence in the theorem is over the randomness of the design matrix $\mathbf{A}$, of the noise vector $\mathbf{z}$ and of the unknown signal $\mathbf{x}_0$. (ii) As was discussed in Section 4.1 the result applies to a properly defined sequence of M-Estimators of growing dimensions m and n such that $m/n \to \delta \in (0,\infty)$. (We have dropped the dependence of $\hat{\mathbf{x}}$ and $\mathbf{x}_0$ on n to simplify notation.) (iii) The terms involving division by α and β are understood as taking their limiting values when $\alpha = 0$ and $\beta = 0$, i.e. $\mathcal{D}(0,\tau_g,\beta,\tau_h) = \lim_{\alpha\to 0^+}\mathcal{D}(\alpha,\tau_g,\beta,\tau_h)$ and $\mathcal{D}(\alpha,\tau_g,0,\tau_h) = \lim_{\beta\to 0^+}\mathcal{D}(\alpha,\tau_g,\beta,\tau_h)$.

Before proceeding with a further discussion of the result, let us state Assumption 4.2.2 on the functionals L and F as required by Theorem 4.2.1.

Assumption 4.2.2 (Properties of L and F)**.** *We say that Assumption 4.2.2 holds if all the following are true.*

(a) $\lim_{\tau\to 0^+} F(\tau,\tau) = 0$ *and* $\lim_{c\to+\infty}\left\{\frac{c^2}{2\tau} - F(c,\tau)\right\} = +\infty$ *for all* $\tau > 0$.

(b) $\lim_{\tau\to 0^+} L(\alpha,\tau) < +\infty$, $\lim_{\tau\to 0^+} L(0,\tau) = 0$, *and*, $-\infty < L_{2,+}(0,0) := \lim_{\tau\to 0^+} L_{2,-}(0,\tau) \leq 0$.

(c) $\frac{1}{m}\mathcal{L}(\mathbf{z}) \xrightarrow{P} L_0 \in [0,\infty]$. *Also,* $L_0 = -\lim_{\tau\to+\infty} L(c,\tau) \geq -L(c,\tau')$ *for all* $c \in \mathbb{R}$, $\tau' > 0$.

(d) If $L_0 = +\infty$, *then* $\lim_{\tau\to+\infty} \frac{L(c,\tau)}{\tau} = 0$, *for all* $c \in \mathbb{R}$.

A few remarks on the notation used in Assumption 4.2.2 follow. In (b), $L_{2,-}(0,\tau)$ denotes the left derivative of L with respect to its second argument evaluated at $(0,\tau)$. In (d), L_0 can take the value $+\infty$. For a sequence of random variables $\{\mathcal{X}^{(n)}\}_{n\in\mathbb{N}}$, we write $\mathcal{X}^{(n)} \xrightarrow{P} +\infty$, iff for all $M > 0$, $\lim_{n\to\infty} \mathbb{P}\left(\mathcal{X}^{(n)} > M\right) = 1$.

Separable M-estimators

A special yet popular family of M-estimators involves separable loss/regularizer functions and iid noise/signal distributions. We refer to such instances as "*separable* M-estimators". To be concrete, consider solving

$$\min_{\mathbf{x}} \sum_{j=1}^{m} \ell\left(\mathbf{y}_j - \mathbf{a}_j^T\mathbf{x}\right) + \lambda \sum_{i=1}^{n} f(\mathbf{x}_i), \tag{4.5}$$

where additionally, $\mathbf{z}_j \overset{\text{iid}}{\sim} p_z$ and $\mathbf{x}_{0i} \overset{\text{iid}}{\sim} p_x$. Popular choices for the (scalar) loss function $\ell(v)$ above, include v^2, $|v|$, Huber-loss, etc.. In the separable case, the generic Assumptions 4.2.1 and 4.2.2 translate to very primitive and naturally interpretable conditions. Also, the functionals L and F take here an explicit form, which we call the "Expected Moreau envelope". The *Expected Moreau envelope* associated with the loss function is given by

$$L(c,\tau) = \mathbb{E}_{\substack{G\sim\mathcal{N}(0,1)\\ Z\sim p_Z}} \left[\mathrm{e}_\ell\left(cG + Z;\tau\right) - \ell(Z)\right].$$

The function L, above, has the following remarkable properties: (i) it is smooth regardless of the smoothness of ℓ, and, (ii) it is strictly convex regardless of whether ℓ is itself strictly convex or not. In particular, the second property can be used to show that the uniqueness condition of Theorem 4.2.1 regarding the minimizer α_* of (4.4) is satisfied.

In order to get a better understanding of those issues before discussing Theorem 4.2.1 in its greatest generality, we state below a summary of the main result regarding separable M-estimators. (The formal statement will be given later in Section 4.3, which includes a detailed treatment of separable M-estimators.)

Summary of result for separable M-estimators . *Let $\ell, f : \mathbb{R} \to \mathbb{R}$ be convex non-negative functions, and, $Z \sim p_Z$, $X_0 \sim p_x$ such that for all $c \in \mathbb{R}$:*

$$\mathbb{E}\left[|\ell'_+(cG + Z)|^2\right] < \infty \quad \textit{and} \quad \mathbb{E}\left[|f'_+(cH + X_0)|^2\right] < \infty. \tag{4.6}$$

Further assume $\mathbb{E}X_0^2 < \infty$, and, that either $\mathbb{E}Z^2 < \infty$ or $\sup_v \frac{|\ell(v)|}{|v|} < \infty$. Then, any minimizer $\hat{\mathbf{x}}$ of (4.5) satisfies in probability,

$$\lim_{n\to\infty} \frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \alpha_*^2,$$

where α_* is the unique minimizer to the (SPO) problem in (4.4) with

$$L(c,\tau) = \mathbb{E}\left[\mathrm{e}_\ell\left(cG + Z;\tau\right) - \ell(Z)\right] \quad \textit{and} \quad F(c,\tau) = \mathbb{E}\left[\mathrm{e}_f\left(cH + X_0;\tau\right) - \ell(X_0)\right].$$

We defer most of the discussions to Section 4.3. We only note here that there is *no* smoothness or strict convexity assumption imposed on ℓ or f. Neither is the noise distribution required to have bounded moments. For example, $\ell(v) = |v|$ with z distributed iid Cauchy satisfies all the conditions. The main condition of the theorem is the one in (4.6), which is very primitive, and, easy to check. It essentially guarantees that $\mathrm{e}_\ell\left(cG + Z;\tau\right) - \ell(Z)$ is absolutely integrable, thus L is well-defined. It turns out that this also suffices for all requirements of Assumption 4.2.2 to be satisfied.

Remarks

On Assumption 4.2.1

We have made an effort to identify technical assumptions required for the statement of Theorem 4.2.1 which are as *generic* and *minimal* as possible. Assumption 4.2.1 summarizes those technical conditions that are essential for our result to hold in its most general form. In later sections, when we discuss special cases (e.g. separable M-estimators in Section 4.3), we show that these conditions translate to more *primitive* sufficient conditions that are often easier to check.

Remark 4.2.0.9 (WLLN and Robust Statistics). The most natural setting where Assumption 4.2.1(a) can be easily interpreted is that of separable functions. For instance, if $\mathcal{L}(\mathbf{v}) = \sum_{j=1}^m \ell(\mathbf{v}_j)$ and $\mathbf{z}_j \overset{\text{iid}}{\sim} p_Z(Z)$, then, in view of the WLLN, the natural candidate for $L(c,\tau)$ is $\mathbb{E}[\mathrm{e}_\ell\left(cG + Z;\tau\right) - \ell(Z)]$. Of course, this requires the argument under the expectation be aboslutely integrable. This is naturally satisfied for most loss functions in the case of noise distributions with bounded moments.

On the other hand, when the noise is (say) heavy-tailed, some extra caution is required on the choice of the loss function; this leads to (4.6). As a warning to this discussion, Assumption 4.2.1 does *not* require separability. For example, we use Theorem 4.2.1 to analyze the error performance of the square-root lasso (for which $\mathcal{L}(\mathbf{v}) = \|\mathbf{v}\|_2$) in Section 6.5, and, that of another instance with a non-separable regularizer function in Section 6.3.

Remark 4.2.0.10 (Convexity of L and F). We remark that if Assumption 4.2.1 holds, then both the functions F and L defined therein are *jointly convex* in their arguments. This follows from the facts that (a) the Moureau envelope of a convex function is jointly convex in its arguments (cf. Lemma B.4.1(ii)), (b) taking limits preserves convexity. In that sense, the continuity requirement of the assumption on L and F is rather mild, since convex functions are continuous on the interior of their domain [Roc97, Thm. 10.1].

Remark 4.2.0.11 (Robust Statistics). Assumption 4.2.1(b) is tailored to scenarios in which the noise distribution has unbounded moments (e.g. mean, variance); in this case $\|\mathbf{z}\|_2/\sqrt{n}$ is not bounded with high probability. It is not hard to see that condition 4.2.1(b) implies $\sup_{\mathbf{v}} \frac{\|\mathcal{L}(\mathbf{v})\|_2}{\|\mathbf{v}\|_2} < \infty$; such a requirement that $\mathcal{L}$ grows at most linearly at infinity is natural in the context of robust statistics.

On Assumption 4.2.2

Remark 4.2.0.12 (Continuity). Conditions (a), (b) and (c) impose continuity and growth requirements on L and F. Those are rather naturally inherited by corresponding properties of the Moreau-envelope functions. In Appendix B.4 we have gathered such relevant and useful properties of Moreau-envelopes, which we use extensively throughout the text. For an illustration, it is not hard to see[3] that $\lim_{\tau\to 0^+} \mathrm{e}_{\mathcal{L}}(\mathbf{z};\tau) = \mathcal{L}(\mathbf{z})$. This, of course is in line with Assumption 4.2.2(b) that $\lim_{\tau\to 0^+} L(0,\tau) = 0$.

Remark 4.2.0.13 (Robust Statistics). Assumption 4.2.2(d) is meant to deal with cases of noise with unbounded moments (this will often translate to $L_0 = +\infty$). In such cases, we require that $L(c,\tau)$ grows sub-linearly in τ. Once more, this property is essentially inherited without any extra effort by the corresponding property of the Moreau-envelope.

[3] Formally, this is a well-known continuity result on Moreau-envelopes. see Lemma B.4.1(ix)

On the Theorem

Remark 4.2.0.14 (Limits). In evaluating the objective function $\mathcal{D}$ of the (SPO) at $\alpha = 0$ and $\beta = 0$, Assumptions 4.2.2(a)-(b) turn out to be useful, giving

$$\lim_{\beta \to 0^+} L\left(\alpha, \frac{\tau_g}{\beta}\right) = -L_0 \quad \text{and} \quad \lim_{\alpha \to 0^+} F\left(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h}\right) = 0.$$

Remark 4.2.0.15 (Convexity). An important property of the (SPO) is that it is *convex*: its objective function $\mathcal{D}(\alpha, \tau_g, \beta, \tau_h)$ is (jointly) convex in α, τ_g and concave in β, τ_h. As is well known, convexity translates to the ability to efficiently solve the optimization; see also Remark 4.2.0.18 below.

Remark 4.2.0.16 (Uniqueness of α_*). Theorem 4.2.1 assumes that the (SPO) problem has a unique minimizer α_*. In most cases discussed in this paper, the uniqueness property is a consequence of the fact that the function $L(c, \tau)$ turns out to be (jointly) *stricly* convex in its arguments. In the separable case, this translates to the strict convexity of the expected Moreau envelope function $\mathbb{E}[\mathrm{e}_\ell(cG + Z; \tau) - \ell(Z)]$, cf Remark 4.3.0.26.

Further Discussions

Remark 4.2.0.17 (The role of the parameters). The role of the normalized number of measurement $m/n \to \delta$ and that of the regularizer parameter λ are explicit in (4.4). On the other hand, the structure of $\mathbf{x}_0$ and the choice of the regularizer f are implicit through F. Similarly, any prior knowledge on the noise vector $\mathbf{z}$ and the effect of the loss function $\mathcal{L}$ are also implicit in (4.4) through L. In the separable case, the role of those summary parameters is played by the Expected Moreau envelope function.

Remark 4.2.0.18 (An alternative characterization). The (SPO) problem in (4.4) is *convex-concave* and only involves four *scalar* variables. Thus, the optimal α_* can, in principle, be efficiently numerically computed. Equivalently, α_* can be expressed as the solution to the corresponding first-order optimality conditions, which offers an alternative to the current statement of Theorem 4.2.1. In Section 4.3 we explicitly derive the system of stationary equations for the case of separable M-estimators. It is often possible to solve the stationary equations by means of simple iterative schemes (cf. Remark 4.3.0.29). Furthermore, this alternative formulation might be easier to work with when deriving analytic properties of α_*. As an example, in Sections 6.1–6.3 for specific instances of M-estimators, we start from the stationary equations, combine them in an appropriate way and derive insightful and practically

useful properties, such as lower bounds on α_*, necessary conditions on the problem parameters such that α_* (correspondingly, the equated error) be bounded, etc.

Remark 4.2.0.19 (Optimal cost). Although not stated as part of our main result, the analysis that leads to Theorem 4.2.1 further characterizes the limiting behavior of the optimal cost, say C_*, of the M-Estimator in (4.3). Let Γ_* be the optimal cost of the (SPO), then

$$\frac{1}{n}\min_{\mathbf{x}}\{\mathcal{L}(\mathbf{y}-\mathbf{A}\mathbf{x})-\mathcal{L}(\mathbf{z})+\lambda(f(\mathbf{x})-f(\mathbf{x}_0))\}\xrightarrow{P}\Gamma_*. \tag{4.7}$$

Remark 4.2.0.20 (Asymptotics). The statement of the theorem holds under an asymptotic setup in which the problem dimensions m and n grow to infinity. In Chapter 6.10 we examine via simulations the validity of the prediction for finite values of m and n. The results indicate that the asymptotic prediction becomes accurate for values of the problem parameters ranging on a few hundreds, and, in cases even on a few tens.

Remark 4.2.0.21 (Proof). The fundamental tool behind our analysis is the CGMT (Theorem 3.3.1). As seen in Section 3.3, the CGMT associates with a primary optimization (PO) problem a simplified auxiliary optimization (AO) problem from which we can tightly infer properties of the original (PO), such as the optimal cost, the optimal solution, etc.. We manage to write the general M-estimator in (4.3) as a (PO) problem so that CGMT is applicable. This leads to a corresponding (AO) problem. Next, we analyze the error of the (AO) and translate the result to the (PO) thanks to the CGMT. These ideas form the basic mechanics of the proof and are rather simple to explain; see Chapter 5 for an outline.

Remark 4.2.0.22 (Why “Master"?). All existing results in the literature on the performance of specific instances of M-estimators can be seen as special cases of Theorem 4.2.1. Beyond those, the theorem can be used to derive a wide range of novel results, including instances where the loss function and the regularizer may be non-smooth and non-separable, and where the noise distribution may have unbounded moments. We discuss several examples in Chapter 6.

Remark 4.2.0.23 (Premises/Opportunities). Theorem 4.2.1 paves the way to answering optimality questions regarding the performance of M-estimators under different scenarios. The first fundamental step in answering such optimality questions (see Chapter 1) is characterizing the squared error in terms of the problem design parameters, i.e. f, ℓ, λ and δ. And, of course, this is exactly what Theorem 4.2.1 achieves. Since the characterization differs from the corresponding results of classical statistics (where n is considered fixed), the questions will not in general admit

the same answers. In the high-dimensional regime, our knowledge of those issues is rather limited and there is an exciting potential for exploring new phenomena and providing answers that are both of theoretical and of practical interest. We provide a few preliminary results towards this direction in Chapter 6.

4.3 Separable M-estimators

We specialize the general result of Section 4.2 to the popular case where the loss function $\mathcal{L}$ and the regularizer f are both separable and the noise vector and signal $\mathbf{x}_0$ both have entries iid. To make things concrete, assume[4]

$$\mathcal{L}(\mathbf{v}) = \sum_{j=1}^{m} \ell(\mathbf{v}_j) \qquad \text{and} \qquad \mathbf{z}_j \overset{\text{iid}}{\sim} p_Z,\ j = 1, \dots, m,$$

$$f(\mathbf{x}) = \sum_{i=1}^{n} f(x_i) \qquad \text{and} \qquad \mathbf{x}_{0i} \overset{\text{iid}}{\sim} p_x,\ i = 1, \dots, n.$$

Henceforth, both ℓ and f are proper closed convex functions. Also, it is further assumed

$$\ell(0) = 0 = \min_v \ell(v) \quad \text{and} \quad f(0) = 0. \tag{4.8}$$

Satisfying Assumptions 4.2.1 and 4.2.2

To apply Theorem 4.2.1, we first need to verify that Assumptions 4.2.1 and 4.2.2 hold for both the loss function and the noise distribution, and for the regularizer and the signal distribution.

Loss Function and Noise Distribution

In the separable case Assumptions 4.2.1 and 4.2.2 essentially translate to the following requirement on ℓ and p_Z:

$$\mathbb{E}\left[|\ell'_+(cG + Z)|^2\right] < \infty, \qquad \text{for all } c \in \mathbb{R}, \tag{4.9}$$

where the expectation is over $Z \sim p_Z$ and $G \sim \mathcal{N}(0, 1)$. This is shown in Lemma 4.3.1 below.

Lemma 4.3.1 (Expected Moreau envelope–Loss fcn)**.** *If ℓ and p_Z satisfy (4.9), then, Assumptions 4.2.1(a) and 4.2.2(b)-(d) hold with*

$$L(c, \tau) = \mathbb{E}\left[\mathrm{e}_\ell\left(cG + Z; \tau\right) - \ell(Z)\right]. \tag{4.10}$$

[4]Note the slight abuse of notation here in using f to denote both the vector-valued and scalar regularizer function.

The condition in (4.9) is very primitive and is, in general, easy to check. It essentially guarantees that $\mathrm{e}_\ell(cG + Z; \tau) - \ell(Z)$ is absolutely integrable (for a proof see Appendix B.3). Hence, L in Lemma 4.3.1 is well-defined and it satisfies Assumption 4.2.1(a) as a result of applying the WLLN. A few examples for which (4.9) is satisfied include:

1. $\ell(v) = v^2$ and $\mathbb{E}Z^2 < \infty$,
2. (4.9) is trivially satisfied if $\ell(v) = |v|$ for any noise distribution p_Z,
3. Huber-loss and $Z \sim \text{Cauchy}(0, 1)$.

Apart from (4.9), we also need to satisfy Assumption 4.2.1(b), which here translates to the following requirement:

$$\mathbb{E}Z^2 < \infty \qquad \text{or} \qquad \sup_{v \in \mathbb{R}} |\ell'_+(v)| < \infty. \tag{4.11}$$

The second condition above on boundedness of the sub-differential is equivalent to $\sup_v \frac{|\ell(v)|}{|v|} < \infty$. That is, if Z has unbounded second moments then ℓ needs to grow to infinity at most linearly, e.g. $|\cdot|$, Huber-loss, etc.

Regularizer and Signal Distribution

Not surprisingly, following the results of Section 4.3, the required condition on f and p_x becomes

$$\mathbb{E}\left[|f'_+(cH + X_0)|^2\right] < \infty, \qquad \text{for all } c \in \mathbb{R}, \tag{4.12}$$

where the expectation is over $X_0 \sim p_x$ and $H \sim \mathcal{N}(0, 1)$. Additionally, the following mild assumptions are required:

$$\exists\, x_+ > 0,\ x_- < 0 \ \text{ such that } \ 0 \le f(x_\pm) < \infty \qquad \text{and} \qquad \mathbb{E}X_0^2 < \infty. \tag{4.13}$$

Lemma 4.3.2 (Expected Moreau Envelope–Regularizer fcn)**.** *If f and p_x satisfy* (4.12) *and* (4.13)*, then, Assumptions 4.2.1(a) and 4.2.2(a) hold with*

$$F(c, \tau) = \mathbb{E}\left[\mathrm{e}_f(cH + X_0; \tau) - f(X_0)\right]. \tag{4.14}$$

The Expected Moreau Envelope

If conditions (4.9), (4.11) and (4.12) are satisfied, then Theorem 4.2.1 is applicable with L and F given as in (4.10) and (4.14), respectively. We call those functions the *Expected Moreau envelopes*. The important role they play in determining the error performance of the corresponding M-estimator is apparent from Theorem 4.2.1. In this section, we discuss two key features that they possess, namely, *smoothness* and *strict convexity*.

Lemma 4.3.3 (Smoothness)**.** *Suppose ℓ is a closed proper convex function and p_Z a noise density such that (4.9) holds. Then, the function $L(c,\tau) := \mathbb{E}\left[\mathrm{e}_\ell\left(cG+Z;\tau\right)-\ell(Z)\right]$ is differentiable in $\mathbb{R}\times\mathbb{R}_{>0}$ with*

$$\frac{\partial L}{\partial c} = \mathbb{E}\left[\mathrm{e}'_\ell\left(cG+Z;\tau\right)G\right] \quad \text{and} \quad \frac{\partial L}{\partial \tau} = -\frac{1}{2}\mathbb{E}\left[\left(\mathrm{e}'_\ell\left(cG+Z;\tau\right)\right)^2\right].$$

Remark 4.3.0.24. Note that L is smooth, regardless of any non-smoothness of ℓ. This is a well-known fact about Moreau envelope approximations and also one of the primal reasons behind the important role those functions play in convex analysis [RW09]. The property is naturally inherited to the *Expected* Moreau envelopes as revealed by the lemma above.

Lemma 4.3.4 (Strict Convexity)**.** *Suppose ℓ is a closed proper convex function and p_Z a noise density such that (4.9) holds and the following are satisfied:*

(a) Either there exists $x\in\mathbb{R}$ at which ℓ is not *differentiable, or, there exists interval $\mathcal{I}\subset\mathbb{R}$ where ℓ is differentiable with a strictly increasing derivative,*

(b) $\mathrm{Var}(Z)\neq 0$ [5]*, and, at each $z\in\mathbb{R}$, $p_Z(z)$ is either a Dirac delta function or it is continuous.*

Then, $L(c,\tau) := \mathbb{E}\left[\mathrm{e}_\ell\left(cG+Z;\tau\right)-\ell(Z)\right]$ is jointly strictly convex in $\mathbb{R}_{>0}\times\mathbb{R}_{>0}$.

Remark 4.3.0.25. The function L is strictly convex, without requiring any strong or strict convexity assumption on ℓ. Interestingly, this property is not in general true for Moreau envelope approximations but it turns out to be the case for the *Expected* Moreau envelope L. The fact that the latter further involves taking an expectation over $cG+Z$, with G having a nonzero density on the entire real line, turns out to be critical.

[5] We require that there exist at least two values of $z\in\mathbb{R}$ for which $p_Z(z)>0$. In particular, there is *no* requirement that $\mathrm{Var}(Z)$ be defined, e.g. Cauchy distribution is allowed.

Remark 4.3.0.26 (Strict convexity $\implies$ Uniqueness of α_*). The strict convexity property of L is critical because it guarantees uniqueness of the minimizer α_* of the (SPO) problem in Theorem 4.2.1. This implication is proved in Lemma B.3.3 in Appendix B.3.

Error Prediction

We are now ready to state the main result of this section which characterizes the squared error of separable M-estimators. This is essentially a corollary of Theorem 4.2.1.

Theorem 4.3.1 (Separable M-estimators)**.** *Suppose ℓ and p_Z satisfy* (4.9)*,* (4.11)*, and, the two conditions of Lemma 4.3.4. Further assume that f, p_x satisfy* (4.12) *and* (4.13)*. Let $\hat{\mathbf{x}}$ be any minimizer of the separable M-estimator and consider the (SPO) problem in* (4.4) *with L and F given as in* (4.10) *and* (4.14)*, respectively. If the set of minimizers of the (SPO) over α is bounded, then there is a unique such minimizer α_* for which it holds in probability that*

$$\lim_{n\to\infty} \frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \alpha_*^2.$$

Remark 4.3.0.27 (Boundedness). Applying Theorem 4.3.1 requires a few primitive and easy to check assumptions on ℓ, Z, f and X_0. In contrast to the general case in Theorem 4.2.1, here, the uniqueness of α_* is guaranteed if the set of minimizers of the (SPO) over α is bounded. The boundedness condition is essentially in one to one correspondence, with the squared error of the M-estimator being (stochastically) bounded or not. We expect the boundedness assumption, which is generic in nature, to translate to necessary and sufficient primitive conditions on ℓ, f, p_Z, p_x and δ. For example, in Remark 6.1.0.33 we show that in the case of un-regularized M-estimators, a necessary such condition is that the normalized number of measurements be larger than 1, i.e. $\delta > 1$ [6]. Identifying such conditions that would guarantee bounded error is an important design issue, since it provides guarantees and guidelines on how the loss function, the regularizer and the number of measurements ought to be chosen. In the general case, this remains an open question. We expect that Theorem 4.3.1 itself and the proof ideas behind it (in particular, see

[6] Besides, in Remark 6.3.0.41, we show that with appropriate regularization, the necessary condition on the number of measurements becomes $\delta > \overline{D}_{f,\mathbf{x}_0}$, where $\overline{D}_{f,\mathbf{x}_0} = \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))/n$ is the normalized Gaussian squared distance to the cone of subdifferential defined in 2.10. Recall that for many useful examples $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \ll n$ implying that with an appropriate regularizer the signal $\mathbf{x}_0$ can be robustly estimated with a number of measurements that is less than the dimension of the signal. Of course, this is in agreement with the phase-transition result of Theorem 2.2.1.

Lemma B.1.5(b)) can be used to answer this question. Since this is not the main focus of the paper, we leave the rest for future work.

As a system of nonlinear equations

Theorem 4.3.1 predicts the error of the M-estimator as the optimizer α_* to a convex-concave optimization problem with four optimization variables. Equivalently, α_* can be expressed via the first-order optimality conditions (stationary equations) corresponding to this optimization. Recall from Lemma 4.3.3 that L and F are differentiable (irrespective of smoothness of ℓ and f). The error of the M-estimator is then the unique $\alpha_* \geq 0$ for which there exist $\tau_{g_*} \geq 0, \beta_* \geq 0$ and $\tau_{h*} \geq 0$ satisfying

$$\begin{aligned} &\frac{\partial \mathcal{D}}{\partial \alpha}\Big|_{p_*}(\alpha - \alpha_*) \geq 0, \qquad &&\frac{\partial \mathcal{D}}{\partial \tau_g}\Big|_{p_*}(\tau_g - \tau_{g_*}) \geq 0, \\ &\frac{\partial \mathcal{D}}{\partial \beta}\Big|_{p_*}(\beta - \beta_*) \leq 0, \qquad &&\frac{\partial \mathcal{D}}{\partial \tau_h}\Big|_{p_*}(\tau_h - \tau_{h*}) \leq 0, \end{aligned} \tag{4.15}$$

for all $\alpha, \beta \geq 0, \tau_g, \tau_h > 0$ and $p_* = (\alpha_*, \tau_{g_*}, \beta_*, \tau_{h*})$. A similar remark to the one that follows Theorem 4.2.1 is in place regarding the values $\alpha = 0$ and $\beta = 0$. At these, the derivatives above should be interpreted as the corresponding (upper) limits as $\alpha \to 0^+$ and $\beta \to 0^+$. The continuity properties of the Moreau envelope (see Lemma B.4.1) guarantee that those limits are well-defined

When $\alpha_* > 0$ and there also exist optimal values β, τ_g, τ_h, all of them strictly positive, then (4.15) holds with equalities. In this case, a little bit of algebra and an appropriate change of variables from τ_g, τ_h to κ, ν, show that the optimality conditions can be expressed as follows:

$$\begin{cases} \alpha^2 = \mathbb{E}\left[\left(\frac{\lambda}{\nu} \cdot \mathrm{e}_f'\left(\frac{\beta}{\nu}H + X_0; \frac{\lambda}{\nu}\right) - \frac{\beta}{\nu}H\right)^2\right], \\ \beta^2 = \delta \cdot \mathbb{E}\left[\left(\mathrm{e}_\ell'(\alpha G + Z, \kappa)\right)^2\right], \\ \nu\alpha = \delta \cdot \mathbb{E}\left[\mathrm{e}_\ell'(\alpha G + Z, \kappa) \cdot G\right], \\ \kappa\beta = \frac{\beta}{\nu} - \frac{\lambda}{\nu} \cdot \mathbb{E}\left[\mathrm{e}_f'\left(\frac{\beta}{\nu}H + X_0; \frac{\lambda}{\nu}\right) \cdot H\right]. \end{cases} \tag{4.16}$$

Here, e_f' and e_ℓ', denote the first derivatives of the Moureau envelopes with respect to their first argument.

Remarks

Remark 4.3.0.28 (Reformulations). The system of equations in (4.16) can be easily reformulated in terms of the proximal operator of f and ℓ, using

$$\mathrm{e}'_\ell(\chi,\tau) = \frac{1}{\tau}(\chi - \mathrm{prox}_\ell(\chi;\tau)),$$

and similar for f (see Lemma B.4.1(iii)). In the case of additional smoothness assumptions on the loss function and/or the regularizer, further reformulations are possible. For example, if ℓ is two times differentiable, then using Stein's formula for normal random variables we can make the following substitution in (4.16):

$$\mathbb{E}\left[\mathrm{e}'_\ell(\alpha G + Z,\kappa)\cdot G\right] = \alpha\cdot\mathbb{E}\left[\mathrm{e}''_\ell(\alpha G + Z,\kappa)\right], \tag{4.17}$$

where the double-prime superscript denotes the second derivative with respect to the first argument. Such reformulations are often convenient for analysis purposes; see for example Remark 6.1.0.34.

Remark 4.3.0.29 (Numerical Evaluations). The system of equations in (4.16) comprises four nonlinear equations in four unknowns. Setting $\mathbf{t} = (\alpha,\beta,\nu,\kappa)$ for the vector of unknowns, the system of equations in (4.16) can be written as $\mathbf{t} = S(\mathbf{t})$, for appropriately defined $S:\mathbb{R}^4\to\mathbb{R}^4$. We have empirically observed that a simple recursion $\mathbf{t}_{k+1} = S(\mathbf{t}_k), k = 0,1,\ldots$ converges to a solution $\mathbf{t}_*$ satisfying $\mathbf{t}_* = S(\mathbf{t}_*)$. This observation is particularly useful since it allows for efficient numerical experimentations, cf. Section 6.10. It is certainly an interesting and practically useful subject of future work to identify analytic conditions under which such simple recursive schemes provide efficient means of solving (4.16).

Remark 4.3.0.30 (Extensions). The results of this section extend naturally, and without any extra effort, to the case of "block-seperable" loss functions and/or regularizers. A popular example that falls in this category is $\ell_{1,2}$-regularization, which is typically used for the recovery of block-sparse signals. In such a case $f(\mathbf{x}) = \sum_{i=1}^{b}\|[\mathbf{x}]_i\|_2$, where $[\mathbf{x}]_i = [\mathbf{x}_{(i-1)t+1},\mathbf{x}_{(i-1)t+2},\ldots,\mathbf{x}_{(i-1)t+t}]$, $i = 1,\ldots,b$ is the i^{th} block of $\mathbf{x}$. Here, b is the number of blocks and t is the length of each block. In the proportional high-dimensional regime, one would assume b growing linearly with n with a constant ratio of $1/t$.

4.4 Survey of Relevant Literature

There is a very long list of early results on the error performance of regularized M-estimators which derive "order-wise" bounds that involve unknown scaling con-

stants (e.g. [CT07; BCW11; BRT09; Neg+12; Wai14; Ver14; Ban+14; LHC15] and references therein). Nevertheless, in this discussion we focus entirely on more recent results that derive precise characterizations rather than loose bounds.

Unless otherwise stated, the literature that we describe below takes the random measurement matrix $\mathbf{A}$ to have independent Gaussian entries (but, see Remark 4.4.0.31). Also, it studies the high-dimensional asymptotic regime where m and n grow to infinity at a proportional rate.

Chronologically, the first such results were derived using the AMP framework by Bayati, Donoho, Maleki and Montanari [DMM11; BM12]. Both references consider a least-squares loss function with ℓ_1-regularization (a.k.a. LASSO) and Gaussian noise distribution: [DMM11] developed formal expressions for the reconstruction error at high-SNR under optimal tuning of the regularizer parameter $\lambda > 0$; [BM12] explicitly characterizeed the reconstruction error for all values of λ and all values of SNR. Subsequent works [Mal+13; Tae+13; DJM13] involve extensions of the results to other separable regularizers (e.g. $\ell_{1,2}$-norm). In late 2013, Donoho and Montanari [DM13] introduced an extension of the AMP framework to analyze the error performance of loss functions other than least-squares. Their analysis applies to separable, strongly-convex and smooth loss functions, to iid signal statistics, and to iid noise statistics with bounded second moments. Donoho & Montanari consider no regularization, hence their analysis restricts the normalized number of measurements to $\delta = m/n > 1$. Very recently, Bradic & Chen [BC15] built upon the framework of [DM13] and extended the analysis to sparse signal recovery and ℓ_1-regularization, under more general (but somewhat stringent) conditions on the loss function and on the noise and signal statistics. Our work raises the assumptions on separability, smoothness and strong convexity of the loss function and considers general convex regularizers and more general signal and noise statistics. Also, our analytic approach via the CGMT framework is somewhat more direct and potentially more powerful. The AMP framework involves two steps of analysis: (a) it analyzes the error performance of the AMP algorithm based on a state evolution framework inspired by statistical physics; (b) it shows that the AMP algorithm has the same error performance as the M-estimator. This way it concludes about the behavior of the latter. In contrast, our approach directly analyzes the error behavior of the original M-estimator. Nevertheless, we remark on the algorithmic advantage of the AMP framework which (whenever applicable) comes with a fast(er) iterative algorithm with the same error performance guarantees as the convex M-estimator.

Also, the AMP framework has been used for the analysis of other problems beyond noisy signal recovery from linear measurements (see [Mon15] and references therein). It remains an open and potentially interesting question to study deeper connections between the two different frameworks of analysis, namely the CGMT and the AMP frameworks.

Our approach, which is based on Gaussian process methods, is inspired and motivated by the work of Stojnic [Sto13a]. Stojnic considered an ℓ_1-constrained version of the LASSO under Gaussian noise distribution in the high-SNR regime. Under this setting, he was the first to note that Gordon's GMT could be combined with a convex duality argument to yield bounds that are tight. Shortly after, in [OTH13b] we extended Stojnic's results to the regularized case by deriving tight high-SNR bounds for the square-root LASSO with general convex regularizers. Since then, in a series of works [TOH14; TH14; Thr+15; TPH15; TH15; TOH15; TAH16] we have built upon these early results to obtain the powerful and transparent framework presented in this thesis. A critical element of the framework is the CGMT Theorem 3.3.1, which is a refined, clear, and extended version of Stojnic's idea. Beyond that, a plethora of new ideas and techniques have been blended that lead to the final results as presented here.

Finally, a third approach to analyze the mean-squared error performance of high-dimensional M-estimators has been undertaken by El Karoui in [Kar13; EK15]. El Karoui uses leave-one-out and martingale ideas from statistics and ideas from random matrix theory to accurately predict the squared error of ridge-regularized (a.k.a. $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$) M-estimators. The analysis can handle noise distributions with unbounded moments, but it requires a smooth and separable loss function. In our work, we drop both these assumptions and extend the results to general convex regularizers. In comparing the two works, we note that El Karoui's proof technique can deal with more general assumptions on the design matrix $\mathbf{A}$. (Nevertheless, please see Remark 4.4.0.31.) Beyond matrices with iid entries, El Karoui [EK15] further considers elliptical models. Even though we do not explicitely consider such an extension in the current paper, our proof technique is readily applicable to this more general scenario.

Remark 4.4.0.31 (On Universality). Since the works [Sto09b; Cha+12; Ame+13] we now have a very clear understanding of the phase transitions of non-smooth convex signal recovery methods with iid Gaussian measurements. Under the same measurement model, the current paper extends this clear picture to the noisy setting

by precisely characterizing the reconstruction error. Here, we briefly discuss relevant results that prove the universal behavior of iid Gaussian measurements over a wider class of distributions.

As discussed in Section 2.2, [OT15] have established the universality of the Guassian design for the phase-transition results of noiseless CS (see [OT15, Prop. 5.1], for an exact statement). Oymak & Tropp further derive conclusions for the noisy setting: they prove the universality of the error bounds that we derived in [OTH13b] (presented here in Chapter 7) for the constrained LASSO. It remains an open challenge to extend these results to the general setting of the arbitrary loss and regularizer functions of the current paper. We remark that the results of [OT15] use some of the ideas that were developed in [OTH13b; TOH15] and in the current paper. Also, note that the results of El Karoui [EK15] on the ridge-regularized M-estimators hold for matrices with iid entries beyond Gaussian.

From this discussion we have excluded random measurement models beyond ones with iid entries. An important example includes design matrices with orthogonal rows, e.g. Isotropically Random Orthogonal (IRO) matrices, randomly subsampled Fourier and Hadamard matrices, etc.. While the universality of phase transition appears to extend to such designs, this is not the case for the reconstruction error. We will pursue this study in Chapter 8.

Remark 4.4.0.32 (Heuristic results). In parallel to the works referenced above, there have been a number of works that studied the same questions, mixing heuristic-based arguments and extended simulations. For example, [GBS09; KWT10; RGF09; VKC14] use the replica method from statistical physics, which provides a powerful tool for tackling hard analytical problems, but still lacks mathematical rigor in some parts. Closer to the setting of our work, the high-dimensional error performance of regularized M-estimators has been previously considered via heuristic arguments and simulations in [EK+13; Bea+13]. In particular, Bean et. al. [Bea+13] shows that maximum likelihood estimators are in general inefficient in high-dimension and initiate the study of optimal loss functions. It is worth revisiting and extending those results in connection to the mathematically rigorous approach of the current paper.

Chapter 5

ANALYSIS FRAMEWORK

The goal of this chapter is to present an analytical framework that uses the CGMT (Theorem 4.2.1) for analyzing the estimation performance of regularized M-estimators under Gaussian measurement matrices. Recall from Section 3.3 that the CGMT associates with a primary optimization (PO) problem (cf. Eqn. (3.11a)) a simplified auxiliary optimization (AO) problem (cf. Eqn. (3.11b)) from which we can tightly infer properties of the original (PO), such as the optimal cost, the optimal solution, etc.. The prescribed framework includes four major steps:

1. *Identify the (PO)*: Express the regularized M-estimator in (1.2) as a (PO) problem that satisfies the convexity and compactness assumptions of the CGMT. Derive the corresponding (AO).

2. *"Scalarization" of the (AO)*: Treat the (AO) as a deterministic optimization and simplify it with the goal of reducing it to an optimization problem involving only scalar optimization variables.

3. *Convergence analysis of the (AO)*: Identify the converging limit of the (AO). Typically, this limit is itself a (deterministic) min-max optimization problem, which we term the Scalar Performance Optimization (SPO).

4. *Analysis of the (SPO)*: Prove that conditions (a)–(c) of Thm. 3.3.1(iii) hold for the (SPO), which often translates to strict-convexity requirements on its objective function.

5.1 How it Works

In essence, Step 1 can be accomplished by expressing the (convex) loss function in (1.2) in a variational form through its Fenchel conjugate, bringing (1.2) to the appropriate min-max form of the (PO). Convexity is guaranteed by the imposed convexity assumptions on the loss function and the regularizer. On the technical side the necessary compactness conditions of the CGMT need also to be guaranteed.

The premise of the CGMT is that the (AO) is (significantly) simpler to analyze than the corresponding (PO). It is shown here that this is indeed the case in the analysis

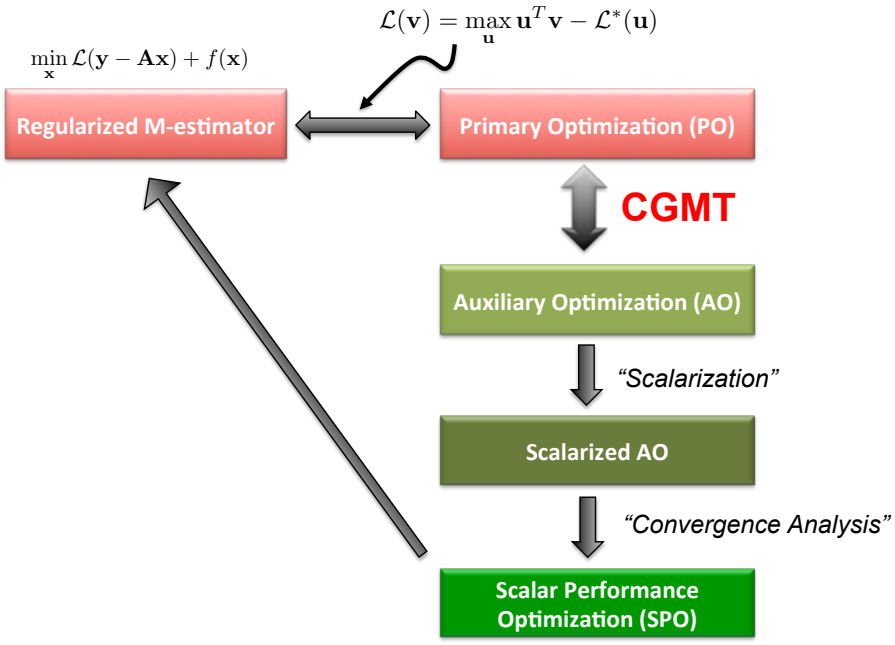

Figure 5.1: Schematic representation of the CGMT framework. The first step involves equivalently expressing the regularized M-estimator as a min-max Primary Optimization (PO) (cf. (3.11a)). (These problems are hard to directly analyze and are thus shown in red.) The CGMT Theorem 3.3.1 associates with the (PO) an Auxiliary Optimization (AO) problem that is simpler to analyze (hence, depicted in green). The second step of the framework involves simplifying the (AO) into an optimization problem that only involves scalar variables. This makes possible the convergence analysis that follows as a third step and leads to a deterministic Scalar Performance Optimization (SPO). The last step involves using the (SPO) to conclude about the original regularized M-estimator.

of regularized M-estimators. We split the analysis of the (AO) into two stages: Steps 2 and 3 above.

In Step 2, we treat the (AO) as a deterministic optimization and reduce it into a scalar optimization, i.e., one that involves only scalar optimization variables). Besides the trivial cases, directly optimizing over the original vector variables in (3.11b) is impossible. Instead, we introduce techniques that break this into steps. As an example, by appropriately arranging terms it is often possible to perform the optimization over the direction of the original vector optimization variable while keeping its magnitude fixed and having it play the role of the remaining scalar optimization variable. At the end, we manage to rewrite the (AO) in (3.11b) that involves a min-max optimization over two vector variables of sizes n and m respectively, as an optimization over, at most, four scalar variables. Importantly, this

optimization is convex-concave. Observe that this is in contrast to the original (AO) problem in (3.11b), which itself is not necessarily convex-concave.

Once the (AO) is written as a scalar optimization, we perform a convergence analysis of it. Using standard concentration arguments and convergence results, such as the Gaussian concentration of Lipschitz functions, or the WLLN, it is easy to compute the converging limit of the objective function when the optimization variables are considered fixed. The technical work involved here shows that the min-max cost (and, potentially the corresponding optimizers) of the random objective function converges to the min-max value of the converging limit of the objective that was previously identified. We call the converging limit of the the (AO), which is itself a (deterministic) min-max optimization, the Scalar Performance Optimization (SPO). As we show, convexity is again crucial here.

Identifying the converging limit of the (AO) in Step 3 facilitates the proof of conditions (a)–(c) of Theorem 3.3.1 (or, Corollary 3.3.2 in an asymptotic setup). It is shown that these conditions essentially translate to a strict-convexity requirement on the objective function of the (SPO). In its turn, we show that the strict convexity property holds because the Expected Moreau Envelope functions that naturally arise in the analysis are strictly convex themselves. Once the conditions are guaranteed to hold, the (asymptotic) CGMT applies and we can conclude about properties of the original (PO) problem (and correspondingly, the regularized M-estimator).

5.2 An Example

In Chapter 4 we characterized the squared-error of regularized M-estimators under noisy linear measurements and a Gaussian measurement model. The analysis is based on the CGMT framework as outlined above. Later in the thesis, we use the same framework to extend the results of Chapter 4 to the Isotropically Random Orthogonal measurement model in Chapter 8, to general metrics of performance beyond squared-error in Chapter 9, and, to non-linear measurements in Chapter 11. Despite the different features present in the analysis corresponding to each one of those problems, the core ideas are as outlined earlier in this chapter. To better illustrate those and to keep things concrete, we provide here an outline of the proof of Theorem 4.2.1. Leaving some technical challenges aside (addressed in Appendix B), the mechanics are easy to explain and provide valuable intuition regarding both the assumptions required and the flavor of the final result. For instance, we will be able to show without much effort, how the Moreau envelope functions $\mathrm{e}_{\mathcal{L}}\left(c\mathbf{g}+\mathbf{z};\tau\right)$

and $\mathrm{e}_f\,(c\mathbf{h}+\mathbf{x}_0;\tau)$ appear in the final result.

Introductory idea

Our goal is to characterize the nontrivial limiting behavior of $\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2$, where $\hat{\mathbf{x}}$ is any solution to the following minimization,

$$\min_{\mathbf{x}} \mathcal{L}(\mathbf{y}-\mathbf{A}\mathbf{x})+\lambda f(\mathbf{x}).$$

To get a direct handle on the error term, it is convenient to change the optimization variable to $\mathbf{w}:=\mathbf{x}-\mathbf{x}_0$, so then $\hat{\mathbf{w}}:=\hat{\mathbf{x}}-\mathbf{x}_0$ is a solution to (recall $\mathbf{y}=\mathbf{A}\mathbf{x}_0+\mathbf{z}$)

$$\hat{\mathbf{w}}:=\arg\min_{\mathbf{w}} \mathcal{L}(\mathbf{z}-\mathbf{A}\mathbf{w})+\lambda f(\mathbf{x}_0+\mathbf{w}) =: M(\mathbf{w}). \tag{5.1}$$

There is a simple but standard argument that is in the heart of most analyses of such minimization estimators, and comes as follows. Suppose we knew that the error $\|\hat{\mathbf{w}}\|_2$ converges eventually to some deterministic value, call it α_*. This is equivalent to $\hat{\mathbf{w}}$ belonging in the following set

$$\mathcal{S}_\epsilon=\{\mathbf{w}\mid |\|\mathbf{w}\|_2-\alpha_*|<\epsilon\}, \tag{5.2}$$

with probability one (w.p.1) for all $\epsilon>0$. Letting $\mathcal{S}_\epsilon^c$ denote the complement of that set, observe, that if w.p. 1,

$$M(\hat{\mathbf{w}})<\inf_{\mathbf{w}\in\mathcal{S}_\epsilon^c} M(\mathbf{w}), \tag{5.3}$$

then $\hat{\mathbf{w}}$ must lie in $\mathcal{S}_\epsilon$. Note that with this standard trick we have translated a question on the optimal solution of the minimization problem in (5.1) to one regarding its optimal cost. One possible approach in comparing the two random processes in (5.3) would be to first identify the converging limits of both. If say

$$M(\hat{\mathbf{w}})\xrightarrow{P}\overline{M} \qquad \text{and} \qquad \inf_{\mathbf{w}\in\mathcal{S}_\epsilon^c} M(\mathbf{w})\xrightarrow{P}\overline{M}_{\mathcal{S}_\epsilon^c}, \tag{5.4}$$

then, (5.3) holds as long as

$$\overline{M}<\overline{M}_{\mathcal{S}_\epsilon^c}, \tag{5.5}$$

which is just a comparison between two deterministic quantities.

This is exactly the approach we want to take here: show (5.4) and (5.5). Unfortunately, directly working with the objective function M and proving (5.4) turns out to be rather challenging. Instead, we prove the desired indirectly, via working with

an *auxiliary* objective function which is simpler to analyze. What justifies this idea is the *Convex Gaussian Min-max Theorem* (Theorem 3.3.1).

Recall that the CGMT associates with a primary optimization (PO) problem (cf. (3.11a)) a simplified auxiliary optimization (AO) (cf. (3.11b)) problem from which we can tightly infer properties of the original (PO), such as the optimal cost, the optimal solution, etc..

Identifying a (PO) and the Corresponding (AO) Problem

The M-estimator optimization in (5.1) will play the role of the (PO), and we need to identify the corresponding (AO). To do so, we first need to bring (5.1) in the form of (3.11a) as required by the CGMT.

The idea here is to use duality[1]. Specifically, we can equivalently view the minimization in (5.1) as follows:

$$\min_{\mathbf{w},\mathbf{v}} \mathcal{L}(\mathbf{v}) + \lambda f(\mathbf{x}_0 + \mathbf{w}) \quad \text{sub.to} \quad \mathbf{v} = \mathbf{z} - \mathbf{A}\mathbf{w}.$$

Then, associating a dual variable $\mathbf{u}$ with the equality constraint above, we have

$$\min_{\mathbf{w},\mathbf{v}} \max_{\mathbf{u}} \mathbf{u}^T\mathbf{A}\mathbf{w} \underbrace{-\mathbf{u}^T\mathbf{z} + \mathbf{u}^T\mathbf{v} + \mathcal{L}(\mathbf{v}) + \lambda f(\mathbf{x}_0 + \mathbf{w})}_{\psi(\mathbf{w},\mathbf{v},\mathbf{u})}. \tag{5.6}$$

Clearly, this is now in the desired format: we can identify the bilinear form $\mathbf{u}^T\mathbf{A}\mathbf{w}$ and a function $\psi(\mathbf{w},\mathbf{v},\mathbf{u})$ which is convex in $(\mathbf{w},\mathbf{v})$ and concave in $\mathbf{u}$. Thus, immediately, the corresponding (AO) problem becomes[2]:

$$\min_{\mathbf{w},\mathbf{v}} \max_{\mathbf{u}} -\|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} - \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} - \mathbf{u}^T\mathbf{z} + \mathbf{u}^T\mathbf{v} + \mathcal{L}(\mathbf{v}) + \lambda f(\mathbf{x}_0 + \mathbf{w}). \tag{5.7}$$

Now that we have identified the (AO) problem, we wish to apply Corollary 3.3.2 for the set $\mathcal{S}_\epsilon$ of (5.2). Applying the corollary amounts to analyzing the convergence of the (AO) problem (and that of its "restricted" counterpart). This will be performed in two stages. The first involves a deterministic analysis, in which the optimization in (5.7) is simplified and reduced to one which only involves scalar random variables. In the second stage, we analyze the convergence properties of this scalar optimization.

[1] A preliminary version of this idea first appeared in [TPH15], in which the authors analyzed the error performance of the Generalized-LASSO. We have extended the idea here to apply to any convex loss function $\mathcal{L}$.

[2] When compared to (3.11b) it is more convenient in (5.7) to write the two terms $\|\mathbf{w}\|\mathbf{g}^T\mathbf{u}$ and $\|\mathbf{u}\|\mathbf{h}^T\mathbf{w}$ with a minus sign instead. We can do this, since $\mathbf{g}$ and $\mathbf{h}$ are Gaussian vectors; thus, their distribution is sign independent.

Before proceeding with these, in all the above we have been silent regarding any compactness requirements of Theorem 3.3.1. These technicalities are carefully handled in Appendix B. (In particular, this is where Assumption 4.2.1(b) becomes useful.)

Scalarization of the (AO)

A key idea that facilitates the analysis of the (AO) in (5.7) is to reduce the optimization into one that only involves scalar optimization variables. The objective function of the (AO) is tailored towards this direction, and the only modification required is to express $f(\mathbf{x}_0+\mathbf{w})$ via its variational form as $\sup_{\mathbf{s}} \mathbf{s}^T(\mathbf{x}_0+\mathbf{w}) - f^*(\mathbf{s})$, where f^* is the Fenchel conjugate function.

This way, the variables $\mathbf{u}$ and $\mathbf{w}$ appear in the objective only through either linear terms or through their magnitudes. This observation suggests that one can easily optimize over their directions while fixing the magnitudes. To illustrate this, fixing the magnitude of $\mathbf{u}$ as $\|\mathbf{u}\|_2=\beta\geq 0$, we can optimize over its direction by aligning it with $-\|\mathbf{w}\|_2\mathbf{g}-\mathbf{z}+\mathbf{v}$. Then (5.7) simplifies to the following,

$$\min_{\mathbf{w},\mathbf{v}} \max_{\beta\geq 0,\mathbf{s}} \beta\|\|\mathbf{w}\|_2\mathbf{g}+\mathbf{z}+\mathbf{v}\|_2-\beta\mathbf{h}^T\mathbf{w}+\mathcal{L}(\mathbf{v})+\lambda\mathbf{s}^T(\mathbf{x}_0+\mathbf{w})-f^*(s). \tag{5.8}$$

Suppose we could switch the order of min-max above. Then it would be possible to do the same trick with $\mathbf{w}$, i.e. fix $\|\mathbf{w}\|_2=\alpha\geq 0$ and minimize over its direction to get

$$\max_{\beta\geq 0,\mathbf{s}} \min_{\alpha\geq 0,\mathbf{v}} \beta\|\alpha\mathbf{g}+\mathbf{z}+\mathbf{v}\|_2+\mathcal{L}(\mathbf{v})-\alpha\|\beta\mathbf{h}-\lambda\mathbf{s}\|_2+\lambda\mathbf{s}^T\mathbf{x}_0-f^*(s). \tag{5.9}$$

Justifying that flipping in the order of min-max is not straightforward since the objective function in (5.8) is *not* convex-concave; thus, what would otherwise be the arguments to be called upon, namely the Minimax Theorems (e.g. [Sio+58]), are not directly applicable here. Yet, in Appendix B, we show that such a minimax property holds asymptotically in the problem dimensions; thus, (5.9) is (for our purposes) equivalent to (5.8). We leave the details aside for the moment, and, proceed with the simplification of (5.9).

In (5.9), we have reduced the optimization over $\mathbf{w}$ and $\mathbf{u}$ to scalars α and β. Next, we wish to simplify the optimization over $\mathbf{s}$ and $\mathbf{v}$. However, the same trick as the one we applied for the former two variables won't work. The new idea that we need here is to write the terms $\|\|\mathbf{w}\|_2\mathbf{g}+\mathbf{z}+\mathbf{v}\|_2$ and $\|\beta\mathbf{h}-\lambda\mathbf{s}\|_2$ using

$$\|\mathbf{t}\|_2=\inf_{\tau>0}\ \frac{\tau}{2}+\frac{\|\mathbf{t}\|_2^2}{2\tau}.$$

What we achieve with this is that the corresponding terms become now *separable* over the entries of the vectors $\mathbf{v}$ and $\mathbf{s}$, which makes the optimization over them easier. The only price we have to pay is introducing just two more scalar optimization variables. That is (5.9) becomes

$$\sup_{\substack{\beta\geq 0\\ \tau_h>0}} \inf_{\substack{\alpha\geq 0\\ \tau_g>0}} \frac{\beta\tau_g}{2} + \min_{\mathbf{v}} \left\{ \frac{\beta}{2\tau_g}\|\alpha\mathbf{g}+\mathbf{z}+\mathbf{v}\|_2^2 + \mathcal{L}(\mathbf{v}) \right\} \\ - \frac{\alpha\tau_h}{2} - \min_{\mathbf{s}} \left\{ \frac{\alpha}{2\tau_h}\|\beta\mathbf{h}-\lambda\mathbf{s}\|_2^2 - \lambda\mathbf{s}^T\mathbf{x}_0 + f^*(s) \right\}.$$

It can be readily seen that the minimization over $\mathbf{v}$ gives rise to the Moreau envelope function of $\mathcal{L}$ evaluated at $\mathbf{z}+\alpha\mathbf{g}$ with index τ_g/β. A rather straightforward completion of squares arguments and a call upon the relation between the Moreau envelopes of conjugate pairs, leads to a similar conclusion regarding the minimization over $\mathbf{s}$, as well. Deferring the details to the appendix, we have reached the following scalar optimization

$$\sup_{\substack{\beta\geq 0\\ \tau_h>0}} \inf_{\substack{\alpha\geq 0\\ \tau_g>0}} \frac{\beta\tau_g}{2} + \mathrm{e}_{\mathcal{L}}\left(\alpha\mathbf{g}+\mathbf{z};\frac{\tau_g}{\beta}\right) - \frac{\alpha\tau_h}{2} - \frac{\alpha\beta^2}{2\tau_h}\|\mathbf{h}\|_2^2 + \lambda\cdot\mathrm{e}_f\left(\frac{\beta\alpha}{\tau_h}\mathbf{h}+\mathbf{x}_0;\frac{\alpha\lambda}{\tau_h}\right). \tag{5.10}$$

Convergence analysis of the (AO)

Once we have simplified the (AO), it is now possible to analyze the convergence of its optimal cost. We start with the objective function of (5.10), which we shall denote $\mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h)$ for convenience. Fix[3] $\alpha,\tau_g,\beta,\tau_h$, then,

$$\frac{1}{n}\mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h) \xrightarrow{P} \frac{\beta\tau_g}{2} + L\left(\alpha,\frac{\beta}{\tau_g}\right) - \frac{\alpha\tau_h}{2} - \frac{\alpha\beta^2}{2\tau_h} + \lambda\cdot F\left(\frac{\alpha\beta}{\tau_h},\frac{\alpha\lambda}{\tau_h}\right) =: \mathcal{D}(\alpha,\tau_g,\beta,\tau_h), \tag{5.11}$$

where we have assumed that L and F above are such that

$$\frac{1}{n}\mathrm{e}_{\mathcal{L}}\left(c\mathbf{g}+\mathbf{z};\tau\right) \xrightarrow{P} L(c,\tau) \qquad \text{and} \qquad \frac{1}{n}\mathrm{e}_f\left(c\mathbf{h}+\mathbf{x}_0;\tau\right) \xrightarrow{P} F(c,\tau).$$

This corresponds to Assumption 4.2.1(a), except that in the latter we have $\mathrm{e}_{\mathcal{L}}\left(c\mathbf{g}+\mathbf{z};\tau\right)-\mathcal{L}(\mathbf{z})$ instead of just $\mathrm{e}_{\mathcal{L}}\left(c\mathbf{g}+\mathbf{z};\tau\right)$, and similarly for f. The reason for this slight tweak is to account for noise vectors $\mathbf{z}$ with unbounded moments. For those, $\mathrm{e}_{\mathcal{L}}\left(c\mathbf{g}+\mathbf{z};\tau\right)$ might not converge, but $\mathrm{e}_{\mathcal{L}}\left(c\mathbf{g}+\mathbf{z};\tau\right)-\mathcal{L}(\mathbf{z})$ will. We handle these issues in the Appendix.

[3] To be precise, an appropriate rescaling is required here. See Appendix B.

Our next step is to use the *point-wise* convergence of (5.11) in order to prove the following result:

$$\inf_{\substack{\alpha\geq 0\\ \tau_g>0}}\sup_{\substack{\beta\geq 0\\ \tau_h>0}}\frac{1}{n}\mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h)\xrightarrow{P}\inf_{\substack{\alpha\geq 0\\ \tau_g>0}}\sup_{\substack{\beta\geq 0\\ \tau_h>0}}\mathcal{D}(\alpha,\tau_g,\beta,\tau_h)=:\overline{\phi}. \tag{5.12}$$

This statement is of course much stronger than the one in (5.11). The proof requires two main ingredients: (i) translating the point-wise convergence into a *uniform* one over compact sets, (ii) proving that $\mathcal{D}$ is level-bounded with respect to its arguments, thus, the sets of optimizers in (5.12) are bounded. For the first point, convexity turns out to be critical, while the latter can be shown if Assumption 4.2.2 holds.

Concluding

The analysis of the (AO) problem led us to (5.12). The same arguments also show that

$$\inf_{\substack{|\alpha-\alpha_*|\geq\epsilon\\ \tau_g>0}}\sup_{\substack{\beta\geq 0\\ \tau_h>0}}\frac{1}{n}\mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h)\xrightarrow{P}\inf_{\substack{|\alpha-\alpha_*|\geq\epsilon\\ \tau_g>0}}\sup_{\substack{\beta\geq 0\\ \tau_h>0}}\mathcal{D}(\alpha,\tau_g,\beta,\tau_h)=:\phi_{\mathcal{S}_\epsilon^c}. \tag{5.13}$$

Recall from Section 5.2 that the variable α plays the role of the magnitude of $\mathbf{w}$, hence the random optimization in the LHS of (5.13) corresponds to the restricted (AO) problem $\phi_{\mathcal{S}_\epsilon^c}(\mathbf{g},\mathbf{h})$ of Corollary 3.3.2. What remains for the corollary to apply is showing that $\overline{\phi}_{\mathcal{S}_\epsilon^c}>\overline{\phi}$. This follows by assumption of the theorem that the minimizer over α in the RHS of (5.12) is unique. Applying the corollary shows the desired and concludes the proof.

Chapter 6

SPECIFIC EXAMPLES

Theorem 4.2.1 establishes the asymptotic limit of the squared error of convex regularized M-estimators. As was detailed in Section 4.1, this family of estimators includes many popular instances, such as regularized least-squares (a.k.a. generalized-LASSO), the square-root LASSO, regularized LAD, and so on. When the general results of Chapter 4 are applied to specific problem instances, they yield simplified expressions for the error performance. Analyzing those, it is possible to answer a number of interesting questions, such as the following.

- What is the minimum number of measurements required for stable estimation? How does this number depend on regularization?
- What is a lower bound on the squared-error performance of M-estimators?
- Are there problem instances for which specific choices of loss and regularizer functions achieve the MMSE performance?

We provide answers to such questions for a number of popular estimators in Sections 6.1–6.9. Moreover, in Section 6.10 we present numerical simulation results that illustrate the validity of those theoretical predictions.

6.1 M-estimators without Regularization

Consider an M-estimator without regularization, i.e.,

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \sum_{j=1}^{m} \ell(\mathbf{y}_j - \mathbf{a}_j^T \mathbf{x}_j). \tag{6.1}$$

For simplicity, we consider $\mathbf{z}_j \overset{\text{iid}}{\sim} p_Z$ and a separable loss function. Assuming that ℓ and p_Z satisfy the assumptions of Theorem 4.3.1, and, noting that $f = 0 \implies F(c,\tau) = 0$, the squared error of (6.1) is predicted by the minimizer α_* of the following (SPO) problem

$$\inf_{\substack{\alpha\geq 0\\ \tau_g>0}} \sup_{\beta\geq 0} \frac{\beta\tau_g}{2} + \delta L(\alpha, \frac{\tau_g}{\beta}) - \alpha\beta, \tag{6.2}$$

where we have performed the (straightforward) optimization over τ_h: $\inf_{\tau_h>0} \frac{\tau_h}{2} + \frac{\beta^2}{2\tau_h} = \beta$. We may equivalently express α_* as the solution to the first-order optimality conditions of (6.2). In particular, the stationary equations (see (4.16)) simplify in this case to the following system of two equations in two unknowns:

$$\begin{cases} \alpha^2 = \delta\kappa^2 \mathbb{E}\left[\left(\mathrm{e}'_\ell(\alpha G + Z, \kappa)\right)^2\right], \\ \alpha = \delta\kappa \cdot \mathbb{E}\left[\mathrm{e}'_\ell(\alpha G + Z, \kappa) \cdot G\right]. \end{cases} \tag{6.3}$$

Starting from (6.3), some interesting conclusions can be drawn regarding the performance of M-estimators without regularization, which we gather in the following remarks.

Remark 6.1.0.33 (Stable recovery). It follows from (6.3) that in the absence of regularization it is required that the number of measurements m is at least as large as the dimension of the ambient space n ($\delta \geq 1$), in order for the recovery to be *stable*, i.e. the error be finite. To see this, assume stable recovery, then there exists (α_*, κ_*) satisfying (6.3). Starting from the second equation, applying the Cauchy-Schwarz inequality and substituting back the first equation we find:

$$\alpha_* = \delta\kappa_* \cdot \mathbb{E}\left[\mathrm{e}'_\ell(\alpha_* G + Z, \kappa_*) \cdot G\right] \leq \delta\kappa_* \cdot \sqrt{\mathbb{E}\left[\left(\mathrm{e}'_\ell(\alpha_* G + Z, \kappa_*)\right)^2\right]} = \delta\kappa_* \frac{\alpha_*}{\sqrt{\delta}\kappa_*},$$

from where it follows that $\delta \geq 1$.

Remark 6.1.0.34 (Stein's Formula). Assume e_ℓ is two times differentiable (e.g., this is the case if ℓ is two times differentiable). Then, applying Stein's formula (4.17), a simple rearrangement of (6.3) shows that

$$\alpha_*^2 = \frac{1}{\delta} \frac{\mathbb{E}\left[\left(\mathrm{e}'_\ell(\alpha_* G + Z, \kappa_*)\right)^2\right]}{\left(\mathbb{E}\left[\mathrm{e}''_\ell(\alpha_* G + Z, \kappa_*)\right]\right)^2}. \tag{6.4}$$

Remark 6.1.0.35 (Least-Squares). The simplest instance of the general M-estimator is the Least-squares, i.e. $\hat{\mathbf{x}} := \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$. Of course, in this case, $\hat{\mathbf{x}}$ has a closed form expression which can be directly used to predict the error behavior. However, for illustration purposes, we show how the same result can be also obtained from (6.3). This is also one of the few cases where α_* can be expressed in closed form. Assume $\delta > 1$ and $\mathbf{z}_j \overset{\text{iid}}{\sim} p_Z$ with bounded second moment, i.e. $0 < \mathbb{E}Z^2 = \sigma^2 < \infty$. Then, it can be readily checked that all assumptions hold for $\frac{1}{2}(\cdot)^2, p_z$. Also,

$e'_{\frac{1}{2}(\cdot)^2}(\chi;\tau) = \frac{\chi}{1+\tau}$ and $e''_{\frac{1}{2}(\cdot)^2}(\chi;\tau) = \frac{1}{1+\tau}$. Solving for the second equation in (6.3) gives $\kappa_* = \frac{1}{\delta-1}$. Substituting this into the first, we recover the well-known formula

$$\alpha_*^2 = \sigma^2 \frac{1}{\delta - 1}. \tag{6.5}$$

Remark 6.1.0.36 (Related work). M-estimators of the form (6.1), under the additional regularity assumptions of ℓ being strongly convex and smooth, had been previously analyzed by Donono & Montanari [DM13]; their proof technique is based on the AMP framework [DMM09]. In particular, the formula in (6.4) coincides with the corresponding expression in [DM13, Thm. 4.1]; only here, ℓ need not be smooth or strongly convex (and, in fact not necessarily separable).

6.2 Ridge Regularzation

A popular regularizer in the machine learning and statistics literature is the ridge regularizer (also known as Tikhonov regularizer), i.e.

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \sum_{j=1}^{m} \ell(\mathbf{y}_j - \mathbf{a}_j^T \mathbf{x}_j) + \lambda \frac{\|\mathbf{x}\|_2^2}{2}. \tag{6.6}$$

We specialize Theorem 4.2.1 to this case. For simplicity, we assume a separable loss function, and, $\mathbf{z}_j \overset{\text{iid}}{\sim} p_Z$ and $\mathbf{x}_{0,i} \overset{\text{iid}}{\sim} p_X$.

We will apply Theorem 4.3.1. Suppose that ℓ satisfies the assumptions. Also, assume $\mathbb{E}X_0^2 = \sigma_x^2 < \infty$. Then, for $f = \frac{1}{2}(\cdot)^2$, it is easily verified that $\mathbb{E}[(f'(cH + X_0))^2] = \mathbb{E}[(cH + X_0)^2] < \infty$. Hence, the squared-error of (6.6) is predicted by α_*, the unique minimizer to the (SPO) in (4.4) with

$$F(c,\tau) = \frac{c^2 + \sigma_x^2}{2(\tau+1)} - \sigma_x^2.$$

The first-order optimality conditions (see (4.16)) of this problem simplify after some algebra to the following two equations in two unknowns:

$$\begin{cases} \alpha^2 = \delta\kappa^2 \cdot \mathbb{E}\left[\mathrm{e}'_\ell(\alpha G + Z, \kappa)^2\right] + \lambda^2\kappa^2\sigma_x^2, \\ \alpha\,(1 - \lambda\kappa) = \delta\kappa \cdot \mathbb{E}\left[\mathrm{e}'_\ell(\alpha G + Z, \kappa) \cdot G\right]. \end{cases} \tag{6.7}$$

Remark 6.2.0.37 (Stein's Formula). Assume $\mathrm{prox}_\ell(x;\tau)$ is two times differentiable with respect to c (e.g., this is the case if ℓ is two times differentiable), and write $\mathrm{prox}'_\ell(x,\tau)$ for the derivative with respect to x. Applying (4.17), a simple rearrangement of (6.7) yields the following equivalent system of equations

$$\begin{cases} \delta - 1 + \kappa\lambda = \delta \cdot \mathbb{E}\left[\mathrm{prox}'_\ell(\alpha G + Z;\kappa)\right], \\ \alpha^2 = \delta\mathbb{E}\left[(\alpha G + Z - \mathrm{prox}_\ell(\alpha G + Z;\kappa))^2\right] + \lambda^2\kappa^2\sigma_x^2. \end{cases} \tag{6.8}$$

Remark 6.2.0.38 (Least-squares loss). Consider a least-squares loss function where $\ell(x) = \frac{1}{2}x^2$ and a noise distribution of variance $\mathbb{E}Z^2 = \sigma_z^2 < \infty$. Then $\text{prox}_\ell(x;\tau) = \frac{x}{1+\tau}$ and $\text{prox}'_\ell(x;\tau) = \frac{1}{1+\tau}$. Substituting in (6.8) gives

$$\begin{cases} 1 - \kappa\lambda = \dfrac{\delta\kappa}{1+\kappa}, \\ \alpha^2(1 - \delta \cdot \dfrac{\kappa^2}{(1+\kappa)^2}) = \delta \cdot \dfrac{\kappa^2}{(1+\kappa)^2}\sigma_z^2 + \lambda^2\kappa^2\sigma_x^2. \end{cases} \tag{6.9}$$

Now, we can solve these to get the following closed form expression for α^*:

$$\alpha^2 = \left(\delta \cdot \frac{\kappa^2}{(1+\kappa)^2} \cdot \sigma_z^2 + \lambda^2\sigma_x^2\kappa^2\right) \cdot \left(1 - \delta \cdot \frac{\kappa^2}{(1+\kappa)^2}\right)^{-1}, \tag{6.10}$$

where

$$\kappa = \frac{1 - \delta - \lambda + \sqrt{(1-\delta-\lambda)^2 + 4\lambda}}{2\lambda}. \tag{6.11}$$

Observe that letting $\lambda \to 0$ (which would correspond to ordinary least-squares) and assuming $\delta > 1$, κ in (6.11) approaches $1/(\delta - 1)$ and the optimal α^2 in (6.10) becomes $\sigma_z^2/(\delta - 1)$, which agrees with (6.5), as expected.

Remark 6.2.0.39 (Achieving the MMSE). Let a Gaussian input distribution $\mathbf{x}_{0,i} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and any noise distribution of power $\mathbb{E}Z^2 = \sigma_z^2 < \infty$. We show that a ridge-regularized M-estimator with a least-squares loss function and optimally tuned λ achieves asymptotically the Minimum Mean-Squared Error (MMSE) of estimating $\mathbf{x}_0$ from $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$.

First, we use the results of Remark 6.2.0.38 to calculate the achieved error of the M-estimator optimized over the values of the regularizer parameter:

$$o_* := \inf_{\lambda>0} \lim_{n\to\infty} \frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \inf_{\lambda>0} \left\{\alpha^2(\kappa(\lambda), \lambda) \text{ as in (6.10)} \mid \kappa(\lambda) \text{ satisfies (6.11)}\right\}. \tag{6.12}$$

The optimization over λ is possible as follows. From (6.9), we find

$$\delta\left(\frac{\kappa}{\kappa+1}\right)^2 = \frac{(1-\kappa\lambda)^2}{\delta}. \tag{6.13}$$

Substituting this in (6.10), and denoting $x = \kappa\lambda$, gives

$$\alpha^2 = \frac{\delta x^2 + \sigma^2(1-x)^2}{\delta - (1-x)^2}. \tag{6.14}$$

Minimizing α^2 over $\lambda > 0$ in (6.12) is equivalent to minimizing the fraction above over $0 < x < 1$, since there always exist κ, λ satisfying $x = \kappa\lambda$ and (6.13). Thus, performing the optimization over $0 < x < 1$ in (6.14) we find

$$o_* = \frac{1}{2}\left(1 - \sigma^2 - \delta + \sqrt{(1-\delta)^2 + 2\sigma^2(\delta+1) + \sigma^4}\right). \tag{6.15}$$

Next, Wu and Verdu have shown in [WV12, Thm. 8, Eqn. (56)] that the MMSE is given by the expression in the right-hand side above as well. This completes the proof of the claim.

Remark 6.2.0.40 (Related work). Ridge-regularized M-estimators have been also analyzed by El Karoui in [Kar13]. In particular, the formula in (6.8) coincides with the corresponding expression in [Kar13, Thm. 2.1][1]. The result in [Kar13] requires additional smoothness assumptions on ℓ. Our result holds under relaxed assumptions and has been derived as a corollary of Theorem 4.2.1. On the other hand, [Kar13, Thm. 2.1] is shown to be true for design matrices $\mathbf{A}$ with iid entries beyond Gaussian, e.g. sub-Gaussian.

6.3 Cone-constrained M-estimators

When introducing regularized M-estimators in (4.1) we captured prior knowledge on the structure (or distribution) of the unknown signal $\mathbf{x}_0$ via the regularizer function f. It might be the case, depending on the application, that access to this information is instead provided in the form of $\mathbf{x}_0$ belonging to some set $C \subset \mathbb{R}^n$. Then, it is natural to obtain an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ by solving

$$\min_{\mathbf{x}\in C} \sum_{j=1}^{m} \ell(\mathbf{y}_j - \mathbf{a}_j^T \mathbf{x}). \tag{6.16}$$

We call such an estimator a *constrained* M-estimator. Typically, C can be described in the form $C = \{\mathbf{x} \mid g(\mathbf{x}) \leq c\}$ for some function g and a constant $c \in \mathbb{R}$. Henceforth, we assume that g is convex, and so the optimization above is convex. Note then that there exists by Lagrangian duality a value of λ for which the regularized M-estimator with $f(x) = g(x)$ is equivalent to (6.16).

The analysis framework of Chapter 5 is readily applicable to constrained M-estimators and can lead to a statement similar to Theorem 4.3.1 regarding constrained M-estimators. Instead, here we rather focus on a "benchmark analysis" of (6.16) as described below.

[1] In comparing (6.8) to [Kar13, Eqn. (4)], due to some differences in normalizations the following "dictionary" needs to be used to match the results: $\alpha \leftrightarrow r_\rho(\kappa)$, $\kappa \leftrightarrow c_\rho(\kappa)$, $\delta^{-1}\lambda \leftrightarrow \tau$ and $\delta^{-1} \leftrightarrow \kappa$.

We assume additional prior information on $\mathbf{x}_0$ via knowledge of the value $g(\mathbf{x}_0)$, in which case $C = \{\mathbf{x} \mid g(\mathbf{x}) \leq g(\mathbf{x}_0)\}$, i.e. the set of descent directions of g at $\mathbf{x}_0$. In some sense (which in cases can be made precise, e.g. see Chapter 7), this additional prior information corresponds to a genie choosing the optimal value of the regularizer parameter in the corresponding regularized version.

Furthermore, we relax the constraint by substituting C with its conic hull, the tangent cone of g at $\mathbf{x}_0$ (cf. Defn. 2.2.1). We call the resulting program, a *cone-constrained* M-estimator. Essentially, this corresponds to studying the performance of 6.16 in the regime of "high-SNR". Intuitively, this is the case as follows: At high-SNR, we expect the solution $\hat{\mathbf{x}}$ to be in the close neighborhood of the true signal $\mathbf{x}_0$, and inside this small neighborhood the tangent cone is a good approximation of the cone of descent directions (see Figure 5.1 for an illustration). This intuition is made precise in Chapter 7.

In what follows we analyze the squared-error performance of cone-constrained M-estimators. The results are insightful and yield connections to the theory of noiseless linear inverse problems presented in Section 2.2. For the special case of a least-squares loss function, a more complete analysis along with further discussions on the relevance of cone-constrained estimators to the more realistic regularized and constrained versions is given in Chapter 7.

Error Performance

We consider

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}\in C} \sum_{j=1}^{m} \ell(\mathbf{y}_j - \mathbf{a}_j^T\mathbf{x}), \tag{6.17}$$

where

$$C = \mathcal{T}_g(\mathbf{x}_0) + \mathbf{x}_0 := \{\lambda\mathbf{h} \mid \lambda \geq 0, g(\mathbf{x}_0 + \mathbf{h}) \leq g(\mathbf{x}_0)\} + \mathbf{x}_0$$

and g a proper, closed, convex function. Here, $\mathcal{T}_g(\mathbf{x}_0)$ is the tangent cone of g at $\mathbf{x}_0$, which is assumed fixed. The constrained minimization above can be written in the general form of regularized M-estimators in (4.1) by choosing the regularizer to be the indicator function for the cone, i.e. $f(\mathbf{x}) = \boldsymbol{\delta}_{\{\mathbf{x}\in C\}}$ [2]. Recall that $\text{dist}(\mathbf{v}, C)$ denotes the distance of a vector $\mathbf{v}$ to a set C. We have,

$$\mathrm{e}_{\boldsymbol{\delta}_{\{\mathbf{x}\in C\}}}(c\mathbf{h} + \mathbf{x}_0; \tau) = \frac{1}{2\tau} \min_{\mathbf{v}\in C - \mathbf{x}_0} \|c\mathbf{h} - \mathbf{v}\|_2^2 = \frac{1}{2\tau}\text{dist}^2\left(c\mathbf{h}, \mathcal{T}_g(\mathbf{x}_0)\right) = \frac{c^2}{2\tau}\text{dist}^2\left(\mathbf{h}, \mathcal{T}_g(\mathbf{x}_0)\right).$$

[2] Note that this is a non-separable regularizer function.

In the last equality above we have used the homogeneity of the cone $\mathcal{T}_g(\mathbf{x}_0)$. Let $(\mathcal{T}_g(\mathbf{x}_0))^\circ$ denote the polar cone of $\mathcal{T}_g(\mathbf{x}_0)$, and

$$\mathbf{D}(\text{cone}(\partial g(\mathbf{x}_0))) := \mathbb{E}\left[\text{dist}^2\left(\mathbf{h}, (\mathcal{T}_g(\mathbf{x}_0))^\circ\right)\right] = \mathbb{E}\left[\|\mathbf{h}\|_2^2 - \text{dist}^2\left(\mathbf{h}, \mathcal{T}_g(\mathbf{x}_0)\right)\right].$$

Recall from Section 2.2 that this is the *Gaussian distance squared* to the cone of subdifferential. Here, we assume that

$$\frac{\mathbf{D}(\text{cone}(\partial g(\mathbf{x}_0)))}{n} \to \overline{\mathbf{D}}_{g,\mathbf{x}_0} \in (0,1). \tag{6.18}$$

This translates to an assumption on the degrees of freedom of the structured signal $\mathbf{x}_0$ being proportional to its dimension. For example, for a k-sparse $\mathbf{x}_0$ and $g(\mathbf{x}) = \|\mathbf{x}\|_1$, it can be readily checked by inspection of (2.14) that (6.18) is satisfied in the linear regime of sparsity: $k = \rho n$, $\rho \in (0,1)$.

With (6.18), Assumption 4.2.1(a) holds with $F(c,\tau) = \frac{c^2}{2\tau}(1 - \overline{\mathbf{D}}_{g,\mathbf{x}_0})$. For this, it is straightforward to check that Assumption 4.2.2(a) is also satisfied. Overall, if ℓ, p_Z satisfy the conditions of Theorem 4.3.1 and $g, \mathbf{x}_0$ are such that (6.18) holds, then Theorem 4.2.1 applies. Then, the squared error of the cone-constrained M-estimator in (6.17) is predicted by the unique minimizer α_* of the (SPO) problem below:

$$\inf_{\substack{\alpha\geq 0\\ \tau_g>0}} \sup_{\beta\geq 0} \frac{\beta\tau_g}{2} + \delta\cdot\mathbb{E}\left[\mathrm{e}_\ell\left(\alpha G + Z; \tau_g/\beta\right) - \ell(Z)\right] - \alpha\beta\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}. \tag{6.19}$$

Compared to (4.4), we have performed the (straightforward) optimization over τ_h: $\inf_{\tau_h>0} \frac{\tau_h}{2} + \frac{\beta^2\overline{\mathbf{D}}_{g,\mathbf{x}_0}^2}{2\tau_h} = \beta\overline{\mathbf{D}}_{g,\mathbf{x}_0}$.

Remarks

Remark 6.3.0.41 (Stable recovery). Starting from (6.19) we can conclude on the minimum number of measurements required for stable recovery. We show that the normalized number of measurements δ need to be at least as large as $\overline{\mathbf{D}}_{g,\mathbf{x}_0}$, in order for the error to be finite. This is to be compared with the case where no regularization is used, which required $\delta \geq 1 > \overline{\mathbf{D}}_{g,\mathbf{x}_0}$ (see Remark 6.1.0.33). To prove the claim, assume finite error, then the value where it converges is predicted

by (6.19). Standard first-order optimality conditions give[3]

$$\beta - \frac{\delta}{\beta}\mathbb{E}\left[\left(\mathrm{e}'_\ell(\alpha G + Z;\tau_g/\beta)\right)^2\right] \geq 0, \tag{6.20a}$$

$$\delta\mathbb{E}[\mathrm{e}'_\ell(\alpha G + Z;\tau_g/\beta)\cdot G] - \beta\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}} \geq 0, \tag{6.20b}$$

$$\frac{\tau}{2} + \frac{\delta\tau}{2\beta^2}\mathbb{E}\left[\left(\mathrm{e}'_\ell(\alpha G + Z;\tau_g/\beta)\right)^2\right] - \alpha\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}} \leq 0. \tag{6.20c}$$

Starting from the second equation, applying the Cauchy-Schwarz inequality and substituting back the first equation we conclude as follows:

$$\beta\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}} \leq \delta\mathbb{E}[\mathrm{e}_\ell(\alpha G + Z;\tau_g/\beta)\cdot G] \leq \delta\sqrt{\mathbb{E}[\left(\mathrm{e}'_\ell(\alpha G + Z;\tau_g/\beta)\right)^2]} \leq \delta\frac{\beta}{\sqrt{\delta}} \Rightarrow \delta \geq \overline{\mathbf{D}}_{g,\mathbf{x}_0}.$$

Remark 6.3.0.42. (Least-squares loss) Consider a least-squares loss function and a noise distribution of variance $\mathbb{E}Z^2 = \sigma^2 < \infty$. Then, the solution to (6.19) admits an insightful closed form expression. First, in (6.19) perform the optimization over τ_g. Equating (6.20a) to 0, gives $\tau_g = \sqrt{\delta}\sqrt{\alpha^2+\sigma^2} - \beta$. Substituting this in (6.19), we are left to solve for

$$\inf_{\alpha\geq 0}\sup_{\beta\geq 0}\beta\left(\sqrt{\delta}\sqrt{\alpha^2+\sigma^2} - \alpha\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}\right) - \frac{\beta^2}{2}.$$

It can be easily checked that if $\delta > \overline{\mathbf{D}}_{g,\mathbf{x}_0}$, then the optimal α_* is

$$\alpha_*^2 = \sigma^2\frac{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}{\delta - \overline{\mathbf{D}}_{g,\mathbf{x}_0}}. \tag{6.21}$$

It is insightful to compare this with (6.5), the corresponding error formula for least-squares: the only difference is that 1 is substituted with the statistical dimension $\overline{\mathbf{D}}_{g,\mathbf{x}_0}$. Also, verifying the conclusion of the previous remark, we now require $\delta > \overline{\mathbf{D}}_{g,\mathbf{x}_0}$ instead of $\delta > 1$, implying that *robust* recovery is in general possible with fewer measurements than the dimension of the signal. In Chapter 7 we obtain a *non-asymptotic* version of (6.21).

Remark 6.3.0.43. (Lower Bound) In (6.20b) apply Stein's inequality and combine it with (6.20a) to yield

$$\alpha^2 \geq \frac{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}{\delta}\frac{\beta^2/\delta}{\mathbb{E}\left[\mathrm{e}''_\ell(\alpha G + Z;\tau_g/\beta)\right]} \geq \frac{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}{\delta}\frac{\mathbb{E}\left[\left(\mathrm{e}'_\ell(\alpha G + Z;\tau_g/\beta)\right)^2\right]}{\mathbb{E}\left[\mathrm{e}''_\ell(\alpha G + Z;\tau_g/\beta)\right]}. \tag{6.22}$$

[3] The three equations in (6.20) correspond to differentiation of the objective of (6.19) with respect to τ, α and β, respectively. If any of the variables is zero at the optimal, then, the corresponding equation holding with an inequality is necessary and sufficient. On the other hand, if the optimal is strictly positive, then the equation should hold with equality.

For the first inequality above, we have assumed that at the optimal, $\mathbb{E}\left[\mathrm{e}''_\ell(\alpha G + Z; \tau_g/\beta)\right] < \infty$. When this holds, (see Remark 6.3.0.44 for an instance where this is not the case) we can use the above to lower bound the error performance in terms of the Fisher information of the noise. Based on a result of [MB07], Donoho and Montanari prove in [DM13, Lem. 3.4,3.5] that the right-hand side in (6.22) is further lower bounded by $I(Z)/(1 + \alpha^2 I(Z))$, where $I(Z) = \mathbb{E}\left(\frac{\partial}{\partial z} \log p_Z(z)\right)^2$ denotes the Fisher information of the random variable Z, which is assumed to have a differentiable density. Using this and solving for α^2, we conclude with

$$\alpha^2 \geq \frac{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}{\delta - \overline{\mathbf{D}}_{g,\mathbf{x}_0}} \frac{1}{I(Z)}. \tag{6.23}$$

For Gaussian noise of variance σ^2, we have $1/I(Z) = \sigma^2$. In this case the lower bound in (6.23) coincides with the error formula of the least-squares loss function, which then proves optimality of the latter.

Remark 6.3.0.44. (Consistent Estimators) The lower bound in (6.23) only holds if the optimal α_* in (6.19) is strictly positive. This is not always the case: under some circumstances, it is possible to choose the loss function such that the resulting cone-constrained M-estimator is consistent. Theorem 4.2.1 is the starting point to identifying such interesting scenarios.

Here, we illustrate this through an example: we assume a sparse Gaussian-noise model and use a Least Absolute Deviations (LAD) loss function. More precisely, $p_Z(Z) = \bar{s}\delta_0(Z) + (1 - \bar{s})\frac{1}{\sqrt{2\pi}} \exp(-Z^2/2)$, $\bar{s} \in (0, 1)$ and $\ell(v) = |v|$. In Section C.1 we prove that when $\bar{s}, \delta$ and $\overline{\mathbf{D}}_{g,\mathbf{x}_0}$ are such that

$$\delta \geq \overline{\mathbf{D}}_{g,\mathbf{x}_0} + \min_{\kappa > 0} \left\{ \bar{s}(1 + \kappa^2) + (\delta - \bar{s})\sqrt{\frac{2}{\pi}} \int_\kappa^\infty (G - \kappa)^2 \exp(-G^2/2) \mathrm{d}G \right\}, \tag{6.24}$$

then the first-order optimality conditions in (6.20) are satisfied for $\alpha \to 0, \tau_g \to 0$ and some $\beta > 0$. Thus, when the number of measurements is large enough such that (6.24) holds, then $\alpha_* = 0$, and, $\mathbf{x}_0$ is perfectly recovered[4]. See Figure 6.1 for

[4]The problem is very closely related to the demixing problem in which one aims to extract two (or more) constituents from a mixture of structured vectors [McC+14]. In that context, recovery conditions like the one in (6.24) have been generalized to other kinds of structures beyond sparsity [MT14; McC+14; FM14]. Our purpose here has been to illustrate how Theorem 4.2.1 can be used to derive such results. Besides, the generality of the paper's setup offers the potential of extending such consistency-type results beyond cone-constrained M-estimators and beyond fixed signals $\mathbf{x}_0$. This is an interesting direction of future research.

an illustration. The vertical dashed line corresponds to the the sparsity level $\bar{s}$ for which (6.24) holds with equality. The estimation error of the LAD is zero below that level, as predicted by (6.24).

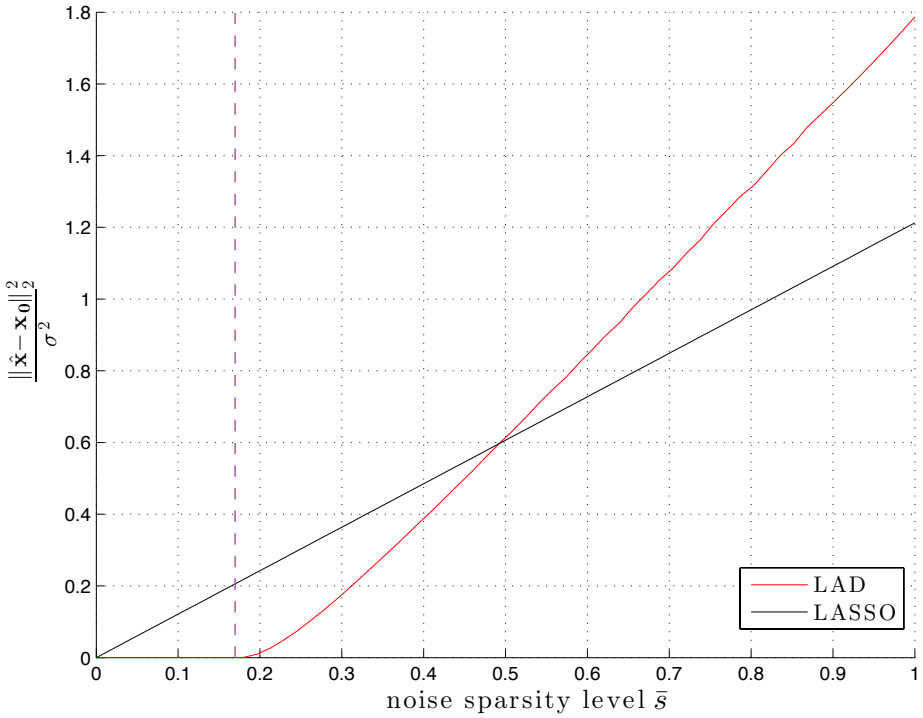

Figure 6.1: Using the predictions of Theorem 4.2.1 to analytically compare the performance of different instances of M-estimators. Here, we compare a least-absolute deviations (LAD) to a least-squares (LASSO) loss function in (6.16) for sparse signal estimation under sparse noise. The normalized squared error is plotted as a function of the sparsity-level $\bar{s}$ of the noise at the high-SNR regime. The noise is sparse with sparsity level $\bar{s}$ and nonzero entries are i.i.d $\mathcal{N}(0, \sigma^2)$ and $\sigma^2 \to 0$. Also, the sparsity level of the unknown signal is fixed to be 0.1 and the normalized number of measurements is $\delta = 3/5$.

Remark 6.3.0.45 (LASSO vs LAD). The precise error predictions can be used to analytically and accurately compare the performance between different instances of M-estimators. For an illustration, we may use the results of this section to compare the squared error performance of a least-absolute deviations (LAD) loss function to a least-squares (LASSO) loss function. We assume sparse noise and we use (6.16) with $C = \{\mathbf{x} \mid g(\mathbf{x}) \leq g(\mathbf{x}_0)\}$. Let $\bar{s}$ denote the noise sparsity and the non-zero entries of the noise vector be iid $\mathcal{N}(0, \sigma^2)$. Figure 6.1 compares the normalized-squared error (NSE) performance $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2/\sigma^2$ of the LAD to that of the LASSO

at high-SNR (i.e., $\sigma^2 \to 0$), when $\mathbf{x}_0$ is sparse and $g(\mathbf{x}) = \|\mathbf{x}\|_1$. As discussed earlier (please see also Section 7.4), the NSE at high-SNR is an upper bound on the NSE at arbitrary values of the SNR. Also, at high-SNR, the performance of (6.16) is equivalent to that of the cone-constrained M-estimator in (6.17). The predictions follow from (6.19). More precisely, for the least-squares loss function, a slight modification of the derivations presented in Remark 6.3.0.42 to account for the sparsity of the noise shows that the NSE behaves as $\bar{s}\frac{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}{\delta-\overline{\mathbf{D}}_{g,\mathbf{x}_0}}$ (compare to (6.21)). This is plotted in black in Figure 6.1. On the other hand, the NSE of the LAD is plotted in red. The corresponding formula has been derived in [TH14, Thm. 3.1]; we refer the interested reader to the original reference for the details.

It is seen that the LAD method outperforms the LASSO when the noise sparsity level is up to around 50%. For less sparse noise vectors, the LASSO error is smaller. The dashed vertical line identifies the sparsity level $\bar{s}$ for which (6.24) holds with equality. It was shown analytically in Remark 6.3.0.44 that the estimation error of the LAD is zero below that sparsity level. Thus, in this regime of very sparse noise the LAD significantly outperforms the LASSO.

It is interesting to evaluate how the two methods compare in the other extreme of non-sparse noise (i.e. $\bar{s} = 1$). Starting from (6.19) it can be shown (the interested reader is referred to [TH14, Cor. 3.2] for the detailed derivations) that when $\bar{s} = 1$, then the NSE of the LAD at high-SNR behaves as $\frac{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}{\delta-\overline{\mathbf{D}}_{g,\mathbf{x}_0}-\delta\omega(\overline{\mathbf{D}}_{g,\mathbf{x}_0}/\delta)} - 1$, where $\omega(\eta) := 2(1-\eta)\phi^2(\eta) - \frac{2}{\sqrt{\pi}}\phi(\eta)e^{-\phi^2(\eta)} - \eta + 1$ for all $\eta \in (0,1)$, and ϕ satisfies $\eta = \frac{2}{\sqrt{\pi}}\int_0^{\phi(\eta)} e^{-t^2/2}\mathrm{d}t$. Hence, when compared to (6.21), it can be shown that the NSE of the LAD is larger than the NSE of the LASSO by no more than $\pi/2$ times for all values of $\overline{\mathbf{D}}_{g,\mathbf{x}_0} \in (0,1)$ [5].

6.4 Generalized-LASSO

The generalized LASSO solves

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda f(\mathbf{x}). \tag{6.25}$$

For simplicity, suppose that f is separable and satisfies the assumptions of Theorem 4.3.1. Also, assume $\mathbf{z}_j \overset{\text{iid}}{\sim} p_Z$ such that $0 < \mathbb{E}Z^2 =: \sigma^2 < \infty$. Then, for $\ell = \frac{1}{2}(\cdot)^2$, it

[5] Results similar to the ones discussed here can be interpreted as extensions of corresponding results in classical statistics in which n is assumed fixed and no regularization is used. For example, it is interesting to compare our conclusion that when the noise is Gaussian then the penalized-LAD is no more than $\pi/2$ times worse than the generalized-LASSO, with a similar result in [CM73, pp. 839], which corresponds to no regularization and $\delta \to \infty$.

is easily verified that $\mathbb{E}[(\ell'(cG+Z))^2] = \mathbb{E}[(cG+Z)^2] < \infty$. Hence, the squared-error of (6.25) is predicted by α_*, the unique minimizer to the (SPO) in (4.4) with $L(c,\tau) = \frac{c^2+\sigma^2}{2(\tau+1)} - \sigma^2$.

Equivalently, the error is predicted by the solution to the stationary equations in (4.16) with $e'_{\frac{1}{2}(\cdot)^2}(\chi;\tau) = \frac{\chi}{1+\tau}$. The second and third equations in (4.16) give

$$\begin{aligned}\beta^2(1+\kappa)^2 &= \delta(\alpha^2+\sigma^2),\\ \nu(1+\kappa) &= \delta.\end{aligned}$$

Solving these for κ and ν, and substituting them in the remaining two equations results in the following system of two nonlinear equations in two unknowns

$$\begin{cases}\delta\frac{\alpha^2}{\alpha^2+\sigma^2} = \mathbb{E}\left[\left(\frac{\lambda}{\beta}e'_f\left(\frac{\sqrt{\alpha^2+\sigma^2}}{\sqrt{\delta}}H + X_0, \lambda\frac{\sqrt{\alpha^2+\sigma^2}}{\beta\sqrt{\delta}}\right) - H\right)^2\right]\\ \beta(1-\delta) + \beta^2\frac{\sqrt{\delta}}{\sqrt{\alpha^2+\sigma^2}} = \lambda\mathbb{E}\left[e'_f\left(\frac{\sqrt{\alpha^2+\sigma^2}}{\sqrt{\delta}}H + X_0, \lambda\frac{\sqrt{\alpha^2+\sigma^2}}{\beta\sqrt{\delta}}\right)\cdot H\right].\end{cases} \tag{6.26}$$

For the special case of ℓ_1-regularization, the result above was proved by Bayati and Montanari [BM12] using the AMP framework. In the generality presented here, the result appears to be novel.

Remark 6.4.0.46. (Not consistent) An interesting observation from (6.26) is that the generalized LASSO cannot achieve perfect recovery, irrespective of the choice of the regularizer function. To see this, the first equation in (6.26) for $\alpha = 0$ gives $\mathbb{E}\left[\left(\frac{\lambda}{\beta}e'_f(\frac{\sigma}{\sqrt{\delta}}H + X_0, \frac{\lambda\sigma}{\beta\sqrt{\delta}}) - H\right)^2\right] = 0$. Then, it must hold, almost surely, that the argument under the expectation sign be equal to zero. Evaluating the derivative of the envelope function as in Lemma B.4.1(iii), this becomes equivalent to $X_0 = \operatorname{prox}_f\left(\frac{\sigma}{\sqrt{\delta}}H + X_0; \frac{\lambda\sigma}{\beta\sqrt{\delta}}\right)$. This, when combined with the optimality conditions for the Moreau envelope (see (B.86)), gives that almost surely $\frac{\sigma}{\sqrt{\delta}}H \in \partial f(X_0)$. Thus, we have reached a contradiction because H can take any real value as a Gaussian random variable.

6.5 Square-root LASSO

The generalized Square-root LASSO

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \sqrt{n}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x}). \tag{6.27}$$

In contrast, to the other examples in this section, the square-root LASSO is an instance of (4.1) with a *non*-separable loss function. Observe the normalization

of the loss function with a $\sqrt{n}$-factor. This is to satisfy our condition that $(\forall c > 0)(\exists C > 0)\ \left[\|\mathbf{v}\|_2 \le c\sqrt{n} \implies \frac{1}{\sqrt{n}} \sup_{\mathbf{s}\in\partial\mathcal{L}(\mathbf{v})} \|\mathbf{s}\|_2 \le C\right]$.

In Appendix C we show that when $\mathcal{L}(\mathbf{v}) = \sqrt{n}\|\mathbf{v}\|_2$ and $\mathbf{z} \sim p_{\mathbf{z}}$ with $\mathbb{E}\left[\|\mathbf{z}\|_2^2/m\right] = \sigma^2 \in (0,\infty)$, then Assumption 4.2.1(a) holds with

$$L(\alpha,\tau) = \begin{cases} \frac{1}{\sqrt{\delta}}(\sqrt{\alpha^2+\sigma^2}-\sigma) - \frac{\tau}{2\delta} & \text{, if } \sqrt{\delta}\sqrt{\alpha^2+\sigma^2} \ge \tau, \\ \frac{1}{2\tau}(\alpha^2+\sigma^2) - \frac{\sigma}{\sqrt{\delta}} & \text{, otherwise.} \end{cases} \tag{6.28}$$

Also, Assumption 4.2.1(b) is trivially satisfied and Section C.2 shows the same for Assumptions 4.2.2(b)-(d). Thus, considering any regularizer that satisfies Assumptions 4.2.1(a) and 4.2.2(a), Theorem 4.2.1 applies, and predicts the squared error of (6.27) as the unique minimizer α_* to the following optimization:

$$\inf_{\alpha\ge 0} \sup_{\substack{\beta\ge 0 \\ \tau_h>0}} -\frac{\alpha\tau_h}{2} - \frac{\alpha\beta^2}{2\tau_h} + \lambda\cdot F\left(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h}\right) + \begin{cases} \beta\sqrt{\delta}\sqrt{\alpha^2+\sigma^2} & \text{, if } \beta \le 1 \\ \sqrt{\delta}\sqrt{\alpha^2+\sigma^2} & \text{, otherwise} \end{cases}. \tag{6.29}$$

To arrive to (6.29) starting from (4.4), we have replaced L with (6.28) and have performed the minimization over τ_g as shown below:

$$\inf_{\tau_g\ge 0} \begin{cases} \frac{\beta\tau_g}{2} - \frac{\tau_g}{2\beta} + \sqrt{\delta}\sqrt{\alpha^2+\sigma^2} & \text{, if } \delta(\alpha^2+\sigma^2) \ge \frac{\tau_g^2}{\beta^2} \\ \frac{\beta\delta}{2\tau_g}(\alpha^2+\sigma^2) + \frac{\beta\tau_g}{2} & \text{, otherwise.} \end{cases} = \begin{cases} \beta\sqrt{\delta}\sqrt{\alpha^2+\sigma^2} & \text{, if } \beta \le 1 \\ \sqrt{\delta}\sqrt{\alpha^2+\sigma^2} & \text{, otherwise} \end{cases}. \tag{6.30}$$

The optimization in (6.30) can be simplified one step further. It is shown in Appendix C that $-\frac{\alpha\beta^2}{2\tau_h} + \lambda F\left(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h}\right)$ is a non-increasing function of β for $\beta > 0$. Therefore, the (SPO) becomes equivalent to the following

$$\inf_{\alpha\ge 0} \sup_{\substack{0\le\beta\le 1 \\ \tau_h\ge 0}} \beta\sqrt{\delta}\sqrt{\alpha^2+\sigma^2} - \frac{\alpha\tau_h}{2} - \frac{\alpha\beta^2}{2\tau_h} + \lambda\cdot F\left(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h}\right). \tag{6.31}$$

In the next sections, we specialize the result to the cases of sparse, group-sparse and low-rank signal recovery.

6.6 Sparse Recovery via the LASSO

Assume each entry $\mathbf{x}_{0,i}, i = 1,\dots,n$ is sampled i.i.d. from a distribution

$$p_{X_0}(x) = (1-\rho)\cdot\delta_0(x) + \rho\cdot q_0(x), \tag{6.32}$$

where δ_0 is the delta Dirac function, $\rho \in (0,1)$ and q_{X_0} a probability density function with second moment normalized to $1/\rho$ so that (without loss of generality):

$$n^{-1}\|\mathbf{x}_0\|_2^2 \xrightarrow{P} \sigma_x^2 = 1. \tag{6.33}$$

Then, $\mathbf{x}_0$ is ρn-sparse on average and we solve (6.27) with ℓ_1-regularization. The Fenchel's conjugate of the ℓ_1-norm is simply the indicator function of the ℓ_∞ unit ball. Hence, without much effort, for $c_1, c_2 \in \mathbb{R}$ and $\tau > 0$,

$$\begin{aligned} e_{f^*}(c_1\mathbf{h} + c_2\mathbf{x}_0; \tau) &= \frac{1}{2\tau}\sum_{i=1}^{n} \min_{|v_i|\le 1} (v_i - (c_1\mathbf{h}_i + c_2\mathbf{x}_{0,i}))^2 \\ &= \frac{1}{2\tau}\sum_{i=1}^{n} \eta^2(c_1\mathbf{h}_i + c_2\mathbf{x}_{0,i}; 1), \end{aligned} \tag{6.34}$$

where we have denoted

$$\eta(x;\tau') := (x/|x|)\,(|x| - \tau')_+ \tag{6.35}$$

for the soft thresholding operator with parameter $\tau' > 0$. By Lemma B.2.5 it follows then that

$$\begin{aligned} e_f(\mathbf{x}_0 + c\mathbf{h}; \tau) &= \frac{\|\mathbf{x}_0 + c\mathbf{h}\|_2^2}{2\tau} - \frac{\tau}{2}\sum_{i=1}^{n} \eta^2(\frac{1}{\tau}\mathbf{x}_{0,i} + \frac{c}{\tau}\mathbf{h}_i; 1) \\ &= \frac{\|\mathbf{x}_0 + c\mathbf{h}\|_2^2}{2\tau} - \frac{1}{2\tau}\sum_{i=1}^{n} \eta^2(\mathbf{x}_{0,i} + c\mathbf{h}_i; \tau). \end{aligned}$$

Consequently, an application of the weak law of large numbers shows that Assumption 4.2.1 is satisfied for

$$F(c,\tau) = \frac{1}{2\tau} + \frac{c^2}{2\tau} - \frac{1}{2\tau}\mathbb{E}[\eta^2(X_0 + cH; \tau)] \tag{6.36}$$

where the expectation is over $h \sim \mathcal{N}(0,1)$ and $X_0 \sim p_{X_0}$. F above further satisfies[6] Assumption 4.2.2. Thus, substituting this in (6.31) yields a prediction of the squared error as the solution to the following minimization problem:

$$\inf_{\alpha \ge 0} \sup_{\substack{0 \le \beta \le 1 \\ \tau_h \ge 0}} \beta\sqrt{\delta}\sqrt{\alpha^2 + \sigma^2} - \frac{\alpha\tau_h}{2} + \frac{\tau_h}{2\alpha} - \frac{\alpha}{2\tau}\mathbb{E}[\eta^2(\frac{\tau}{\alpha}X_0 + \beta H; \lambda)]. \tag{6.37}$$

We have appleed extra effort in order to obtain the following equivalent but more insightful characterization of the error, as stated below. The result follows by analyzing and massaging the first-order optimality conditions of (6.37); see Appendix C.3 for a proof.

[6] This can be readily checked from (6.36), but note, here, f is separable and $\mathbf{x}_0$ is distributed iid; Thus one can more easily check the more primitive conditions of Section 4.3.

Theorem 6.6.1 (Sparse Recovery with Square-root LASSO)**.** *If $\delta > 1$, then define $\lambda_{crit} = 0$. Otherwise, let $\lambda_{crit}, \kappa_{crit}$ be the unique pair of solutions to the following set of equations:*

$$\begin{cases} \kappa^2\delta = \sigma^2 + \mathbb{E}\left[(\eta(\kappa H + X_0; \kappa\lambda) - X_0)^2\right], & (6.38) \\ \kappa\delta = \mathbb{E}[(\eta(\kappa H + X_0; \kappa\lambda) \cdot h)], & (6.39) \end{cases}$$

where $h \sim \mathcal{N}(0,1)$ and is independent of $X_0 \sim p_{X_0}$ (cf. (6.32)). Then, for any $\lambda > 0$, with probability one,

$$\lim_{n\to\infty} \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \begin{cases} \delta\kappa_{crit}^2 - \sigma^2 & , \lambda \le \lambda_{crit}, \\ \delta\kappa_*^2(\lambda) - \sigma^2 & , \lambda \ge \lambda_{crit}, \end{cases}$$

where $\kappa_*^2(\lambda)$ is the unique solution to (6.38).

Figure 11.2 validates the prediction of the theorem for q_{X_0} being Gaussian[7]. Later in Figure 11.1 we present simulation results for q_{X_0} distributed iid Bernoulli. For the case of compressed ($\delta < 1$) measurements, observe the two different regimes of operation, one for $\lambda \le \lambda_{\text{crit}}$ and the other for $\lambda \ge \lambda_{\text{crit}}$, precisely as they are predicted by the theorem. A further detailed discussion on the distinct regions of operation of the square-root LASSO and their role is included in Section 7.6.

Remark 6.6.0.47. This is the first precise analysis result for the ℓ_2-LASSO stated in that generality. An analogous result, but via different analysis tools, has only been known before for the ℓ_2^2-LASSO as it appears in [BM12].

6.7 Group-Sparse Recovery via the Group-LASSO

Let $\mathbf{x}_0 \in \mathbb{R}^n$ be composed of t non-overlapping blocks of constant size b each such that $n = t \cdot b$. Each block $[\mathbf{x}_0]_i, i = 1, \dots, t$ is sampled i.i.d. from a probability density in $\mathbb{R}^b$: $p_{X_0}(\mathbf{x}) = (1-\rho) \cdot \delta_0(\mathbf{x}) + \rho \cdot q_{X_0}(\mathbf{x}), \mathbf{x} \in \mathbb{R}^b$, where $\rho \in (0,1)$. Thus, $\mathbf{x}_0$ is ρt-block-sparse on average. We operate in the regime of linear measurements $m/n = \delta \in (0, \infty)$. As is common we use in (6.27) the $\ell_{1,2}$-norm to induce block-sparsity, i.e., $f(\mathbf{x}) = \sum_{i=1}^t \|[\mathbf{x}_0]_i\|_2$; this version of the LASSO is often referred to as group-LASSO in the literature [YL06b]. It is not hard to show that $\frac{1}{n} e_{f^*}(c_1\mathbf{h} + c_2\mathbf{x}_0; \tau) \xrightarrow{P} \frac{1}{2b\tau}\mathbb{E}\left[\|\vec{\eta}(c_1\mathbf{h} + c_2 X_0; 1)\|_2^2\right]$, where $\vec{\eta}(\mathbf{x}; \tau') = \mathbf{x}/\|\mathbf{x}\| \, (\|\mathbf{x}\|_2 - \tau')_+, \mathbf{x} \in \mathbb{R}^b$ is the vector soft thresholding operator and

[7] This is known as the "sparse-Gaussian" model. For it, the system of equations in (6.38)–(6.39) obtains an even more explicit formulation which significantly simplifies the numerical evaluation of the solution. We refer the interested reader to [Thr+15] for the details.

$h \sim \mathcal{N}(0, \mathbf{I}_b)$, $X_0 \sim p_{X_0}$ and are independent. From this, the functional F can be easily calculated and substituting that in (6.31) yields a prediction of the squared error as the solution to the following minimization problem:

$$\inf_{\alpha \geq 0} \sup_{\substack{0 \leq \beta \leq 1 \\ \tau_h \geq 0}} \beta\sqrt{\delta}\sqrt{\alpha^2 + \sigma^2} - \frac{\alpha\tau_h}{2} + \frac{\tau_h}{2\alpha} - \frac{\alpha}{2\tau_h b}\mathbb{E}\left[\|\vec{\eta}(\beta\mathbf{h} + \frac{\tau_h}{\alpha}X_0; \lambda)\|_2^2\right]. \quad (6.40)$$

Figure 11.3 illustrates the accuracy of the prediction.

6.8 Low-rank Matrix Recovery via the Trace-LASSO

Let $\mathbf{X}_0 \in \mathbb{R}^{d \times d}$ be an unknown matrix of rank r, in which case, $\mathbf{x}_0 = \text{vec}(\mathbf{X}_0)$ with $n = d^2$. Assume $m/d^2 = \delta \in (0, \infty)$ and $r/d = \rho \in (0, 1)$. As usual in this setting, we consider nuclear-norm regularization; in particular, we choose $f(\mathbf{x}) = \sqrt{d}\|\mathbf{X}\|_*$. Furthermore, for this choice of regularizer, we have

$$\begin{aligned}\frac{1}{n}e_{f^*}(c_1\mathbf{H} + c_2\mathbf{X}_0; \tau) &= \frac{1}{2d^2\tau}\min_{\|\mathbf{V}\|_2 \leq \sqrt{d}}\|\mathbf{V} - (c_1\mathbf{H} + c_2\mathbf{X}_0)\|_F^2 \\ = \frac{1}{2d\tau}\min_{\|\mathbf{V}\|_2 \leq 1}\|\mathbf{V} - d^{-1/2}(c_1\mathbf{H} + c_2\mathbf{X}_0)\|_F^2 &= \frac{1}{2d\tau}\sum_{i=1}^{d}\eta^2\left(\mathrm{s}_i\left(d^{-1/2}(c_1\mathbf{H} + c_2\mathbf{X}_0)\right); 1\right),\end{aligned} \quad (6.41)$$

where $\eta(\cdot;\cdot)$ is as in (6.35), $\mathrm{s}_i(\cdot)$ denotes the i^{th} singular value of its argument and $\mathbf{H} \in \mathbb{R}^{d \times d}$ has entries $\mathcal{N}(0, 1)$. If conditions are met such that the empirical distribution of the singular values of (the sequence of random matrices) $c_1\mathbf{H} + c_2\mathbf{X}_0$ converges asymptotically to a limiting distribution, say $q(c_1, c_2)$, then 6.41 converges to $\frac{1}{2}\mathbb{E}_{x \sim q(c_1,c_2)}\left[\eta^2(x; 1)\right]$. From this, the functional F can be computed similar to Sections 6.6 & 6.7 and substituted in (6.31). For instance, this will be the case if $d^{-1/2}\mathbf{X}_0 = \mathbf{U}\mathbf{S}\mathbf{V}^t$, where $\mathbf{U}, \mathbf{V}$ unitary matrices and $\mathbf{S}$ is a diagonal matrix whose entries have a given marginal distribution with bounded moments (in particular, independent of d). We leave the details and the problem of (numerically) evaluating F for future work.

6.9 Robust Estimators

In this section, we investigate instances where the noise distribution has unbounded moments. In the presence of (say) heavy-tailed noise, it is a common practice to use a loss function that grows to infinity no faster than linearly. This is also suggested by Assumption 4.2.1(b) (cf. (4.11) for the separable case), as has already been discussed.

For illustration, we assume $\mathbf{z} \overset{\text{iid}}{\sim} \text{Cauchy}(0,1)$ and consider two examples of loss functions for which we show that Theorem 4.2.1 is applicable.

LAD

As a first example, consider the regularized-LAD estimator:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1 + \lambda f(\mathbf{x}). \tag{6.42}$$

The loss function is separable, with $\ell(v) = |v|$. Easily, for all $c \in \mathbb{R}$

$$\mathbb{E}\left[|\ell'_+(cG+Z)|^2\right] = \mathbb{E}\left[|\text{sign}(cG+Z)|^2\right] = 1 < \infty,$$

satisfying Assumption (4.9). Also, $\mathbb{E}Z^2$ is undefined, but, $\sup_v \frac{\ell(v)}{|v|} = 1 < \infty$, thus, (4.11) holds. Finally, $|\cdot|$ is not differentiable at zero satisfying the conditions of Lemma 4.3.4. With these, Theorem 4.3.1 is applicable.

Huber-loss

The Huber-loss function with parameter $\rho > 0$ is defined as

$$h_\rho(v) = \begin{cases} \frac{v^2}{2} & , |v| \le \rho, \\ \rho|v| - \frac{\rho^2}{2} & , \text{otherwise.} \end{cases} \tag{6.43}$$

Consider a regularized M-estimator with $\ell(v) = h_\rho(v)$. We show here that this choice satisfies the Assumptions of Theorem 4.3.1. Indeed, for all $c \in \mathbb{R}$

$$\mathbb{E}\left[|\ell'_+(cG+Z)|^2\right] \le \mathbb{E}\left[|cG+Z| \,\middle|\, |cG+Z| \le \rho\right] + \mathbb{E}\left[\rho \,\middle|\, |cG+Z| > \rho\right] < \infty,$$

satisfying Assumption (4.9). Also, $\sup_v \frac{\ell(v)}{|v|} = \rho < \infty$, thus, (4.11) holds. Finally, h_ρ is differentiable with a strictly increasing derivative in the interval $[-\rho, \rho]$. With these, Theorem 4.3.1 is applicable. Figure 6.4 illustrates the validity of the prediction via numerical simulations.

6.10 Numerical Simulations

We have performed a few numerical simulations on specific instances of M-estimators discussed in previous sections. The purpose is to illustrate both the validity of the prediction of Theorem 4.2.1, as well as that of the remarks that followed as a consequence of it.

Figure 6.2 . We consider the regularized LAD estimator of (6.42) under an iid sparse-Gaussian noise model. The unknown signal is also considered sparse, which

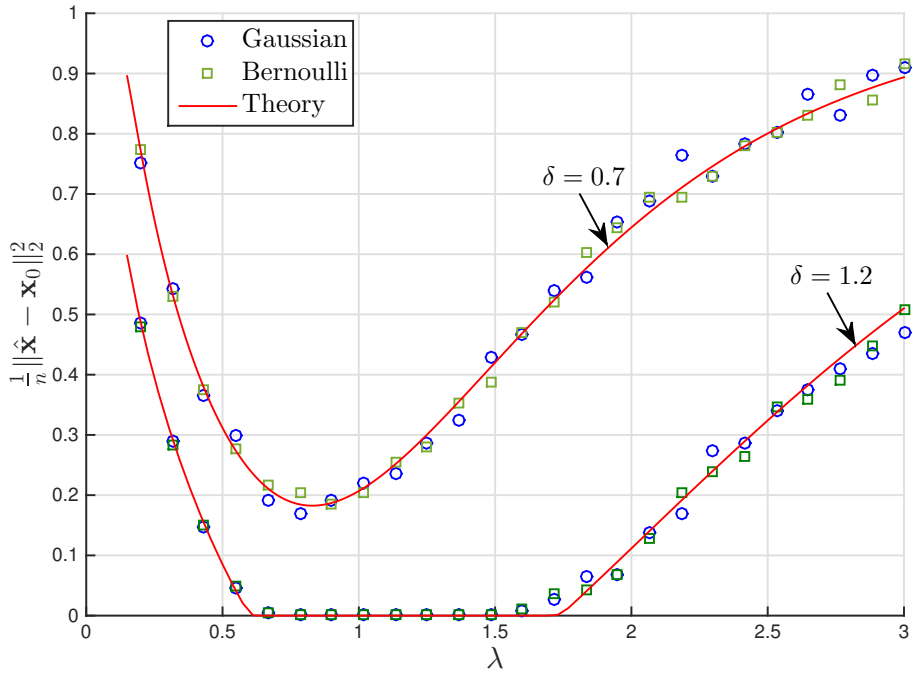

Figure 6.2: Squared error of the l_1-Regularized LAD with Gaussian ($\circ$) and Bernoulli ($\square$) measurements as a function of the regularizer parameter λ for two different values of the normalized number of measurements, namely $\delta = 0.7$ and $\delta = 1.2$. Also, $\mathbf{x}_{0,i} \overset{\text{iid}}{\sim} p_x(x) = 0.9\delta_0(x) + 0.1\phi(x)/\sqrt{0.1}$ and $\mathbf{z}_j \overset{\text{iid}}{\sim} p_z(z) = 0.7\delta_0(z) + 0.3\phi(z)$ for $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. For the simulations, we used $n = 768$ and the data were averaged over five independent realizations.

leads to the natural choice of ℓ_1 regularization, i.e. $f(x) = \|\mathbf{x}\|_1$. Apart from the very close agreement of the theoretical prediction of Theorem 4.2.1 to the simulated data, the following facts are worth observing.

- When the number of measurements m gets large enough, then, for an appropriate range of values of the regularizer parameter, the estimator is consistent, i.e. the unknown signal $\mathbf{x}_0$ is perfectly recovered. This is relevant to Remark 6.3.0.44 where we proved this to be the case for the closely related cone-constrained LAD estimator. For that, we were able to quantify how large m should be as a function of the sparsities of the noise and of the signal, see (6.24).

- The prediction of Theorem 4.2.1 remains accurate when the measurement matrix has entries iid Bernoulli ($\{\pm 1\}$), which supports the universality claim.

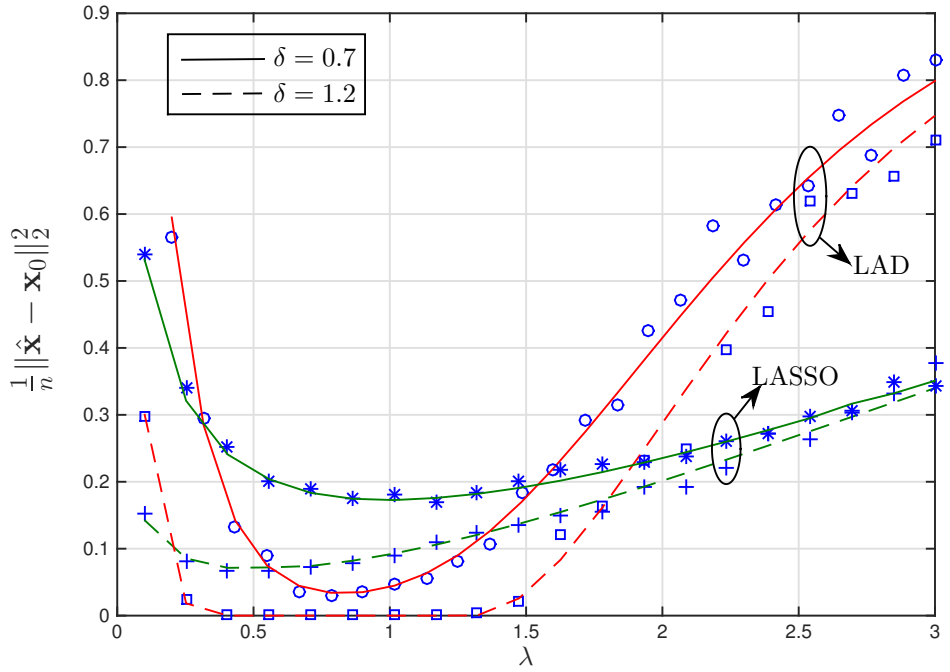

Figure 6.3: Comparing the squared error of the ℓ_1-Regularized LAD with the corresponding error of the LASSO. Both are plotted as functions of the regularizer parameter λ, for two different values of the normalized measurements, namely $\delta = 0.7$ and $\delta = 1.2$. The noise and signal are iid sparse-Gaussian as follows: $\mathbf{x}_{0,i} \overset{\text{iid}}{\sim} p_x(x) = 0.9\delta_0(x) + 0.1\phi(x)/\sqrt{0.1}$ and $\mathbf{z}_j \sim p_z(z) = 0.9\delta_0(z) + 0.1\phi(z)$ with $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. For the simulations, we used $n = 768$ and the data were averaged over five independent realizations.

Figure 6.3 . The model for both the noise and for the unknown signal is here the same as in Figure 6.2, i.e. both are iid sparse. We use ℓ_1-regularization and two different loss functions, namely, a least-absolute-deviations one and a least-squares one, corresponding to a LAD and a LASSO estimator, respectively. The figure aims to compare the performance of the two. Intuition suggests that the LAD is more appropriate for a sparse noise model, since ℓ_1 promotes sparsity. This is indeed the case, in the sense that for good choices of the regularizer parameter λ, the LAD outperforms by far the LASSO. (In the extreme of a large enough number of measurements, the LAD is consistent and this is not the case for the LASSO.) However, it is worth observing that for a different and relatively big range of values of λ, the LASSO performs better. This indicates the importance of the tuning of the regularizer parameter, to which the predictions of Theorem 4.2.1 can offer valuable guidelines and insights.

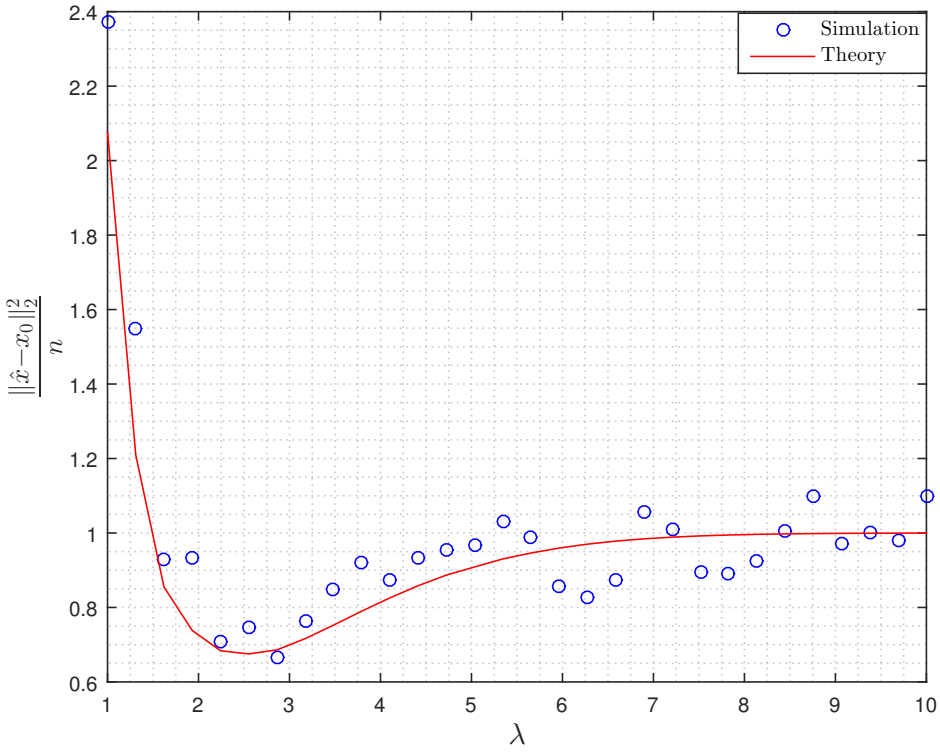

Figure 6.4: Squared error of the ℓ_1-Regularized M-Estimator with Huber-loss as a function of the regularizer parameter λ. Here, $\delta = 0.7$, $\mathbf{x}_0 \overset{\text{iid}}{\sim} p_x(x) = 0.9\delta_0(x) + 0.1\phi(x)/\sqrt{0.1}$ and $p_z(z) = 0.9\delta(z) + 0.1\eta(z)$ with $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $\eta(z) = \frac{1}{\pi(1+z^2)}$. For the simulations, we used $n = 1024$ and the data are averaged over 5 independent realizations.

Figure 6.4 . For this figure, we have assumed an ℓ_1-regularized estimator with Huber-loss $\ell(v) = H_1(v)$. The noise is iid Cauchy$(0, 1)$. In Section 6.9 it was shown that all the Assumptions of Theorem 4.3.1 are satisfied in this setting. The figure validates the prediction. To obtain the prediction we numerically solved the corresponding system of nonlinear equations (see (4.16)) using the efficient iterative scheme described in Remark 4.3.0.29.

Chapter 7

NOISE SENSITIVITY OF THE GENERALIZED-LASSO

Chapter 4 derives a precise and general characterization of the squared-error of regularized M-estimators. In Sections 6.5, we showed how this general result specializes to regularized least-squares (aka *Generalized LASSO*), and we obtained precise error expressions for arbitrary values of the noise distribution and of the rest of the involved parameters (e.g. number of measurements, regularizer). Here, we study the *worst-case* error behavior over the noise variance. Our main focus is on the Generalized-LASSO algorithm. In particular, we measure the reconstruction fidelity by the *Normalized Squared Error* (NSE) (also, referred to as *noise sensitivity*), defined as the ratio between the square reconstruction error and the noise variance, and we obtain *tight* upper bounds on it.

The bounds are tight in the sense that they are attained for the worst-case noise distribution. In fact, we will see that this happens when the noise variance approaches zero; hence, the derived formulae can also be interpreted as precise error predictions in the *high-SNR* regime. A particularly appealing feature of the derived bounds is that they come in *closed-form* and are *geometric* in nature. More specifically, we show that they only depend on a first-order information on the regularizer function f and on the signal $\mathbf{x}_0$, which is captured by the subdifferential $\partial f(\mathbf{x}_0)(\mathbf{x}_0)$; this is in contrast to the general results of Chapters 4 & 6, which require higher-order information on f and on the signal distribution $p_{\mathbf{x}_0}$ in order to account for arbitrary values of noise level. Among others, this particular nature of the results allows for insightful interpretations and for establishing valuable connections with classical results on Ordinary Least-Squares (OLS) and with the related problems of noiseless compressed sensing and of proximal denoising. Finally, we will see that the majority of the results in this chapter are *non-asymptotic*.

We start with introducing three variations of the Generalized LASSO in Section 7.1. In Section 7.2, we study the trivial case of no regularization, corresponding to OLS; classical results on its error performance are revisited under three different noise models: (i) Gaussian noise, (ii) arbitrary fixed noise, and (iii) adversarial noise. In Section 7.3 we derive corresponding bounds for the Generalized LASSO, and show that they very much resemble those of OLS. The formal statement of these results,

along with proofs and further discussions on the implications, follow in the rest of the Chapters. Sections 7.4–7.7 study the case of Gaussian noise; Arbitrary fixed noise and adversarial noise are studied in Section 7.8 and 7.9, respectively.

7.1 Introduction

The idea of ordinary least-squares (OLS) for recovering an unknown signal $\mathbf{x}_0 \in \mathbb{R}^n$ from a vector $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$ of noisy linear observations is very old and can be traced back to the works of Gauss and Legendre. OLS have been classically studied in the statistics literature in the regime of large number of observations but only a few variables to be estimated. The Generalized LASSO is a natural extension of OLS in the modern high-dimensional inference regime; by introducing a regularization term it aims to promote prior information on the structure of the unknown signals.

Generalized LASSO

We distinguish a total of three variations of regularized Least-squares. Although these have been discussed earlier in Chapter 6, we repeat the terminology here for the reader's convenience.

- ⋆ **C-LASSO**[1]:

$$\hat{\mathbf{x}}_c(\mathbf{A}, \mathbf{z}) = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad f(\mathbf{x}) \leq f(\mathbf{x}_0). \tag{7.1}$$

- ⋆ **ℓ_2-LASSO**[2]:

$$\hat{\mathbf{x}}_{\ell_2}(\lambda, \mathbf{A}, \mathbf{z}) = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \frac{\lambda}{\sqrt{m}} f(\mathbf{x}) \right\}. \tag{7.2}$$

- ⋆ **ℓ_2^2-LASSO**:

$$\hat{\mathbf{x}}_{\ell_2^2}(\tau, \mathbf{A}, \mathbf{z}) = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \sigma\lambda f(\mathbf{x}) \right\}. \tag{7.3}$$

The compressed nature of observations in modern inference problems, poses the following urgent questions: *What is the minimum number of measurements required*

[1] C-LASSO in (7.1) stands for "Constrained LASSO". The algorithm assumes *a priori* knowledge of $f(\mathbf{x}_0)$. The acronym "LASSO" was introduced by Tibshirani in 1996 [Tib96] (essentially) referring to the C-LASSO with ℓ_1-regularization.

[2] In the statistics literature the variant of the LASSO algorithm in (7.2) is mostly known as the "square-root LASSO" [BCW11]. Throughout the thesis, we have used both acronyms; however, in this chapter we stick to the more compact term "ℓ_2-LASSO".

to recover $\mathbf{x}_0$ robustly*, that is with error proportional to the noise level? When recovery is robust, can we explicitly characterize how good the estimate is? Can we do so with bounds that are simple and in closed-form?* We will address these questions in this chapter.

Ordinary Least-Squares

It is insightful and instructive to start by considering the simplest case of all, i.e. the case of no regularization. Of course, this corresponds to OLS, which solves

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2. \tag{7.4}$$

OLS has a long history originating in the early 1800s due to works by Gauss and Legendre [Sti81; Mer77], and its behavior is by now very well understood. In particular, in the classical setting $m > n$, assuming $\mathbf{A}$ is full column-rank, (7.4) has a unique solution which is famously given by

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{7.5}$$

The squared error-loss of the OLS estimate in (7.5) is thus expressed as

$$\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \mathbf{z}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\mathbf{z}. \tag{7.6}$$

Starting from (7.6) and imposing certain generic assumptions on the measurement matrix $\mathbf{A}$ and/or the noise vector $\mathbf{z}$, it is possible to conclude precise and simple formulae characterizing the estimation error $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2$. As an example, when the entries of $\mathbf{z}$ are drawn iid normal of zero-mean and variance σ^2, then $\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \sigma^2\text{trace}((\mathbf{A}^T\mathbf{A})^{-1})$. Furthermore, when the entries of $\mathbf{A}$ are drawn i.i.d. normal of zero-mean and variance $1/m$, $\mathbf{A}^T\mathbf{A}$ is a Wishart matrix whose asymptotic eigendistribution is well known. Using this, and letting m, n grow, we find that the squared error concentrates around

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \approx \frac{n}{m-n}. \tag{7.7}$$

Such expressions serve as valuable insights regarding the behavior of the OLS estimator and are meant to provide guidelines for the effectiveness of the estimator in practical situations.

Structured Signals

Gauss and Legendre proposed the OLS method in the context of traditional statistical data analysis. In today's inference problems, the signals of interest are structured, i.e. they often have few few degrees of freedom relative to their ambient

dimension. To appreciate how knowledge of the structure of the unknown signal $\mathbf{x}_0$ can help alleviate the ill-posed nature of the problem, consider a desired signal $\mathbf{x}_0$ which is k-sparse i.e., has only $k < n$ (often $k \ll n$) non-zero entries. Suppose we make m noisy measurements of $\mathbf{x}_0$ using the $m \times n$ measurement matrix $\mathbf{A}$ to obtain $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, and further suppose each set of m columns of $\mathbf{A}$ be linearly independent. Then, as long as $m > k$, we can always find the sparsest solution to

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2,$$

via exhaustive search of $\binom{n}{k}$ such least-squares problems. Under the same assumptions that led to (7.7), this gives a normalized squared error

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \approx \frac{k}{m-k}. \tag{7.8}$$

The catch here, of course, is that the computational complexity of the estimation procedure that was just described is exponential in the ambient dimension n, thus, making it intractable.

On the other hand, the Generalized-LASSO allows estimating the *structured* signal $\mathbf{x}_0$ in a computationally efficient way and is thus appealing. It then becomes crucial to provide answers to the following questions regarding its statistical performance:

- *How many measurements m are needed?*
- *How does the normalized squared error behave and how does it compare to* (7.7) *and* (7.8)*?*
- *Can we provide generic answers that will hold for the general class of signal structures beyond sparsity?*

In particular, we are interested in bounds on the error performance that are sharp and simple, similar to those that characterize the OLS. Under same assumptions on the distribution of the measurement matrix and the noise vector, we ask whether it is possible to derive bounds that resemble (7.7) and (7.8)? It turns out that we can and this chapter is dedicated to providing a full performance analysis and computation of such bounds.

7.2 Revisiting Least Squares

We start by briefly reviewing the OLS equations and derive performance bounds under the generic assumption that the entries of $\mathbf{A}$ are i.i.d. zero-mean normal with variance $1/m$. We examine three different noise models: (i) Gaussian noise, (ii) arbitrary noise, but independent of $\mathbf{A}$, and (iii) adversarial noise

Recall that the OLS solves (7.4). It is clear that when $m < n$, (7.4) is ill posed. However, when $m > n$ and $\mathbf{A}$ has i.i.d. normal entries, $\mathbf{A}$ is full column rank with high probability. The solution of (7.4) is then unique and given by $\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. Recalling that $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, this gives a squared error-loss as in (7.6).

Gaussian Noise

For the purposes of this section, further assume that the entries of $\mathbf{z}$ are i.i.d. zero-mean normal with variance σ^2 and independent of the entries of $\mathbf{A}$. In this case, the normalized mean-squared-error takes the form,

$$\begin{aligned}\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 &= \mathbb{E}[\mathbf{z}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\mathbf{z}] \\ &= \sigma^2\mathbb{E}[\mathrm{trace}(\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T)] \\ &= \sigma^2\mathbb{E}[\mathrm{trace}((\mathbf{A}^T\mathbf{A})^{-1})].\end{aligned}$$

$\mathbf{A}^T\mathbf{A}$ is a Wishart matrix and the distribution of its inverse is well studied. In particular, when $m > n + 1$, we have $\mathbb{E}[(\mathbf{A}^T\mathbf{A})^{-1}] = \frac{m}{m-n-1}\mathbf{I}_n$ [HS]. Hence,

$$\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = m\sigma^2\frac{n}{(m-1)-n}.$$

Noting that $\mathbb{E}\|\mathbf{z}\|_2^2 = m\sigma^2$ and letting m, n be large enough we conclude with the stronger concentration result on the squared-error of OLS:

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \approx \frac{n}{(m-1)-n}. \tag{7.9}$$

Fixed Noise

Fix any noise vector $\mathbf{z}$, with the only assumption being that it is chosen independently of the measurement matrix $\mathbf{A}$. Denote the projection of $\mathbf{z}$ onto the range space of $\mathbf{A}$ by $\mathrm{Proj}(\mathbf{z}, \mathrm{Range}(\mathbf{A})) := \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{z}$ and the minimum singular value of $\mathbf{A}$ by $\sigma_{min}(\mathbf{A})$. Then,

$$\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2 \leq \frac{\|\mathbf{A}(\hat{\mathbf{x}} - \mathbf{x}_0)\|_2}{\sigma_{min}(\mathbf{A})} = \frac{\|\,\mathrm{Proj}(\mathbf{z}, \mathrm{Range}(\mathbf{A}))\|_2}{\sigma_{min}(\mathbf{A})}. \tag{7.10}$$

It is well known that, when $\mathbf{A}$ has entries i.i.d. zero-mean normal with variance $1/m$, then $\sigma_{min}(\mathbf{A}) \approx 1 - \sqrt{\frac{n}{m}}$, [Ver10a]. Also, since $\mathbf{z}$ is independent of $\mathbf{A}$, and the range space of $\mathbf{A}$ is uniformly random subspace of dimension n in $\mathbb{R}^m$, it can be shown that $\| \text{Proj}(\mathbf{z}, \text{Range}(\mathbf{A}))\|_2^2 \approx \frac{n}{m}\|\mathbf{z}\|_2^2$ (e.g. [CR09, p. 13]). With these, we conclude that with high probability on the draw of $\mathbf{A}$,

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim \frac{n}{(\sqrt{m} - \sqrt{n})^2}. \tag{7.11}$$

Worst-Case Squared-Error

Next, assume no restriction at all on the noise vector $\mathbf{z}$. In particular, this includes the case of *adversarial* noise, i.e., noise that has information of the sensing matrix A and can adapt itself accordingly. As expected, this can cause the reconstruction error to be, in general, significantly worse than the guarantees in (7.9) and (7.11). In more detail, we can write,

$$\begin{aligned} \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2 &\leq \frac{\|\mathbf{A}(\hat{\mathbf{x}} - \mathbf{x}_0)\|_2}{\sigma_{min}(\mathbf{A})} = \frac{\|\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{z}\|_2}{\sigma_{min}(\mathbf{A})} \\ &\leq \|\mathbf{z}\|_2 \frac{\|\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\|_2}{\sigma_{min}(\mathbf{A})} \leq \|\mathbf{z}\|\sigma_{min}^{-1}(\mathbf{A}), \end{aligned} \tag{7.12}$$

where $\|\mathbf{M}\|_2$ denotes the spectral norm of a matrix $\mathbf{M}$ and we used the fact that the spectral norm of a symmetric projection matrix is upper bounded by 1. It is not hard to show that equality in (7.12) is achieved when $\mathbf{z}$ is equal to the left singular value of $\mathbf{A}$ corresponding to its minimum singular value. Using the fact that $\sigma_{min}(\mathbf{A}) \approx 1 - \sqrt{\frac{n}{m}}$, we conclude that,

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim \frac{m}{(\sqrt{n} - \sqrt{m})^2}. \tag{7.13}$$

7.3 Least-squares Meets Compressed Sensing

Motivating Examples

In Section 7.2 and in particular in Equations (7.9)–(7.13), we reviewed classical bounds on the normalized square-error of the OLS, which corresponds to the LASSO in the trivial case $f(\cdot) = 0$. How do those results change when a nontrivial convex function $f(\cdot)$ is introduced? What is a precise and simple upper bound on the NSE of the LASSO when the unknown signal is sparse and $f(\cdot) = \|\cdot\|_1$? What if the unknown signal is low-rank and nuclear norm is chosen as the regularizer?

Is it possible to generalize such bounds to arbitrary structures and corresponding convex regularizers?

We provide explicit answers to all these questions. While the formal statement of the results is deferred to Section 7.4–7.9, we provide an overview of them and highlight their implications here. Throughout, we assume that the entries of the sensing matrix $\mathbf{A}$ are i.i.d. zero-mean normal with variance $1/m$.

Sparse Signal Estimation

Assume $\mathbf{x}_0 \in \mathbb{R}^n$ has k nonzero entries. We estimate $\mathbf{x}_0$ via the LASSO with f being the ℓ_1-norm. First, suppose that the noise vector has i.i.d. zero-mean normal entries with variance σ^2. Then, the NSE of the C-LASSO admits the following sharp upper bound[3], which is attained in the limit as the noise variance σ^2 goes to zero:

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim \frac{2k(\log \frac{n}{k} + 1)}{m - 2k(\log \frac{n}{k} + 1)}. \tag{7.14}$$

Compare this to the formula (7.9) for the OLS. (7.14) is obtained from (7.9) after simply replacing the ambient dimension n in the latter with $2k(\log \frac{n}{k} + 1)$. Also, while (7.9) requires $m > n$, (7.14) relaxes this requirement to $m > 2k(\log \frac{n}{k} + 1)$. This is to say that any number of measurements greater than $2k(\log \frac{n}{k} + 1) \ll n$ are sufficient to guarantee robust recovery. Note that this coincides with the classical phase-transition threshold in the noiseless compressed sensing discussed in Section 2.2, see (2.15).

If instead of the C-LASSO, one uses the ℓ_2-LASSO with $\lambda \geq \sqrt{2 \log \frac{n}{k}}$, then

$$\frac{\|\hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim \frac{(\lambda^2 + 3)k}{m - (\lambda^2 + 3)k}. \tag{7.15}$$

Again, observe how (7.15) is obtained from (7.9) after simply replacing the ambient dimension n with $(\lambda^2 + 3)k$. The role of the regularizer parameter λ is explicit in (7.15). Also, substituting $\lambda \approx \sqrt{2 \log \frac{n}{k}}$ in (7.15) (almost) recovers (7.14). This suggests that choosing this value of the regularizer parameter is optimal in that it results in the regularized LASSO performing as well as the constrained version. Note that this value for the optimal regularizer parameter only depends on the sparsity level k of the unknown signal $\mathbf{x}_0$ and *not* the unknown signal iteself.

[3] The statements in this section hold true with high probability in $\mathbf{A}, \mathbf{z}$ and under mild assumptions. See Section 7.4 for the formal statement of the results.

Next, consider the more general case in which the noise vector $\mathbf{z}$ can be anything but is drawn independently of the sensing matrix $\mathbf{A}$. If ones uses the C-LASSO to estimate $\mathbf{x}_0$, then the estimation error is bounded as follows[4] :

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|^2} \lesssim \frac{2k(\log \frac{n}{k} + 1)}{(\sqrt{m} - \sqrt{2k(\log \frac{n}{k} + 1)})^2}. \tag{7.16}$$

Accordingly, the ℓ_2-LASSO for $\lambda \geq \sqrt{2 \log \frac{n}{k}}$ gives:

$$\frac{\|\hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|^2} \lesssim 2\frac{(\lambda^2 + 3)k}{(\sqrt{m} - \sqrt{(\lambda^2 + 3)k})^2}. \tag{7.17}$$

Once more, (7.16) and (7.17) resemble the corresponding formula describing OLS in (7.11). The only difference is that the ambient dimension n is substituted with $2k(\log \frac{n}{k} + 1)$ and $(\lambda^2 + 3)k$, respectively[5].

Low-rank Matrix Estimation

Assume $\mathbf{X}_0 \in \mathbb{R}^{\sqrt{n} \times \sqrt{n}}$ is a rank-r matrix, and let $\mathbf{x}_0 = \mathrm{vec}(\mathbf{X}_0) \in \mathbb{R}^n$ be the vectorization of $\mathbf{X}_0$. We use the generalized LASSO with $f(\mathbf{x}) = \|\mathrm{vec}^{-1}(\mathbf{x})\|_\star$. The nuclear norm of a matrix (i.e. sum of singular values) is known to promote low-rank solutions [Faz02; RFP10].

As previously, suppose first that $\mathbf{z}$ has i.i.d. zero-mean normal entries with variance σ^2. Then, the NSE of the C-LASSO and that of the ℓ_2-LASSO for $\lambda \geq 2n^{1/4}$ are bounded as follows:

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim \frac{6\sqrt{n}r}{m - 6\sqrt{n}r}, \tag{7.18}$$

and

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim \frac{\lambda^2 r + 2\sqrt{n}(r + 1)}{m - (\lambda^2 r + 2\sqrt{n}(r + 1))}. \tag{7.19}$$

Just like in the estimation of sparse signals in Section 7.3, it is clear from the bounds above that they can be obtained from the OLS bound in (7.9) after only substituting the dimension of the ambient space n with $6\sqrt{n}r$ and $\lambda^2 r + 2\sqrt{n}(r + 1)$, respectively. And again, $6\sqrt{n}r$ is exactly the phase transition threshold for the noiseless compressed sensing of low-rank matrices, see (2.16).

[4] The formula below is subject to some simplifications meant to highlight the essential structure. See Section 7.8 for the details.

[5] It is conjectured in [TOH14] that the factor of 2 in (7.17) is not essential and that it only appears as an artifact of the proof technique therein. See, also, Section 7.8.

Moving to the case where $\mathbf{z}$ is arbitrary but independent of $\mathbf{A}$, we find that

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim \frac{6\sqrt{n}r}{(\sqrt{m} - 6\sqrt{n}r)^2}, \tag{7.20}$$

and

$$\frac{\|\hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0\|_2^2}{\|\mathbf{z}\|_2^2} \lesssim 2\frac{\lambda^2 r + 2\sqrt{n}(r+1)}{(\sqrt{m} - \sqrt{\lambda^2 r + 2\sqrt{n}(r+1)})^2}. \tag{7.21}$$

General Structures

From the discussion in Sections 7.3 and 7.3, it is becoming clear that the error bounds for the OLS admit nice and simple generalizations to error bounds for the generalized LASSO. What changes in the formulae bounding the NSE of the OLS when considering the NSE of the LASSO is only that the ambient dimension n is substituted by a specific summary parameter.

This parameter depends on the particular structure of the unknown signal, but not the signal itself. For example, in the sparse case, it depends only on the sparsity of $\mathbf{x}_0$, not $\mathbf{x}_0$ itself, and in the low-rank case, it only depends on the rank of $\mathbf{X}_0$, not $\mathbf{X}_0$ itself. Furthermore, it depends on the structure-inducing function $f(\cdot)$ that is being used. Finally, it is naturally dependent on whether the constrained or the regularized LASSO is being used. In the case of regularized LASSO, it also depends on the value λ of the regularizer parameter. Interestingly, the value of this parameter corresponding to the NSE of the constrained LASSO is exactly the phase-transition threshold of the corresponding noiseless CS problem.

The general result of this chapter shows that this summary parameter is nothing but (i) the *statistical dimension* $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ (recall (2.7) and (2.10)) for the constrained LASSO (ii) the *Gaussian distance squared to the scaled subdifferential* $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ (recall (2.12)) for the regularized LASSO. To see that this is consistent with the bounds presented above on the specific instances of sparse and low-rank recovery, recall for example from Section 2.2 that for $f(\mathbf{x}) = \ell_1$ and $\mathbf{x}_0$ a k-sparse vector, $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \approx 2k\log\frac{n}{k}$. Hence, one obtains (7.14) and (7.16) when substituting the n with $2k\log\frac{n}{k}$ in (7.9) and (7.11), respectively.

To conclude this section, we repeat once more: *the classical and well-known error analysis of the NSE of the OLS can be naturally extended to describe the NSE of the generalized LASSO*. In particular, when the entries of $\mathbf{A}$ are i.i.d. normal, then an error bound on the NSE of the OLS translates to a bound on the NSE of the

generalized (constrained or regularized) LASSO after (almost) only substituting the ambient dimension n by either $\mathbf{D}(\mathrm{cone}(\partial f(\mathbf{x}_0)))$ or $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$. $\mathbf{D}(\mathrm{cone}(\partial f(\mathbf{x}_0)))$ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ are *summary parameters* that capture the geometry of the LASSO problem.

7.4 The NSE of Generalized LASSO in Gaussian Noise

In this Section we assume that the noise vector $\mathbf{z}$ has entries distributed i.i.d. normal $\mathcal{N}(0,\sigma^2)$ and derive sharp upper bounds on the NSE of the generalized LASSO.

First, we formally define the performance measures of interest. Then, we describe the main steps of the required technical analysis, which is again based on the CGMT framework of Chapter 5. Sections 7.5, 7.6 and 7.7 are each devoted to upper-bounding the NSE of the C-LASSO, ℓ_2-LASSO and ℓ_2^2-LASSO respectively.

The NSE in High-SNR

Define the Normalized Squared Error as

$$\mathrm{NSE}(\sigma) := \frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{m\sigma^2}. \tag{7.22}$$

Further define the *worst-case* NSE as

$$\mathrm{wNSE} := \sup_{\sigma>0} \mathrm{NSE}(\sigma).$$

We say that recovery of $\mathbf{x}_0$ is *robust* whenever $\mathrm{wNSE} < \infty$. Also, consider the *asymptotic* NSE,

$$\mathrm{aNSE} := \lim_{\sigma\to 0} \mathrm{NSE}(\sigma).$$

In this section we derive precise closed-form expressions for the aNSE of the generalized LASSO. We conjecture that

$$\mathrm{wNSE} := \sup_{\sigma>0} \mathrm{NSE}(\sigma) = \lim_{\sigma\to 0} \mathrm{NSE}(\sigma) =: \mathrm{aNSE}. \tag{7.23}$$

This highlights the significance of studying the performance at high-SNR, since it leads to the following implication

$$\mathrm{aNSE} \approx \eta \;\Rightarrow\; \mathrm{NSE}(\sigma) \lesssim \eta,$$

i.e. the formulae characterizing the NSE at high-SNR are in fact *tight upper bounds* on the NSE for arbitrary values of the SNR.

In fact, we have proved in [OTH13b, Sec. 10] that (7.38) is true for the C-LASSO. For the regularized LASSO, the conjecture is supported by extended empirical observations. Besides, the same phenomenon has been observed and proved to be true for related estimation problems. Examples include the proximal denoising problem (7.33) in [OH15; DJM13; DGM13] and the LASSO problem with ℓ_1 penalization [DMM11].

Remark 7.4.0.48. (Proving (7.38)) In 6.5 we obtained exact expressions for the NSE of the regularized LASSO that can be evaluated for arbitrary values of the SNR by solving (6.31). Hence, an alternative to the approach presented below to recover the results of this section is by evaluating the former in the limit of $\sigma \to 0$. More importantly, it is possible in principle to evaluate the worst-case NSE by computing $\sup_{\sigma>0} \mathrm{NSE}(\sigma)$. The challenge lies in the fact that the $\mathrm{NSE}(\sigma)$ for arbitrary values of the noise variance is expressed not in closed form but rather as the minimizer to (6.31). However, this offers a systematic way to a rigorous proof of the fact that $\sup_{\sigma>0} \mathrm{NSE}(\sigma) = \lim_{\sigma\to 0} \mathrm{NSE}(\sigma)$.

Analysis

The analysis is based on the CGMT framework of Chapter 5. In addition to the general characteristics of the analysis in Chapter 5, two additional features are important for the results of this section and are worth emphasizing. First, before applying the CGMT framework, we introduce a "first-order approximation" of the LASSO minimization which is shown to be tight in the high-SNR regime. The approximated LASSO problem leads to an Auxiliary Optimization (AO) that is very simple to analyze (in particular, much simpler than the generic (AO) in (5.7)). As part of the analysis, it becomes clear why the statistical dimension and the Gaussian distance squared appear in the derived error formulae. Second, the bounds derived here are *non-asymptotic*. The analysis in Chapter 5 leads to an asymptotic version of the results. However, the CGMT Theorem 3.3.1 is *non-asymptotic* and after some some extra work in the "*convergence analysis of the (AO)*" step of the CGMT framework (see Section 5.1.) non-asymptotic results are also possible.

Below, we present some of these key ideas. We provide specific references to either [OTH13b; Oym15] [6] or the appendix for the details of the proofs.

For the purposes of exposition we use the ℓ_2-LASSO. The analysis for the con-

[6] When referring to [OTH13b] keep in mind the following: a) in [OTH13b] the entries of $\mathbf{A}$ have variance 1 and not $1/m$ as here, b) [OTH13b] uses slightly different notation for $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ and $\mathbf{D}(\mathrm{cone}(\partial f(\mathbf{x}_0)))$ ($\mathbf{D}_f(\mathbf{x}_0, \lambda)$ and $\mathbf{D}_f(\mathbf{x}_0, \mathbf{R}^+)$, respectively).

strained version C-LASSO is to a large extent similar. In fact, [OTH13b] treats those two under a common framework.

Remark 7.4.0.49. The results presented here on the aNSE of the C-LASSO and of the ℓ_2-LASSO, were proved in [OTH13b]. Even though the proof is based on the same ideas as those underlying the CGMT framework, at the time of writing we were missing the clean formulation of both Theorem 3.3.1 and of the framework described in 5. As a result, the analysis in [OTH13b] is put in a somewhat different language and is slightly more convoluted and lengthier than the now available framework would allow. In fact, it is only thanks to this general framework that we were later able to extend the analysis to the ℓ_2^2-LASSO in [TPH15] (proving an earlier conjecture of [OTH13b]). Since the focus of this thesis is on results that are far more general than the aNSE performance of the generalized LASSO, we have decided not to include the proof details of Theorems 7.5.1 and 7.6.1 here. Besides, the interested reader can find these not only in [OTH13b], but also in [Oym15].

First-Order Approximation. Recall the ℓ_2-LASSO problem introduced in (7.2):

$$\hat{\mathbf{x}}_{\ell_2} = \arg\min_{\mathbf{x}} \ \|\mathbf{y}-\mathbf{A}\mathbf{x}\| + \frac{\lambda}{\sqrt{m}} f(\mathbf{x}). \tag{7.24}$$

A key idea behind our approach is using the linearization of the convex structure inducing function f around the vector of interest $\mathbf{x}_0$ [Roc97; BL10]. From convexity of f, for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{s} \in \partial f(\mathbf{x}_0)$, we have $f(\mathbf{x}) \geq f(\mathbf{x}_0) + \mathbf{s}^T(\mathbf{x}-\mathbf{x}_0)$. In particular,

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \sup_{\mathbf{s}\in\partial f(\mathbf{x}_0)} \mathbf{s}^T(\mathbf{x}-\mathbf{x}_0) =: \hat{f}(\mathbf{x}), \tag{7.25}$$

and approximate equality holds when $\|\mathbf{x}-\mathbf{x}_0\|$ is "small". Recall that $\partial f(\mathbf{x}_0)$ denotes the subdifferential of f at $\mathbf{x}_0$ and is always a compact and convex set [Roc97]. We also assume that $\mathbf{x}_0$ is not a minimizer of f, hence, $\partial f(\mathbf{x}_0)$ does not contain the origin.

We substitute f in (7.24) by its first-order approximation $\hat{f}$ to get a corresponding "*Approximated LASSO*" problem. To write the approximated problem in an easy-to-work-with format, recall that $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} = \mathbf{A}\mathbf{x}_0 + \sigma\mathbf{v}$, for $\mathbf{v} \sim \mathcal{N}(0, \mathbf{I}_m)$ and change the optimization variable from $\mathbf{x}$ to $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$:

$$\tilde{\mathbf{w}}_{\ell_2} = \arg\min_{\mathbf{w}} \left\{ \|\mathbf{A}\mathbf{w}-\sigma\mathbf{v}\| + \frac{1}{\sqrt{m}} \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}. \tag{7.26}$$

We denote $\tilde{\mathbf{w}}_{\ell_2}$ the optimal solution of the approximated problem in (7.26) and $\hat{\mathbf{w}}_{\ell_2} = \hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0$ for the optimal solution of the original problem in (7.24)[7]. Also, denote the optimal cost achieved in (7.26) by $\tilde{\mathbf{w}}_{\ell_2}$, as $\Phi_{\ell 2}(\mathbf{A}, \mathbf{v})$. Finally, note that the approximated problem corresponding to C-LASSO can be written as in (7.26), with only $\lambda \partial f(\mathbf{x}_0)$ being substituted by $\text{cone}(\partial f(\mathbf{x}_0))$.

Taking advantage of the simple characterization of $\hat{f}$ via the subdifferential $\partial f(\mathbf{x}_0)$, we are able to *precisely* analyze the optimal cost and the normalized squared error of the resulting approximated problem. The approximation is tight when $\|\hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0\| \to 0$ and we later show that this is the case when the noise level $\sigma \to 0$. This fact allows us to translate the results obtained for the Approximated LASSO problem to corresponding *precise* results for the original LASSO problem, in the *small noise variance regime*.

We follow the steps of the CGMT framework as prescribed in Chapter 5.

Determining the (AO). The (approximated) LASSO problem in (7.26) is simpler than the original one in (7.2), yet, still hard to directly analyze. It should come as no surprise at this point that, in view of the CGMT, we analyze instead a corresponding Auxiliary Optimization (AO) problem.

First, using the fact that

$$\sqrt{m}\|\mathbf{A}\mathbf{w} - \sigma\mathbf{v}\|_2 = \max_{\|\mathbf{u}\|=1} \mathbf{u}^T\mathbf{G}\mathbf{w} - \sqrt{m}\sigma\mathbf{u}^T\mathbf{v}, \tag{7.27}$$

we bring the minimization in (7.26) in the appropriate format of a (PO) as in (3.11a). Here, we have used the assumption that $\mathbf{A}$ has entries of variance $1/m$ and $\mathbf{G}$ denotes a matrix with iid standard normal entries. Then, we find that the (AO) becomes:

$$\phi_{\ell_2}(\mathbf{g}, \mathbf{h}) = \min_{\mathbf{w}} \max_{\|\mathbf{u}\|_2 \le 1} \left\{ \|\mathbf{w}\|_2 \mathbf{u}^T\mathbf{g} - \|\mathbf{u}\|_2 \mathbf{h}^T\mathbf{w} - \sigma\sqrt{m}\mathbf{u}^T\mathbf{v} + \max_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}.$$

Moreover, since $\mathbf{v}$ is iid Gaussian, $\|\mathbf{w}\|_2\mathbf{g} - \sigma\sqrt{m}\mathbf{v}$ is distributed $\mathcal{N}(\mathbf{0}, (\|\mathbf{w}\|_2^2 + m\sigma^2)\mathbf{I}_m)$. Therefore, it is equivalent to analyze the following (AO) instead:

$$\phi_{\ell_2}(\mathbf{g}, \mathbf{h}) = \min_{\mathbf{w}} \max_{\|\mathbf{u}\|_2 \le 1} \left\{ \sqrt{\|\mathbf{w}\|_2^2 + m\sigma^2}\mathbf{g}^T\mathbf{u} - \|\mathbf{u}\|_2 \mathbf{h}^T\mathbf{w} + \max_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}, \tag{7.28}$$

where we have abused some notation and $\mathbf{g}$ is still used to denote an iid Gaussian vector in $\mathbb{R}^m$.

[7] We follow this convention throughout: we use the symbol "~" over variables that are associated with the approximated problems. To distinguish, we use the symbol "^" for the variables associated with the original problem .

Scalarization of the (AO). Here, we perform the deterministic analysis of $\phi_{\ell_2}(\mathbf{g}, \mathbf{h})$ for fixed $\mathbf{g} \in \mathbb{R}^m$ and $\mathbf{h} \in \mathbb{R}^n$.

First, we can easily maximize over the *direction* of $\mathbf{u}$ to equivalently express the optimization as

$$\phi_{\ell_2}(\mathbf{g}, \mathbf{h}) = \min_{\mathbf{w}} \max_{\substack{0\le\beta\le 1 \\ \mathbf{s}\in\partial f(\mathbf{x}_0)}} \left\{ \beta\sqrt{\|\mathbf{w}\|_2^2 + m\sigma^2}\|\mathbf{g}\| - (\beta\mathbf{h} - \lambda\mathbf{s})^T\mathbf{w} \right\}.$$

The objective is now convex in $\mathbf{w}$ and (jointly) concave in $\beta, \mathbf{s}$, and the constraint sets over which maximization occurs are bounded. Thus, as in [Roc97, Corollary 37.3.2] we can flip the order of min-max. Then, it is easy to minimize over the direction of $\mathbf{w}$ to find

$$\phi_{\ell_2}(\mathbf{g}, \mathbf{h}) = \max_{\substack{0\le\beta\le 1 \\ \mathbf{s}\in\partial f(\mathbf{x}_0)}} \min_{\alpha\ge 0} \left\{ \beta\sqrt{\alpha^2 + m\sigma^2}\|\mathbf{g}\| - \alpha\|\beta\mathbf{h} - \lambda\mathbf{s}\|_2 \right\}. \tag{7.29}$$

As a last step, we flip the order of min-max once more. Maximization over $\mathbf{s}$ results in the distance term below, (defined as $\text{dist}(\mathbf{v}, \lambda\partial f(\mathbf{x}_0)) := \min_{\mathbf{s}\in\partial f(\mathbf{x}_0)} \|\mathbf{v} - \lambda\mathbf{s}\|_2$):

$$\max_{0\le\beta\le 1} \min_{\alpha\ge 0} \{\sqrt{\alpha^2 + \sigma^2}\ \|\mathbf{g}\|_2\beta - \alpha \cdot \text{dist}(\beta\mathbf{h}, \lambda\partial f(\mathbf{x}_0))\}. \tag{7.30}$$

In just a few lines we were able to reduce the (AO) problem to an equivalent optimization in (7.30) that now only involves two *scalar* variables, out of which α plays the role of $\|\mathbf{w}\|_2$. Also, the objective is strongly convex with respect to α (this can be used to satisfy the conditions of Theorem 3.3.1).

Convergence Analysis of the (AO). One may now proceed following the asymptotic convergence analysis framework of the (AO) prescribed in Chapter 5, which leads to asymptotic bounds of the aNSE of the LASSO (see for example [TOH15, Sec. 3.3.3] or (6.21)). Instead, here we obtain bounds that are *non-asymptotic*. It is shown in [OTH13b] that when $\lambda \in \mathcal{R}_{\text{ON}}$ then the optimal value of β in (7.30) is 1 with high probability. We will formally define $\mathcal{R}_{\text{ON}}$ later in Section 7.6; for now, it suffices to mention that this regime is in a sense (that will soon made precise) the "interesting" regime of values of the regularizer parameter. When this is the case, (7.30) simplifies to a minimization problem over α and is trivial to solve for its optimal value.

The result is summarized in Lemma 7.4.1 below.

Lemma 7.4.1 (Deterministic Result)**.** *Let* $\mathbf{w}_\phi(\mathbf{g},\mathbf{h})$ *be a minimizer of the problem in* (7.28)*. If* $\|\mathbf{g}\| > \text{dist}(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))$*, then,*

$$a) \quad \|\mathbf{w}_\phi(\mathbf{g},\mathbf{h})\|^2 = m\sigma^2 \frac{\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))}{\|\mathbf{g}\|^2 - \text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))},$$

$$b) \quad \phi_{\ell_2}(\mathbf{g},\mathbf{h}) = \sqrt{m}\sigma\sqrt{\|\mathbf{g}\|^2 - \text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))}.$$

Of interest is making probabilistic statements about $\phi_{\ell_2}(\mathbf{g},\mathbf{h})$ and the norm of its minimizer $\|\mathbf{w}_\phi(\mathbf{g},\mathbf{h})\|$. Lemma 7.4.1 provides us with closed-form deterministic solutions for both of them, which only involve the quantities $\|\mathbf{g}\|^2$ and $\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))$. The ℓ_2-norm and the distance function to a convex set are 1-Lipschitz functions. Application of Proposition 3.1.1 shows that $\|\mathbf{g}\|^2$ and $\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))$ concentrate nicely around their means, $\mathbb{E}\left[\|\mathbf{g}\|^2\right] = m$ and $\mathbb{E}\left[\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))\right] = \mathbf{D}(\lambda\partial f(\mathbf{x}_0))$, respectively. Combining this with Lemma 7.4.1, we conclude with Lemma 7.4.2 below.

Lemma 7.4.2 (Probabilistic Result)**.** *Assume that* $(1-\epsilon_L)m \geq \mathbf{D}(\lambda\partial f(\mathbf{x}_0)) \geq \epsilon_L m$ *for some constant* $\epsilon_L > 0$*. Define*[8]*,*

$$\eta = \sqrt{m - \mathbf{D}(\lambda\partial f(\mathbf{x}_0))} \quad \text{and} \quad \gamma = \frac{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))}{m - \mathbf{D}(\lambda\partial f(\mathbf{x}_0))}.$$

Then, for any $\epsilon > 0$*, there exists a constant* $c > 0$ *such that, for sufficiently large* m*, with probability* $1 - \exp(-cm)$*,*

$$\left|\phi_{\ell_2}(\mathbf{g},\mathbf{h}) - \sqrt{m}\sigma\eta\right| \leq \epsilon\sqrt{m}\sigma\eta \quad \text{and} \quad \left|\frac{\|\mathbf{w}_\phi(\mathbf{g},\mathbf{h})\|^2}{m\sigma^2} - \gamma\right| \leq \epsilon\gamma.$$

Remark: In Lemma 7.4.2, the condition "$(1-\epsilon_L)m \geq \mathbf{D}(\lambda\partial f(\mathbf{x}_0))$" ensures that $\|\mathbf{g}\| > \text{dist}(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))$ (cf. Lemma 7.4.1) with high probability over the realizations of $\mathbf{g}$ and $\mathbf{h}$.

Applying the CGMT. Before proceeding, let us recap. Application of Gordon's Lemma to the approximated LASSO problem in (7.26) introduced the simpler (AO) (7.28). Without much effort, we found in Lemma 7.4.2 that its cost, $\phi_{\ell_2}(\mathbf{g},\mathbf{h})$, and the normalized squared norm of its minimizer, $\frac{\|\tilde{\mathbf{w}}(\mathbf{g},\mathbf{h})\|^2}{m\sigma^2}$, concentrate around $\sqrt{m}\sigma\eta$ and γ, respectively. Now, it remains to apply the CGMT Theorem 3.3.1(ii) & (iii) to conclude that the same same results translate to $\Phi_{\ell 2}(\mathbf{A},\mathbf{v})$ and $\tilde{\mathbf{w}}_{\ell 2}(\mathbf{A},\mathbf{v})$. The

[8] Observe that the dependence of η and γ on λ, m and $\partial f(\mathbf{x}_0)$, is implicit in this definition.

conditions of statement (iii) of the theorem are shown to be satisfied using the strong convexity of (7.28) over $\mathbf{w}$ (see [OTH13b] or [TOH15]).

From the Approximated LASSO Back to the Original The final step requires us to translate this bound on the NSE of the Approximated LASSO to a bound on the NSE of the original one. We choose σ small enough such that $\|\tilde{\mathbf{w}}_{\ell_2}\|$ is small and so $f(\mathbf{x}_0 + \tilde{\mathbf{w}}_{\ell_2}) \approx \hat{f}(\mathbf{x}_0 + \tilde{\mathbf{w}}_{\ell_2})$. Using this and combining the results above we show that $\|\hat{\mathbf{w}}_{\ell_2}\|^2/(m\sigma^2)$ concentrates with high probability around γ (see Section 9.1.2 in [OTH13b]).

7.5 Constrained LASSO

Theorem 7.5.1. *Assume there exists a constant $\epsilon_L > 0$ such that, $(1-\epsilon_L)m \geq \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \geq \epsilon_L m$ and m is sufficiently large. For any $\epsilon > 0$, there exists a constant $C = C(\epsilon, \epsilon_L)$ such that, with probability $1 - \exp(-Cm)$,*

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|^2}{m\sigma^2} \leq (1+\epsilon)\frac{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}{m - \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}. \tag{7.31}$$

Furthermore, there exists a deterministic number $\sigma_0 > 0$ (i.e. independent of $\mathbf{A}, \mathbf{v}$) such that, if $\sigma \leq \sigma_0$, with the same probability,

$$\left| \frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|^2}{m\sigma^2} \times \frac{m - \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} - 1 \right| < \epsilon. \tag{7.32}$$

Observe in Theorem 7.5.1 that as m approaches $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$, the NSE increases and when $m = \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$, NSE $= \infty$. This behavior is not surprising as when $m < \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$, one cannot even recover $\mathbf{x}_0$ from noiseless observations via (2.1) hence it is futile to expect noise robustness.

Example (sparse signals): Figure 7.1 illustrates Theorem 7.5.1 when $\mathbf{x}_0$ is a k-sparse vector and f is the ℓ_1 norm. In this case, $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ is only a function of k and n and can be exactly calculated, [Don06a]. The dark-blue region corresponds to the unstable region $m < \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$. The dashed gray line obeys $m = 1.4 \times \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ and yields a constant (worst-case) NSE of 2.5 as sparsity varies. We note that for ℓ_1 minimization, the NSE formula was first proposed by Donoho et al. in [DMM11].

Relation to Proximal Denoising

It is interesting to compare the NSE of the C-LASSO to the MSE risk of the constrained proximal denoiser.

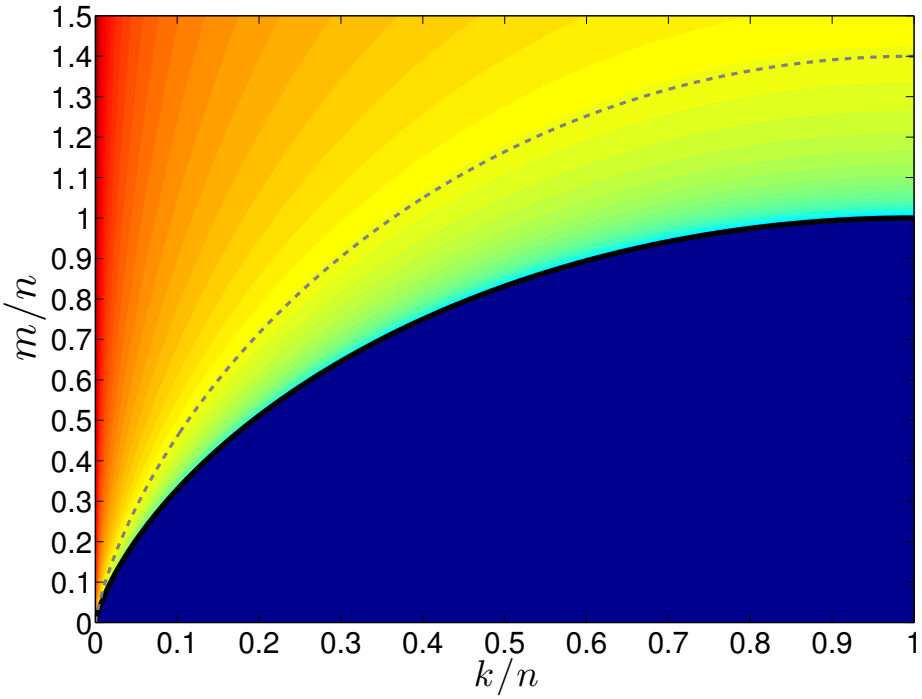

Figure 7.1: NSE heatmap for ℓ_1 minimization based on Theorem 7.5.1. The x and y axes are the sparsity and measurements normalized by the ambient dimension. To obtain the figure, we plotted the heatmap of the function $-\log \frac{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}{m-\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}$ (clipped to ensure the values are between $[-10, 5]$).

The proximal denoising problem tries to estimate $\mathbf{x}_0$ from noisy but uncompressed observations $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$, where the entries of $\mathbf{z}$ are i.i.d. zero-mean Gaussian with variance σ^2. In particular, it solves,

$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\sigma f(\mathbf{x}) \right\}. \tag{7.33}$$

A closely related approach to estimate $\mathbf{x}_0$, which requires prior knowledge $f(\mathbf{x}_0)$ about the signal of interest $\mathbf{x}_0$, is solving the constrained denoising problem:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 \quad \text{subject to} \quad f(\mathbf{x}) \leq f(\mathbf{x}_0). \tag{7.34}$$

The natural question to be posed in both cases is how well can one estimate $\mathbf{x}_0$ via (7.33) (or (7.34)) [Don95; DJM13; CJ13; OH15]? The minimizer $\hat{\mathbf{x}}$ of (7.33) (or (7.34)) is a function of the noise vector $\mathbf{z}$ and the common measure of performance, is the normalized mean-squared-error which is defined as $\frac{\mathbb{E}\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2}{\sigma^2}$. It has been shown that the normalized MSE of (7.34) is upper bounded by $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$

[OH15; CJ13]. Furthermore, this bound is attained asymptotically as $\sigma \to 0$. From Theorem 7.5.1 we find that the corresponding quantity $\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|^2/\sigma^2$ is upper bounded by

$$\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))\frac{m}{m - \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))},$$

and is again attained asymptotically as $\sigma \to 0$. We conclude that the NSE of the LASSO problem is amplified compared to the corresponding quantity of proximal denoising by a factor of $\frac{m}{m-\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} > 1$. This factor can be interpreted as the penalty paid in the estimation error for observing noisy linear measurements of the unknown signal instead of just noisy measurements of the signal itself.

7.6 ℓ_2-LASSO

Characterization of the NSE of the ℓ_2-LASSO is more involved than that of the NSE of the C-LASSO. For this problem, choice of λ naturally plays a critical role.

Background

Before introducing the main result, it is required to introduce some further notation.

Let $C \subset \mathbb{R}^n$ be a closed and nonempty convex set. For any vector $\mathbf{v} \in \mathbb{R}^n$, we denote its (unique) projection onto C as $\text{Proj}(\mathbf{v}, C)$, i.e.

$$\text{Proj}(\mathbf{v}, C) := \text{argmin}_{\mathbf{s}\in C} \|\mathbf{v} - \mathbf{s}\|.$$

The distance of $\mathbf{v}$ to the set C can then be written as

$$\text{dist}(\mathbf{v}, C) := \|\mathbf{v} - \text{Proj}(\mathbf{v}, C)\|.$$

Recall the definition of the Gaussian distance squared to the scaled subdifferential in (2.12):

$$\mathbf{D}(\lambda\partial f(\mathbf{x}_0)) = \mathbb{E}\left[\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))\right].$$

Accordingly, define the Gaussian correlation as:

$$\mathbf{C}(\lambda\partial f(\mathbf{x}_0)) := \mathbb{E}\left[(\mathbf{h} - \text{Proj}(\mathbf{h}, \lambda\partial f(\mathbf{x}_0)))^T \text{Proj}(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))\right].$$

Further recall from (2.13) the deep relation between and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ and $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ (one that is stronger than the obvious fact that $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \le \min_{\lambda\ge 0} \mathbf{D}(\lambda\partial f(\mathbf{x}_0))$):

$$\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \approx \min_{\lambda\ge 0} \mathbf{D}(\lambda\partial f(\mathbf{x}_0)). \tag{7.35}$$

Moreover, as the next lemma shows, the minimum of $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ in (7.35) is uniquely attained. The lemma also reveals an interesting relation between $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ and $\mathbf{C}(\lambda\partial f(\mathbf{x}_0))$.

Lemma 7.6.1 ([Ame+13])**.** *Suppose $\partial f(\mathbf{x}_0)$ is nonempty and does not contain the origin. Then,*

1. $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ *is a strictly convex function of* $\lambda \geq 0$*, and is differentiable for* $\lambda > 0$*.*
2. $\frac{\partial \mathbf{D}(\lambda\partial f(\mathbf{x}_0))}{\partial\lambda} = -\frac{2}{\lambda}\mathbf{C}(\lambda\partial f(\mathbf{x}_0))$*.*

NSE

Definition 7.6.1 ($\mathcal{R}_{\text{ON}}$)**.** Suppose $m > \min_{\lambda\geq 0}\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$. Define $\mathcal{R}_{\text{ON}}$ as follows,

$$\mathcal{R}_{\text{ON}} = \{\lambda > 0 \mid m - \mathbf{D}(\lambda\partial f(\mathbf{x}_0)) > \max\{0, \mathbf{C}(\lambda\partial f(\mathbf{x}_0))\}\}\,.$$

Theorem 7.6.1 (non-asymptotic)**.** *Assume there exists a constant $\epsilon_L > 0$ such that $(1-\epsilon_L)m \geq \max\{\mathbf{D}(\lambda\partial f(\mathbf{x}_0)), \mathbf{D}(\lambda\partial f(\mathbf{x}_0)) + \mathbf{C}(\lambda\partial f(\mathbf{x}_0))\}$ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0)) \geq \epsilon_L m$. Further, assume that m is sufficiently large. Then, for any $\epsilon > 0$, there exist a constant $C = C(\epsilon, \epsilon_L)$ and a deterministic number $\sigma_0 > 0$ (i.e. independent of $\mathbf{A}, \mathbf{v}$) such that, whenever $\sigma \leq \sigma_0$, with probability $1 - \exp(-C\min\{m, \frac{m^2}{n}\})$,*

$$\left|\frac{\|\hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0\|^2}{m\sigma^2} \times \frac{m - \mathbf{D}(\lambda\partial f(\mathbf{x}_0))}{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))} - 1\right| < \epsilon.$$

Regions Of Operation

First, we identify the regime in which the ℓ_2-LASSO can robustly recover $\mathbf{x}_0$. In this direction, the number of measurements should be large enough to guarantee at least noiseless recovery in (2.1), which is the case when $m > \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ from Theorem 2.2.1. To translate this requirement in terms of $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$, recall (7.35) and Lemma 7.6.1, and define λ_{best} to be the *unique* minimizer of $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ over $\lambda \in \mathbb{R}^+$. We then write the regime of interest as $m > \mathbf{D}(\lambda_{\text{best}}\cdot\partial f(\mathbf{x}_0)) \approx \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$.

Next, we identify three important values of the penalty parameter, λ, needed to describe the distinct regions of operation of the estimator.

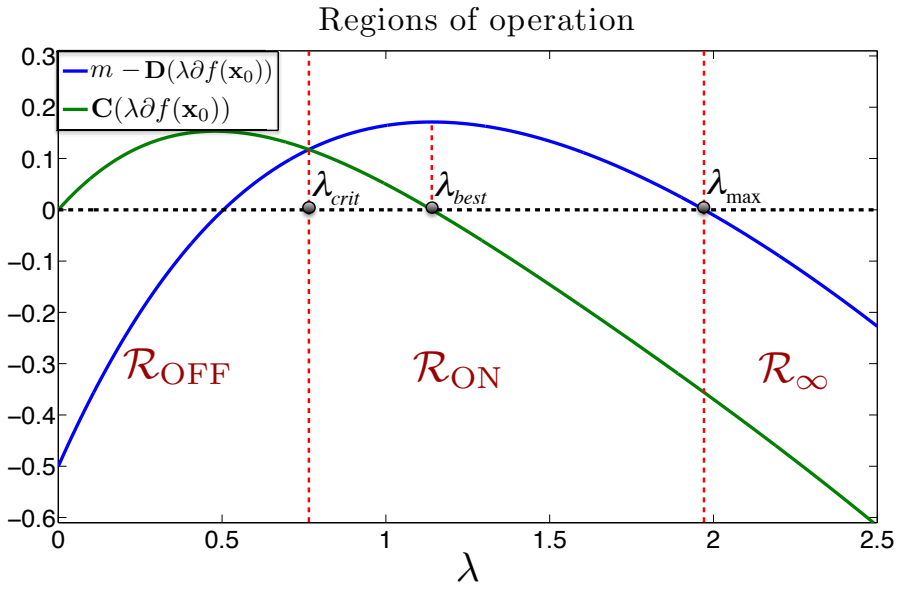

Figure 7.2: Regions of operation of the ℓ_2-LASSO.

1. λ_{best} : λ_{best} is optimal in the sense that the NSE is minimized for this particular choice of the penalty parameter (see Section 7.6). This also explains the term “best" we associate with it.

2. $\lambda_{\max}$: Over $\lambda \geq \lambda_{\text{best}}$, the equation $m = \mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ has a unique solution. We denote this solution by $\lambda_{\max}$. For values of λ larger than $\lambda_{\max}$, we have $m \leq \mathbf{D}(\lambda\partial f(\mathbf{x}_0))$.

3. λ_{crit} : Over $0 \leq \lambda \leq \lambda_{\text{best}}$, if $m \leq n$, the equation $m - \mathbf{D}(\lambda\partial f(\mathbf{x}_0)) = \mathbf{C}(\lambda\partial f(\mathbf{x}_0))$ has a unique solution which we denote λ_{crit}. Otherwise, it has no solution and $\lambda_{\text{crit}} := 0$.

Based on the above definitions, we recognize the three distinct regions of operation of the ℓ_2-LASSO, as follows:

1. $\mathcal{R}_{\text{ON}} = \{\lambda \in \mathbb{R}^+ \big| \lambda_{\text{crit}} < \lambda < \lambda_{\max}\}$.

2. $\mathcal{R}_{\text{OFF}} = \{\lambda \in \mathbb{R}^+ \big| \lambda \leq \lambda_{\text{crit}}\}$.

3. $\mathcal{R}_{\infty} = \{\lambda \in \mathbb{R}^+ \big| \lambda \geq \lambda_{\max}\}$.

See Figure 7.2 for an illustration of the definitions above and Section 8 in [OTH13b] for the detailed proofs of the statements.

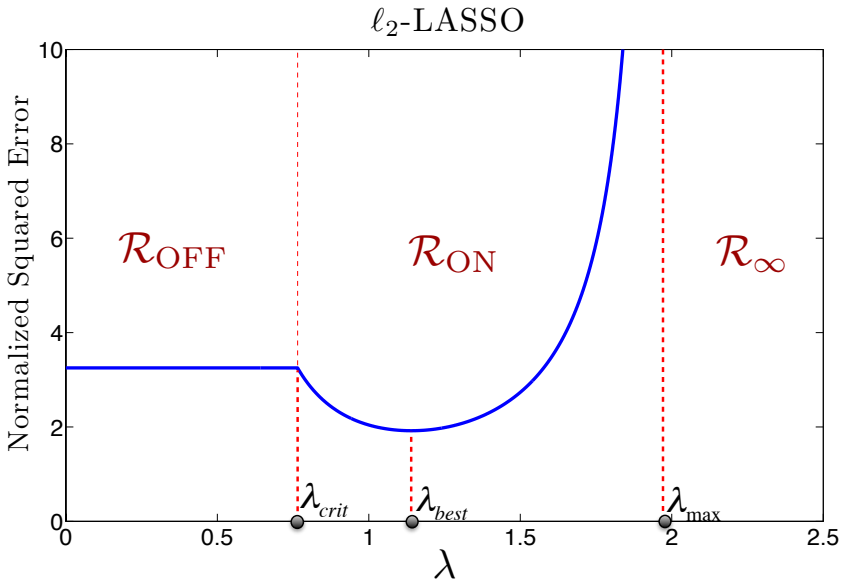

Figure 7.3: We consider the ℓ_1-penalized ℓ_2-LASSO problem for a k sparse signal in $\mathbb{R}^n$. For $\frac{k}{n} = 0.1$ and $\frac{m}{n} = 0.5$, we have $\lambda_{\text{crit}} \approx 0.76$, $\lambda_{\text{best}} \approx 1.14$, $\lambda_{\max} \approx 1.97$.

Characterizing the NSE in each Region

Theorem 7.6.1 upper bounds the NSE of the ℓ_2-LASSO in $\mathcal{R}_{ON}$. Here, we also briefly discuss some observations that can be made regarding $\mathcal{R}_{\text{OFF}}$ and $\mathcal{R}_{\infty}$:

- $\mathcal{R}_{\text{ON}}$: Begin with observing that $\mathcal{R}_{\text{ON}}$ is a nonempty and open interval. In particular, $\lambda_{\text{best}} \in \mathcal{R}_{\text{ON}}$ since $m > \mathbf{D}(\lambda_{\text{best}} \cdot \partial f(\mathbf{x}_0))$. Theorem 7.6.1 proves that for all $\lambda \in \mathcal{R}_{\text{ON}}$ and for σ sufficiently small,

$$\frac{\|\hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0\|}{m\sigma^2} \approx \frac{\mathbf{D}(\lambda \partial f(\mathbf{x}_0))}{m - \mathbf{D}(\lambda \partial f(\mathbf{x}_0))}. \tag{7.36}$$

  Also, empirical observations suggest that (7.36) holds for arbitrary σ when $\approx$ replaced with $\lesssim$. Finally, we should note that the NSE formula $\mathbf{D}(\lambda \partial f(\mathbf{x}_0))/(m - \mathbf{D}(\lambda \partial f(\mathbf{x}_0)))$ is a convex function of λ over $\mathcal{R}_{\text{ON}}$.

- $\mathcal{R}_{\text{OFF}}$: For $\lambda \in \mathcal{R}_{\text{OFF}}$, the LASSO estimate $\hat{\mathbf{x}}_{\ell_2}$ satisfies $\mathbf{y} = \mathbf{A}\hat{\mathbf{x}}_{\ell_2}$ and the optimization (7.2) reduces to the standard ℓ_1 minimization (2.1) of noiseless CS. This suggests that when $\sigma \to 0$,

$$\frac{\|\hat{\mathbf{x}}_{\ell_2} - \mathbf{x}_0\|}{m\sigma^2} \approx \frac{\mathbf{D}(\lambda_{\text{crit}} \cdot \partial f(\mathbf{x}_0))}{(m - \mathbf{D}(\lambda_{\text{crit}} \cdot \partial f(\mathbf{x}_0)))}, \text{ for all } \lambda \in \mathcal{R}_{\text{OFF}}. \tag{7.37}$$

  In [OTH13b, Lem. 9.2], we prove this only for sufficiently small values of λ; no complete rigorous proof of the non asymptotic statement in (7.37) is

available. However, we have shown in [TOH15, Thm. 6] that the statement is true asymptotically. We omit the details; also see the remark following.

Remark 7.6.0.50. Essentially, the technical reason why we have only been able to characterize the NSE in $\mathcal{R}_{\text{ON}}$ is the following. In that regime the optimal value of β in the (AO) in (7.30) is 1 [OTH13b]. As we saw in Lemma 7.4.1 this allows evaluation of the optimal value of the (AO) in closed-form. On the other hand, when $\lambda \in \mathcal{R}_{\text{OFF}}$ it is no more straightforward how to optimize (7.30) over β. However, following the asymptotic framework of Chapter 5 the analysis becomes possible and leads to an asymptotic version of (7.37). We omit the details for brevity, but refer the interested reader to [TOH15, Thm. 6].

- $\mathcal{R}_{\infty}$: Empirically, we observe that the stable recovery of $\mathbf{x}_0$ is not possible for $\lambda \in \mathcal{R}_{\infty}$.

Optimal Tuning of the Penalty Parameter

It is not hard to see that the formula in (7.36) is strictly increasing in $\mathbf{D}(\lambda \partial f(\mathbf{x}_0))$. Thus, when $\sigma \to 0$, the NSE achieves its minimum value when the penalty parameter is set to λ_{best}. Recall from (7.35) that $\mathbf{D}(\lambda_{\text{best}} \cdot \partial f(\mathbf{x}_0)) \approx \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ and compare the formulae in Theorems 7.5.1 and 7.6.1, to conclude that the C-LASSO and ℓ_2-LASSO can be related by choosing $\lambda = \lambda_{\text{best}}$. In particular, we have,

$$\frac{\|\hat{\mathbf{x}}_{\ell_2}(\lambda_{\text{best}}) - \mathbf{x}_0\|^2}{m\sigma^2} \approx \frac{\mathbf{D}(\lambda_{\text{best}} \cdot \partial f(\mathbf{x}_0))}{m - \mathbf{D}(\lambda_{\text{best}} \cdot \partial f(\mathbf{x}_0))} \approx \frac{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}{m - \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} \approx \frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|^2}{m\sigma^2}.$$

It is important to note that deriving λ_{best} does not require knowledge of any properties (e.g. variance) of the noise vector neither does it require knowledge of the unknown signal $\mathbf{x}_0$ itself. All it requires is knowledge of the particular structure of the unknown signal. For example, in the ℓ_1-case, λ_{best} depends only on the sparsity of $\mathbf{x}_0$, not $\mathbf{x}_0$ itself, and in the nuclear norm case, it only depends on the rank of $\mathbf{x}_0$, not $\mathbf{x}_0$ itself.

7.7 ℓ_2^2-LASSO

An Early Conjecture

In [OTH13b] we proposed a mapping between the penalty parameters λ of the ℓ_2-LASSO program (7.2) and τ of the ℓ_2^2-LASSO program (7.3), for which the NSE of the two problems behaves the same. The mapping function is defined as follows.

Definition 7.7.1 (Mapping Function)**.** For any $\lambda \in \mathcal{R}_{\mathrm{ON}}$, define

$$\mathrm{map}(\lambda) = \lambda \, \frac{m - \mathbf{D}(\lambda \partial f(\mathbf{x}_0)) - \mathbf{C}(\lambda \partial f(\mathbf{x}_0))}{\sqrt{m - \mathbf{D}(\lambda \partial f(\mathbf{x}_0))}}.$$

Observe that $\mathrm{map}(\lambda)$ is well-defined over the region $\mathcal{R}_{\mathrm{ON}}$, since $m > \mathbf{D}(\lambda \partial f(\mathbf{x}_0))$ and $m - \mathbf{D}(\lambda \partial f(\mathbf{x}_0)) > \mathbf{C}(\lambda \partial f(\mathbf{x}_0))$ for all $\lambda \in \mathcal{R}_{\mathrm{ON}}$. It can be proven that $\mathrm{map}(\cdot)$ defines a bijective mapping from $\mathcal{R}_{\mathrm{ON}}$ to $\mathbb{R}^+$ [OTH13b, Theorem 3.3] .

Theorem 7.7.1 (Properties of $\mathrm{map}(\cdot)$)**.** *Assume* $m > \min_{\lambda \geq 0} \mathbf{D}(\lambda \partial f(\mathbf{x}_0))$*. The function* $\mathrm{map}(\cdot) : \mathcal{R}_{ON} \to \mathbb{R}^+$ *is strictly increasing and continuous. Thus, its inverse function* $\mathrm{map}^{-1}(\cdot) : \mathbb{R}^+ \to \mathcal{R}_{ON}$ *is well defined.*

Some other useful properties of the mapping function include the following:

- $\mathrm{map}(\lambda_{\mathrm{crit}}) = 0$,
- $\lim_{\lambda \to \lambda_{\max}} \mathrm{map}(\lambda) = \infty$.

Based on this mapping, we conjectured translating the results on the NSE of the ℓ_2-LASSO over $\mathcal{R}_{\mathrm{ON}}$ to corresponding results on the ℓ_2^2-LASSO for $\tau \in \mathbb{R}^+$ as follows: When $m > \mathbf{D}(\lambda_{\mathrm{best}} \cdot \partial f(\mathbf{x}_0)))$, it had been conjectured in [OTH13b] that, for any $\tau > 0$,

$$\frac{\mathbf{D}(\mathrm{map}^{-1}(\lambda) \cdot \partial f(\mathbf{x}_0))}{m - \mathbf{D}(\mathrm{map}^{-1}(\lambda) \cdot \partial f(\mathbf{x}_0))} \tag{7.38}$$

accurately characterizes the NSE $\|\hat{\mathbf{x}}_{\ell_2^2} - \mathbf{x}_0\|^2/(m\sigma^2)$ for sufficiently small σ, and upper bounds it for arbitrary σ. The claim was supported by extended numerical simulations (see Section 13 in [OTH13b]).

Proving the Conjecture

We were able to establish the validity of the conjecture in [TPH15] in an asymptotic setting. We formally state the result here.

For convenience denote the *normalized* Gaussian distance squared and Gaussian correlation as

$$\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)(\lambda) := \mathbf{D}(\lambda \partial f(\mathbf{x}_0))/n \quad \text{and} \quad \overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau)(\lambda) := \mathbf{C}(\lambda \partial f(\mathbf{x}_0))/n. \tag{7.39}$$

To familiarize with these definitions, it is instructive to specialize to the case where $f = \|\cdot\|_1$ and $\mathbf{x}_0$ is a k-sparse vector, with $k/n = \rho \in (0,1)$. Then, $\partial f(\mathbf{x}_0)$ has a simple characterization and $\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$, $\overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau)$ admit simple closed-form expressions in terms of the tail distribution $Q(\tau)$ of a standard Gaussian (e.g.,[OTH13b, App. H]):

$$\begin{aligned}\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau) &= \rho(1+\tau)^2 + (1-\rho)(2(1+\tau^2)Q(\tau) - \sqrt{2/\pi}\tau e^{-\frac{\tau^2}{2}}),\\ \overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau) &= -\rho\tau^2 + (1-\rho)(2\tau^2 Q(\tau) - \sqrt{2/\pi}\tau e^{-\frac{\tau^2}{2}}).\end{aligned} \tag{7.40}$$

Our results hold in the linear asymptotic regime as in Chapter 4. In particular, we assume (i) $m/n = \delta \in (0,\infty)$, with δ a constant, and, (ii) $\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau) \in (0,1)$ is a constant for any constant $\tau > 0$. Here and onwards, "constant" indicates a number that is independent of the problem dimensions. For example, in the case of sparse recovery, choosing $f = \|\cdot\|_1$ and $\mathbf{x}_0$ to be k-sparse, with $k/n = \rho \in (0,1)$, it follows from (7.40) that $\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$ is constant independent of n and k for all $\tau > 0$.

We have the following asymptotic versions of Definition 7.7.1 and Theorem 7.7.1

Definition 7.7.2 (map)**.** Let $\mathcal{R}_{\text{ON}} := \{\tau > 0 | \delta - \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau) > \max\{0, \overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau)\}\}$ and define map : $\mathcal{R}_{\text{ON}} \to (0,\infty)$:

$$\text{map}(\tau) := \tau \frac{\delta - \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau) - \overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau)}{\sqrt{\delta}\sqrt{\delta - \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)}}. \tag{7.41}$$

Lemma 7.7.1 shows that the inverse of map is well defined.

Lemma 7.7.1 (map^{-1})**.** *Assume* $\delta > \min_{\tau>0} \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$. *Then,* $\mathcal{R}_{ON}$ *is a nonempty open interval and* map *is strictly increasing, continuous and bijective. In particular, its inverse function* map$^{-1} : (0,\infty) \to \mathcal{R}_{ON}$ *is well defined.*

Theorem 7.7.1 characterizes the limiting behavior of the asymptotic normalized squared error of (7.3).

Theorem 7.7.1. *Fix any* $\lambda > 0$ *in* (7.3) *and let*

$$aNSE := \lim_{\sigma\to 0} NSE(\sigma) = \lim_{\sigma\to 0} \frac{\|\hat{\mathbf{x}}_{\ell_2^2} - \mathbf{x}_0\|_2^2}{m\sigma^2}.$$

Assume a linear asymptotic regime in which $m/n \to \delta \in (0,1)$ *and* $\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$, $\overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau)$ *are also constants. If* $\delta > \min_{\tau>0} \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$*, then, the following limit holds in probability*

$$\lim_{n\to\infty} aNSE = \frac{\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\mathrm{map}^{-1}(\lambda))}{\delta - \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\mathrm{map}^{-1}(\lambda))} =: \eta(\lambda).$$

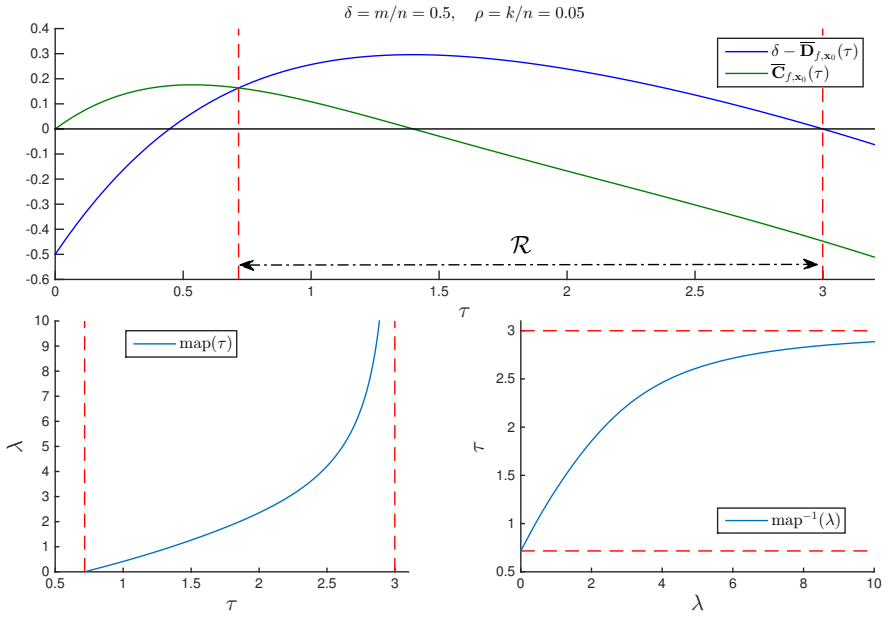

Figure 7.4: Illustration of the region $\mathcal{R}_{\mathrm{ON}}$ and of the map function (Defn. 7.7.2) for $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^n$ a k-sparse vector. map^{-1} maps the value of the regularizer λ in (7.3) to a value in $\mathcal{R}_{\mathrm{ON}}$. $\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$ and $\overline{\mathbf{C}}_{f,\mathbf{x}_0}(\tau)$ are computed as in (7.40).

.

Remarks

Remark 7.7.0.51 (The mapping). The theorem maps $\lambda > 0$ to some value $\tau \in \mathcal{R}_{\mathrm{ON}}$ through map^{-1}. Note that $\mathcal{R}_{\mathrm{ON}}$ is *nonempty* as long as $\frac{m}{n} > \min_\tau \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$ (Lemma 7.7.1). Figure 7.4 illustrates the action of map^{-1} for an instance of a sparse recovery problem.

Remark 7.7.0.52 (Optimal tuning). Thm. 7.7.1 suggests a simple recipe for finding the optimal value λ_{best} of the regularizer parameter.

Lemma 7.7.2. *Recall $\eta(\lambda)$ as defined in Theorem 7.7.1. Let $\lambda_{best} := \arg\min_{\lambda\geq 0} \eta(\lambda)$ and $\tau_{best} := \arg\min_{\tau\geq 0} \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$. Then, $\lambda_{best} = \tau_{best}\sqrt{1 - \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau_{best})/\delta}$.*

The proof of the lemma is not involved and is omitted for brevity. Recall from 7.6.1 that $\overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$ is strictly convex. Thus, τ_{best} can be efficiently calculated as the unique solutions to a convex program. This determines λ_{best}. Note that even though calculating λ_{best} does not require explicit knowledge of $\mathbf{x}_0$ itself, it does assume

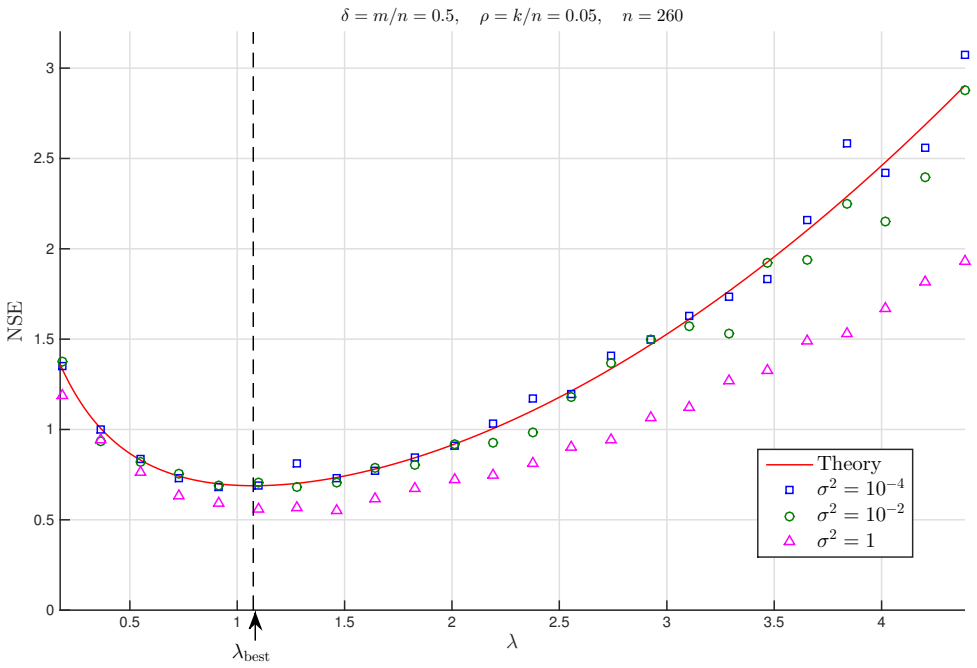

Figure 7.5: Numerical validation of Theorem 7.7.1 for $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^n$ a k-sparse vector. Measured values of the NSE(σ) are averages over 50 realizations of $\mathbf{A}, \mathbf{v}$. The theorem accurately predicts NSE(σ) as $\sigma \to 0$. The results support our claim that aNSE = wNSE. λ_{best} is the value of the optimal regularizer as predicted by Lemma 2.2.

knowledge of the particular structure. For instance, in sparse recovery we need to know the sparsity level k (see Fig. 8.1).

Remark 7.7.0.53 (Phase-transitions). Combining Theorem 7.7.1 with Lemma 7.7.2 it holds with probability one that,

$$\lim_{\sigma\to 0} \min_{\lambda>0} \frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{m\sigma^2} = \frac{\min_\tau \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)}{\delta - \min_\tau \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)}.$$

In view of the wNSE conjecture in (7.38), the quantity in the left hand side can be viewed as the *minimax* NSE of G-LASSO for a *fixed* signal $\mathbf{x}_0$. While $\delta > \min_\tau D(\tau)$, we can always tune (7.3) to guarantee robust recovery. However, as the normalized number of measurements δ approaches $\min_\tau \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$, then, even after optimal tuning, the NSE grows to ∞. This phase-transition characterizing the robustness of (7.3) is identical to (2.5) of noiseless Compressed Sensing.

Remark 7.7.0.54 (Robustness). Theorem 7.7.1 reveals the following interesting feature of (7.3). Given sufficient number of measurements $m/n > \min_\tau \overline{\mathbf{D}}_{f,\mathbf{x}_0}(\tau)$, the

recovery is *robust* for all choices of the regularizer parameter $\lambda > 0$. In particular, this is in contrast to the ℓ_2-LASSO in (7.2). Recall from Section 7.5 that the NSE of the later becomes unbounded if the regularizer parameter is larger than some $\mu_{\max}$.

Remark 7.7.0.55 (Proof). The proof of Theorem 7.7.1 follows from the CGMT framework of Chapter 5 when applied to the corresponding Approximated LASSO problem (see Section 7.4). The details are omitted for brevity and the interested reader is referred to [TPH15].

It is worth mentioning that at the time of writing of [OTH13b], it wasn't clear to us how to leverage the objective function in (7.3) and bring it in the required min-imax format of the (PO) in (3.11a). Recall (7.27) allowed this in the case of the ℓ_2-LASSO. As discussed, we were only able to *conjecture* a formula for the *a*NSE of the ℓ_2^2-LASSO based on an "educated guess" on a mapping between the ℓ_2^2-LASSO and the ℓ_2-LASSO. Later, in [TPH15] we *rigorously* established the conjecture raised. Instead of worrying about the mapping function between (7.3) and (7.2) and translating the results from the latter to the former, we followed a direct approach. The simple but key observation was that the objective function in (7.3) can be appropriately linearized for the purpose of using the GMT, and be written equivalently as:

$$\min_{\mathbf{x}} \max_{\mathbf{u}} \mathbf{u}^T(\mathbf{y} - \mathbf{A}\mathbf{x}) - (1/2)\|\mathbf{u}\|^2 + \lambda\sigma f(\mathbf{x}).$$

This same idea of expressing the loss function (here, least-squares) in a dual form through its convex conjugate function led to generalization of this type of analysis to other convex loss functions (see Section 5.2).

7.8 The NSE of Generalized LASSO with Arbitrary Fixed Noise

Here, we relax the assumption of Section 7.4 that the entries of $\mathbf{z}$ are i.i.d. normal. Instead, assume that the noise vector $\mathbf{z}$ is arbitrary, but still independent of the sensing matrix $\mathbf{A}$. Under this assumption, we derive simple and non-asymptotic upper bounds on the NSE of the C-LASSO and of the ℓ_2-LASSO. Those upper bounds can be interpreted as generalizations of the bound on the error of the OLS as was discussed in Section 7.2. Compared to the bounds of Section 7.4, the bounds derived here not only hold under more general assumption on the noise vector, but they are also non-asymptotic.

C-LASSO

Recall the generalized C-LASSO in (7.1). We introduce an upper bound on its NSE for arbitrary fixed noise vector that is independent of $\mathbf{A}$ and compare this bound to the result of Theorem 7.5.1. We further provide an overview of the proof technique.

Theorem 7.8.1. *[OTH13a] Assume* $m \geq 2$ *and* $0 < t \leq \sqrt{m-1} - \sqrt{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}$. *Then, with probability,* $1 - 6\exp(-t^2/26)$,

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|}{\|\mathbf{z}\|} \leq \frac{\sqrt{m}}{\sqrt{m-1}} \frac{\sqrt{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} + t}{\sqrt{m-1} - \sqrt{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} - t}.$$

Comparison to Theorem 7.5.1. It is interesting to see how the bound of Theorem 7.8.1 compares to the result of Theorem 7.5.1 in the case $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$. Of course, when this is the case the bound of Theorem 7.5.1 is tight and our intention is to see how loose is the bound of Theorem 7.8.1. Essentially[9], the only difference appears in the denominators of the two bounds; $\sqrt{m - \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} \geq \sqrt{m} - \sqrt{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}$ for all regimes of $0 \leq \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) < m$. The contrast becomes significant when $m \approx \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$. In particular, setting $m = (1+\epsilon)^2\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$, we have,

$$\frac{\sqrt{m - \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))}}{\sqrt{m} - \mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} = \frac{\sqrt{2\epsilon + \epsilon^2}}{\epsilon} = \sqrt{\frac{2}{\epsilon} + 1}.$$

Thus, when ϵ is large, the bound of Theorem 7.8.1 is arbitrarily tight. On the other hand, when ϵ is small, it can be arbitrarily worse. Simulation results (see Figure 7.6) verify that the error bound of Theorem 7.8.1 becomes sharp as the number of measurements m increases. Besides, even if tighter, the bound of Theorem 7.5.1 requires stronger assumptions namely, an i.i.d.. Gaussian noise vector $\mathbf{z}$ and an asymptotic setting where m and $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ are large enough.

Proof Overview. We only provide an overview of the proof. The details can be found in [OTH13a]. We begin with introducing some useful notation. $\mathbf{A}\mathcal{T}_f(\mathbf{x}_0)$ will denote the cone obtained by multiplying elements of $\mathcal{T}_f(\mathbf{x}_0)$ by $\mathbf{A}$., i.e.,

$$\mathbf{A}\mathcal{T}_f(\mathbf{x}_0) = \{\mathbf{A}\mathbf{v} \in \mathbb{R}^m \mid \mathbf{v} \in \mathcal{T}_f(\mathbf{x}_0)\}.$$

The lemma below derives a deterministic upper bound on the squared error of the C-LASSO. It is interesting to compare this to the corresponding bound (7.10) for the OLS. Recall the notions of “tangent cone" and “restricted minimum singular value" introduced in Section 2.2.

[9] Precisely: assuming $m \approx m - 1$ and ignoring the t's in the bound of Theorem 7.8.1.

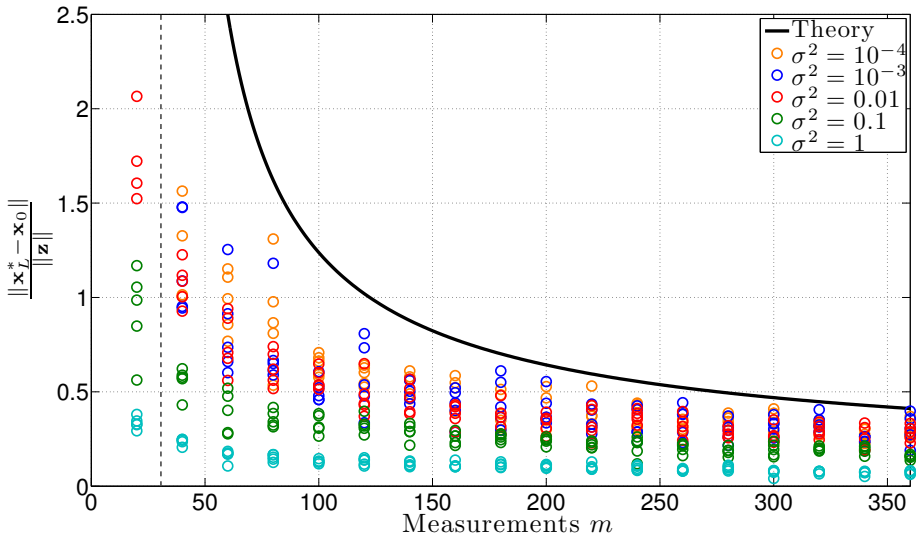

Figure 7.6: NSE of the C-LASSO with ℓ_1-regularization. The unknown signal $\mathbf{x}_0 \in \mathbb{R}^{500}$ is 5-sparse. The number of measurements m varies from 0 to 360. We plot the empirical NSE assuming $\mathbf{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m)$ for several values of σ. The solid black line corresponds to the bound of Theorem 7.8.1. The dashed line corresponds to the phase transition line of noiseless CS, namely $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ (cf. Theorem 2.2.1).

Lemma 7.8.1 (Deterministic error bound)**.**

$$\|\hat{\mathbf{x}}_c - \mathbf{x}_0\| \leq \frac{\|\operatorname{Proj}(\mathbf{z}, \mathbf{A}\mathcal{T}_f(\mathbf{x}_0))\|}{\sigma_{min}(\mathbf{A}, \mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1})}.$$

Proof. From first-order optimality conditions (e.g. [Roc97, p. 270-271]),

$$\left\langle \mathbf{A}^T(\mathbf{A}\hat{\mathbf{x}}_c - \mathbf{y}), \hat{\mathbf{x}}_c - \mathbf{x}_0 \right\rangle \leq 0.$$

Writing $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, and rearranging terms, we find that,

$$\begin{aligned} \|\mathbf{A}(\hat{\mathbf{x}}_c - \mathbf{x}_0)\| &\leq \left\langle \mathbf{z}, \frac{\mathbf{A}(\hat{\mathbf{x}}_c - \mathbf{x}_0)}{\|\mathbf{A}(\hat{\mathbf{x}}_c - \mathbf{x}_0)\|} \right\rangle \\ &\leq \sup_{\mathbf{v} \in \mathbf{A}\mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1}} \langle \mathbf{z}, \mathbf{v} \rangle \qquad (7.42) \\ &\leq \sup_{\mathbf{v} \in \mathbf{A}\mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{B}^{n-1}} \langle \mathbf{z}, \mathbf{v} \rangle \\ &= \|\operatorname{Proj}(\mathbf{z}, \mathbf{A}\mathcal{T}_f(\mathbf{x}_0))\|. \qquad (7.43) \end{aligned}$$

(7.42) follows since $\hat{\mathbf{x}}_c - \mathbf{x}_0 \in \mathcal{T}_f(\mathbf{x}_0)$. For (7.43), we applied Moreau's decomposition Theorem [Roc97, Theorem 31.5]. To conclude with the desired result it remains to invoke the definition of the restricted singular values $\sigma_{min}(\mathbf{A}, \mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1})$. □

To prove Theorem 7.8.1 we will translate the deterministic bound of Lemma 7.8.1 to a probabilistic one. For this, we need a high-probability lower bound for $\sigma_{min}(\mathbf{A}, \mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1})$ and a high-probability upper bound for $\| \operatorname{Proj}(\mathbf{z}, \mathbf{A}\mathcal{T}_f(\mathbf{x}_0))\|$. The former is given by a direct application of the "escape through a mesh" Proposition 2.2.2. The latter requires some more effort to derive. The result is summarized in Lemma 7.8.2 below and is the main technical contribution of [OTH13a]. The proof makes use of Gordon's original GMT Theorem 3.2.1. For a vector $\mathbf{g} \in \mathbb{R}^m$ with independent $\mathcal{N}(0,1)$ entries, we define $\gamma_m := \mathbb{E}[\|\mathbf{g}\|]$. It is well known (e.g. [Gor88]) that $\gamma_m = \sqrt{2}\frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}$ and $\sqrt{m} \geq \gamma_m \geq \frac{m}{\sqrt{m+1}}$.

Lemma 7.8.2 (Restricted correlation)**.** *Let $\mathcal{K} \in \mathbb{R}^n$ be a convex and closed cone, $\mathbf{G} \in \mathbb{R}^{m\times n}$ have independent standard normal entries, $m \geq 2$ and $\mathbf{z} \in \mathbb{R}^m$ be arbitrary and independent of $\mathbf{G}$. For any $t > 0$, pick $\alpha \geq \frac{\sqrt{\mathbf{D}(\mathcal{K}^\circ)}+t}{\gamma_{m-1}}\|\mathbf{z}\|$. Then,*

$$\sup_{\mathbf{v}\in\mathcal{K}\cap\mathcal{S}^{n-1}} \{\mathbf{z}^T\mathbf{G}\mathbf{v} - \alpha\|\mathbf{G}\mathbf{v}\|\} \leq 0, \tag{7.44}$$

with probability $1 - 5\exp(-\frac{t^2}{26})$.

We may now complete the proof of Theorem 7.8.1. Suppose $0 \leq t < \gamma_m - \mathbf{D}(\operatorname{cone}(\partial f(\mathbf{x}_0)))$. First apply Proposition 2.2.2, to find that with probability $1 - \exp(-\frac{t^2}{2})$,

$$\sigma_{min}(\mathbf{A}, \mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1}) \geq \frac{\gamma_m - \sqrt{\mathbf{D}(\operatorname{cone}(\partial f(\mathbf{x}_0)))} - t}{\sqrt{m}}. \tag{7.45}$$

Next, apply Lemma 7.8.2 with $\mathbf{G} = \sqrt{m}\mathbf{A}$ and $\mathcal{K} = \mathcal{T}_f(\mathbf{x}_0)$. With probability $1 - 5\exp(-\frac{t^2}{26})$,

$$\begin{aligned}\| \operatorname{Proj}(\mathbf{z}, \mathbf{A}\mathcal{T}_f(\mathbf{x}_0))\| = \mathbf{z}^T\frac{\operatorname{Proj}(\mathbf{z}, \mathbf{A}\mathcal{T}_f(\mathbf{x}_0))}{\| \operatorname{Proj}(\mathbf{z}, \mathbf{A}\mathcal{T}_f(\mathbf{x}_0))\|} &\leq \sup_{\mathbf{v}\in\mathcal{T}_f(\mathbf{x}_0)\cap\mathcal{S}^{n-1}} \frac{\mathbf{z}^T\mathbf{A}\mathbf{v}}{\|\mathbf{A}\mathbf{v}\|} \\ &\leq \frac{\sqrt{\mathbf{D}(\operatorname{cone}(\partial f(\mathbf{x}_0)))} + t}{\gamma_{m-1}}\|\mathbf{z}\|. \end{aligned} \tag{7.46}$$

Theorem 7.8.1, now follows after substituting (7.45) and (7.46) in Lemma 7.8.1 and using the following: $\gamma_m\gamma_{m-1} = m - 1$ and $\gamma_{m-1} \leq m - 1$.

ℓ_2-LASSO

Consider now the generalized ℓ_2-LASSO in (7.2).

Theorem 7.8.2 ([TOH14]). *Assume $m \geq 2$. Fix the regularizer parameter in (7.2) to be $\lambda \geq 0$ and let $\hat{\mathbf{x}}_{\ell_2}$ be a minimizer of (7.2). Then, for any $0 < t \leq (\sqrt{m-1} - \sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))})$, with probability $1 - 5\exp(-t^2/32)$,*

$$\|\hat{\mathbf{x}} - \mathbf{x}_0\| \leq 2\|\mathbf{z}\| \frac{\sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))} + t}{\sqrt{m-1} - \sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))} - t}.$$

Theorem 7.8.2 provides a simple, general, non-asymptotic and (rather) sharp upper bound on the error of the regularized LASSO estimator (7.2), which also takes into account the specific choice of the regularizer parameter $\lambda \geq 0$. It is non-asymptotic and is applicable in any regime of m, λ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$. Also, the constants involved in it are small making it rather tight[10].

For the bound of Theorem 7.8.2 to be at all meaningful, we require $m > \min_{\lambda\geq 0} \mathbf{D}(\lambda\partial f(\mathbf{x}_0)) = \mathbf{D}(\lambda_{\text{best}}\partial f(\mathbf{x}_0))$. Recall that this translates to the number of measurements being large enough to at least guarantee noiseless recovery Also, similar to the discussion in Section 7.6 there exists a unique λ_{max} satisfying $\lambda_{max} > \lambda_{\text{best}}$ and $\sqrt{\mathbf{D}(\lambda_{max}\partial f(\mathbf{x}_0))} = \sqrt{m-1}$, and, when $m \leq n$, there exists unique $\lambda_{min} < \lambda_{\text{best}}$ satisfying $\sqrt{\mathbf{D}(\lambda_{min}\partial f(\mathbf{x}_0))} = \sqrt{m-1}$. From this, it follows that $\sqrt{m-1} > \sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))}$ if and only if $\lambda \in (\lambda_{min}, \lambda_{max})$. This is a superset of $\mathcal{R}_{\text{ON}}$ (recall the definition in Section 7.6) and is exactly the range of values of the regularizer parameter λ for which the bound of Theorem 7.8.2 is meaningful.

As a superset of $\mathcal{R}_{\text{ON}}$, $(\lambda_{min}, \lambda_{max})$ contains λ_{best} for which, the bound of Theorem 7.8.2 achieves its minimum value since it is strictly increasing in $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$. Recall from Section 7.6 that deriving λ_{best} does not require knowledge of any properties (e.g. variance) of the noise vector neither does it require knowledge of the unknown signal $\mathbf{x}_0$ itself.

As a final remark, comparing Theorem 7.8.2 to Theorem 7.8.1 reveals the similar nature of the two results. Apart from a factor of 2, the upper bound on the error of the regularized LASSO (7.2) for fixed λ, is essentially the same as the upper bound on the error of the constrained LASSO (7.1), with $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ replaced by $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$. This when combined with (7.35) suggests that setting $\lambda = \lambda_{\text{best}}$ in (7.2) achieves performance almost as good as that of the constrained LASSO (7.1).

Comparison to Theorem 7.6.1. To start with, Theorem 7.8.2 is advantageous to Theorem 7.6.1 in that it holds in a more general setting than standard Gaussian noise

[10] It is conjectured in [TOH14] and supported by simulations (e.g. Figure 7.7) that the factor of 2 in Theorem 7.8.2 is an artifact of the proof technique and not essential.

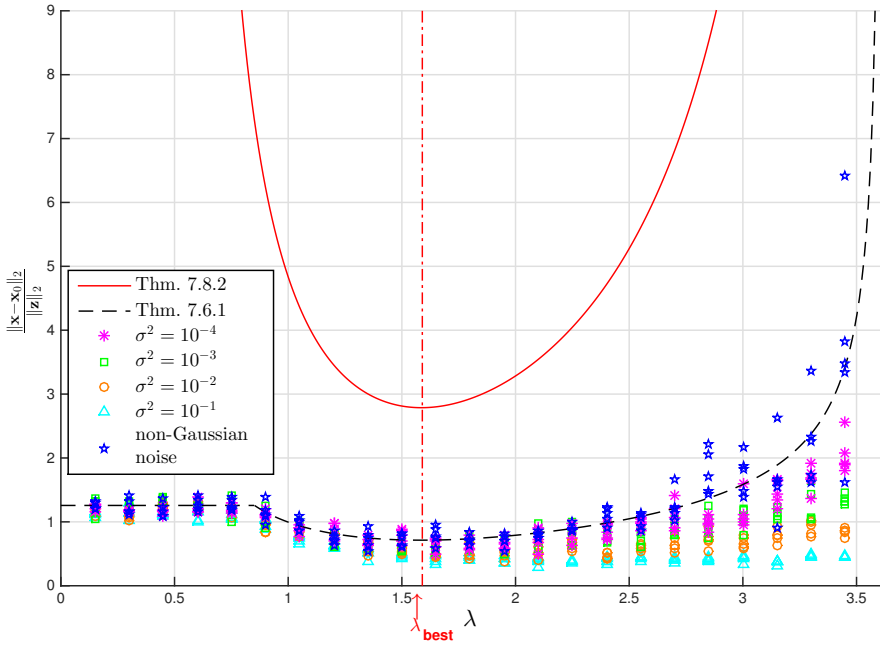

Figure 7.7: Figure 7.7 illustrates the bound of Theorem 7.8.2, which is given in red for $n = 340$, $m = 140$, $k = 10$ and for $\mathbf{A}$ having $\mathcal{N}(0, \frac{1}{m})$ entries. The upper bound of Theorem 7.6.1, which is asymptotic in m and only applies to i.i.d. Gaussian $\mathbf{z}$, is given in black. In our simulations, we assume $\mathbf{x}_0$ is a random unit norm vector over its support and consider both i.i.d. $\mathcal{N}(0, \sigma^2)$ as well as non-Gaussian noise vectors $\mathbf{z}$. We have plotted the realizations of the normalized error for different values of λ and σ. As noted, the bound of Theorem 7.6.1 is occasionally violated since it requires very large m, as well as, i.i.d. Gaussian noise. On the other hand, the bound of Theorem 7.8.2 always holds.

and, also, characterizes a superset of $\mathcal{R}_{\text{ON}}$. Furthermore, it is non-asymptotic, while Theorem 7.6.1 requires $m, \mathbf{D}(\lambda \partial f(\mathbf{x}_0))$ to be large enough. On the other side, when $\mathbf{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m)$, then, Theorem 7.6.1 offers clearly a tighter bound on the NSE. Yet, apart from a factor of 2, this bound only differs from the bound of Theorem 7.8.2 in the denominator, where instead of $\sqrt{m-1} - \sqrt{\mathbf{D}(\lambda \partial f(\mathbf{x}_0))}$ we have the larger quantity $\sqrt{m - \mathbf{D}(\lambda \partial f(\mathbf{x}_0))}$. This difference becomes insignificant and indicates that our bound is rather tight when m is large. Finally, the bound of Theorem 7.6.1 is only conjectured in [OTH13b] to upper bound the estimation error for arbitrary values of the noise variance σ^2. In contrast, Theorem 7.8.2 is a fully rigorous upper bound on the estimation error of (7.2).

Proof Overview. It is convenient to rewrite the generalized ℓ_2-LASSO in terms of

the error vector $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$ as follows:

$$\min_{\mathbf{w}} \|\mathbf{A}\mathbf{w} - \mathbf{z}\| + \frac{\lambda}{\sqrt{m}}(f(\mathbf{x}_0 + \mathbf{w}) - f(\mathbf{x}_0)). \tag{7.47}$$

Denote the solution of (7.47) by $\hat{\mathbf{w}}$. Then, $\hat{\mathbf{w}} = \hat{\mathbf{x}} - \mathbf{x}_0$ and we want to bound $\|\hat{\mathbf{w}}\|$. To simplify notation, for the rest of the proof, we denote the value of that desired upper bound as

$$\ell(t) := 2\|\mathbf{z}\| \frac{\sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))} + t}{\sqrt{m-1} - \sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))} - t}. \tag{7.48}$$

It is easy to see that the optimal value of the minimization in (7.47) is no greater than $\|\mathbf{z}\|$. Observe that $\mathbf{w} = \mathbf{0}$ achieves this value. However, Lemma 7.8.3 below shows that if we constrain the minimization in (7.47) to be only over vectors $\mathbf{w}$ whose norm is greater than $\ell(t)$, then the resulting optimal value is (with high probability on the measurement matrix $\mathbf{A}$) strictly greater than $\|\mathbf{z}\|$. Combining those facts yields the desired result, namely $\|\hat{\mathbf{w}}\| \leq \ell(t)$. The fundamental technical tool in the proof of Lemma 7.8.3 is (not surprisingly at this point) Gordon's Lemma 3.2.1.

Lemma 7.8.3. *Fix some $\lambda \geq 0$ and $0 < t \leq (\sqrt{m-1} - \sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))})$. Let $\ell(t)$ be defined as in (7.48). Then, with probability $1 - 5\exp(-t^2/32)$, we have,*

$$\min_{\|\mathbf{w}\| \geq \ell(t)} \{\|\mathbf{A}\mathbf{w} - \mathbf{z}\| + \frac{\lambda}{\sqrt{m}}(f(\mathbf{x}_0 + \mathbf{w}) - f(\mathbf{x}_0))\} > \|\mathbf{z}\|. \tag{7.49}$$

Proof. Fix λ and t, as in the statement of the lemma. From the convexity of $f(\cdot)$, $f(\mathbf{x}_0 + \mathbf{w}) - f(\mathbf{x}_0) \geq \max_{\mathbf{s} \in \partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w}$. Hence, it suffices to prove that w.h.p. over $\mathbf{A}$,

$$\min_{\|\mathbf{w}\| \geq \ell(t)} \{\sqrt{m}\|\mathbf{A}\mathbf{w} - \mathbf{z}\| + \max_{\mathbf{s} \in \lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w}\} > \sqrt{m}\|\mathbf{z}\|.$$

We begin with applying Gordon's Lemma 3.2.1 to the optimization problem in the expression above. Define, $\overline{\mathbf{z}} = \sqrt{m}\mathbf{z}$, rewrite $\|\mathbf{A}\mathbf{w} - \mathbf{z}\|$ as $\max_{\|\mathbf{a}\|=1}\{\mathbf{a}^T\mathbf{A}\mathbf{w} - \mathbf{a}^T\mathbf{z}\}$ and then apply Lemma 3.2.1 with $\mathcal{S} = \{\mathbf{w} \mid \|\mathbf{w}\| \geq \ell(t)\}$ and $\psi(\mathbf{w}, \mathbf{a}) = -\mathbf{a}^T\overline{\mathbf{z}} + \max_{\mathbf{s} \in \lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w}$. This leads to the following statement:

$$\mathbb{P}\left(\text{ (7.49) is true }\right) \geq 2 \cdot \mathbb{P}\left(\ \mathcal{L}(t; \mathbf{g}, \mathbf{h}) > \|\overline{\mathbf{z}}\|\ \right) - 1,$$

where, $\mathcal{L}(t; \mathbf{g}, \mathbf{h})$ is defined as

$$\min_{\|\mathbf{w}\| \geq \ell(t)} \max_{\|\mathbf{a}\|=1} \{(\|\mathbf{w}\|\mathbf{g} - \overline{\mathbf{z}})^T\mathbf{a} - \min_{\mathbf{s} \in \lambda\partial f(\mathbf{x}_0)} (\mathbf{h} - \mathbf{s})^T\mathbf{w}\}. \tag{7.50}$$

In the remaining, we analyze the simpler optimization problem defined in (7.50), and prove that $\mathcal{L}(t;\mathbf{g},\mathbf{h}) > \|\overline{\mathbf{z}}\|$ holds with probability $1-\frac{5}{2}\exp(-t^2/32)$. We begin with simplifying the expression for $\mathcal{L}(t;\mathbf{g},\mathbf{h})$, as follows:

$$\begin{aligned}\mathcal{L}(t;\mathbf{g},\mathbf{h}) &= \min_{\|\mathbf{w}\|\geq\ell(t)}\{\|\|\mathbf{w}\|\mathbf{g}-\overline{\mathbf{z}}\| - \min_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)}(\mathbf{h}-\mathbf{s})^T\mathbf{w}\}\\ &= \min_{\alpha\geq\ell(t)}\{\|\alpha\mathbf{g}-\overline{\mathbf{z}}\| - \alpha\,\mathrm{dist}(\mathbf{h},\lambda\partial f(\mathbf{x}_0))\}\\ &= \min_{\alpha\geq\ell(t)}\{\sqrt{\alpha^2\|\mathbf{g}\|^2+\|\overline{\mathbf{z}}\|^2-2\alpha\mathbf{g}^T\overline{\mathbf{z}}} - \alpha\,\mathrm{dist}(\mathbf{h},\lambda\partial f(\mathbf{x}_0))\}. \end{aligned} \tag{7.51}$$

The first equality above follows after performing the trivial maximization over $\mathbf{a}$ in (7.50). Next, we show that $\mathcal{L}(t;\mathbf{g},\mathbf{h})$ is strictly greater than $\|\overline{\mathbf{z}}\|$ with the desired high probability over realizations of $\mathbf{g}$ and $\mathbf{h}$. Consider the event $\mathcal{E}_t$ of $\mathbf{g}$ and $\mathbf{h}$ satisfying all three conditions listed below,

$$1.\ \|\mathbf{g}\| \geq \gamma_m - t/4, \tag{7.52a}$$

$$2.\ \mathrm{dist}(\mathbf{h},\lambda\partial f(\mathbf{x}_0)) \leq \sqrt{\mathbf{D}(\lambda\partial f(\mathbf{x}_0))} + t/4, \tag{7.52b}$$

$$3.\ \mathbf{g}^T\overline{\mathbf{z}} \leq (t/4)\|\overline{\mathbf{z}}\|. \tag{7.52c}$$

The conditions in (7.52) hold with high probability. In particular, the first two hold with probability no less than $1-\exp(-t^2/32)$. This is because the ℓ_2-norm and the distance function to a convex set are both 1-Lipschitz functions and, thus, Proposition 3.1.1 applies. The third condition holds with probability at least $1-(1/2)\exp(-t^2/32)$, since $\mathbf{g}^T\overline{\mathbf{z}}$ is statistically identical to $\mathcal{N}(0,\|\overline{\mathbf{z}}\|^2)$. Union bounding yields,

$$\mathbb{P}(\mathcal{E}_t) \geq 1-(5/2)\exp(-t^2/32). \tag{7.53}$$

Furthermore, it can be shown (see Lemma 4.2 in [TOH14]) that if $\mathbf{g}$ and $\mathbf{h}$ are such that $\mathcal{E}_t$ is satisfied, then $\mathcal{L}(t;\mathbf{g},\mathbf{h}) > \|\overline{\mathbf{z}}\|$. This, when combined with (7.53) shows that $\mathbb{P}(\mathcal{L}(t;\mathbf{g},\mathbf{h}) > \|\overline{\mathbf{z}}\|) \geq 1-(5/2)\exp(-t^2/32)$, completing the proof of Lemma 7.8.3. □

7.9 The Worst-Case NSE of Generalized LASSO

Here, we assume no restriction at all on the distribution of the noise vector $\mathbf{z}$. In particular, this includes the case of *adversarial* noise, i.e., noise that has information on $\mathbf{A}$ and can adapt itself accordingly. We compute the resulting worst-case NSE of the C-LASSO in the next section.

Theorem 7.9.1. *Assume* $0 < t \leq \sqrt{m} - \sqrt{\mathbf{D}(\lambda \partial f(\mathbf{x}_0))}$. *Then, with probability* $1 - \exp(-t^2/2)$,

$$\frac{\|\hat{\mathbf{x}}_c - \mathbf{x}_0\|}{\|\mathbf{z}\|} \leq \frac{\sqrt{m}}{\gamma_m - \sqrt{\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))} - t}.$$

Recall in the statement of Theorem 7.9.1 that $\gamma_m = \mathbb{E}[\|\mathbf{g}\|]$, with $\mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_m)$. For large m, $\gamma_m \approx \sqrt{m}$ and according to Theorem 7.9.1, the worst-case NSE of the C-LASSO can be as large as 1. Contrast this to Theorem 7.8.1 and the case where $\mathbf{z}$ is not allowed to depend on $\mathbf{A}$. There, the NSE is approximately $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))/m$ for large m.

Proof. The proof of Theorem 7.9.1 follows easily from Lemma 7.8.1:

$$\|\hat{\mathbf{x}}_c - \mathbf{x}_0\| \leq \frac{\|\mathbf{A}(\hat{\mathbf{x}}_c - \mathbf{x}_0)\|}{\sigma_{min}(\mathbf{A}, \mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1})} \leq \frac{\|\mathbf{z}\|}{\sigma_{min}(\mathbf{A}, \mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1})}.$$

Apply (7.45) to the above, to conclude with the desired upper bound.

Chapter 8

BEYOND IID ENSEMBLES: ISOTROPICALLY RANDOM ORTHOGONAL MATRICES

The results of all previous chapters are proved under the assumption of the entries of the measurement matrix being iid Gaussian. Beyond the Gaussian assumption, numerical evidence suggests the same error predictions holding true for design matrices with entries iid drawn from a wider class of probability distributions. These are further partially supported by recent rigorous theoretical justifications [OT15; EK15].

In this chapter, we take a step further by relaxing the iid-ness assumption on the distribution of the measurement matrix and considering Isotropically Random Orthogonal (IRO) matrices. IRO matrices are sampled uniformly from the manifold of row-orthogonal matrices satisfying $\mathbf{A}\mathbf{A}^T = \mathbf{I}_m$, and are occasionally referred to as being “Haar distributed". Matrices with orthogonal rows are often preferred in practice because their condition number is one and they do not amplify the noise. As a result they have superior noise performance, something we shall also observe. Furthermore, certain classes of orthogonal matrices, such as Fourier, discrete-cosine and Hadamard allow for fast multiplication and reduced complexity.

While in the noiseless case, the performance of IRO matrices is, as discussed in Section 2.2) no different than that of iid matrices, we show things are different in the noisy setting. We precisely characterize the error performance of the generalized-LASSO under IRO matrices and show that it is superior to the error performance of Gaussian designs. Interestingly, we empirically observe the following universality property of IRO matrices: the derived error formulae for IRO matrices hold true for random DCT and Hadamard matrices.

We begin in Section 8.1 with an overview of the results, emphasizing on the difference in the performance of IRO matrices to Gaussian ones. The formal statements of the results are presented in Section 8.2 and a proof outline is included in Section 8.3.

8.1 Introduction

We assume compressed measurements ($\delta := \frac{m}{n} < 1$) and iid Gaussian noise of variance σ^2. For the signal recovery we use the Constrained LASSO[1]:

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \ \text{ subject to } \ f(\mathbf{x}) \leq f(\mathbf{x}_0). \tag{8.1}$$

and measure its performance with the Normalized Squared Error (NSE):

$$\text{NSE}(\sigma) := \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2/\sigma^2. \tag{8.2}$$

Noiseless Case

In the noiseless case recall by Theorem 2.2.1 that $\arg\min_{\{\mathbf{x}|\mathbf{y}=\mathbf{A}\mathbf{x}\}} f(\mathbf{x})$ is the true vector $\mathbf{x}_0$ with probability approaching one if and only if the compression rate δ satisfies

$$\delta > \overline{\omega}_{f,\mathbf{x}_0}^2. \tag{8.3}$$

Here, $\overline{\omega}_{f,\mathbf{x}_0}$ is the normalized Guassian-width, i.e.

$$\overline{\omega}_{f,\mathbf{x}_0} = \omega(\mathcal{T}_f(\mathbf{x}_0))/\sqrt{n}.$$

This result is *universal* over the measurement matrix $\mathbf{A}$ over both the Gaussian and the IRO ensemble: $\mathbf{A}$ appears in the optimality conditions only through its nullspace, which in both cases is an isotropically random subspace in $\mathbb{R}^n$ of dimension $n - m$.

Noisy Case

Gaussian Ensemble: Theorem 7.5.1 proves that the NSE of (8.1) under Gaussian measurements is upper bounded by[2]

$$\frac{\overline{\omega}_{f,\mathbf{x}_0}^2}{\delta - \overline{\omega}_{f,\mathbf{x}_0}^2}. \tag{8.4}$$

[1] For simplicity and concreteness, we focus here on the Constrained LASSO. However, note that all the results in this chapter can be readily extended to the regularized versions. Furthermore, we only present results for the high-SNR regime (similar to those in Chapter 7 for Gaussian matrices) but, these are readily extendible to arbitrary SNR values.

[2] Formally, (8.4) is an asymptotic version of Theorem 7.5.1. The Gaussian distance squared is replaced here by the Gaussian width. This is allowed by (2.8). Also, note that in contrast to Chapter 7, here we assume the entries of $\mathbf{A}$ have unit variance; this explains the difference in scaling with m between (8.2) and (7.22).

The bound is *precise* (or *asymptotically tight*) since it is shown to be achieved with equality in the limit $\sigma \to 0$.

IRO Ensemble : Unlike the noiseless case, in the noisy setting IRO matrices exhibit different recovery performance than that of Gaussians. Using the replica method from statistical physics and through extensive simulation results, [VKC13; Wen+14b] derive expressions that characterize the NSE of (8.1) and report that orthogonal constructions provide a *superior* performance compared to their Gaussian counterparts. As mentioned in [VKC13], even though it provides a powerful tool for tackling hard analytical problems, the replica method still lacks mathematical rigor in some parts [VKC13]. As a follow up to these reports, and also driven by the fact that orthogonal constructions are easier to implement in practical applications [Wen+14b], it is of interest to prove precise bounds on the achieved NSE; ones that would resemble those of Section 7.5 for Gaussian constructions. Towards this direction, Oymak and Hassibi showed in [OH14] that the noisy performance of IRO matrices is at *at least as* good as that of Gaussians. To conclude this, they proved that the *minimum conic singular value* (mCSV) of the former can be no smaller than that of the latter. Recall from Section 2.2 that mCSVs appear naturally as a measure of noise robustness performance (e.g.[Cha+12, Cor. 3.3]), thus, the achieved NSE of IRO can be no worse than that of Gaussians. Adding to this, [OH14] *conjectures* a formula to bound the NSE of (8.1) when $\mathbf{A}$ is IRO.

Contribution

We prove in Theorem 8.2.1 that when the measurement matrix $\mathbf{A}$ is IRO, then the NSE of (8.1) in the high-SNR regime ($\sigma \to 0$) behaves precisely as[3]:

$$(1 - \overline{\omega}^2_{f,\mathbf{x}_0}) \left(\frac{\overline{\omega}^2_{f,\mathbf{x}_0}}{\delta - \overline{\omega}^2_{f,\mathbf{x}_0}} \right). \tag{8.5}$$

As is the case for the Gaussian ensemble (cf. (8.4)), we conjecture this to be the worst-case value of the NSE over all σ. Since $1 - \overline{\omega}^2_{f,\mathbf{x}_0} < 1$, when compared to (8.4), our result implies the superiority in performance of the IRO ensemble when compared to the Gaussian one. In particular, this establishes rigorously the conjecture raised in [OH14]. Our second result in Theorem 8.2.2 derives a high-probability lower bound on the mCSV of IRO matrices. The bound is seen to exceed the corresponding well-known bound for Gaussian matrices.

[3]The formula in (8.5) holds for IRO matrix $\mathbf{A}$ scaled such that $\mathbf{A}\mathbf{A}^T = n\mathbf{I}_m$. This is to allow for a fair comparison with i.i.d. standard Gausian matrices for which $\mathbb{E}[\mathbf{A}\mathbf{A}^T] = n\mathbf{I}_m$.

Approach

The set of techniques available for dealing with IRO matrices is limited compared to the variety of methods available for working with Gaussian matrices. Nonetheless, we are able to prove (8.5) based on a modification of the CGMT framework. Recall from Section 7.4 that in order to apply the CGMT Theorem 3.3.1 for the analysis of the LASSO, we used the fact that $\|\mathbf{a}\|_2 = \max_{\|\mathbf{u}\|_2 \leq 1} \mathbf{u}^T \mathbf{a}$ to write (8.1) as:

$$\min_{\mathbf{x}} \max_{\|\mathbf{u}\|_2 \leq 1} \mathbf{u}^T(\mathbf{y} - \mathbf{A}\mathbf{x}) \text{ subject to } f(\mathbf{x}) \leq f(\mathbf{x}_0), \tag{8.6}$$

to which CGMT is directly applicable. In contrast, when $\mathbf{A}$ is IRO, it is not at all obvious how to use CGMT. To start with, there is no Gaussian matrix. The key idea here is to equivalently express an IRO matrix as:

$$(\mathbf{G}\mathbf{G}^T)^{-1/2}\mathbf{G},$$

with $\mathbf{G} \in \mathbb{R}^{m \times n}$ having entries i.i.d. standard Gaussian and where $(\mathbf{G}\mathbf{G}^T)^{-1/2}$ is the inverse of the square-root of the positive definite (with probability one) $m \times m$ matrix $\mathbf{G}\mathbf{G}^T$. Substituting this expression in (8.1), the LASSO objective is closer but not yet quite of the form required by the CGMT. In particular, the slick trick that led to (8.6) is not enough here and additional ideas are required. Using these we are able to bring (8.1) into the desired format; the argument is sketched in Section 8.3. Once this is done, what remains is to apply the framework of Chapter 5 to conclude with the desired.

8.2 Results

Setup

The matrix $\mathbf{A} \in \mathbb{R}^{m \times n}, m < n$ is modeled to have orthogonal rows $\mathbf{A}\mathbf{A}^T = \mathbf{I}_m$, and the joint probability density of its elements remains unchanged when $\mathbf{A}$ is pre- and post- multiplied by any orthogonal matrices $\mathbf{\Phi} \in \mathbb{R}^{m \times m}, \mathbf{\Theta} \in \mathbb{R}^{n \times n}$, i.e., $p(\mathbf{\Phi}\mathbf{A}\mathbf{\Theta}) = p(\mathbf{A})$. We say that $\mathbf{A}$ is *IRO*[4]. The noise vector $\mathbf{v}$ has entries i.i.d. standard normal $\mathcal{N}(0, 1)$, $f : \mathbb{R}^n \to \mathbb{R}$ is assumed convex and continuous, and, $\mathbf{x}_0$ is not a minimizer of f.

Our results hold in the asymptotic *linear regime*, where m, n and $\omega_{f,\mathbf{x}_0}$ all grow to infinity such that

[4] Different terminologies that appear in the literature to describe the same distribution include "random m-frames in $\mathbb{R}^n$" and "distributed according to the Haar measure on the Stiefel manifold," see [Tro12].

$$m/n \to \delta \in (0,1) \quad \text{and} \quad \omega_{f,\mathbf{x}_0}/\sqrt{n} \to \overline{\omega}_{f,\mathbf{x}_0} \in (0,1),$$

where $\delta, \overline{\omega}_{f,\mathbf{x}_0}$ are constants independent of the problem dimensions, e.g. m,n.

Theorem 8.2.1 (C-LASSO for IRO matrices)**.** *Consider* (8.1) *with* $\mathbf{A}$ *being an IRO matrix and let*

$$\text{aNSE} := \lim_{\sigma \to 0} \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2/\sigma^2.$$

If $\delta > \overline{\omega}_{f,\mathbf{x}_0}^2$*, then, the following limit holds in probability*

$$\lim_{n\to\infty} \frac{\text{aNSE}}{n} = (1 - \overline{\omega}_{f,\mathbf{x}_0}^2)\left(\frac{\overline{\omega}_{f,\mathbf{x}_0}^2}{\delta - \overline{\omega}_{f,\mathbf{x}_0}^2}\right).$$

Theorem 8.2.2 (Minimum Conic Singular Value of IRO matrices)**.** *Assume* $\delta > \overline{\omega}_{f,\mathbf{x}_0}^2$*. Denote*

$$\chi := \sqrt{\delta}\sqrt{\frac{1 - \overline{\omega}_{f,\mathbf{x}_0}^2}{1-\delta}} - \overline{\omega}_{f,\mathbf{x}_0}$$

and $\rho := \overline{\omega}_{f,\mathbf{x}_0}/\chi + 1 - \delta$*. For all* $\zeta > 0$*, with probability 1 in the limit* $n \to \infty$*,* $\sigma_{\min}(\mathbf{A}; \mathcal{T}_f(\mathbf{x}_0))$ *is lower bounded by*

$$\sqrt{\frac{\delta + \rho^2\chi^2 - 2\rho\chi\overline{\omega}_{f,\mathbf{x}_0} - \rho\chi^2(1-\delta)}{\delta + \rho}} - \zeta.$$

Remarks

C-LASSO

Comparison to Gaussian case: For an i.i.d Gaussian matrix with entries of variance $1/n$, it was shown in Section 7.5 that

$$\frac{\text{aNSE}}{n} \approx \frac{\overline{\omega}_{f,\mathbf{x}_0}^2}{\delta - \overline{\omega}_{f,\mathbf{x}_0}^2}.$$

This is strictly greater than the expression of Theorem 8.2.1, proving that the IRO ensemble has strictly superior noise performance. Note that when $\overline{\omega}_{f,\mathbf{x}_0}^2 < \delta \ll 1$, the two formulae are close to each other. This agrees with the fact that the entries of a very “short" IRO matrix are effectively independent for many practical purposes [Jia+06]. Finally, observe that both bounds approach infinity as the compression rate δ approaches $\overline{\omega}_{f,\mathbf{x}_0}^2$. Of course, this agrees with the phase transition in the noiseless case (cf. (8.3)) which is same for both ensembles.

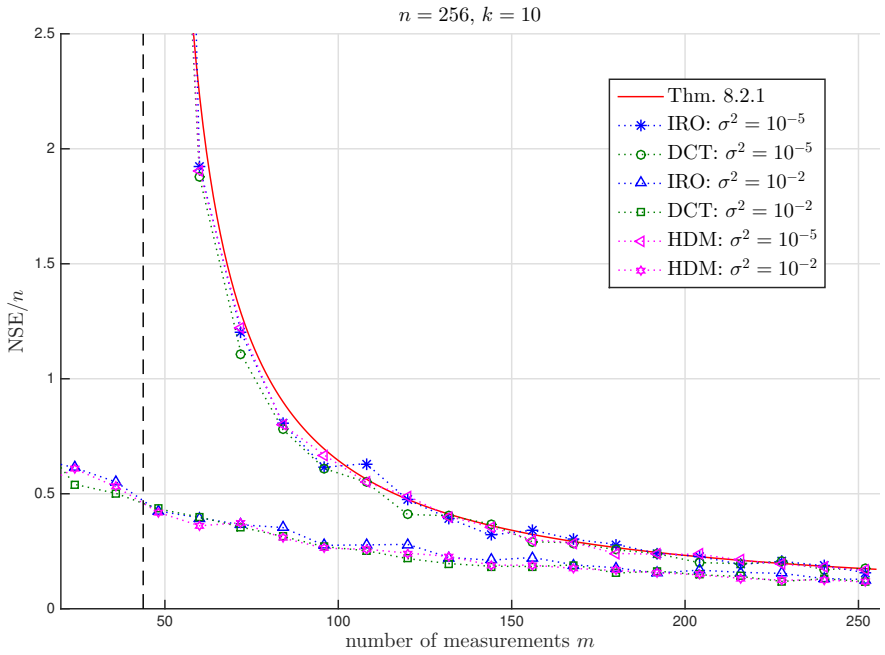

Figure 8.1: Illustration of Theorem 8.2.1 for $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^{256}$ a 10-sparse vector. Simulation results support the claim that aNSE = wNSE. Furthermore, randomly sampled Discrete Cosine Transform (DCT) and Hadamard (HDM) matrices appear to have same NSE performance as IRO matrices. Measured values of the NSE are averages over 25 realizations.

Interpretation: As seen the formula of Theorem 8.2.1 closely resembles the corresponding results for the Gaussian case. Thus, most of the remarks made for the Gaussian case regarding the role of the involved parameters, the geometric nature of the bound and its generality directly transfer to our case. It is useful to recall that $\overline{\omega}^2_{f,\mathbf{x}_0}$ admits precise high-dimensional approximations either in closed-form, or ones that are numerically tractable, for a number of useful instances of f and $\mathbf{x}_0$, e.g. if $f = \|\cdot\|_1$ and $\mathbf{x}_0$ a k-sparse signal with $k/n \to \rho$, then $\overline{\omega}^2_{f,\mathbf{x}_0} \leq 2\rho(\log(1/\rho)+1)$ (cf. Section 2.15).

wNSE: Similar to (7.38), we conjecture that wNSE = aNSE here, as well. In this case, Theorem 8.2.1 would prove a tight upper bound on NSE(σ) for any σ. Simulation results in Figure 8.1 support the claim.

Universality: Our simulations in Figure 8.1 suggest that partial Discrete Cosine Transform (DCT) matrices obtained by *randomly* sampling m rows of the DCT matrix without replacement, and similarly sampled Hadamard (HDM) matrices exhibit

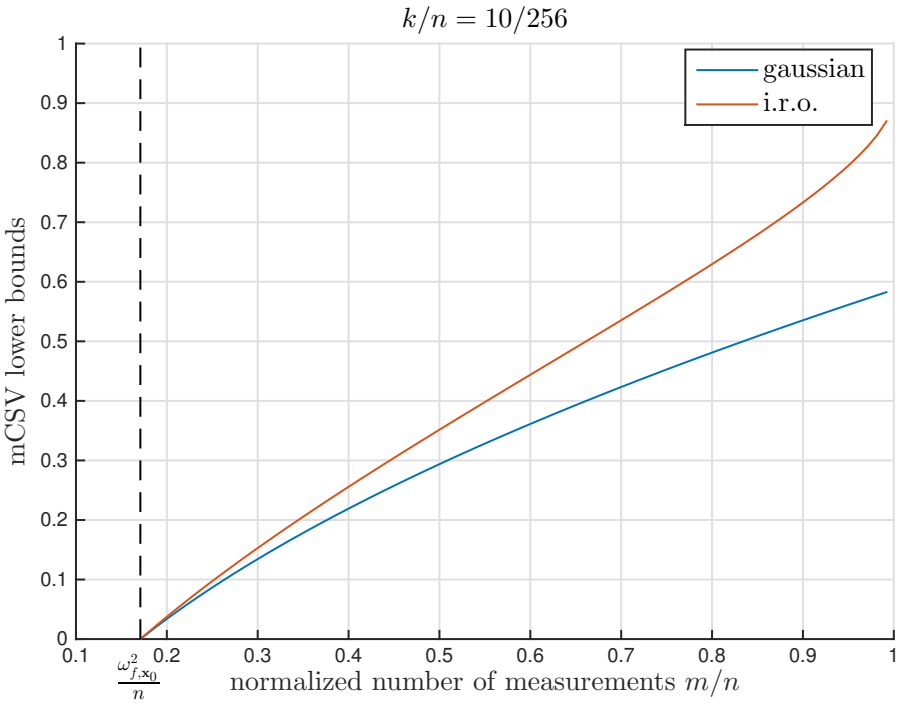

Figure 8.2: Illustration of Theorem 8.2.2. The bound exceeds the corresponding bound for Gaussian matrices. We have chosen $f = \|\cdot\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^n$, a k-sparse vector.

the same NSE performance as the IRO ensemble. Thus, Theorem 8.2.1 appears to predict the NSE of random DCT and HDM matrices as well. Understanding of the behavior of such ensembles is of great practical importance due to their favorable attributes [Wen+14b].

Minimum Conic Singular Value

Comparison to Gaussian case: Recall from the escape through a mesh Proposition 2.2.2 that the mCSV of a matrix with i.i.d. entries $\mathcal{N}(0, 1/n)$ is lower bounded by $\sqrt{\delta} - \overline{\omega}_{f,\mathbf{x}_0}$. The bound of Theorem 8.2.2 exceeds that, which is a strong indication that IRO matrices are *strictly* better conditioned than corresponding Gaussian ones. See Fig. 8.2 for an illustration.

Sanity test: When $\overline{\omega}^2_{f,\mathbf{x}_0} < \delta \ll 1$, the entries of the IRO behave almost as if they are independent [Jia+06]. As expected, then, in this regime the bound of Theorem 8.2.2 approaches $\sqrt{\delta} - \overline{\omega}_{f,\mathbf{x}_0}$, which coincides with the bound on Gaussians. On the

other hand, when $\delta = 1$, it can be seen that, as expected, the expression of Theorem 8.2.2 approaches one.

Tightness: Theorem 8.2.2 provides no guarantees on the exactness of the derived lower bound. This is also the case for the corresponding result on the mCSV of Gaussian matrices. Proving (or disproving) the exactness of the bounds is an open research problem.

General cones: Of course, the bound of Theorem 8.2.2 holds for the minimum singular value of $\mathbf{A}$ with respect to *any* cone, not necessarily a tangent cone or even a convex cone. One just needs to replace $\overline{\omega}_{f,\mathbf{x}_0}$ with the Gaussian width of the corresponding cone. Also, a non-asymptotic version of Theorem 8.2.2 is possible with only few adjustments in the proof presented here.

8.3 Proof Outline

Here, we outline the main steps of the proof. We focus on Theorem 8.2.1. The proof of Theorem 8.2.2 is similar (see Appendix E). We limit our attention to showing the steps and modifications required to apply the CGMT in the case of IRO matrices. In contrast to this part of the proof, which involves several new ideas, afterwards we have transformed the problem into one where the CGMT framework is applicable, then the rest is along the lines of Chapter 5. This latter part and some technical details not discussed here are deferred to Appendix E. We re-write (8.1) by changing the decision variable to be the error vector $\mathbf{w} := \mathbf{x} - \mathbf{x}_0$:

$$\hat{\mathbf{w}} := \min_{\mathbf{w} \in \mathcal{D}_f(\mathbf{x}_0)} \|\mathbf{A}\mathbf{w} - \sigma \mathbf{v}\|_2. \tag{8.7}$$

We evaluate the limiting behavior $\lim_{\sigma \to 0} \|\hat{\mathbf{w}}\|^2/\sigma^2$. Throughout, we write $\|\cdot\|$ instead of $\|\cdot\|_2$.

Formulation in terms of Gaussians

We begin with a simple Lemma that provides a simple characterization of IRO matrices in terms of Gaussians. Let $\mathbf{X}^{1/2}$ denote a square-root of a matrix $\mathbf{X} \in \mathbb{R}^{m \times m}$, and $\mathbf{X}^{-1/2}$ its inverse (if it exists). Also, for random variables x and y with the same distribution, we write $x \sim y$.

Lemma 8.3.1 (IRO matrices)**.** *Let $\mathbf{G} \in \mathbb{R}^{m \times n}$ have entries i.i.d. $\mathcal{N}(0,1)$. Then the matrix $\mathbf{A} = (\mathbf{G}\mathbf{G}^T)^{-1/2}\mathbf{G}$ is a $m \times n$ IRO matrix.*

Proof. It can be readily confirmed that $\mathbf{A}\mathbf{A}^T = \mathbf{I}_m$. We need to prove that the distribution of $\mathbf{A}$ remains invariant after pre- and post- multiplication with orthogonal

matrices of appropriate sizes. Let $\mathbf{\Phi} \in \mathbb{R}^{n\times n}$, $\mathbf{\Theta} \in \mathbb{R}^{m\times m}$ be any orthogonal matrices. First, $\mathbf{A\Theta} \sim (\mathbf{GG}^T)^{-1/2}\mathbf{G\Theta} = ((\mathbf{G\Theta})(\mathbf{G\Theta})^T)^{-1/2}\mathbf{G\Theta}$. Recall that the Gaussian distribution is invariant under orthogonal transformations, i.e. $\mathbf{G} \sim \mathbf{G\Theta}$, to conclude from the above that $\mathbf{A\Theta} \sim \mathbf{A}$. Next, $\mathbf{G} \sim \mathbf{\Phi G}$. Also, it can be directly verified that $\mathbf{\Phi}(\mathbf{GG}^T)^{-1/2}\mathbf{\Phi}$ is the inverse of a square-root of of $\mathbf{\Phi GG}^T\Phi$. With these, $\mathbf{A} \sim ((\mathbf{\Phi G})(\mathbf{\Phi G})^T)^{-1/2}\mathbf{\Phi G} = \mathbf{\Phi}(\mathbf{GG}^T)^{-1/2}\mathbf{G} = \mathbf{\Phi A}$. □

Next, we use Lemma 8.3.1 to write the objective function in (8.7) in terms of Gaussian matrices.

Lemma 8.3.2 (LASSO Objective)**.** *Assume* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *is IRO and* $\mathbf{v} \in \mathbb{R}^m$ *is standard Gaussian, independent of each other. Then, for any* $\mathbf{w} \in \mathbb{R}^n$*,*

$$(\mathbf{Aw} - \sigma\mathbf{v}) \sim (\mathbf{GG}^T)^{-1/2}\mathbf{G}(\sigma\mathbf{q} - \mathbf{w}),$$

where $\mathbf{G} \in \mathbb{R}^{m\times n}$ *and* $\mathbf{q} \in \mathbb{R}^n$ *have entries i.i.d.* $\mathcal{N}(0,1)$ *and are independent of each other.*

Proof. Let $\mathbf{A}, \mathbf{G}, \mathbf{v}, \mathbf{q}$ as in the statement of the Lemma. For any row-orthogonal $\mathbf{Q} \in \mathbb{R}^{m\times n}$, $\mathbf{v} \sim \mathbf{Qq}$. Furthermore, provided that $\mathbf{q}$ is independent of the distribution of $\mathbf{Q}$, the same is then true for $\mathbf{v}$. Hence, letting $\mathbf{Q} = \mathbf{A}$, we have $(\mathbf{Aw} - \sigma\mathbf{v}) \sim \mathbf{A}(\mathbf{w} - \sigma\mathbf{q})$. Apply Lemma 8.3.1 to conclude with the desired. □

Preparing the grounds for applying the CGMT

Using Lemma 8.3.2, we work with the following (probabilistically) equivalent formulation of (8.7):

$$\hat{\mathbf{w}} := \min_{\mathbf{w}\in\mathcal{D}_f(\mathbf{x}_0)} \|(\mathbf{GG}^T)^{-1/2}\mathbf{G}(\mathbf{w} - \sigma\mathbf{q})\|_2, \tag{8.8}$$

This brings a step closer to the CGMT framework, but not yet quite to the point that we can identify the desired format of the (PO) problem described in (3.11a). The goal of this section is to complete this step. We start by using the fact that for any $\mathbf{a} \in \mathbb{R}^m$: $\|\mathbf{a}\| = \max_{\|\mathbf{b}\|\le 1} \mathbf{b}^T\mathbf{a}$. In particular, the objective function in (8.8) can be expressed as follows:

$$\max_{\|\mathbf{b}\|\le 1} \mathbf{b}^T(\mathbf{GG}^T)^{-1/2}(\mathbf{w} - \sigma\mathbf{q}) =$$

$$\max_{\|(\mathbf{GG}^T)^{1/2}\mathbf{b}\|\le 1} \mathbf{b}^T\mathbf{G}(\mathbf{w} - \sigma\mathbf{q}) = \max_{\|\mathbf{G}^T\mathbf{b}\|\le 1} \mathbf{b}^T\mathbf{G}(\mathbf{w} - \sigma\mathbf{q}).$$

It can be checked that the above is equivalent to:

$$\max_{\mathbf{b}} \min_{\ell} \mathbf{b}^T \mathbf{G}(\mathbf{w} - \sigma\mathbf{q} - \ell) + \|\ell\|.$$

Now, we flip the order of max-min [Roc97, Cor. 37.3.2][5]:

$$\hat{\mathbf{w}} = \min_{\mathbf{w}\in\mathcal{D}_f(\mathbf{x}_0),\ell} \max_{\mathbf{b}} \mathbf{b}^T \mathbf{G}(\mathbf{w} - \sigma\mathbf{q} - \ell) + \|\ell\|,$$

or, re-defining $\ell := \mathbf{w} - \sigma\mathbf{q} - \ell$:

$$\hat{\mathbf{w}} = \min_{\mathbf{w}\in\mathcal{D}_f(\mathbf{x}_0),\ell} \max_{\mathbf{b}} \mathbf{b}^T \mathbf{G}\ell + \|\mathbf{w} - \sigma\mathbf{q} - \ell\|. \tag{8.9}$$

This brings (8.7) in the desired format (cf.(3.11a)) for the application of the GMT framework[6]. In particular, the (AO) problem as it corresponds to (8.9) becomes:

$$\tilde{\mathbf{w}}(\mathbf{g},\mathbf{h},\mathbf{q}) = \arg \min_{\mathbf{w}\in\mathcal{D}_f(\mathbf{x}_0),\ell} \max_{\mathbf{b}} \|\ell\|\mathbf{g}^T\mathbf{b} - \|\mathbf{b}\|\mathbf{h}^T\ell + \|\mathbf{w} - \sigma\mathbf{q} - \ell\|. \tag{8.10}$$

The rest of the proof analyzes (8.10) with the goal of determining the limiting behavior of $\|\tilde{\mathbf{w}}\|$ and is included in Appendix E. We just remark here on the assumption of the theorem that $\sigma \to 0$; this also provides a hint on the presence of the Gaussian width of the tangent cone in the final result. When $\sigma \to 0$, it suffices to analyze a "first-order approximation" to problem (8.10) in which the feasible set $\mathcal{D}_f(\mathbf{x}_0)$ is substituted by its conic hull, i.e. $\mathcal{T}_f(\mathbf{x}_0)$. Since the tangent cone captures the local behavior in the neighborhood of $\mathbf{x}_0$, the relaxation will be tight in the limit as $\|\hat{\mathbf{w}}\|_2 \to 0$. The idea is that in the limit $\sigma \to 0$, $\|\hat{\mathbf{w}}\|$ is sufficiently small and the approximation tight. Of course, this approximation is the same as the one we already saw in Section 7.5 when Gaussian matrices were considered.

[5](i) the objective function above is continuous, convex in ℓ, and concave in $\mathbf{b}$, (ii) the constraint sets are convex. We only need to worry about *boundedness* of the constraint sets. Such steps require proper attention in general and are handled rigorously in the Appendix.

[6]To be precise, this requires a trivial modification of (8.9) since $\mathbf{w}$ does not appear in the bilinear form as in (3.11a). This can be handled easily and similar extension can also be found in [FM14, Lem. 5]. In particular, comparing to (3.11a), we can identify in (8.9): $\psi([\ell,\mathbf{w}],\mathbf{b}) := \|\mathbf{w} - \mathbf{q} - \ell\|$, which is continuous and convex in $[\ell,\mathbf{w}]$, as desired. Also, the constraint sets are convex. Please refer to the Appendix for compactness issues.

Chapter 9

BEYOND SQUARED-ERROR: GENERAL PERFORMANCE METRICS

All the results presented up to this point of the thesis consider specifically the squared-error reconstruction performance of (1.2). This section extends the scope of the CGMT framework and derives precise results for other performance metrics. For concreteness, we focus on the problem of sparse recovery under ℓ_1-regularized least-squares (a.k.a LASSO), but also discuss possible extensions to other instances of (1.2). We establish accurate predictions of a wide range of performance metrics that have a Lipschitz property. For illustration, this result can be used to accurately predict the probability that the LASSO successfully identifies the non-zero entries of the unknown signal; specializing the result to the high-SNR regime yields bounds that are geometric in nature and admit insightful interpretations.

In Section 9.1 we motivate the need to study other performance metrics besides the squared-error. We briefly review the known result on the squared-error performance in Section 9.2, but put in a language that is easy to generalize. The main result of the chapter on the LASSO performance under general Lipschitz performance metrics is presented in Section 9.3. For an illustration, we use it to predict the probability of correct support recovery in Section 9.4. Its proof occupies the last Section, 9.5.

9.1 Introduction

Consider recovering a k-sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$ from noisy linear measurements (cf. (1.1)) using the square-root LASSO:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \frac{\lambda}{\sqrt{n}}\|\mathbf{x}\|_1. \tag{9.1}$$

(The normalization with $\sqrt{n}$ is for convenience in the analysis). We remark that the analysis presented here also applies to the ℓ_2^2-LASSO, but we focus on the version in (9.1) for concreteness. Also, for convenience, we shall often refer to (9.1) simply as the LASSO.

Measuring Performance

A “good estimate" might translate to a variety of different desired attributes associated with $\hat{\mathbf{x}}$. This translates to a variety of different *performance metrics*, which we

discuss here.

ℓ_2-reconstruction (or squared-) error: This refers to the standard and somewhat generic measure of performance on which this thesis has focused thus far. It measures the deviation of $\hat{\mathbf{x}}$ from the true signal $\mathbf{x}_0$ in the ℓ_2-norm. Formally, the metric acts on the *reconstruction error vector* $\hat{\mathbf{w}} := \hat{\mathbf{x}} - \mathbf{x}_0$ and returns its Euclidean norm, i.e., $\Psi_{\ell_2}(\hat{\mathbf{w}}) := \|\hat{\mathbf{w}}\|_2 = \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2$. The ℓ_2-error in estimating the coefficients of $\mathbf{x}_0$ also controls the mean squared prediction error, i.e. the error in predicting a (future) response to a fresh (random) measurement (e.g. [OT15, Sec. 8.1]).

Lipschitz Metrics: Beyond the ℓ_2-reconstruction error, we consider performance metrics $\Psi : \mathbb{R}^n \to \mathbb{R}$ that act on the error vector $\hat{\mathbf{w}} := \hat{\mathbf{x}} - \mathbf{x}_0$ and which satisfy a Lipschitz property, i.e. $|\Psi(\mathbf{x}) - \Psi(\mathbf{y})| \leq L \cdot \|\mathbf{x} - \mathbf{y}\|_2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and some L. One common such metric is $\Psi(\mathbf{w}) = \|\mathbf{w}\|_1$, [Neg+12].

Support Recovery: In the problem of sparse recovery a natural performance metric that arises in a variety of contexts (e.g. subset selection in regression, structure estimation in graphical models, sparse approximation [Wai09]) is that of support recovery, i.e. identifying whether an entry of the unknown signal $\mathbf{x}_0$ is on the support (aka is non-zero), or it is off the support (aka is zero). We take a decision based on the solution $\hat{\mathbf{x}}$ of the LASSO: declare the i^{th} entry to be on the support iff $|\hat{\mathbf{x}}_i| \geq \epsilon$. Here $\epsilon > 0$ is a user-defined threshold imposed on $\hat{\mathbf{x}}$; such a hard-thresholding operation is practical due to machine precision inaccuracies in solving (9.1). In Theorem 9.4.1 we accurately predict the *(per-entry) rate of successful on-support and off-support recovery*. Formally, let

$$\Phi_{\epsilon,\text{on}}(\hat{\mathbf{x}}) = \frac{1}{k} \sum_{i \in S(\mathbf{x}_0)} \mathbb{1}_{\{|\hat{\mathbf{x}}_i| \geq \epsilon\}}, \tag{9.2a}$$

$$\Phi_{\epsilon,\text{off}}(\hat{\mathbf{x}}) = \frac{1}{n-k} \sum_{i \notin S(\mathbf{x}_0)} \mathbb{1}_{\{|\hat{\mathbf{x}}_i| \leq \epsilon\}}, \tag{9.2b}$$

where $\mathbb{1}_{\mathcal{A}}$ is the indicator function of a set $\mathcal{A}$. The metric $\Phi_{\epsilon,\text{on}}(\hat{\mathbf{x}})$ (resp. $\Phi_{\epsilon,\text{off}}(\hat{\mathbf{x}})$) measures the ratio of the non-zero (resp. zero) entries of $\mathbf{x}_0$ that are properly identified to be on (resp. off) the support. An equivalent way to interpret the metrics defined above is to consider their expectation. For instance, $\mathbb{E}[\Phi_{\epsilon,\text{on}}(\hat{\mathbf{x}})] = (1/k) \sum_{i \in S(\mathbf{x}_0)} \mathbb{P}(|\hat{\mathbf{x}}_i| \geq \epsilon)$ measures the *average* probability that a single non-zero entry of $\mathbf{x}_0$ is correctly identified to be on the support. In particular, if the entries of $\hat{\mathbf{x}}$ are iid, then in the limit $\Phi_{\epsilon,\text{on}}(\hat{\mathbf{x}})$ converges to the probability that a single on-support entry is correctly identified. Our proof hints that this is indeed the case (cf. Section 9.5).

Working Hypothesis

The unknown signal $\mathbf{x}_0 \in \mathbb{R}^n$ is k-sparse: its first k entries are sampled iid from a distribution p_{X_0} of zero mean and of unit variance ($\mathbb{E}[X_0^2] = 1$), and the rest of them are zero. (Alternatively, we could have assumed all its entries be iid from a distribution as in 6.32). The measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ has entries iid zero mean Gaussian random variables with variance $\frac{1}{n}$. The noise vector $\mathbf{z} \in \mathbb{R}^m$ has entries iid $\mathcal{N}(0, \sigma^2)$. We study the linear asymptotic regime in which the problem dimensions n, m and k all grow to infinity at proportional rates [1]:

$$k/n \to \rho \in (0, 1) \quad \text{and} \quad m/n \to \delta \in (0, \infty).$$

Also, the regularizer parameter λ in (9.1) is considered to be constant, in particular independent of n. Under the current setting, the Signal to Noise Ratio (SNR) becomes SNR $:= \rho/\sigma^2$.

Generalizations

One of the primal purposes of the current chapter is showing how the CGMT framework can be applied to characterize performance measures beyond the squared-error. For simplicity and concreteness, we focus entirely on the LASSO method. However, we remark that the results can in principle be extended to the general class of regularized M-estimators in (1.2) under a setting similar to that of Chapter 4. The essence of the proof is the same as the one here (cf. Section 9.5), but several technical details need to be taken care of. We leave the details and a result on the performance of regularized M-estimators under Lipschitz-like metrics of the generality of Theorem 4.2.1 for future work.

9.2 Review: ℓ_2-reconstruction Error

The ℓ_2-reconstruction error of (9.1) was precisely characterized in Section 6.5. We repeat the result below only this time introducing some new notation that will prove convenient for the statement of the more general result to follow in the next section.

ψ-distance functional: For a function $\psi : \mathbb{R} \to \mathbb{R}$, let $\text{Dist}_{\psi(.)}(\cdot, \cdot) : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$ be defined as

$$\text{Dist}_{\psi(.)}(\kappa, \lambda) := \rho \cdot \mathbb{E}[\psi(X_0 - \eta(\kappa H + X_0, \kappa\lambda))] + (1 - \rho) \cdot \mathbb{E}[\psi(\eta(\kappa H, \kappa\lambda))], \tag{9.3}$$

[1] As usual, the results apply on a sequence of problem instances $\{\mathbf{x}_0, \mathbf{A}, \mathbf{z}, m, k\}_n$ indexed by $n \in \mathbb{N}$ such that the properties mentioned hold for all members of the sequence for all n. To keep notation clear we do not explicitly use the subscript n for symbols of the sequence.

where the expectation is over both $X_0 \sim p_{X_0}$ and $H \sim \mathcal{N}(0,1)$, and $\eta(X,\tau) = (X/|X|)\max\{|X| - \tau, 0\}$ denotes the soft-thresholding operator. The function returns the distance, with respect to the function $\psi(.)$, between a r.v. X_0 and the soft threshold operator applied to the random variable itself after adding a Gaussian noise to it. This motivates the terminology used. Also, note the implicit dependence of the functional on the rest of the problem parameters, namely ρ, δ and σ.

λ_{crit}: There exists a critical value of the regularizer parameter, namely λ_{crit}, such that the error behavior is different when $\lambda \leq \lambda_{\text{crit}}$ compared to $\lambda > \lambda_{\text{crit}}$. Define the pair $(\alpha_{\text{crit}}, \lambda_{\text{crit}})$ as the solution to the following system of equations:

$$\begin{cases} \alpha_{\text{crit}}^2 = \text{Dist}_{(.)^2}(\kappa_{\text{crit}}, \lambda_{\text{crit}}), \\ \delta = \rho \cdot \mathbb{P}\left\{|\kappa H + X_0| \geq \lambda_{\text{crit}}\kappa_{crit}\right\} + 2(1-\rho)Q(\lambda_{\text{crit}}), \end{cases} \tag{9.4}$$

where $\kappa_{crit} = \sqrt{(\alpha_{\text{crit}}^2 + \sigma^2)/\delta}$ and $Q(\cdot)$ is the standard Q-function. Recall by Theorem 6.6.1 that if $\delta \leq 1$, then (9.4) has a unique solution. Otherwise, define $\lambda_{\text{crit}} = 0$. With these we can restate Theorem 6.6.1 as follows.

Lemma 9.2.1 (re-stament of Theorem 6.6.1)**.** *Under the working hypothesis of Section 9.1 and for any fixed $\lambda > 0$, define $\alpha := \alpha(\lambda)$ as the unique solution to the equation $\alpha^2 = Dist_{(.)^2}(\sqrt{(\alpha^2+\sigma^2)/\delta}, \lambda)$ if $\lambda \geq \lambda_{crit}$, and as $\alpha = \alpha_{crit}$ otherwise. Then, it holds in probability that $\lim_{n\to\infty} \frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \alpha^2$.*

9.3 Lipschitz Performance Metrics

Theorem 9.3.1 below generalizes Lemma 9.2.1 to metrics that attain a Lipschitz property. Assumption 9.3.1 below formally defines the required properties of such metrics.

Assumption 9.3.1 (Lipschitz metrics)**.** *We say Assumption 9.3.1 holds for the Lipschitz function $\Psi : \mathbb{R}^n \to \mathbb{R}$ if:*

- *For all constants $c > 0$, there exists a constant $C > 0$ such that for all $\mathbf{x} \in \mathbb{R}^n$ that $\|\mathbf{x}\| \leq c\sqrt{n}$, we have $|\Psi(\mathbf{x})| \leq C\sqrt{n}$.*
- *For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $|\Psi(\mathbf{x}) - \Psi(\mathbf{y})| \leq \frac{L}{\sqrt{n}}\|\mathbf{x} - \mathbf{y}\|_2$, for a constant L independent on n.*
- *For all $\alpha, \lambda > 0$ and $\mathbf{h} \sim \mathcal{N}(0, \mathbf{I}_n)$, there exists function $\Gamma : \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}$ such that*

$$\Psi(\mathbf{x}_0 - \overrightarrow{\eta}(\kappa\mathbf{h} + \mathbf{x}_0, \lambda\kappa)) \xrightarrow{P} \Gamma(\kappa, \lambda). \tag{9.5}$$

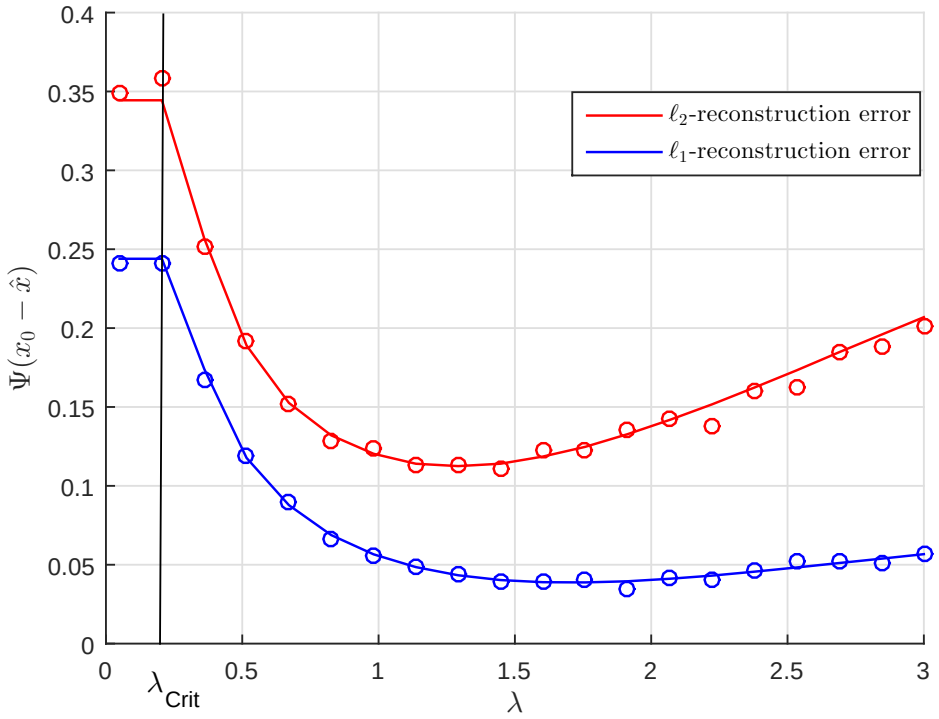

Figure 9.1: Performance of the Square-root Lasso with respect to $\Psi(\mathbf{x}) = \frac{1}{\sqrt{n}}\|\mathbf{x}\|_2$ (Red Line) and $\Psi(\mathbf{x}) = \frac{1}{n}\|\mathbf{x}\|_1$ (Blue line) as a function of λ. The theoretical prediction follows from Theorem 9.3.1. For the simulations, we used $n = 256$, $\delta = 0.8$, $\rho = 0.1$, SNR=0.5 and the data are averaged over five independent realizations.

Here, $\overrightarrow{\eta}$ is the "vector" soft-threshold operator acting element-wise on the entries of its first argument.

The first is a simple scaling requirement such that $\Psi(\mathbf{x}) = O(1)$. The second imposes a growth condition on the Lipschitz constant with respect to n (this is necessary for the asymptotic analysis but can potentially be relaxed). The third requirement of Assumption 9.3.1 is easier to interpret in the "separable-case" in which $\Psi(\mathbf{x}) = (1/n)\sum_i \psi(\mathbf{x}_i)$ for some L-Lipschitz scalar function ψ. Then, condition (9.5) holds by the WLLN for $\Gamma(\kappa, \lambda) = \text{Dist}_\psi(\kappa, \lambda)$ (recall (9.3)).

Theorem 9.3.1 (Lipschitz performance of LASSO)**.** *Under the working hypothesis of Section 9.1 and with α and λ_{crit} defined as in Lemma 9.2.1, fix $\lambda > 0$, let $\hat{\lambda} = max\{\lambda, \lambda_{crit}\}$ and $\kappa = \sqrt{\alpha^2 + \sigma^2}/\sqrt{\delta}$. Then, for any Lipschitz function $\Psi(x)$ that satisfies Assumption 9.3.1, it holds in probability that, $\lim_{n\to\infty} \Psi(\hat{\mathbf{x}} - \mathbf{x}_0) = \Gamma(\kappa, \hat{\lambda})$.*

Evaluating the prediction only involves identifying the function Γ as per Assumption 9.3.1, and calculating the parameters α and λ_{crit} as per Lemma 9.2.1. Of course, Lemma 9.2.1 follows from Theorem 9.3.1 when applied for $\Psi(\hat{\mathbf{x}} - \mathbf{x}_0) = \frac{1}{\sqrt{n}}\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2$, since the latter is easily shown to satisfy Assumption 9.3.1 for $\Gamma(\kappa, \lambda) = \sqrt{\text{Dist}_{(.)^2}(\kappa, \lambda)}$. A different Lipschitz performance metric that is often of interest in practice is the ℓ_1-reconstruction error $\Psi(\hat{\mathbf{x}}-\mathbf{x}_0) = (1/n)\|\hat{\mathbf{x}}-\mathbf{x}_0\|_1$. This is an example of a separable metric, thus it satisfies Assumption 9.3.1 for $\Gamma(\kappa, \lambda) = \text{Dist}_{|.|}(\kappa, \lambda)$. See Figure 9.1 for an illustration. Observe that the prediction of the theorem (although asymptotic) is accurate for problem dimensions of only a few hundreds. Also, the precise nature of the predictions allows optimal tuning of the regularizer parameter λ, the number of measurements δ, etc..

9.4 Support Recovery

Theorem 9.4.1 below characterizes the support recovery metrics introduced in (9.2). Recall that $\epsilon > 0$ is a fixed hard threshold imposed on the entries of the solution $\hat{\mathbf{x}}$ to the LASSO in order to decide whether an entry is on or off the support.

Theorem 9.4.1 (Probability of support recovery)**.** *Under the working hypothesis of Section 9.1 and with α and λ_{crit} defined as in Lemma 9.2.1, fix $\lambda > 0$, let $\kappa = \sqrt{(\alpha^2 + \sigma^2)/\delta}$ and $\hat{\lambda} = \max\{\lambda_{crit}, \lambda\}$. Then, for any $\epsilon > 0$, it holds in probability that*

$$\lim_{n\to\infty} \Phi_{\epsilon,on}(\hat{\mathbf{x}}) = \mathbb{P}\{|\kappa H + X_0| \geq \epsilon + \hat{\lambda}\kappa\}$$

and

$$\lim_{n\to\infty} \Phi_{\epsilon,off}(\hat{\mathbf{x}}) = \mathbb{P}\{|\kappa H| \leq \epsilon + \hat{\lambda}\kappa\}.$$

The metrics in (9.2) are not Lipschitz. Hence, they don't satisfy all requirements of Assumption 9.3.1 of Section 9.3, and Theorem 9.3.1 is not directly applicable. The core idea behind the proof of the theorem is similar to that of Theorem 9.3.1, but requires a few extra arguments (see Section F.1). Figure 9.2 illustrates the validity of the prediction.

Remark 9.4.0.56 (Off-support). When $\epsilon \ll \hat{\lambda}\kappa$, the formula of the theorem for $\Phi_{\epsilon,\text{off}}(\hat{\mathbf{x}})$ reduces to $\mathbb{P}\{|\kappa H| \leq \epsilon + \hat{\lambda}\kappa\} \sim \mathbb{P}\{|H| \leq \hat{\lambda}\}$, which is independent of the problem parameters δ, ρ and SNR. This simple observation is verified in Figure 9.2: the off-support recovery probability is the same for different values of undersampling parameter δ as long as $\lambda \geq \lambda_{\text{crit}}$.

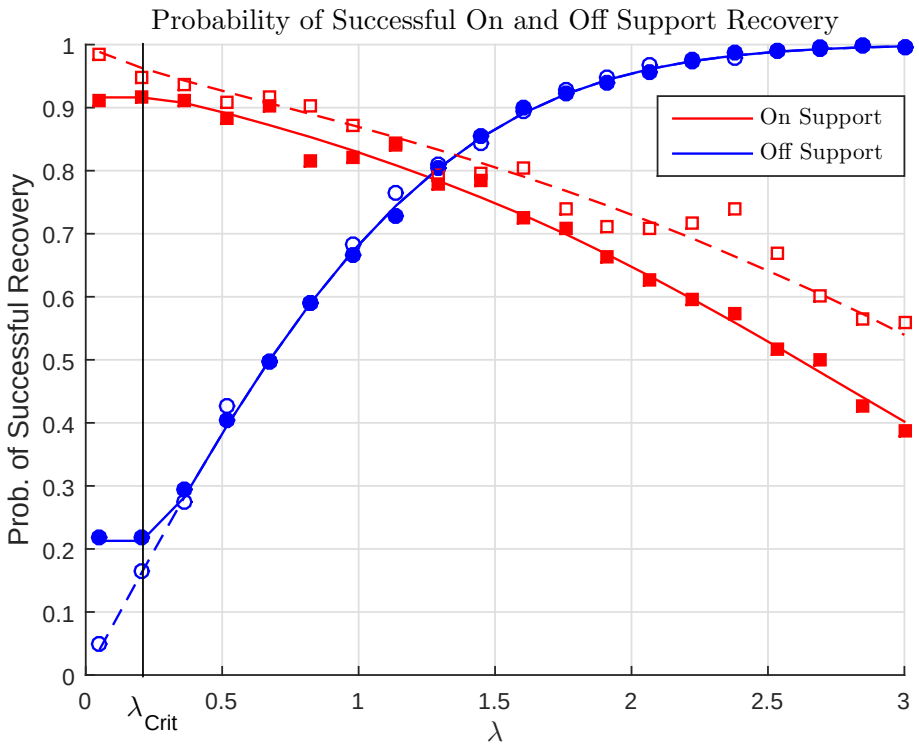

Figure 9.2: Probability of successful recovery of the on-support and of the off-support entries as a function of λ for two different values of the normalized measurements, namely $\delta = 0.8$ (solid) and $\delta = 1.2$ (dashed). The theoretical prediction (shown in solid/dashed lines) follows from Theorem 9.4.1. For the simulation points (shown with squares and circles), we used $n = 256$, SNR= 0.5, $\epsilon = 10^{-3}$, $\rho = 0.1$ and the data are averaged over five independent realizations of the problem.

Remark 9.4.0.57 (Large/Small λ). It is easy to conclude from Theorem 9.4.1 that as λ becomes large $\Phi_{\epsilon,\text{off}}$ (reps. $\Phi_{\epsilon,\text{on}}$) converge to one (resp. zero). Of course, this behavior is expected since large values for the regularizer parameter put more emphasis on the ℓ_1-regularization term in (9.1), thus promoting sparser solutions. Reversed behavior is observed when λ takes values close to zero. An interesting observation is the following: in the under-sampling regime ($\delta < 1$), even when $\lambda \to 0$ there is a non-vanishing probability that an off-support entry is correctly identified as such. This is because λ_{crit} is non-zero in this case. See also Figure 9.2.

Remark 9.4.0.58 (Optimal λ). A natural question becomes that of determining the optimal value of the regularizer parameter. In order to balance between on- and off- support recovery probabilities a reasonable performance metric becomes $\Phi_\epsilon = \omega\Phi_{\epsilon,\text{on}} + (1-\omega)\Phi_{\epsilon,\text{off}}$ for $\omega \in [0,1]$. Theorem 9.4.1 precisely characterizes the behavior of this as a function of λ; thus, it determines the optimal value of λ that

minimizes Φ_ϵ.

Remark 9.4.0.59 (High-SNR Regime). Here, we analyze the probability of support recovery at SNR $\gg 1$ (eqv. $\sigma^2 \to 0$). In this regime, λ_{crit} takes a simple form: if $\delta < 1$, then $\lambda_{\text{crit}} = Q^{-1}(\frac{\delta-\rho}{2(1-\rho)})$ where Q^{-1} is the inverse Q-function, otherwise, $\lambda_{\text{crit}} = 0$ (also, see Section 7.6). Let us first examine the behavior of "off-support" recovery probability. When $\sigma^2 \ll 1$, the formula of Theorem 9.4.1 reduces to the following simpler one:

$$\lim_{n\to\infty} \Phi_{\epsilon,\text{off}}(\hat{\mathbf{x}}) \sim 1 - 2Q\left(\hat{\lambda} + \frac{\epsilon}{\sigma}\sqrt{\delta - \overline{\mathbf{D}}_{\ell_1,\mathbf{x}_0}(\hat{\lambda})}\right), \tag{9.6}$$

for λ such that $\delta > \overline{\mathbf{D}}_{\ell_1,\mathbf{x}_0}(\hat{\lambda})$, $\hat{\lambda} = \max\{\lambda, \lambda_{\text{crit}}\}$, and

$$\overline{\mathbf{D}}_{\ell_1,\mathbf{x}_0}(\lambda) = \rho \cdot \mathbb{E}[(H-\lambda)^2 \big| H > 0] + (1-\rho)\cdot \mathbb{E}[\eta^2(H,\lambda)]. \tag{9.7}$$

Several remarks are in place here.

First, when the threshold ϵ does not scale with σ and $\sigma \to 0$, then naturally the probability converges to one. The same is true as λ grows large, which is again expected. The term $\overline{\mathbf{D}}_{\ell_1,\mathbf{x}_0}(\lambda)$ is nothing but the normalized Gaussian squared distance to the scaled subdifferential ((9.7) is same as (2.14) normalized by n). Observe that (9.6) requires $\delta > \min_{\lambda>0} \overline{\mathbf{D}}_{\ell_1,\mathbf{x}_0}(\lambda)$ in line with Theorem 2.2.1. Also, the formula is valid for λ such that $\delta > \overline{\mathbf{D}}_{\ell_1,\mathbf{x}_0}(\hat{\lambda})$.

For the on-support probability, we can show that it behaves as

$$\lim_{n\to\infty} \Phi_{\epsilon,\text{on}}(\hat{\mathbf{x}}) \sim \mathbb{P}\left\{|\hat{\kappa}H + X_0| \geq \epsilon + \hat{\lambda}\hat{\kappa}\right\}$$

for $\hat{\kappa} = \sigma/\sqrt{\delta - \overline{\mathbf{D}}_{\ell_1,\mathbf{x}_0}(\hat{\lambda})}$, and similar remarks can be made.

9.5 Proofs

Proof of Theorem 9.3.1

We follow the standard approach of the CGMT framework as detailed in Chapter 5. Recall that in Chapter 5 we were interested in the ℓ_2-reconstruction error, thus we applied the CGMT Theorem 3.3.1(iii) to the set (cf. (5.2)):

$$\mathcal{S}_\epsilon = \{\mathbf{w} \mid |\|\mathbf{w}\|_2 - \alpha_*| > \epsilon\}.$$

Instead, here we need to apply the CGMT to a different set:

$$\mathcal{S}_\epsilon = \{\mathbf{w} \mid |\Psi(\mathbf{w}) - \Gamma(\kappa,\lambda)| > \epsilon\}. \tag{9.8}$$

To avoid unnecessary repetitions, we only highlight the parts of the proof that are different.

First, specializing the results of Section 5.2 to the LASSO method in (9.1), the (PO) and (AO) problems of interest become (see for example Section 7.4):

$$\min_{\mathbf{w}} \frac{1}{\sqrt{n}}\|\mathbf{z}-\mathbf{A}\mathbf{w}\|_2+\frac{\lambda}{n}\|\mathbf{x}_0+\mathbf{w}\|_1, \tag{9.9}$$

$$\min_{\mathbf{w}} \max_{0<\beta\leq 1} \beta\frac{\|\mathbf{g}\|_2}{\sqrt{n}}\sqrt{\frac{\|\mathbf{w}\|_2^2}{n}+\sigma^2}-\beta\frac{\mathbf{h}^T\mathbf{w}}{\sqrt{n}}+\frac{\lambda}{n}\|\mathbf{x}_0+\mathbf{w}\|_1. \tag{9.10}$$

To prove Theorem 9.3.1, we need to show that $\Psi(\mathbf{w}_\Phi) \xrightarrow{P} \Gamma(\kappa,\lambda) =: d_*$. The CGMT suggests that d_* is the converging limit of $\Psi(\mathbf{w}_\phi)$, the solution to the *Auxiliary Optimization* (AO) in (9.10). The strategy now becomes clear. First, we need to analyze the (AO) problem in (9.10) and find the converging limit of $\Psi(\mathbf{w}_\phi)$, say d_*. The second step consists of showing that the objective function of the (AO) strictly increases when $\mathbf{w}$ is constrained such that $\Psi(\mathbf{w}_\phi)$ is far from d_*.

We start by a "Sclarization" of the (AO) as in Chapter 5. Only now we also need to keep track of the optimal direction of $\mathbf{w}_\phi$ in (B.24). It can be shown that

$$\mathbf{w}_{\phi,i}=\mathbf{x}_{0,i}-\overrightarrow{\eta}\left(\kappa(\mathbf{g},\mathbf{h})\mathbf{h}_i+\mathbf{x}_{0,i}\ ,\ \kappa(\mathbf{g},\mathbf{h})\cdot\lambda\right),$$

where $\kappa(\mathbf{g},\mathbf{h}) := \sqrt{\alpha^2(\mathbf{g},\mathbf{h})+\sigma^2}/\sqrt{\delta}$ and $\alpha(\mathbf{g},\mathbf{h})$ is the minimizer of the random scalar optimization problem in (B.10) (when specialized to the LASSO, (B.10) can be expressed as in [TAH15, eqn. (46)]).

The next step of the CGMT framework, namely the “Convergence analysis" of the (AO) shows that $\alpha(\mathbf{g},\mathbf{h})$ converges to α as this is defined in Lemma 9.2.1.

Thus, we can condition on the hight probability event that $\alpha(\mathbf{g},\mathbf{h})\to\alpha$ to show that

$$\Psi(\mathbf{w}_\phi)\xrightarrow{P}\Psi(\mathbf{x}_0-\overrightarrow{\eta}(\kappa\mathbf{h}+\mathbf{x}_0,\kappa\cdot\lambda)). \tag{9.11}$$

But the latter term converges to $\Gamma(\kappa,\lambda)$ by assumption, thus showing that the formula of Theorem 9.3.1 holds for the solution of the (AO) problem. It is also worth mentioning that the above argument shows the entries of $\mathbf{w}_\phi$ to be asymptotically iid.

To complete the proof of the theorem, we need to verify that the objective function of the (AO) strictly increases when $\mathbf{w}\in\mathcal{S}_\epsilon$ for the set $\mathcal{S}_\epsilon$ defined in (9.8). For any

$\delta > 0$, let $\epsilon = 2\delta > 0$ and consider any $\mathbf{w} \in \mathcal{S}_{2\delta}$. By (9.11), $|\Psi(\mathbf{w}_\phi) - \Gamma(\kappa, \lambda)| < \delta$ w.p.a. 1. Combined, $|\Psi(\mathbf{w}) - \Psi(\mathbf{w}_\phi)| > \delta$ w.p.a. 1. But then, the Lipschitzness property of Ψ implies that

$$n^{-1/2}\|\mathbf{w} - \mathbf{w}_\phi\|_2 > \delta/L,$$

and the desired conclusion follows by showing that the objective function in (9.10) is *strongly convex* in $\mathbf{w}$ and recalling optimality of $\mathbf{w}_\phi$. In particular, $\beta\|\mathbf{g}\|_2\sqrt{\|\mathbf{w}\|_2^2/n + \sigma^2}$ is strongly convex with coefficient τ/n for some constant $\tau > 0$ (independent of n). Thus the objective function of the optimization in (9.10) (call it $F(\cdot)$) satisfies

$$F(\mathbf{w}) \geq F(\mathbf{w}_\phi) + C\frac{\|\mathbf{w} - \mathbf{w}_\phi\|_2^2}{n} \geq F(\mathbf{w}_\phi) + \frac{\tau\delta}{L}.$$

Proof of Theorem 9.4.1

The two metrics defined in (9.2) do not satisfy the Lipschitz property. Nevertheless the proof of Theorem 9.4.1 follows from Theorem 9.3.1 when combined with a weak-convergence argument. Let $\psi : \mathbb{R} \to \mathbb{R}$ be arbitrary L-Lipschitz function. By Theorem 9.3.1, $(1/n)\sum_i \psi(\mathbf{w}_{\Phi,i}) \xrightarrow{P} \text{Dist}_\psi(\kappa, \lambda)$. Since this holds for all Lipschitz functions, the empirical probability measure of $\mathbf{w}_\Phi$ converges [Bil79, Thm. 25.8]. Hence, it follows (almost identically as in [Bil79, Thm. 19]) that $\Psi_{\epsilon,on} \xrightarrow{P} \Gamma(\kappa, \lambda)$, where Γ as in Assumption 9.3.1 for the function $\Psi(\mathbf{w}) = 1/k\sum_{i=1}^k \mathbb{1}_{\{|\mathbf{w}_i - \mathbf{x}_{0,i}| \geq \epsilon\}}$. Simplifying the "$\Gamma(\kappa, \lambda)$-term" yields the statement of Theorem 9.4.1.

Chapter 10

APPLICATION: THE BIT-ERROR RATE OF THE BOX-RELAXATION OPTIMIZATION

Convex relaxation based methods have emerged as popular tools for structured signal recovery. In the previous chapters, we have developed a complete theory characterizing the error performance of such methods under linear noisy measurements and Gaussian design matrices. Here, we apply the results in a specific example, that of data detection in Multiple-Input Multiple-Output (MIMO) communication systems with a large number of antennas at both ends, and further explore the practical implications.

To be concrete, we consider the problem of recovering an n-dimensional signal $\mathbf{x}_0 \in C^n$ from the noisy MIMO input-output relation $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$, where C is a finite constellation (e.g. BPSK, M-ary PAM, etc.) , $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the MIMO channel matrix (assumed to be known), $\mathbf{z} \in \mathbb{R}^m$ is the noise vector [NLM13; Wen+14a; NC14; Cha+15]. A large host of exact and heuristic optimization algorithms have been proposed. Exact algorithms, such as sphere decoding and its variants, become computationally prohibitive as the problem dimension grows.

Heuristic algorithms such as zero-forcing, MMSE, decision-feedback, etc., [GLS12; Fos96; HV05] have inferior performances that are often difficult to precisely characterize. A popular such heuristic is the *Box Relaxation Optimization* (BRO) decoder, which is a convex relaxation of the ML decoder [TRL01; YYU02; Ma+02], and allows one to recover the signal via convex optimization followed by hard thresholding. Despite its popularity, very little has been known about its performance.

Applying the theory developed in this thesis, this chapter characterizes the *bit-wise error rate* (BER) of the BRO in the regime of large dimensions and under Gaussian assumptions on the channel matrix, for the first time. Further implications of the analysis are also discussed.

In order to make ideas concrete, we focus primarily on the case of BPSK signal recovery which then occupies Sections 10.1 and 10.2. In Section 10.3 we include extensions of the results to other signal constellations.

10.1 BPSK Signal Recovery

Setup. Our goal is to recover an n-dimensional BPSK vector $\mathbf{x}_0 \in \{\pm 1\}^n$ from the noisy multiple-input multiple output (MIMO) relation $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the MIMO channel matrix (assumed to be known) and $\mathbf{z} \in \mathbb{R}^m$ is the noise vector. We assume that $\mathbf{A}$ has entries iid $\mathcal{N}(0, 1/n)$ and $\mathbf{z}$ has entries iid $\mathcal{N}(0, \sigma^2)$. The normalization is such that the reciprocal of the noise variance σ^2 is equal to the Signal-to-Noise Ratio, i.e. SNR $= 1/\sigma^2$.

The Maximum-Likelihood (ML) decoder. The ML decoder which maximizes the probability of error (assuming the $\mathbf{x}_{0,i}$ are equally likely) is given by $\min_{\mathbf{x}\in\{\pm1\}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$. Solving the above is often computationally intractable, especially when n is large, and therefore a variety of heuristics have been proposed (zero-forcing, mmse, decision-feedback, etc.) [Ver98].

Box Relaxation Optimization. The heuristic we shall use is referred to as Box Relaxation Optimization (BRO). It consists of two steps. The first one involves solving a convex relaxation of the ML algorithm, where $\mathbf{x} \in \{\pm 1\}^n$ is relaxed to $\mathbf{x} \in [-1, 1]^n$. The output of the optimization is hard-thresholded in the second step to produce the final binary estimate. Formally, the algorithm outputs an estimate $\mathbf{x}^*$ of $\mathbf{x}_0$ given as

$$\hat{\mathbf{x}} = \arg \min_{-1\le \mathbf{x}_i \le 1} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2, \tag{10.1a}$$

$$\mathbf{x}^* = \operatorname{sign}(\hat{\mathbf{x}}), \tag{10.1b}$$

where the sign function returns the sign of its input and acts element-wise on input vectors.

Bit error probability. We evaluate the performance of the detection algorithm by the bit error probability P_e, defined as the expectation of the Bit Error Rate BER . Formally,

$$BER \ := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{x}_i^* \neq \mathbf{x}_{0,i}\}}, \tag{10.2a}$$

$$P_e := \mathbb{E}\left[BER\right] = \frac{1}{n}\sum_{i=1}^{n} \Pr\left(\mathbf{x}_i^* \neq \mathbf{x}_{0,i}\right). \tag{10.2b}$$

Our main result analyzes the BER of the (BRO) in (10.1). We assume a large-system limit where $m, n \to \infty$ at a proportional rate δ. The SNR is assumed

constant; in particular, it does not scale with n. Let $Q(\cdot)$ denote the Q-function associated with the standard normal density $p(h) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-h^2/2}$.

Theorem 10.1.1 (*BER* of the (BRO) for BPSK)**.** *Let BER denote the bit error rate of the detection scheme in* (10.1) *for some fixed but unknown BPSK signal* $\mathbf{x}_0 \in \{\pm 1\}^n$. *For constant* SNR *and* $\frac{m}{n} \to \delta \in (\frac{1}{2}, \infty)$, *it holds:*

$$\lim_{n\to\infty} BER = Q(1/\tau_*),$$

where τ_* *is the unique solution to*

$$\min_{\tau>0} \frac{\tau}{2}\left(\delta - \frac{1}{2}\right) + \frac{1/\mathrm{SNR}}{2\tau} + \frac{\tau}{2}\int_{\frac{2}{\tau}}^{\infty}\left(h - \frac{2}{\tau}\right)^2 p(h)\mathrm{d}h. \tag{10.3}$$

Theorem 10.1.1 derives a *precise* formula for the bit error probability of the (BRO). As is common in the results of this thesis, the formula involves solving a *convex* and deterministic minimization problem in (10.3). The proof of the theorem is of course based on the CGMT framework of Chapter 5. In particular, we rely on the extension of the framework to general performance metrics in Chapter 9. See Appendix F for details.

10.2 Implications

Computing τ_*. It can be shown that the objective function of (10.3) is strictly convex when $\delta > \frac{1}{2}$. When $\delta < \frac{1}{2}$, it is well known that even the noiseless box relaxation fails [Cha+12]. (In fact, $\delta = \frac{1}{2}$ is the recovery threshold for this convex relaxation.) Thus, (10.3) has a unique solution τ_*. Observe that the problem parameters δ and SNR appear explicitly in (10.3); naturally then τ_* is indeed a function of those. The minimization in (10.3) can be efficiently solved numerically. In addition, owing to the strict convexity of the objective function, τ_* can be equivalently expressed as the unique solution to the corresponding first order optimality conditions.

Numerical illustration. Figure 10.1 illustrates the accuracy of the prediction of Theorem 10.1.1. Note that although the theorem requires $n \to \infty$, the prediction is already accurate for n ranging on a few hundreds.

BER **at high-SNR**. It can be shown that when $\mathrm{SNR} \gg 1$, then

$$\tau_* = 1/\sqrt{(\delta - 1/2)SNR}.$$

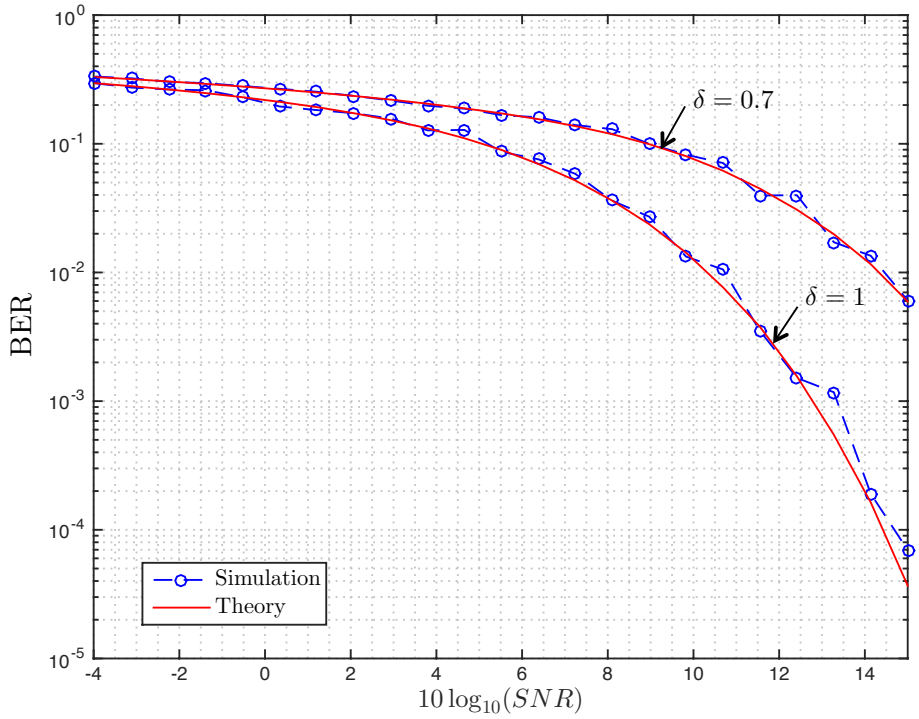

Figure 10.1: Bit error rate performance of the Boxed Relaxation: BER as a function of SNR for different values of the ratio $\delta = \lceil m/n \rceil$ of receive to transmit antennas. The theoretical prediction follows from Theorem 10.1.1. For the simulations, we used $n = 512$. The data are averages over 20 independent realizations of the channel matrix and of the noise vector for each value of the SNR.

This can be intuitively understood as follows: at high-SNR, we expect τ_* to be going to zero (correspondingly BER to be small). When this is the case, the last term in (10.3) is negligible; then τ_* is the solution to $\min_{\tau>0} \frac{\tau}{2}\left(\delta - \frac{1}{2}\right) + \frac{1/\mathrm{SNR}}{2\tau}$ which gives the desired result. Hence, for SNR $\gg 1$,

$$\lim_{n\to\infty} BER \approx Q(\sqrt{(\delta - 1/2)\cdot \mathrm{SNR}}). \tag{10.4}$$

In Figure 10.2 we have plotted this high-SNR expression for the $\log_{10}(BER)$ vs its exact value as predicted by Theorem 10.1.1. It is interesting to observe that the former is actually a very good approximation to the latter even for small practical values of SNR. The range of SNR values for which the approximation is valid becomes larger with increasing δ. Heuristically, for $\delta > 0.7$ the expression in (10.4) is a good proxy for the true probability of error at practical SNR values.

Comparison to the matched filter bound. Theorem 10.1.1 gives us a handle on the

P_e of (BRO) in (10.1) and therefore allows us to evaluate its practical performance. Here, we compare the performance to an idealistic case, where all $n-1$ but 1 bits of $\mathbf{x}_0$ are known to us. As is customary in the field, we refer to the bit error rate of this case as the *matched filter bound* (MFB) and denote it by BER^{MFB}. The (MFB) corresponds to the probability of error in detecting (say) $\mathbf{x}_{0,n} \in \{\pm 1\}$ from: $\tilde{\mathbf{y}} = \mathbf{x}_{0,n}\mathbf{a}_n + \mathbf{z}$, where $\tilde{\mathbf{y}} = \mathbf{y} - \sum_{i=1}^{n-1} \mathbf{x}_{0,i}\mathbf{a}_i$ is assumed known, and, $\mathbf{a}_i$ denotes the i^{th} column of $\mathbf{A}$. The ML estimate is just the sign of the projection of the vector $\tilde{\mathbf{y}}$ to the direction of $\mathbf{a}_n$. Without loss of generality assume that $\mathbf{x}_{0,n} = 1$. Then, the output of the matched filter becomes $\text{sign}(\tilde{X})$, where $\tilde{X} = \|\mathbf{a}_n\|^2 + \sigma^2\nu$, where $\nu \sim \mathcal{N}(0,1)$. When $n \to \infty$, $\|\mathbf{a}_n\|^2 \xrightarrow{P} \delta$. Hence, with probability one,

$$\lim_{n\to\infty} \text{BER}^{MFB} = \lim_{n\to\infty} \mathbb{P}(\tilde{X} < 0) = Q(\sqrt{\delta \cdot \text{SNR}}). \tag{10.5}$$

A direct comparison of (10.5) to (10.4) shows that at high-SNR, the performance of the (BRO) is $10\log_{10}\frac{\delta}{\delta-1/2}$dB off that of the (MFB) . In particular, in the square case ($\delta = 1$), where the number of receive and transmit antennas are the same, the (BRO) is 3dB off the (MFB) . When the number of receive antennas is much larger, i.e. when $\delta \to \infty$, then the performance of the (BRO) approaches the (MFB) .

Improving performance with local algorithms. To prove Theorem 10.1.1 we essentially apply the CGMT Theorem 3.3.1 for the following set[1]:

$$\mathcal{S} = \{\mathbf{v} : \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(\mathbf{v}_i \le -1)} - Q(\frac{1}{\tau^*})\right| < \epsilon\},$$

A study of our analysis of the (AO) reveals that error events for each of the bits in the (AO) are *iid* (see Appendix F). This means that if, for constant k, we define the set:

$$\mathcal{S}_k^* = \{\mathbf{v} : \left|\frac{1}{\binom{n}{k}} \sum_{\substack{T\in\{1,\ldots,n\}\\ |T|=k}} \mathbb{1}_{(\mathbf{v}_{i_1} \le -1,\ldots,\mathbf{v}_{i_k} \le -1)} - Q^k(\frac{1}{\tau^*})\right| < \epsilon\},$$

then $\lim_{n\to\infty} \mathbb{P}\{\mathbf{w}_\phi \in \mathcal{S}_k^*\} = 1$. By Thm. 10.1.1, this implies $\lim_{n\to\infty} \mathbb{P}\{\mathbf{w}_\Phi \in \mathcal{S}_k^*\} = 1$, which means that error events for any fixed k bits in the (PO) are also iid. This fact has significant consequences. For example, it implies that, when a block of data is in error, only a few of its bits are. This means that the output of the (BRO) can be used by various local methods to further reduce the BER. As an example, we can perform a greedy bit-flipping operation on the output of the (BRO) . A possible

[1] Strictly speaking, the performance metric defined by the BER is not Lipschitz, hence a weak approximation argument like the one in Section 9.5 is needed.

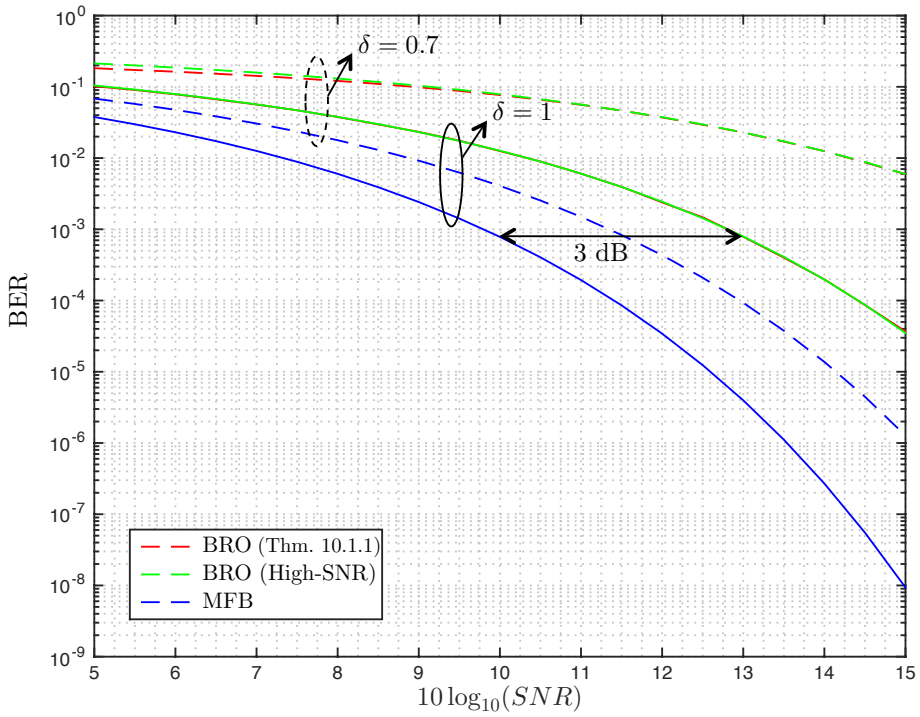

Figure 10.2: Bit error rate of the Box Relaxation Optimization (BRO) in (10.1) in comparison to the Matched Filter Bound (MFB) for $\delta = 0.7$ (dashed lines) and $\delta = 1$ (solid lines). The red curves follow the formula of Thm. 10.1.1, the green ones correspond to (10.4), and, BER^{MFB} of (10.5) is in blue.

implementation of this might be as follows. For $i = 1, 2, \ldots, n$ repeat the following: if $\|\mathbf{y} - \mathbf{A}\mathbf{x}^* - 2 \cdot \mathrm{sign}(\mathbf{x}_i^*) \cdot \mathbf{a}_i\|_2 < \|\mathbf{y} - \mathbf{A}\mathbf{x}^*\|_2$, then $\mathbf{x}_i^{**} = -\mathbf{x}_i^*$; otherwise, $\mathbf{x}_i^{**} = \mathbf{x}_i^*$. Output $\mathbf{x}^{**}$. Here, $\mathbf{a}_i$ denotes the i^{th} column of $\mathbf{A}$. Our preliminary simulation results suggest this simple bit-flipping scheme significantly improves the BER. In future work, we seek to analyze the overall performance, i.e. characterize the BER of $\mathbf{x}^{**}$.

10.3 Extensions

Theorem 10.1.1 precisely computes the BER of the box relaxation method (BRO) to recover BPSK signals in MIMO systems. As the interested reader may expect, similar results can be achieved for higher order constellations (m-PAM, m-QAM, m-PSK, etc.). We discuss such extensions here.

M-PAM consellations

Suppose that each one of the transmit antennas sends a symbol belonging to an M-PAM constellation, i.e. $\mathbf{x}_{0,i} \in C := \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$, for $M = 2k$, $k \in \mathbb{I}$.

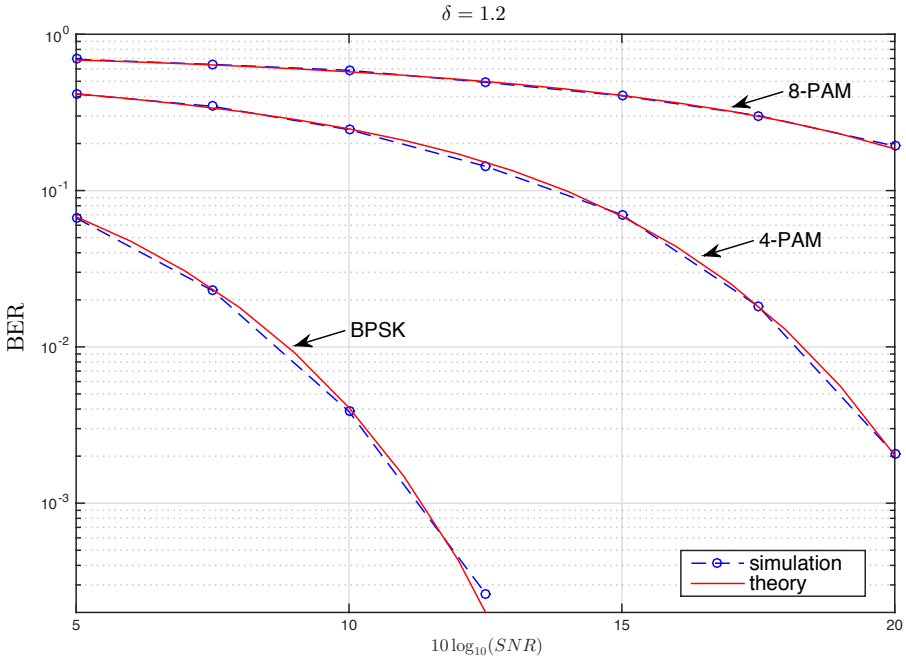

Figure 10.3: Bit error rate of the Box Relaxation Optimization (BRO) in (10.6) as a function of the SNR for BPSK, 4-PAM and 8-PAM signals. The theoretical prediction follows from Theorem 10.3.1. For the simulations, we used n = 512 and $\delta = 1.2$. The data are averages over 20 independent realizations of the channel matrix and of the noise vector for each value of the SNR.

The ML decoder is given by $\min_{\mathbf{x}\in C^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$, which is often computationally intractable. The natural extension of the box-relaxation decoder for BPSK in (10.1) outputs an estimate $\mathbf{x}^*$ of $\mathbf{x}_0$ given as

$$\hat{\mathbf{x}} = \arg \min_{-(M-1)\le \mathbf{x}_i \le (M-1)} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2, \tag{10.6a}$$

$$\mathbf{x}_i^* = \arg \min_{s\in C} |\hat{\mathbf{x}}_i - s|. \tag{10.6b}$$

The theorem below provides a characterization of the BER of the (BRO) scheme in (10.6).

Theorem 10.3.1 (*BER* of the (BRO) for M-ary PAM)**.** *Let BER denote the bit error rate of the detection scheme in* (10.6) *for some unknown signal* $\mathbf{x}_0$ *such that* $\mathbf{x}_{0,i} \overset{iid}{\sim} \text{Uniform}(\{\pm 1, \pm 3, \ldots, \pm(M-1)\})$*, for* $M = 2k$, $k \in \mathbb{I}$. *For constant* SNR *and* $\frac{m}{n} \to \delta \in (1 - \frac{1}{M}, \infty)$*, it holds:*

$$\lim_{n\to\infty} BER = 2(1 - 1/M) \cdot Q(1/\tau_*),$$

where τ_* is the unique solution to

$$\min_{\tau>0} \frac{\tau}{2}\left(\delta - \frac{M-1}{M}\right) + \frac{M^2-1}{3}\cdot\frac{1/\mathrm{SNR}}{2\tau} + \frac{\tau}{M}\sum_{i=1,3,\ldots,M-3}\int_{\frac{M-1-i}{\tau}}^{\infty}\left(h - \frac{M-1-i}{\tau}\right)^2 p(h)\mathrm{d}h + \frac{\tau}{M}\sum_{i=1,3,\ldots,M-1}\int_{\frac{M-1+i}{\tau}}^{\infty}\left(h - \frac{M-1+i}{\tau}\right)^2 p(h)\mathrm{d}h. \tag{10.7}$$

The proof of the theorem is very similar to that of Theorem 10.3.1 (see Appendix F). The theorem assumes a probabilistic model on $\mathbf{x}_0$. In particular, each transmitted symbol $\mathbf{x}_{0,i}$ is uniformly sampled from the M-ary PAM constellation $C = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$. A straightforward extension of the result to other distributions is of course possible. Also, observe that the theorem guarantees meaningful BER performance provided that the ratio of transmit to receive antennas δ is no less than $1 - 1/M < 1$.

Chapter 11

NON-LINEAR MEASUREMENTS

This chapter extends the performance analysis beyond the linear measurement model. Instead of linear measurements $\mathbf{y}_j = \mathbf{a}_j^T \mathbf{x}_0 + \mathbf{z}_j$ we consider estimating an unknown signal vector $\mathbf{x}_0 \in \mathbb{R}^n$ from m measurements taking the following form:

$$\mathbf{y}_j = g_j(\mathbf{a}_j^T \mathbf{x}_0), \quad j = 1, 2, \ldots, m. \tag{11.1}$$

The g_j's are independent copies of a generically *random*, possibly *non-linear* and potentially unspecified link function g. Such measurement functions could arise in applications where the measurement device has nonlinearities and uncertainties. It could also arise *by design*, e.g., $g_j(x) = \text{sign}(x + z_j)$, corresponds to noisy 1-bit quantized measurements.

For the estimation, we use the generalized-LASSO and ask:

What is the recovery performance of the generalized-LASSO with non-linear measurements of the form (11.1)? (Q.3)

While this approach seems to naively ignore the nonlinearities, we will discuss several good reasons motivating question (Q.3). In fact, it turns out that the LASSO solution is a good estimator of the unknown signal up to a constant of proportionality (information about the magnitude of the signal is in general lost in the nonlinearity).

The main result of this chapter answers question (Q.3) in a precise way. In fact, it does so by establishing an interesting connection to already known results on the LASSO performance under linear measurements, that were derived in previous chapters. This has several worth-exploring implications For instance, it encompasses state-of-the art theoretical results of *one-bit Compressed Sensing*, and generalizations to higher levels of quantization Also, it is used to prove that the optimal quantizer of the measurements that minimizes the estimation error of the Generalized LASSO is the celebrated Lloyd-Max quantizer.

We begin in Section 11.1 with motivating Question Q.3. Our answer to the question is simple to state, and is thus summarized in the same section. The section further discusses how this extends relevant results in the literature. A formal statement of our main theorem follows in Section 11.2, where numerical illustrations are

also provided. The chapter concludes in Section 11.3 with an optimal and efficient algorithm for the design of quantized measurements.

11.1 Motivation & Contribution

Non-linear Measurements

We consider recovering a structured signal $\mathbf{x}_0 \in \mathbb{R}^n$ from m measurements taking the form in (11.1). For instance, $g_i(x) = x + z_j$, with (say) z_j being normally distributed, recovers the standard linear regression setup with Gaussian noise. Here, we are particularly interested in scenarios where g is *non-linear*. Notable examples include $g(x) = \text{sign}(x)$ (or $g_j(x) = \text{sign}(x + z_j)$) and $g(x) = (x)_+$, corresponding to 1-bit quantized (noisy) measurements and to the censored Tobit model, respectively. Depending on the situation, g might be known or unspecified. In the statistics and econometrics literature, the measurement model in (11.1) is popular under the name *single-index model* and several aspects of it have been well-studied, e.g. [Bri82; Bri77; Ich93; LD89][1].

Using the Generalized LASSO for Non-linear Measurements?

When the link function is *linear*, i.e. $g_i(x) = x + z_i$, we have previously seen that a good estimate of $\mathbf{x}_0$ is obtained via solving the Generalized LASSO[2] algorithm:

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x}). \tag{11.2}$$

Of course, no one stops us from continuing to use it even in cases where $y_i = g(\mathbf{a}_i^T \mathbf{x}_0)$ with g being *non*-linear[3]. But, the question then becomes: Can there be any guarantees that the solution $\hat{\mathbf{x}}$ of the Generalized LASSO is still a good estimate of $\mathbf{x}_0$?

While the LASSO is by nature tailored to a linear model for the measurements (indeed, the first term of the objective function in (11.2) tries to fit $\mathbf{A}\mathbf{x}$ to the observed vector $\mathbf{y}$ presuming that this is of the form $y_i = \mathbf{a}_i^T \mathbf{x}_0 + \text{noise}$), there are several good reasons motivating this question, as discussed below.

[1] The single-index model is a classical topic and can also be regarded as a special case of what is known as *sufficient dimension reduction* problem. There is extensive literature on both subjects; unavoidably, we only refer to the directly relevant works here.

[2] To be more precise the version considered here is the ℓ_2-LASSO. Our results can be accustomed to the ℓ_2^2-LASSO, but for concreteness, we restrict attention to (11.2) throughout. Also, following the common practice in this thesis, we often drop the term "Generalized" and refer to (11.2) simply as the LASSO.

[3] Note that the Generalized LASSO in (11.2) does not assume knowledge of g. All that is assumed is the availability of the measurements y_i. Thus, the link-function might as well be unknown or unspecified.

1. In the early 80's, Brillinger [Bri82] showed that in the classical statistics regime of large m and of fixed n, when the measurement vectors $\mathbf{a}_j$ are Gaussian, the least-squares estimate of $\mathbf{x}_0$ has the favorable property of being (asymptotically) consistent up to a constant of proportionality. In the modern regime of high-dimensions (both large m and n) and of structured signals, the generalized-LASSO method is replacing ordinary least-squares. Thus, it is natural to ask to what extent does Brillinger's result continue to hold under this different setting?

2. The link function g might be unspecified, and so no additional information about it is available in this case. On the other hand, in many interesting examples the nonlinearity arises by design, and thus, is known. It might then be appealing to perform maximum likelihood (ML) type estimation. However, several issues arise: (i) often this requires knowledge of the distribution of the noise, and in practice one would not expect this to be known, (ii) the ML algorithm might be computationally inefficient; in contrast, the LASSO is appealing because an abundance of efficient, specialized solvers for it are readily available.

3. The question (Q.3) can be interpreted as a study of the *robustness* of the LASSO method to model mismatches. More often than not, the linear model of (1.1) represents reality only approximately by ignoring potential non-linearities in the measurement device. Answering question (Q.3) complements the results on the LASSO performance under linear measurements with robustness guarantees.

Summary of Contributions

Question (Q.3) was first studied back in the early 80's by Brillinger [Bri82] who provided answers in the case of solving (11.2) without a regularizer term. This, of course, corresponds to standard Least Squares (LS). Interestingly, he showed that when the measurement vectors are Gaussian, then the LS solution is a consistent estimate of $\mathbf{x}_0$, up to a constant of proportionality μ, which only depends on the link-function g. The result is sharp, but only under the assumption that the number of measurements m grows large, while the signal dimension n stays fixed, which was the typical setting of interest at the time. In the world of structured signals and high-dimensional measurements, the problem was only very recently revisited by Plan and Vershynin [Ver10b]. They consider a *constrained* version of the Gen-

eralized LASSO, in which the regularizer is essentially replaced by a constraint, and derive upper bounds on its performance. The bounds are not tight (they involve absolute constants), but they demonstrate some key features: i) The solution to the constrained LASSO $\hat{\mathbf{x}}$ is a good estimate of $\mathbf{x}_0$ up to the same constant of proportionality μ that appears in Brillinger's result. ii) Thus, $\|\hat{\mathbf{x}} - \mu\mathbf{x}_0\|_2^2$ is a natural measure of performance. iii) Estimation is possible even with $m < n$ measurements by taking advantage of the structure of $\mathbf{x}_0$.

Inspired by the work of Plan and Vershynin [Ver10b] and motivated by recent advances on the precise analysis of the Generalized LASSO with linear measurements, we extend these latter results to the case of non-linear mesaurements. When the measurement matrix $\mathbf{A}$ has entries i.i.d. Gaussian (henceforth, we assume this to be the case without further reference), and the estimation performance is measured in a mean-squared-error sense, we are able to *precisely* predict the asymptotic behavior of the error. The derived expression accurately captures the role of the link function g, the particular structure of $\mathbf{x}_0$, the role of the regularizer f, and, the value of the regularizer parameter λ. Further, it holds for all values of λ, and for a wide class of functions f and g.

Interestingly, our result shows in a very precise manner that in large dimensions, modulo the information about the magnitude of $\mathbf{x}_0$, the LASSO treats non-linear measurements exactly as if they were scaled and noisy linear measurements with scaling factor μ and noise variance σ^2 defined as

$$\mu := \mathbb{E}[\gamma g(\gamma)], \quad \text{and} \quad \sigma^2 := \mathbb{E}[(g(\gamma) - \mu\gamma)^2], \qquad \text{for } \gamma \sim \mathcal{N}(0,1), \tag{11.3}$$

where the expecation is with respect to both γ and g. In particular, when g is such that $\mu \neq 0$[4], then,

the estimation performance of the Generalized LASSO with measurements of the form $y_i = g_i(\mathbf{a}_i^T\mathbf{x}_0)$ is asymptotically the same as if the measurements were rather of the form $y_i = \mu\mathbf{a}_i^T\mathbf{x}_0 + \sigma z_i$, with μ, σ^2 as in (11.3) and z_i standard Gaussian noise.

Owing to this equivalence, all results established in previous chapters on the the performance of the LASSO under noisy linear measurements can be readily used to

[4]This excludes for example link functions g that are even, but also some other not so obvious cases [GP+13, Sec. 2.2]. For a few special cases, e.g. sparse recovery with binary measurements y_i [Yi+15], different methodologies than the LASSO have been recently proposed that do not require $\mu = 0$.

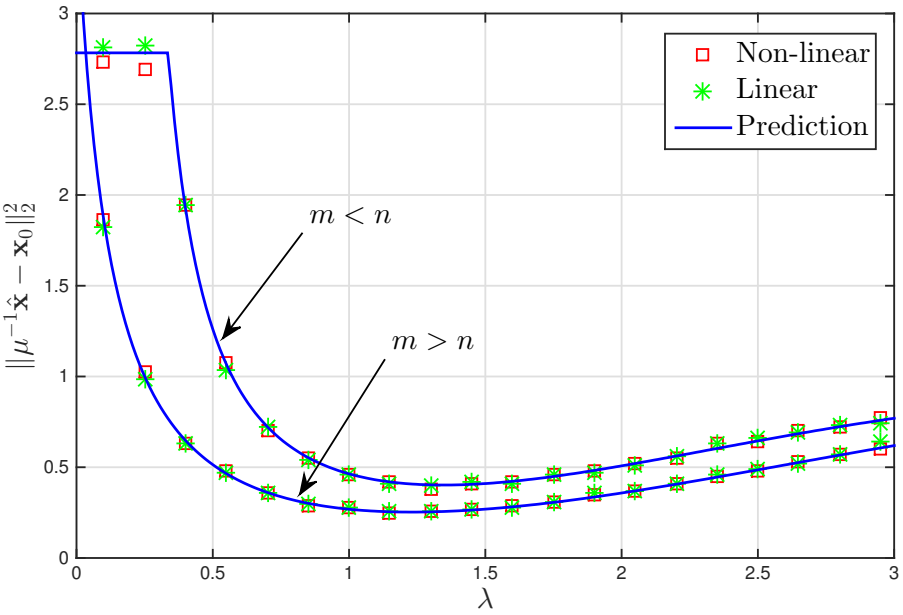

Figure 11.1: Squared error of the ℓ_1-regularized LASSO with non-linear measurements (□) and with corresponding linear ones (⋆) as a function of the regularizer parameter λ; both compared to the asymptotic prediction. Here, $g_i(x) = \text{sign}(x+0.3z_i)$ with $z_i \sim \mathcal{N}(0,1)$. The unknown signal $\mathbf{x}_0$ is of dimension $n = 768$ and has $\lceil 0.15n \rceil$ non-zero entries. The different curves correspond to $\lceil 0.75n \rceil$ and $\lceil 1.2n \rceil$ number of measurements, respectively. Simulation points are averages over 20 problem realizations.

characterize the performance of the LASSO in the presence of nonlinearities. Figure 11.1 serves as an illustration; the error with non-linear measurements matches well with the error of the corresponding linear ones and both are accurately predicted by our analytic expression.

Under the generic model in (11.1), which allows for g to even be unspecified, $\mathbf{x}_0$ can, in principle, be estimated only up to a constant of proportionality [Bri82; LD89; Ver10b]. For example, if g is uknown then any information about the norm $\|\mathbf{x}_0\|_2$ could be absorbed in the definition of g. The same is true when $g(x) = \text{sign}(x)$, even though g might be known here. In these cases, what becomes important is the *direction* of $\mathbf{x}_0$. Motivated by this, and, in order to simplify the presentation, we have assumed throughout that $\mathbf{x}_0$ *has unit Euclidean norm*[5], i.e. $\|\mathbf{x}_0\|_2 = 1$.

[5] In [Ver10b, Remark 1.8], they note that their results can be easily generalized to the case when $\|\mathbf{x}_0\|_2 \neq 1$ by simply redefining $\bar{g}(x) = g(\|\mathbf{x}_0\|_2 x)$ and accordingly adjusting the values of the parameters μ and σ^2 in (11.3). The very same argument is also true in our case.

Discussion of Relevant Literature

Extending an Old Result. Brillinger [Bri82] identified the asymptotic behavior of the estimation error of the LS solution $\hat{\mathbf{x}}_{LS} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. He showed that $\hat{\mathbf{x}}_{LS}$ is a strongly consistent estimate of $\mathbf{x}_0$ up to a constant of proportionality μ and further identified its asymptotic distribution [Bri82, Thm. 1]. From the latter, it can be deduced[6] that when both m and n grow large, but such that $m/n \to +\infty$, then $\sqrt{m/n}\|\hat{\mathbf{x}}_{LS} - \mu\mathbf{x}_0\|_2 \to \sigma$. Here, μ and σ^2 are same as in (11.3). Our result can be viewed as a generalization of the above in several directions. First, we extend Brillinger's result to the regime where $m/n = \delta \in (1, \infty)$ and both grow large by showing that

$$\lim_{n\to\infty} \|\hat{\mathbf{x}}_{LS} - \mu\mathbf{x}_0\|_2 = \frac{\sigma}{\sqrt{\delta - 1}}. \tag{11.4}$$

Second, and most importantly, we consider solving the Generalized LASSO instead, to which LS is only a very special case. This allows versions of (11.4) where the error is finite even when $\delta < 1$ (e.g., see (11.7)). Note the additional challenges faced when considering the LASSO: i) $\hat{\mathbf{x}}$ no longer has a closed-form expression, ii) the result needs to additionally capture the role of $\mathbf{x}_0$, f, and, λ.

Motivated by Recent Work. Plan and Vershynin consider the *constrained* Generalized LASSO (see also Section 6.3)[7] :

$$\hat{\mathbf{x}}_{\text{C-LASSO}} = \arg \min_{f(\mathbf{x}) \le f(\mathbf{x}_0)} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2, \tag{11.5}$$

with $\mathbf{y}$ as in (11.1) . In its simplest form, their result shows that when $m \gtrsim \omega^2(\mathcal{T}_f(\mu\mathbf{x}_0))$ then with high probability,

$$\|\hat{\mathbf{x}}_{\text{C-LASSO}} - \mu\mathbf{x}_0\|_2 \lesssim \frac{\sigma\sqrt{\omega^2(\mathcal{T}_f(\mu\mathbf{x}_0)) + \zeta}}{\sqrt{m}}. \tag{11.6}$$

Recall from Definition 2.2.3 that $\omega^2(\mathcal{T}_f(\mu\mathbf{x}_0))$ is the (squared) Gaussian width and often $\omega^2(\mathcal{T}_f(\mu\mathbf{x}_0))$ is less than n. Thus, estimation is in principle is possible with

[6] Theorem 1 in [Bri82] does not require $n \to \infty$. Here, we have translated the original result to the doubly asymptotic setting that is adapted throughout the thesis, where $m, n \to \infty$. Note however that Brillinger's result is valid only in the classical statistical regime: here, $m/n \to \infty$. Also, to conclude with the stated result starting from Brillinger's theorem, we have silently assumed that $\max_i \mathbf{x}_{0,i} \xrightarrow{P} 0$. I would like to thank Martin Slawski for pointing out the necessity of this assumption.

[7] In fact, Plan and Vershynin consider the slightly more general formulation of the constrained LASSO $\min_{\mathbf{x}\in\mathcal{K}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$, where $\mathcal{K} \subset \mathbb{R}^n$ is some known set (not necessarily convex).

$m < n$ measurements. The parameters μ and σ that appear in (11.6) are the same as in (11.3) and $\zeta := \mathbb{E}[(g(\gamma) - \mu\gamma)^2\gamma^2]$. Observe that, in contrast to Brillinger's result and to our setting, the result in (11.6) is *non-asymptotic*. Also, it suggests the critical role played by μ and σ. On the other hand, (11.6) is only an upper bound on the error, and also it suffers from unknown absolute proportionality constants (hidden in $\lesssim$).

Moving the analysis into an asymptotic setting, our work expands upon the result of [Ver10b]. First, we consider the regularized LASSO instead, which is more commonly used in practice. Most importantly, we improve the loose upper bounds into *precise* expressions. In turn, this proves in an exact manner the role played by μ and σ^2 to which (11.6) is only indicative. For a direct comparison with (11.6) we mention the following result which follows from our analysis (we omit the proof for brevity). Assume f is convex, $m/n = \delta \in (0, \infty)$, $\omega^2(\mathcal{T}_f(\mu\mathbf{x}_0))/n = \rho \in (0, 1]$ and $n \to \infty$. Also, $\delta > \rho$. Then, (11.6) yields an upper bound $C\sigma\sqrt{\rho/\delta}$ to the error, for some constant $C > 0$. Instead, we show

$$\|\hat{\mathbf{x}}_{\text{C-LASSO}} - \mu\mathbf{x}_0\|_2 \leq \sigma \frac{\sqrt{\rho}}{\sqrt{\delta - \rho}}. \tag{11.7}$$

11.2 Results

Modeling Assumptions

Unknown structured signal: We let $\mathbf{x}_0 \in \mathbb{R}^n$ represent the unknown signal vector. We assume that $\mathbf{x}_0 = \bar{\mathbf{x}}_0/\|\bar{\mathbf{x}}_0\|_2$, with $\bar{\mathbf{x}}_0$ sampled from a probability density $p_{\bar{\mathbf{x}}_0}$ in $\mathbb{R}^n$. Thus, $\mathbf{x}_0$ is deterministically of unit Euclidean-norm (this is mostly to simplify the presentation, see Footnote 4). As in Chapter 4, information about the structure of $\bar{\mathbf{x}}_0$ (and correspondingly of $\mathbf{x}_0$) is encoded in $p_{\bar{\mathbf{x}}_0}$.

Regularizer: We consider *convex* regularizers $f : \mathbb{R}^n \to \mathbb{R}$.

Measurement matrix: The entries of $\mathbf{A} \in \mathbb{R}^{m\times n}$ are i.i.d. $\mathcal{N}(0, 1)$.

Measurements and Link-function. We observe $\mathbf{y} = \vec{g}(\mathbf{A}\mathbf{x}_0)$ where $\vec{g}$ is a (possibly random) map from $\mathbb{R}^m$ to $\mathbb{R}^m$ and $\vec{g}(\mathbf{u}) = [g_1(u_1), \ldots, g_m(u_m)]^T$. Each g_i is i.i.d. from a real valued random function g for which μ and σ^2 are defined in (11.3). We assume that μ and σ^2 are nonzero and bounded.

Asymptotics. The regime of study is the linear asymptotic regime of Chapter 4. We repeat here for convenience: we consider a sequence of problem instances $\{\bar{\mathbf{x}}_0^{(n)}, \mathbf{A}^{(n)}, f^{(n)}, m^{(n)}\}_{n\in\mathbb{N}}$ indexed by n such that $\mathbf{A}^{(n)} \in \mathbb{R}^{m\times n}$ has entries i.i.d.

$\mathcal{N}(0,1)$, $f^{(n)} : \mathbb{R}^n \to \mathbb{R}$ is proper convex, and $m := m^{(n)}$ with $m = \delta n, \delta \in (0,\infty)$. We further require that the following conditions hold:

(a) $\overline{\mathbf{x}}_0^{(n)}$ is sampled from a probability density $p_{\overline{\mathbf{x}}_0}^{(n)}$ in $\mathbb{R}^n$ with one-dimensional marginals that are independent of n and have bounded second moments. Furthermore, $n^{-1}\|\overline{\mathbf{x}}_0^{(n)}\|_2^2 \xrightarrow{P} \sigma_x^2 = 1$.

(b) For any $n \in \mathbb{N}$ and any $\|\mathbf{x}\|_2 \le C$, it holds $n^{-1/2} f(\mathbf{x}) \le c_1$ and $n^{-1/2} \max_{\mathbf{s}\in\partial f^{(n)}(\mathbf{x})} \|\mathbf{s}\|_2 \le c_2$, for constants $c_1, c_2, C \ge 0$ independent of n.

The assumption $\sigma_x^2 = 1$ holds without loss of generality and is only necessary to simplify the presentation. In (b), $\partial f(\mathbf{x})$ denotes the subdifferential of f at $\mathbf{x}$. The condition itself is no more than a normalization condition on f.

Every such sequence $\{\overline{\mathbf{x}}_0^{(n)}, \mathbf{A}^{(n)}, f^{(n)}\}_{n\in\mathbb{N}}$ generates a sequence $\{\mathbf{x}_0^{(n)}, \mathbf{y}^{(n)}\}_{n\in\mathbb{N}}$ where $\mathbf{x}_0^{(n)} := \overline{\mathbf{x}}_0^{(n)}/\|\overline{\mathbf{x}}_0^{(n)}\|_2$ and $\mathbf{y}^{(n)} := \vec{g}^{(n)}(\mathbf{A}\mathbf{x}_0)$. When clear from the context, we drop the superscript (n).

Precise Error Prediction

Let $\{\mathbf{x}_0^{(n)}, \mathbf{A}^{(n)}, f^{(n)}, \mathbf{y}^{(n)}\}_{n\in\mathbb{N}}$ be a sequence of problem instances that satisfy all the conditions above. With these, define the sequence $\{\hat{\mathbf{x}}^{(n)}\}_{n\in\mathbb{N}}$ of solutions to the corresponding LASSO problems for fixed $\lambda > 0$:

$$\hat{\mathbf{x}}^{(n)} := \min_{\mathbf{x}} \frac{1}{\sqrt{n}} \left\{ \|\mathbf{y}^{(n)} - \mathbf{A}^{(n)}\mathbf{x}\|_2 + \lambda f^{(n)}(\mathbf{x}) \right\}. \tag{11.8}$$

The main contribution of this paper is a precise evaluation of $\lim_{n\to\infty} \|\mu^{-1}\hat{\mathbf{x}}^{(n)} - \mathbf{x}_0^{(n)}\|_2^2$ with high probability over the randomness of $\mathbf{A}$, of $\mathbf{x}_0$, and of g.

Our main result requires a further assumption on $p_{\overline{\mathbf{x}}_0}^{(n)}$ and $f^{(n)}$. For the readers already familiar with the content of Chapter 4 this should come as no surprise. In particular, recall that Theorem 4.2.1 characterizing the performance under *linear* measurements holds under Assumptions 4.2.1(a) and 4.2.2(a). Due to slight differences on normalization here[8] , the corresponding assumptions here are expressed as follows.

[8]Note that: (i) while here the entries of $\mathbf{A}$ have unit variance, the variance is normalized to $1/n$ in Chapter 4; (ii) there is a difference in the normalization between (11.8) and (6.27). Those modifications are here necessary since we have imposed $\|\mathbf{x}_0\|_2 = 1$ (compare to $\|\mathbf{x}_0\|_2 = O(\sqrt{n})$ in Chapter 4). While proper adjustments are required due to such differences, the results of Chapter 4 are of course still applicable in the current setup (at least when g is affine, but see Theorem 11.2.1).

Assumption 11.2.1. *We say that Assumption 11.2.1 holds for f and $p_{\mathbf{x}_0}$ if there exists $F : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$ such that*

$$\frac{1}{n}\{e_{\tilde{f}}(c\mathbf{h} + \mu\overline{\mathbf{x}}_0; \tau) - \tilde{f}(\mu\mathbf{x}_0)\} \xrightarrow{P} F(c, \tau)$$

and F satisfies Assumption 4.2.2(a).
Here, $\tilde{f}(\mathbf{x}) = \frac{1}{\sqrt{n}} f(\mathbf{x}\sqrt{n})$, μ is as in (11.3) and the convergence is over $\overline{\mathbf{x}}_0 \sim p_{\overline{\mathbf{x}}_0}$ and $\mathbf{h} \sim \mathcal{N}(0, \mathbf{I}_n)$.

All remarks made in Chapter 4 regarding the mild nature of the assumptions therein continue to hold for Assumption 11.2.1. For instance, it is naturally met for separable regularizers (although, this is not necessary). Also observe that Assumption 11.2.1 is exactly the same as Assumptions 4.2.1(a) and 4.2.2(a) for homogeneous regularizers of order 1 (e.g. norms).

Theorem 11.2.1 (Non-linear=Linear)**.** *Consider the asymptotic setup of Section 11.2 and let Assumption 11.2.1 hold. Recall μ and σ^2 as in (11.3) and let $\hat{\mathbf{x}}$ be the minimizer of the Generalized LASSO in (11.8) for fixed $\lambda > 0$ and for measurements given by (11.1). Further let $\hat{\mathbf{x}}^{lin}$ be the solution to the Generalized LASSO when used with* linear *measurements of the form $\mathbf{y}^{lin} = \mathbf{A}(\mu\mathbf{x}_0) + \sigma\mathbf{z}$, where $\mathbf{z}$ has entries i.i.d. standard normal. Then, in the limit of $n \to \infty$, with probability one,*

$$\|\hat{\mathbf{x}} - \mu\mathbf{x}_0\|_2^2 = \|\hat{\mathbf{x}}^{lin} - \mu\mathbf{x}_0\|_2^2.$$

Theorem 11.2.1 relates in a very precise manner the error of the Generalized LASSO under non-linear measurements to the error of the same algorithm when used under appropriately scaled noisy linear measurements. Chapters 4 and 6 derive precise asymptotic expressions for the latter, which may then be translated to the more general setting of nonlinear measurements. The theorem below is an immediate corollary of Theorem 11.2.1 when combined with (6.31), which predicts the error of the LASSO under *linear* measurements.

Corollary 11.2.1 (Precise Error Formula for non-linear measurements)**.** *Under the same assumptions of Theorem 11.2.1 and $\delta := m/n$, it holds, with probability one,*

$$\lim_{n\to\infty} \|\hat{\mathbf{x}} - \mu\mathbf{x}_0\|_2^2 = \alpha_*^2,$$

where α_* is the unique optimal solution to the convex program

$$\inf_{\alpha\geq 0} \sup_{\substack{0\leq\beta\leq 1\\ \tau\geq 0}} \beta\sqrt{\delta}\sqrt{\alpha^2+\sigma^2} - \frac{\alpha\tau}{2} - \frac{\alpha\beta^2}{2\tau} + \lambda \cdot F\left(\frac{\alpha\beta}{\tau}, \frac{\alpha\lambda}{\tau}\right). \tag{11.9}$$

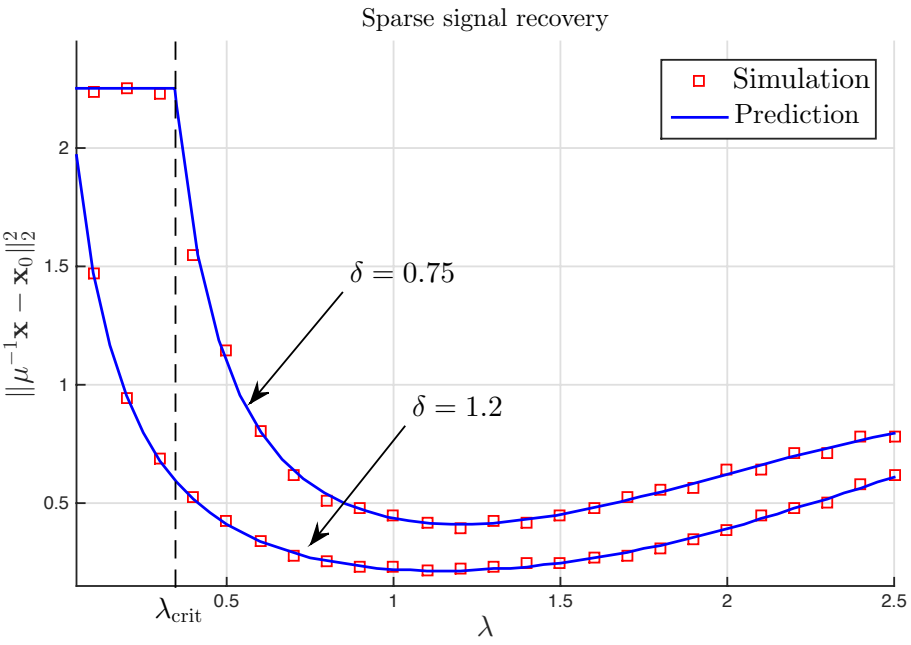

Figure 11.2: Squared error of the ℓ_1-regularized LASSO as a function of the regularizer parameter for noisy 1-bit measurements $g_i(x) = \text{sign}(x + 0.3z_i)$. Here, $\mathbf{x}_0$ is sparse with $p_{X_0}(+1) = p_{X_0}(0) = 0.05$, $p_{X_0}(+1) = 0.9$. The theoretical prediction is obtained by Corollary 11.2.1. Finally, $\delta = 0.75$, $n = 512$, and the simulation points represent averages over 20 realizations.

Note that the two parameters capturing the role of the non-linearity g, namely σ^2 and μ, appear in the objective function in (11.9) explicitly and implicitly through F, respectively. All further remarks that were made in previous chapters regarding the scalar performance optimization (SPO) in (11.9) are also valid here. For instance, it is straightforward to specialize the result to the cases of sparse, group-sparse and low-rank signal recovery to obtain corresponding results to those in Sections 6.6-6.7 (see also [TAH15] for details). Figures 11.1, 11.2 and 11.3 illustrate the accuracy of these predictions.

The proof of Theorem 11.2.1 is deferred to Appendix G.

11.3 Application: Optimal q-bit Quantization

Setup

Consider recovering a *sparse* unknown signal $\mathbf{x}_0 \in \mathbb{R}^n$ from scalar q-bit quantized linear measurements. Let $\mathbf{t} := \{t_0 = 0, t_1, \dots, t_{L-1}, t_L = +\infty\}$ represent a (symmetric with respect to 0) set of decision thresholds and $\boldsymbol{\ell} := \{\pm\ell_1, \pm\ell_2, \dots, \pm\ell_L\}$ the

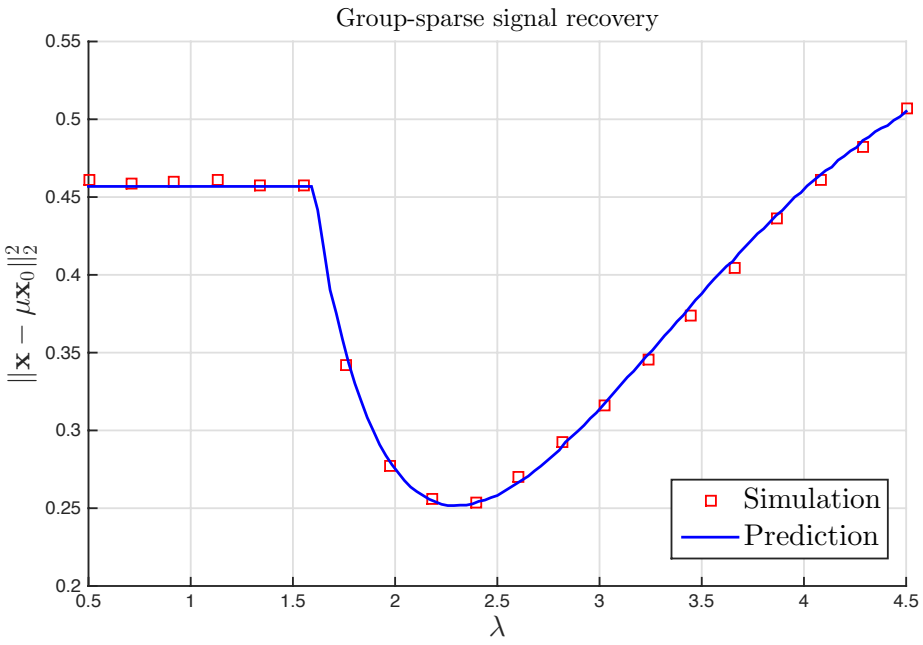

Figure 11.3: Squared error of the group-sparse LASSO as a function of the regularizer parameter compared to the asymptotic predictions for noisy 1-bit measurements $g_i(x) = \text{sign}(x + 0.3z_i)$. Here, $\mathbf{x}_0$ is group-sparse: it is composed of $t = 512$ blocks of block size $b = 3$, and each block is zero with probability 0.95, otherwise its entries are iid $\mathcal{N}(0, 1)$. The theoretical prediction is obtained by Corollary 11.2.1. Finally, $\delta = 0.75$, and the simulation points represent averages over 20 realizations.

corresponding representation points, such that $L = 2^{q-1}$. Then, quantization of a real number x into q-bits can be represented as

$$Q_q(x, \ell, \mathbf{t}) = \text{sign}(x) \sum_{i=1}^{L} \ell_i \mathbf{1}_{\{t_{i-1} \le |x| \le t_i\}},$$

where $\mathbf{1}_{\mathcal{S}}$ is the indicator function of a set $\mathcal{S}$. For example, 1-bit quantization with level ℓ corresponds to $Q_1(x, \ell) = \ell \cdot \text{sign}(x)$. The measurement vector $\mathbf{y} = [y_1, y_2 \ldots, y_m]^T$ takes the form

$$y_i = Q_q(\mathbf{a}_i^T \mathbf{x}_0, \ell, \mathbf{t}), \quad i = 1, 2, \ldots, m, \tag{11.10}$$

where $\mathbf{a}_i^T$'s are the rows of a measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, which is henceforth assumed i.i.d. standard Gaussian. We use the LASSO to obtain an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ as

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda \|\mathbf{x}\|_1. \tag{11.11}$$

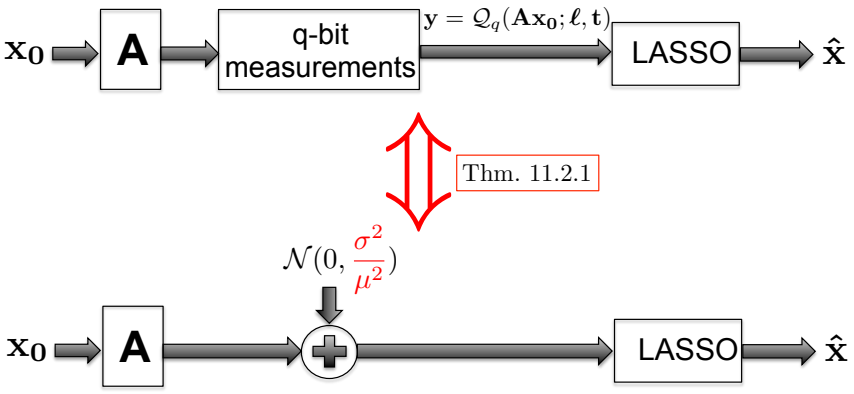

Figure 11.4: Illustration of the equivalence result of Theorem 11.2.1 applied to quantized measurements.

Henceforth, we assume for simplicity that $\|\mathbf{x}_0\|_2 = 1$. Also, in our case, μ is known since $g = Q_q$ is known; thus it is reasonable to scale the solution of (11.11) as $\mu^{-1}\hat{\mathbf{x}}$ and consider the error quantity $\|\mu^{-1}\hat{\mathbf{x}} - \mathbf{x}_0\|_2$ as a measure of estimation performance. Clearly, the error depends (besides others) on the number of bits q, on the choice of the decision thresholds $\mathbf{t}$ and on the quantization levels ℓ. An interesting question of practical importance becomes how to optimally choose these to achieve less error. As a running example for this section, we seek optimal quantization thresholds and corresponding levels

$$(\mathbf{t}_*, \ell_*) = \arg\min_{\mathbf{t},\ell} \|\mu^{-1}\hat{\mathbf{x}} - \mathbf{x}_0\|_2, \tag{11.12}$$

while keeping all other parameters such as the number of bits q and of measurements m fixed.

Consequences of Precise Error Prediction

Theorem 11.2.1 shows that $\|\mu^{-1}\hat{\mathbf{x}} - \mathbf{x}_0\|_2 = \|\hat{\mathbf{x}}^{\text{lin}} - \mathbf{x}_0\|_2$, where $\hat{\mathbf{x}}^{\text{lin}}$ is the solution to (11.11), but only, this time with a measurement vector $\mathbf{y}^{\text{lin}} = \mathbf{A}\mathbf{x}_0 + \frac{\sigma}{\mu}\mathbf{z}$, where μ, σ are as in (11.14) and $\mathbf{z}$ has entries i.i.d. standard normal. See Figure B.1 for an illustration. Thus, lower values of the ratio σ^2/μ^2 correspond to lower values of the error and the design problem posed in (11.12) is equivalent to the following simplified one:

$$(\mathbf{t}_*, \ell_*) = \arg\min_{\mathbf{t},\ell} \frac{\sigma^2(\mathbf{t}, \ell)}{\mu^2(\mathbf{t}, \ell)}. \tag{11.13}$$

To be explicit, μ and σ^2 above can be easily expressed from (11.3) after setting $g = Q_q$ as follows:

$$\mu := \mu(\boldsymbol{\ell}, \mathbf{t}) = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{L} \ell_i \cdot \left(e^{-t_{i-1}^2/2} - e^{-t_i^2/2} \right) \quad \text{and} \quad \sigma^2 := \sigma^2(\boldsymbol{\ell}, \mathbf{t}) := \tau^2 - \mu^2, \tag{11.14}$$

$$\text{where} \quad \tau^2 := \tau^2(\boldsymbol{\ell}, \mathbf{t}) = 2 \sum_{i=1}^{L} \ell_i^2 \cdot (Q(t_{i-1}) - Q(t_i)) \quad \text{and} \quad Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-u^2/2) \mathrm{d}u.$$

An Algorithm for Finding Optimal Quantization Levels and Thresholds

In contrast to the initial problem in (11.12), the optimization involved in (11.13) is explicit in terms of the variables $\boldsymbol{\ell}$ and $\mathbf{t}$, but is still hard to solve in general. Interestingly, we show in Appendix G that the popular Lloyd-Max (LM) algorithm can be an effective algorithm for solving (11.13), since the values to which it converges are stationary points of the objective in (11.13). Note that this is not a directly obvious result since the classical objective of the LM algorithm is minimizing the quantity $\mathbb{E}[\|\mathbf{y} - \mathbf{A}\mathbf{x}_0\|_2^2]$ rather than $\mathbb{E}[\|\mu^{-1}\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2]$.

Chapter 12

CONCLUSIONS AND FUTURE WORK

We will conclude with some brief remarks on various directions for future research that are suggested by the methods and results presented in this dissertation.

High-dimensional Theory of M-estimation

The general and precise results of Chapter 4 (cf., Theorem 4.2.1) can be used to compare performance between different instances of regularized M-estimators under different settings. Figure 6.3 serves as a preliminary numerical illustration: under the specific setting, LAD outperforms the LASSO for appropriate choices of λ. Starting from the error expressions of Theorem 4.2.1 it is interesting to quantify such comparisons and yield such *analytic* conclusions.

Along these lines, one of the most exciting (and at the same time challenging) potential implications of Theorem 4.2.1 is identifying optimal choices for the (convex) loss and regularizer functions under different settings. Since the error characterization differs from the corresponding results of classical statistics (where the signal dimension is fixed), we expect new phenomena to arise and the answers to differ in general. When it comes to the regularizer, the optimality question has been partially considered in the literature. When the structured signal $\mathbf{x}_0$ is considered *fixed*, then a good choice for the regularizer f is one that minimizes the statistical dimension of the tangent cone of f at $\mathbf{x}_0$ (cf. Section 6.3) [Cha+12; Ame+13; OTH13b][1]. The results presented in [Cha+12] and [Ame+13] together, prove that this is indeed the optimal choice in the noiseless case (cf. Section 2.2). The same is true in the high-SNR regime when a least-squares loss function is used (cf. Section 7.4). The more general setting of Chapter 4, will allow revisiting of this question and extension of the results to capture instances where $\mathbf{x}_0$ is associated with a prior distribution $p_{\mathbf{x}_0}$, the loss function differs from a least-squares one, and the noise variance is not necessarily tending to zero. Theorem 4.2.1 suggests the critical role to be played in this effort by the Expected Moreau envelope, which is in fact a generalization of the statistical dimension (cf. Section 6.3). When it comes to the optimal choice

[1] Based on this, Chandrasekaran et. al. have suggested the notion of "atomic-norms" as a principled way of constructing appropriate convex regularizer functions for different kinds of structures [Cha+12].

of the loss function with respect to the noise distribution $p_{\mathbf{z}}$, less is known. Again, the expected Moreau envelope will be central in the optimization, but is yet to be understood how this will translate into practical recipes for the design of optimal loss functions.

Another important question that is also related to the optimal choice of loss/regularizer functions examines the conditions under which the squared-error of (1.2) becomes zero, if this is at all possible. In Remark 6.3.0.44, we discussed an example of an M-estimator that under specific noise and signal distributions becomes consistent provided that the normalized number of measurements is large enough and that the regularizer parameter is chosen on the correct range (also, see Figure 6.3). Answering such questions boils down to identifying conditions under which $\alpha_* = 0$ can be the optimal solution to the (SPO) of Theorem 4.2.1.

Furthermore, Theorem 4.2.1 can be used to provide valuable insights and guidelines regarding optimal choices of the regularizer parameter. In Section 6.10 and Figure 6.3 we presented an example that highlights the importance of selecting λ within the correct range of values, otherwise the performance can be significantly deteriorated. The closed-form formulae of Sections 7.6 and 7.7 suggested simple recipes for optimal tuning of λ. However, they require some prior knowledge on the structure of the signal (e.g. sparsity level, rank) that might not be available in practice (at least in precise form). Thus, it is interesting to study modifications that adapt to these more realistic scenarios, but still take advantage of the precise error formulae.

Beyond Gaussian Designs

All the results of this dissertation, apart from those in Chapter 8, are proved for design matrices that have entries iid Gaussian. Yet, there are potentials of extending the results to other classes of distributions.

Matrices with iid entries. Preliminary numerical results (Figure 6.2 is an example) suggest a *universality* property of the derived error prediction to design matrices with entries iid drawn from a wider class of probability distributions, such as sub-Gaussians. This is similar to the universality of Gaussians in the noiseless setting, which was discussed in Section 2.2 and was recently established by [OT15]. Oymak & Tropp further include preliminary results for the noisy case, where they prove universality of the high-SNR error bounds on the squared error of the constrained LASSO (cf. Section 7.5). Furthermore, El Karoui proves the results to be universal in the case for M-estimators with ridge-regulararization and a twice differ-

entiable loss function [EK15]. Extending these to the general setting of regularized M-estimators of Chapter 4 is of interest.

Isotropically Random Orthogonal (IRO) Matrices. Recall that an IRO matrix $\mathbf{A}$ is sampled uniformly at random from the manifold of row-orthogonal matrices satisfying $\mathbf{A}\mathbf{A}^T = \mathbf{I}_m$. Chapter 8 characterizes the error performance of the Generalized-LASSO under IRO matrices and shows it is different (in fact, superior) to the one under Gaussian designs. The analysis was based on expressing IRO matrices using Gaussians and appropriately massaging the CGMT framework. It is of interest to extend the analysis and corresponding results to loss functions beyond least-squares. Furthermore, numerical simulations in Chapter 8 suggest that the formulae obtained for IRO matrices are also true for random DCT and Hadamard matrices. This is particularly important since the latter designs are often preferred in practice due to reduced computational and storage complexity. Proving what appears to be a universality property of IRO matrices when it comes to recovery performance of regularized M-estimators is an open question.

Elliptical Distributions. Assume $\mathbf{G}$ with entries iid Gaussian, ϵ_i's to be independent and independent of $\mathbf{G}$ and $\mathbf{A} = \text{diag}(\epsilon_1, \ldots, \epsilon_m)\mathbf{G}$. We refer the reader to [EK15] for a motivation on the potential significance of studying such "elliptical-like" distributions. It is rather straightforward to extend the CGMT framework, and consequently the predictions of this thesis, to account for such a class of distributions. It might be worth considering the details in future work.

Graphical LASSO and LP Decoding

The CGMT Theorem 3.3.1 played a key role in the course of this thesis. It let us carry out the analysis to a simple Auxiliary Optimization (AO) instead of the original Primary Optimization (PO) problem. It is conceivable that this key idea is applicable to other problems than the ones considered in this thesis. We discuss two such promising examples next.

The Graphical LASSO is a convex regularized likelihood optimization algorithm that is popularly used to estimate Gaussian graphical models [YL07; FHT08]. There is a long literature on its relevant applications and on related algorithmic issues, but to the best of our knowledge there are no available precise results on its performance.

The LP decoder is a popular linear program for decoding linear codes, especially low density parity check (LDPC) codes. Since the work of Feldman [Fel03], who

was the first to propose the particular relaxation and accompany it with preliminary performance guarantees, and despite a long list of follow-up references [Von], it has remained an open problem to *precisely* quantify its achieved block error probability.

Even though there is no reason to believe that the performance analysis of the two problems ought to have any commonalities (owing to their completely different natures), with some appropriate manipulations we can show that they both boil down to deriving a *matrix analogue version* of the CGMT. To make the question concrete, let $\mathbf{A} \in \mathbb{R}^{m\times n}$ have entries iid Gaussian, $\mathcal{S}_{\mathbf{W}} \subset \mathbb{R}^{m\times m}$, $\mathcal{S}_{\mathbf{U}} \subset \mathbb{R}^{n\times n}$ and $\psi : \mathbb{R}^{n\times n} \times \mathbb{R}^{m\times m} \to \mathbb{R}$. We seek an (AO) that corresponds to the following matrix analogue of the (PO) in (3.11a):

$$\min_{\mathbf{W}\in\mathcal{S}_{\mathbf{W}}} \max_{\mathbf{U}\in\mathcal{S}_{\mathbf{U}}} \operatorname{trace}(\mathbf{U}^T\mathbf{A}\mathbf{W}) + \psi(\mathbf{W},\mathbf{U}).$$

The two desired features for the (AO) are that:

a) It is tightly related to the (PO) in a sense similar to the CGMT, i.e., its optimal cost concentrates to the same value as the value of concentration of the (PO).

b) It is simpler to analyze than the (PO).

This observation is promising, and if there be such a matrix analogue version of the CGMT it is very likely to have applications to other problems as well.

A Complex CGMT

The CGMT Theorem 3.3.1 requires the entries of the matrix $\mathbf{A}$ in the (PO) to be real Gaussians. Is it possible to extend the theorem to matrices that have entries iid from a circularly-symmetric complex normal distribution? The driving motivation behind this question is extending the results of Chapter 10 to signal constellations such as M-QAM and M-PSK when the channel coefficients (corresponding to the entries of $\mathbf{A}$) are modeled as complex Gaussians. We are unaware of a "complex version" of even Gordon's original GMT Theorem 3.2.1.

Simple Denoising

In the simple denoising problem, the goal is to estimate an unknown structured signal $\mathbf{x}_0 \in \mathbb{R}^n$ from noisy but *uncompressed* observations $\mathbf{y} = \mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^n$. A natural estimate is obtained by solving the following minimization problem:

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{y}-\mathbf{x}) + \lambda f(\mathbf{x}), \tag{12.1}$$

for some convex loss function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$, a convex regularizer $f : \mathbb{R}^n \to \mathbb{R}$ and $\lambda > 0$. One is then interested in characterizing the estimation performance of (12.1) (e.g., measured in the squared-norm $\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2$) as a function of the involved problem parameters, and further use this to optimally choose the loss function depending on the noise distribution, the value of the regularizer parameter, etc..

Of course, this setup is very similar to the one that has been the subject of this thesis: the measurements $\mathbf{y} = \mathbf{x} + \mathbf{z}$ follow (1.1) with an identity measurement matrix $\mathbf{A} = \mathbf{I}_n$ and the same is true for (12.1) when compared to (1.2). When $\mathbf{A}$ is iid Gaussian, Theorem 4.2.1 derives a precise characterization of the squared error $\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2$ for a general noise distribution, signal structure, loss function, regularizer function and regularizer parameter. To the best of our knowledge there is no available result for the simple denoising problem that reflects the generality of Theorem 4.2.1. The recent works of Chatterjee [Cha+14] and of Oymak and Hassibi [OH15], only consider a least-squares loss function. Moreover, [OH15] only considers the high-SNR regime, while [Cha+14] only considers the constrained version of (12.1).

While the analysis that leads to Theorem 4.2.1 is not directly applicable here (since it requires $\mathbf{A}$ to be iid Gaussian), it is conceivable that some of the techniques developed in this dissertation might still be applicable. To support this claim, we show in Appendix H how such ideas lead to novel characterizations of the squared error of (12.1) for a least-squares loss function. The obtained results significantly extend the corresponding state of the art results in [OH15; Cha+14]. It is an interesting and promising direction of future work, to generalize the analysis to other loss functions.

Algorithmic Opportunities

(a) Our analysis of the box relaxation optimization in Chapter 10 shows in a precise way that, when a block of data is in error, only a few of its bits are. We believe this suggests that its output can be used by various local methods (e.g. Markov Chain Monte Carlo) to device novel algorithms with even better BER and with *provable* performance guarantees, which might be of impact in numerous applications, such as massive MIMO.

(b) Suppose a device with nonlinear output measurements $y_j = g_j(\mathbf{a}_i^T\mathbf{x}_0)$ of $\mathbf{x}_0$ as in Chapter 11. If the regularized least-squares estimator is employed, what is the optimal function h that should applied on top of g to minimize the resulting estimation error? The performance characterization of Chapter 11 allows a concrete

formulation of this optimization, which seems worth exploring.

(c) The CGMT Theorem 3.3.1 relates the regularized M-estimation optimization program with a seemingly unrelated Auxiliary Optimization (AO) problem. Our work shows that the (AO) is simpler to analyze and it effectively predicts the error performance of the M-estimator. It would be interesting to understand whether, beyond the purposes of analysis, the (AO) problem can be useful in suggesting alternative efficient estimation algorithms. A successful example of such an efficient estimation algorithm in the literature, which has also been used for the analysis of the M-estimators, is the Approximate Message Passing (AMP) and its variants [DMM09]. Is it possible to use the machinery of the CGMT to analyze the (AMP)? Is there a deeper relationship of the (AO) problem to the (AMP)? It is intriguing to attempt putting the two methods, which to date appear to emerge independently of each other, under a unifying framework.

BIBLIOGRAPHY

[AAGM15] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D Milman. *Asymptotic Geometric Analysis, Part I*. Vol. 202. American Mathematical Soc., 2015 (cit. on p. 12).

[AG82] Per Kragh Andersen and Richard D Gill. "Cox's regression model for counting processes: a large sample study". In: *The annals of statistics* (1982), pp. 1100–1120 (cit. on pp. 206, 208, 248, 252, 259).

[Ame+13] Dennis Amelunxen et al. "Living on the edge: A geometric theory of phase transitions in convex optimization". In: *arXiv preprint arXiv:1303.6672* (2013) (cit. on pp. 7, 8, 13, 14, 16, 18, 21, 58, 106, 164, 236, 237, 248).

[Bac10] Francis R Bach. "Structured sparsity-inducing norms through submodular functions". In: *Advances in Neural Information Processing Systems*. 2010, pp. 118–126 (cit. on p. 9).

[Ban+14] Arindam Banerjee et al. "Estimation with norm regularization". In: *Advances in Neural Information Processing Systems*. 2014, pp. 1556–1564 (cit. on p. 57).

[Bar93] Andrew R Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: *Information Theory, IEEE Transactions on* 39.3 (1993), pp. 930–945 (cit. on p. 9).

[BC15] Jelena Bradic and Jiao Chen. "Robustness in sparse linear models: relative efficiency based on robust approximate message passing". In: *arXiv preprint arXiv:1507.08726* (2015) (cit. on p. 57).

[BCW10] Richard G Baraniuk, Volkan Cevher, and Michael B Wakin. "Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective". In: *Proceedings of the IEEE* 98.6 (2010), pp. 959–971 (cit. on p. 9).

[BCW11] Alexandre Belloni, Victor Chernozhukov, and Lie Wang. "Square-root lasso: pivotal recovery of sparse signals via conic programming". In: *Biometrika* 98.4 (2011), pp. 791–806 (cit. on pp. 41, 57, 89).

[Bea+13] Derek Bean et al. "Optimal M-estimation in high-dimensional regression". In: *Proceedings of the National Academy of Sciences* 110.36 (2013), pp. 14563–14568 (cit. on p. 59).

[Bil79] Patrick Billingsley. *Probability and measure*. N.Y.: Wiley, 1979 (cit. on p. 142).

[BL10] Jonathan M Borwein and Adrian S Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Vol. 3. Springer, 2010 (cit. on p. 99).

[BLM+15] Mohsen Bayati, Marc Lelarge, Andrea Montanari, et al. “Universality in polytope phase transitions and message passing algorithms”. In: *The Annals of Applied Probability* 25.2 (2015), pp. 753–822 (cit. on pp. 7, 19, 21).

[BLM13] Stéphane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013 (cit. on p. 29).

[BM11] Mohsen Bayati and Andrea Montanari. “The dynamics of message passing on dense graphs, with applications to compressed sensing”. In: *Information Theory, IEEE Transactions on* 57.2 (2011), pp. 764–785 (cit. on p. 21).

[BM12] Mohsen Bayati and Andrea Montanari. “The LASSO risk for gaussian matrices”. In: *Information Theory, IEEE Transactions on* 58.4 (2012), pp. 1997–2017 (cit. on pp. 43, 57, 79, 82, 232, 234).

[BNO03] Dimitri P Bertsekas, Angelia Nedić, and Asuman E Ozdaglar. *Convex analysis and optimization*. Athena Scientific Belmont, 2003 (cit. on pp. 242, 244).

[Boc+15] Holger Boche et al. *Compressed Sensing and its Applications*. Springer, 2015 (cit. on p. 4).

[Bri77] David R. Brillinger. “The identification of a particular nonlinear time series system”. In: *Biometrika* 64.3 (1977), pp. 509–515 (cit. on p. 152).

[Bri82] David R Brillinger. “A GENERALIZED LINEAR MODEL WITH" GAUSSIAN" REGRESSOR VARIABLES”. In: *A Festschrift For Erich L. Lehmann* (1982), p. 97 (cit. on pp. 152, 153, 155, 156).

[BRT09] Peter J Bickel, Yacov Ritov, and Alexandre B Tsybakov. “Simultaneous analysis of Lasso and Dantzig selector”. In: *The Annals of Statistics* 37.4 (2009), pp. 1705–1732 (cit. on p. 57).

[Bru+14] John J Bruer et al. “Time–Data Tradeoffs by Aggressive Smoothing”. In: *Advances in Neural Information Processing Systems*. 2014, pp. 1664–1672 (cit. on pp. 3, 18).

[BV09] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2009 (cit. on p. 3).

[BY93] ZD Bai and YQ Yin. “Limit of the smallest eigenvalue of a large dimensional sample covariance matrix”. In: *The annals of Probability* (1993), pp. 1275–1294 (cit. on p. 36).

[CDS98] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. “Atomic decomposition by basis pursuit”. In: *SIAM journal on scientific computing* 20.1 (1998), pp. 33–61 (cit. on pp. 3, 7).

[Cha+12] Venkat Chandrasekaran et al. “The convex geometry of linear inverse problems”. In: *Foundations of Computational Mathematics* 12.6 (2012), pp. 805–849 (cit. on pp. 7–9, 13, 14, 16, 18, 21, 32, 58, 125, 145, 164, 236, 237, 239).

[Cha+14] Sourav Chatterjee et al. “A new perspective on least squares under convex constraint”. In: *The Annals of Statistics* 42.6 (2014), pp. 2340–2381 (cit. on pp. 168, 268).

[Cha+15] Jeon Charles et al. “Optimality of Large MIMO Detection via Approximate Message Passing”. In: *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE. 2015 (cit. on p. 143).

[CJ13] Venkat Chandrasekaran and Michael I Jordan. “Computational and statistical tradeoffs via convex relaxation”. In: *Proceedings of the National Academy of Sciences* 110.13 (2013), E1181–E1190 (cit. on pp. 18, 104, 105).

[CM73] Jon F Claerbout and Francis Muir. “Robust modeling with erratic data”. In: *Geophysics* 38.5 (1973), pp. 826–844 (cit. on pp. 3, 8, 78).

[CR09] Emmanuel J Candès and Benjamin Recht. “Exact matrix completion via convex optimization”. In: *Foundations of Computational mathematics* 9.6 (2009), pp. 717–772 (cit. on p. 93).

[CRT06] Emmanuel J Candès, Justin Romberg, and Terence Tao. “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”. In: *Information Theory, IEEE Transactions on* 52.2 (2006), pp. 489–509 (cit. on pp. 2, 3).

[CT06] Emmanuel J Candes and Terence Tao. “Near-optimal signal recovery from random projections: Universal encoding strategies?” In: *Information Theory, IEEE Transactions on* 52.12 (2006), pp. 5406–5425 (cit. on pp. 7, 8, 20).

[CT07] Emmanuel Candes and Terence Tao. “The Dantzig selector: Statistical estimation when p is much larger than n”. In: *The Annals of Statistics* (2007), pp. 2313–2351 (cit. on p. 57).

[CX13] Jian-Feng Cai and Weiyu Xu. “Guarantees of Total Variation Minimization for Signal Recovery”. In: *arXiv preprint arXiv:1301.6791* (2013) (cit. on p. 237).

[DGM13] David L Donoho, Matan Gavish, and Andrea Montanari. “The phase transition of matrix recovery from Gaussian measurements matches the minimax MSE of matrix denoising”. In: *Proceedings of the National Academy of Sciences* 110.21 (2013), pp. 8405–8410 (cit. on p. 98).

[DJ94] David L Donoho and Iain M Johnstone. “Minimax risk overl p-balls forl p-error”. In: *Probability Theory and Related Fields* 99.2 (1994), pp. 277–303 (cit. on p. 43).

[DJM13] David L Donoho, Lain Johnstone, and Andrea Montanari. “Accurate Prediction of Phase Transitions in Compressed Sensing via a Connection to Minimax Denoising”. In: *IEEE transactions on information theory* 59.6 (2013), pp. 3396–3433 (cit. on pp. 43, 57, 98, 104).

[DM13] David Donoho and Andrea Montanari. “High dimensional robust m-estimation: Asymptotic variance via approximate message passing”. In: *arXiv preprint arXiv:1310.7320* (2013) (cit. on pp. 57, 70, 76).

[DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. “Message-passing algorithms for compressed sensing”. In: *Proceedings of the National Academy of Sciences* 106.45 (2009), pp. 18914–18919 (cit. on pp. 21, 70, 169).

[DMM11] David L Donoho, Arian Maleki, and Andrea Montanari. “The noise-sensitivity phase transition in compressed sensing”. In: *Information Theory, IEEE Transactions on* 57.10 (2011), pp. 6920–6941 (cit. on pp. 43, 57, 98, 103, 232, 234).

[Don+00] David L Donoho et al. “High-dimensional data analysis: The curses and blessings of dimensionality”. In: *AMS Math Challenges Lecture* (2000), pp. 1–32 (cit. on p. 2).

[Don06a] David L Donoho. “Compressed sensing”. In: *Information Theory, IEEE Transactions on* 52.4 (2006), pp. 1289–1306 (cit. on pp. 3, 7, 103).

[Don06b] David L Donoho. “High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension”. In: *Discrete & Computational Geometry* 35.4 (2006), pp. 617–652 (cit. on p. 20).

[Don95] David L Donoho. “De-noising by soft-thresholding”. In: *Information Theory, IEEE Transactions on* 41.3 (1995), pp. 613–627 (cit. on p. 104).

[DT05] David L Donoho and Jared Tanner. “Sparse nonnegative solution of underdetermined linear equations by linear programming”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.27 (2005), pp. 9446–9451 (cit. on p. 237).

[DT09a] David Donoho and Jared Tanner. “Counting faces of randomly projected polytopes when the projection radically lowers dimension”. In: *Journal of the American Mathematical Society* 22.1 (2009), pp. 1–53 (cit. on pp. 7, 20).

[DT09b] David Donoho and Jared Tanner. "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367.1906 (2009), pp. 4273–4293 (cit. on p. 19).

[Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010 (cit. on p. 211).

[EK+13] Noureddine El Karoui et al. "On robust regression with high-dimensional predictors". In: *Proceedings of the National Academy of Sciences* 110.36 (2013), pp. 14557–14562 (cit. on p. 59).

[EK12] Yonina C Eldar and Gitta Kutyniok. *Compressed sensing: theory and applications*. Cambridge University Press, 2012 (cit. on p. 4).

[EK15] Noureddine El Karoui. "On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators". In: (2015) (cit. on pp. 58, 59, 123, 166).

[EPP00] Theodoros Evgeniou, Massimiliano Pontil, and Tomaso Poggio. "Regularization networks and support vector machines". In: *Advances in computational mathematics* 13.1 (2000), pp. 1–50 (cit. on p. 41).

[Faz02] Maryam Fazel. "Matrix rank minimization with applications". PhD thesis. 2002 (cit. on p. 95).

[Fel03] Jon Feldman. "Decoding error-correcting codes via linear programming". PhD thesis. Massachusetts Institute of Technology, 2003 (cit. on p. 166).

[FHT08] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9.3 (2008), pp. 432–441 (cit. on p. 166).

[Flo16] Sharon Florentine. *10 fastest-growing tech skills*. 2016. URL: http://www.cio.com/article/3060812/it-skills-training/10-fastest-growing-tech-skills.html#slide5 (visited on 04/26/2016) (cit. on p. 1).

[FM14] Rina Foygel and Lester Mackey. "Corrupted sensing: Novel guarantees for separating structured signals". In: *Information Theory, IEEE Transactions on* 60.2 (2014), pp. 1223–1247 (cit. on pp. 8, 13, 14, 16, 21, 41, 76, 132, 236, 237).

[Fos96] Gerard J Foschini. "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas". In: *Bell labs technical journal* 1.2 (1996), pp. 41–59 (cit. on p. 143).

[FR13] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Vol. 1. 3. Springer, 2013 (cit. on p. 4).

[GBS09] Dongning Guo, Dror Baron, and Shlomo Shamai. "A single-letter characterization of optimal noisy compressed sensing". In: *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE. 2009, pp. 52–59 (cit. on p. 59).

[GLS12] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*. Vol. 2. Springer Science & Business Media, 2012 (cit. on p. 143).

[Gor85] Yehoram Gordon. "Some inequalities for Gaussian processes and applications". In: *Israel Journal of Mathematics* 50.4 (1985), pp. 265–289 (cit. on p. 30).

[Gor88] Yehoram Gordon. *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbb{R}^n$. Springer, 1988 (cit. on pp. 12, 21, 30, 32, 117, 184).

[GP+13] Alexandra L Garnham, Luke A Prendergast, et al. "A note on least squares sensitivity in single-index model estimation and the benefits of response transformations". In: *Electronic Journal of Statistics* 7 (2013), pp. 1983–2004 (cit. on p. 154).

[GRY11] Silvia Gandy, Benjamin Recht, and Isao Yamada. "Tensor completion and low-n-rank tensor recovery via convex optimization". In: *Inverse Problems* 27.2 (2011), p. 025010 (cit. on p. 237).

[HS] Wolfgang Härdle and Léopold Simar. *Applied multivariate statistical analysis*. Vol. 22007. Springer (cit. on p. 92).

[Hub11] Peter J Huber. *Robust statistics*. Springer, 2011 (cit. on pp. 3, 5, 41).

[Hub73] Peter J Huber. "Robust regression: asymptotics, conjectures and Monte Carlo". In: *The Annals of Statistics* (1973), pp. 799–821 (cit. on p. 5).

[HV05] Babak Hassibi and Haris Vikalo. "On the sphere-decoding algorithm I. Expected complexity". In: *Signal Processing, IEEE Transactions on* 53.8 (2005), pp. 2806–2818 (cit. on p. 143).

[Ich93] Hidehiko Ichimura. "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models". In: *Journal of Econometrics* 58.1 (1993), pp. 71–120 (cit. on p. 152).

[Jia+06] Tiefeng Jiang et al. "How many entries of a typical orthogonal matrix can be approximated by independent normals?" In: *The Annals of Probability* 34.4 (2006), pp. 1497–1529 (cit. on pp. 127, 129).

[JL84] William B Johnson and Joram Lindenstrauss. "Extensions of Lipschitz mappings into a Hilbert space". In: *Contemporary mathematics* 26.189-206 (1984), p. 1 (cit. on p. 4).

[Joh06] Iain M Johnstone. "High dimensional statistical inference and random matrices". In: *arXiv preprint math/0611589* (2006) (cit. on p. 19).

[Jon92] Lee K Jones. "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training". In: *The annals of Statistics* (1992), pp. 608–613 (cit. on p. 9).

[Kar13] Noureddine El Karoui. "Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results". In: *arXiv preprint arXiv:1311.2445* (2013) (cit. on pp. 58, 72).

[KSV13] Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. "Low-rank tensor completion by Riemannian optimization". In: *BIT Numerical Mathematics* (2013), pp. 1–22 (cit. on p. 237).

[KWT10] Yoshiyuki Kabashima, Tadashi Wadayama, and Toshiyuki Tanaka. "Statistical mechanical analysis of a typical reconstruction limit of compressed sensing". In: *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE. 2010, pp. 1533–1537 (cit. on p. 59).

[LD89] Ker-Chau Li and Naihua Duan. "Regression analysis under link violation". In: *The Annals of Statistics* (1989), pp. 1009–1052 (cit. on pp. 152, 155).

[LHC15] Yen-Huan Li, Ya-Ping Hsieh, and Volkan Cevher. *A Geometric View on Constrained M-Estimators*. Tech. rep. 2015 (cit. on p. 57).

[LM08] Friedrich Liese and Klaus-J Miescke. *Statistical decision theory: estimation, testing, and selection*. Springer Science & Business Media, 2008 (cit. on pp. 194, 206, 208).

[LT91] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Vol. 23. Springer, 1991 (cit. on pp. 12, 27, 28, 30).

[Ma+02] Wing-Kin Ma et al. "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA". In: *Signal Processing, IEEE Transactions on* 50.4 (2002), pp. 912–922 (cit. on p. 143).

[Mac16] Brian Mack. *HIMSS 16 Buzzword of the Year*. 2016. URL: http://gl-hc.org/himss16-buzzword-of-the-year/ (visited on 04/20/2016) (cit. on p. 1).

[Mal+13] Arian Maleki et al. "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)". In: *Information Theory, IEEE Transactions on* 59.7 (2013), pp. 4290–4308 (cit. on p. 57).

[MB07] Mokshay Madiman and Andrew Barron. “Generalized entropy power inequalities and monotonicity properties of information”. In: *Information Theory, IEEE Transactions on* 53.7 (2007), pp. 2317–2329 (cit. on p. 76).

[McC+14] Michael B McCoy et al. “Convexity in source separation: Models, geometry, and algorithms”. In: *Signal Processing Magazine, IEEE* 31.3 (2014), pp. 87–95 (cit. on p. 76).

[Mer77] Mansfield Merriman. “On the history of the method of least squares”. In: *The Analyst* (1877), pp. 33–36 (cit. on p. 90).

[Mon15] Andrea Montanari. “Statistical estimation: from denoising to sparse regression and hidden cliques”. In: *Statistical Physics, Optimization, Inference and Message-passing Algorithms: Lecture Notes of the Les Houches School of Physics-Special Issue, October 2013* (2015), p. 127 (cit. on p. 58).

[MT14] Michael B McCoy and Joel A Tropp. “Sharp recovery bounds for convex demixing, with applications”. In: *Foundations of Computational Mathematics* 14.3 (2014), pp. 503–567 (cit. on pp. 21, 76).

[NC14] T Lakshmi Narasimhan and Ananthanaryanan Chockalingam. “Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems”. In: *Selected Topics in Signal Processing, IEEE Journal of* 8.5 (2014), pp. 847–860 (cit. on p. 143).

[Neg+12] Sahand N Negahban et al. “A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers”. In: *Statistical Science* 27.4 (2012), pp. 538–557 (cit. on pp. 57, 134).

[NLM13] Hien Quoc Ngo, Erik G Larsson, and Thomas L Marzetta. “Energy and spectral efficiency of very large multiuser MIMO systems”. In: *Communications, IEEE Transactions on* 61.4 (2013), pp. 1436–1449 (cit. on p. 143).

[NM94] Whitney K Newey and Daniel McFadden. “Large sample estimation and hypothesis testing”. In: *Handbook of econometrics* 4 (1994), pp. 2111–2245 (cit. on pp. 194, 248, 252, 259, 263).

[NT09] Deanna Needell and Joel A Tropp. “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples”. In: *Applied and Computational Harmonic Analysis* 26.3 (2009), pp. 301–321 (cit. on p. 10).

[OH10] Samet Oymak and Babak Hassibi. “New null space results and recovery thresholds for matrix rank minimization”. In: *arXiv preprint arXiv:1011.6326* (2010) (cit. on pp. 13, 18, 21, 236).

[OH14] Samet Oymak and Babak Hassibi. "A case for orthogonal measurements in linear inverse problems". In: *Information Theory, 2014. ISIT 2014. Proceedings. International Symposium on* (2014), pp. 3175 – 3179 (cit. on p. 125).

[OH15] Samet Oymak and Babak Hassibi. "Sharp mse bounds for proximal denoising". In: *Foundations of Computational Mathematics* (2015), pp. 1–65 (cit. on pp. 98, 104, 105, 168, 268).

[ORS15] Samet Oymak, Benjamin Recht, and Mahdi Soltanolkotabi. "Sharp Time–Data Tradeoffs for Linear Inverse Problems". In: *arXiv preprint arXiv:1507.04793* (2015) (cit. on pp. 3, 18).

[OT15] Samet Oymak and Joel A Tropp. "Universality laws for randomized dimension reduction, with applications". In: *arXiv preprint arXiv:1511.09433* (2015) (cit. on pp. 4, 19, 59, 123, 134, 165).

[OTH13a] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi. "Simple Bounds for Noisy Linear Inverse Problems with Exact Side Information". In: *arXiv preprint arXiv:1312.0641* (2013) (cit. on pp. 115, 117).

[OTH13b] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi. "The Squared-Error of Generalized LASSO: A Precise Analysis". In: *arXiv preprint arXiv:1311.0830* (2013) (cit. on pp. 8, 31, 41, 58, 59, 98, 99, 101, 103, 107–111, 114, 119, 164, 236, 237).

[Oym+15] Samet Oymak et al. "Simultaneously structured models with application to sparse and low-rank matrices". In: *Information Theory, IEEE Transactions on* 61.5 (2015), pp. 2886–2908 (cit. on p. 237).

[Oym15] Samet Oymak. "Convex Relaxation for Low-Dimensional Representation: Phase Transitions and Limitations". PhD thesis. California Institute of Technology, 2015 (cit. on pp. 98, 99).

[Pel15] Mason Pelt. *"Big Datas an overused buzzword and this Twitter bot proves it*. 2015. URL: http://siliconangle.com/blog/2015/10/26/big-data-is-an-over-used-buzzword-and-this-twitter-bot-proves-it/ (visited on 04/20/2016) (cit. on p. 1).

[PVY14] Yaniv Plan, Roman Vershynin, and Elena Yudovina. "High-dimensional estimation with geometric constraints". In: *arXiv preprint arXiv:1404.3749* (2014) (cit. on p. 255).

[Rec+11] Benjamin Recht et al. "Hogwild: A lock-free approach to parallelizing stochastic gradient descent". In: *Advances in Neural Information Processing Systems*. 2011, pp. 693–701 (cit. on p. 3).

[RFP10] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization”. In: *SIAM review* 52.3 (2010), pp. 471–501 (cit. on pp. 7, 95).

[RGF09] Sundeep Rangan, Vivek Goyal, and Alyson K Fletcher. “Asymptotic analysis of map estimation via the replica method and compressed sensing”. In: *Advances in Neural Information Processing Systems*. 2009, pp. 1545–1553 (cit. on p. 59).

[Roc97] R Tyrell Rockafellar. *Convex analysis*. Vol. 28. Princeton university press, 1997 (cit. on pp. 15, 38, 48, 99, 101, 116, 132, 197, 200, 210, 212, 238, 241, 243, 244, 246, 261).

[RSV12] Emile Richard, Pierre-André Savalle, and Nicolas Vayatis. “Estimation of simultaneously sparse and low rank matrices”. In: *arXiv preprint arXiv:1206.6474* (2012) (cit. on p. 237).

[RT95] Calyampudi Radhakrishna Rao and Helge Toutenburg. *Linear models*. Springer, 1995 (cit. on p. 41).

[RV06] Mark Rudelson and Roman Vershynin. “Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements”. In: *Information Sciences and Systems, 2006 40th Annual Conference on*. IEEE. 2006, pp. 207–212 (cit. on pp. 12, 13, 21).

[RW09] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*. Vol. 317. Springer Science & Business Media, 2009 (cit. on pp. 53, 199, 224–226, 261).

[RXH11] Benjamin Recht, Weiyu Xu, and Babak Hassibi. “Null space conditions and thresholds for rank minimization”. In: *Mathematical programming* 127.1 (2011), pp. 175–202 (cit. on pp. 21, 28).

[Sio+58] Maurice Sion et al. “On general minimax theorems.” In: *Pacific Journal of Mathematics* 8.1 (1958), pp. 171–176 (cit. on pp. 36, 38, 65, 192, 199, 258).

[Sle62] David Slepian. “The One-Sided Barrier Problem for Gaussian Noise”. In: *Bell System Technical Journal* 41.2 (1962), pp. 463–501 (cit. on p. 27).

[SS86] Fadil Santosa and William W Symes. “Linear inversion of band-limited reflection seismograms”. In: *SIAM Journal on Scientific and Statistical Computing* 7.4 (1986), pp. 1307–1330 (cit. on pp. 3, 8).

[Sti81] Stephen M Stigler. “Gauss and the invention of least squares”. In: *The Annals of Statistics* (1981), pp. 465–474 (cit. on p. 90).

[Sto09a] Mihailo Stojnic. “Block-length dependent thresholds in block-sparse compressed sensing”. In: *arXiv preprint arXiv:0907.3679* (2009) (cit. on pp. 21, 236).

[Sto09b] Mihailo Stojnic. "Various thresholds for ℓ_1-optimization in compressed sensing". In: *arXiv preprint arXiv:0907.3666* (2009) (cit. on pp. 7, 13, 14, 18, 21, 58, 237).

[Sto13a] Mihailo Stojnic. "A framework to characterize performance of LASSO algorithms". In: *arXiv preprint arXiv:1303.7291* (2013) (cit. on p. 58).

[Sto13b] Mihailo Stojnic. "Upper-bounding ℓ_1-optimization weak thresholds". In: *arXiv preprint arXiv:1303.7289* (2013) (cit. on pp. 7, 18, 21).

[TA16] Panos Toulis and Edoardo M Airoldi. "Stochastic Gradient Methods for Principled Estimation with Large Datasets". In: *Handbook of Big Data* (2016), p. 241 (cit. on p. 3).

[Tae+13] Armeen Taeb et al. "Maximin Analysis of Message Passing Algorithms for Recovering Block Sparse Signals". In: *arXiv preprint arXiv:1303.2389* (2013) (cit. on p. 57).

[TAH15] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi. "LASSO with Non-linear Measurements is Equivalent to One With Linear Measurements". In: *Advances in Neural Information Processing Systems*. 2015, pp. 3402–3410 (cit. on pp. 141, 160).

[TAH16] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi. "Precise Error Analysis of Regularized M-estimators in High-dimensions". In: *arXiv preprint arXiv:1601.06233* (2016) (cit. on pp. 8, 58).

[Tao12] Terence Tao. *Topics in random matrix theory*. Vol. 132. American Mathematical Soc., 2012 (cit. on p. 19).

[TG07] Joel A Tropp and Anna C Gilbert. "Signal recovery from random measurements via orthogonal matching pursuit". In: *Information Theory, IEEE Transactions on* 53.12 (2007), pp. 4655–4666 (cit. on p. 7).

[TH14] Christos Thrampoulidis and Babak Hassibi. "Estimating structured signals in sparse noise: A precise noise sensitivity analysis". In: *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE. 2014, pp. 866–873 (cit. on pp. 41, 58, 78).

[TH15] Christos Thrampoulidis and Babak Hassibi. "Isotropically random orthogonal matrices: Performance of lasso and minimum conic singular values". In: *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 556–560 (cit. on p. 58).

[Thr+15] Christos Thrampoulidis et al. "Precise error analysis of the lasso". In: *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE. 2015, pp. 3467–3471 (cit. on pp. 58, 82).

[Tib96] Robert Tibshirani. “Regression shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), pp. 267–288 (cit. on pp. 3, 41, 89).

[TOH14] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi. “Simple error bounds for regularized noisy linear inverse problems”. In: *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE. 2014, pp. 3007–3011 (cit. on pp. 58, 95, 118, 121).

[TOH15] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi. “Regularized Linear Regression: A Precise Analysis of the Estimation Error”. In: *Proceedings of The 28th Conference on Learning Theory*. 2015, pp. 1683–1709 (cit. on pp. 58, 59, 101, 103, 109, 198).

[TPH15] Christos Thrampoulidis, Ashkan Panahi, and Babak Hassibi. “Asymptotically exact error analysis for the generalized ℓ_2^2-LASSO”. In: *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 2021–2025 (cit. on pp. 58, 64, 99, 110, 114).

[TRL01] Peng Hui Tan, Lars K Rasmussen, and Teng J Lim. “Constrained maximum-likelihood detection in CDMA”. In: *Communications, IEEE Transactions on* 49.1 (2001), pp. 142–153 (cit. on p. 143).

[Tro12] Joel A Tropp. “A comparison principle for functions of a uniformly random subspace”. In: *Probability Theory and Related Fields* 153.3-4 (2012), pp. 759–769 (cit. on p. 126).

[Tro15] Joel A Tropp. “Convex recovery of a structured signal from independent random linear measurements”. In: *Sampling Theory, a Renaissance*. Springer, 2015, pp. 67–101 (cit. on pp. 12, 14).

[Ver10a] Roman Vershynin. “Introduction to the non-asymptotic analysis of random matrices”. In: *arXiv preprint arXiv:1011.3027* (2010) (cit. on pp. 93, 187).

[Ver10b] Roman Vershynin. “Introduction to the non-asymptotic analysis of random matrices”. In: *arXiv preprint arXiv:1011.3027* (2010) (cit. on pp. 153–155, 157, 196, 255, 256).

[Ver14] Roman Vershynin. “Estimation in high dimensions: a geometric perspective”. In: *arXiv preprint arXiv:1405.5103* (2014) (cit. on p. 57).

[Ver98] Sergio Verdu. *Multiuser detection*. Cambridge university press, 1998 (cit. on p. 144).

[VKC13] Mikko Vehkapera, Yoshiyuki Kabashima, and Saikat Chatterjee. “Analysis of Regularized LS Reconstruction and Random Matrix Ensembles in Compressed Sensing”. In: *arXiv preprint arXiv:1312.0256* (2013) (cit. on p. 125).

[VKC14] Mikko Vehkapera, Yoshiyuki Kabashima, and Saptarshi Chatterjee. "Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing". In: *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE. 2014, pp. 3185–3189 (cit. on p. 59).

[Von] Pascal Vontobel. *www.PseudoCodewords.info*. https://sites.google.com/site/pseudocodewords/papers/papers-on-linear-programming-decoding. Accessed: 2015-11-30 (cit. on p. 167).

[Wai09] Martin J Wainwright. "Sharp thresholds for high-dimensional and noisy sparsity recovery using-constrained quadratic programming (Lasso)". In: *Information Theory, IEEE Transactions on* 55.5 (2009), pp. 2183–2202 (cit. on p. 134).

[Wai14] Martin J Wainwright. "Structured regularizers for high-dimensional problems: Statistical and computational issues". In: *Annual Review of Statistics and Its Application* 1 (2014), pp. 233–253 (cit. on p. 57).

[Wan13] Lie Wang. "The L1 penalized LAD estimator for high dimensional linear regression". In: *Journal of Multivariate Analysis* 120 (2013), pp. 135–151 (cit. on p. 41).

[Wen+14a] Chao-Kai Wen et al. "Message Passing Algorithm for Distributed Downlink Regularized Zero-Forcing Beamforming with Cooperative Base Stations". In: *Wireless Communications, IEEE Transactions on* 13.5 (2014), pp. 2920–2930 (cit. on p. 143).

[Wen+14b] Chao-Kai Wen et al. "On Sparse Vector Recovery Performance in Structurally Orthogonal Matrices via LASSO". In: *arXiv preprint arXiv:1410.7295* (2014) (cit. on pp. 125, 129).

[WM10] John Wright and Yi Ma. "Dense error correction via-minimization". In: *Information Theory, IEEE Transactions on* 56.7 (2010), pp. 3540–3560 (cit. on p. 41).

[Wri+13] John Wright et al. "Compressive principal component pursuit". In: *Information and Inference* 2.1 (2013), pp. 32–68 (cit. on p. 237).

[WV12] Yihong Wu and Sergio Verdú. "Optimal phase transitions in compressed sensing". In: *Information Theory, IEEE Transactions on* 58.10 (2012), pp. 6241–6263 (cit. on p. 72).

[XH11] Weiyu Xu and Babak Hassibi. "Precise stability phase transitions for minimization: A unified geometric framework". In: *Information Theory, IEEE Transactions on* 57.10 (2011), pp. 6894–6919 (cit. on p. 20).

[Yi+15] Xinyang Yi et al. “Optimal linear estimation under unknown nonlinear transform”. In: *arXiv preprint arXiv:1505.03257* (2015) (cit. on p. 154).

[YL06a] Ming Yuan and Yi Lin. “Model selection and estimation in regression with grouped variables”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.1 (2006), pp. 49–67 (cit. on p. 41).

[YL06b] Ming Yuan and Yi Lin. “Model selection and estimation in regression with grouped variables”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.1 (2006), pp. 49–67 (cit. on p. 82).

[YL07] Ming Yuan and Yi Lin. “Model selection and estimation in the Gaussian graphical model”. In: *Biometrika* 94.1 (2007), pp. 19–35 (cit. on p. 166).

[YYU02] Aylin Yener, Roy D Yates, and Sennur Ulukus. “CDMA multiuser detection: A nonlinear programming approach”. In: *Communications, IEEE Transactions on* 50.6 (2002), pp. 1016–1024 (cit. on p. 143).

[ZL15] Qinqing Zheng and John Lafferty. “A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 109–117 (cit. on p. 3).

Appendix A

PROOFS FOR CHAPTER 3

A.1 Proof of the GMT

We begin with using Theorem 3.1.1 to prove an analogue of Theorem 3.2.1 for discrete sets. The proof is almost identical to the proof of Gordon's original Lemma 3.1 in [Gor88]. Nevertheless, we include it here for completeness. Theorem 3.2.1 then follows from Lemma A.1.1 by a compactness argument.

Onwards, we suppress notation and write $\|\cdot\|$ instead of $\|\cdot\|_2$.

Lemma A.1.1 (Gordon's Gaussian Min-max Theorem: Discrete Sets)**.** *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$, $g \in \mathbb{R}$, $\mathbf{g} \in \mathbb{R}^m$ *and* $\mathbf{h} \in \mathbb{R}^d$ *have entries i.i.d.* $\mathcal{N}(0,1)$ *and be independent of each other. Also, let* $\mathcal{I}_1 \subset \mathbb{R}^d$, $\mathcal{I}_2 \subset \mathbb{R}^m$ *be finite sets of vectors and* $\psi(\cdot,\cdot)$ *be a finite function defined on* $\mathcal{I}_1 \times \mathcal{I}_2$. *For all* $c > 0$,

$$\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{I}_1}\max_{\mathbf{u}\in\mathcal{I}_2}\left\{\mathbf{u}^T\mathbf{A}\mathbf{w} + g\|\mathbf{w}\|\|\mathbf{u}\| + \psi(\mathbf{w},\mathbf{u})\right\} \geq c\right) \geq \mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{I}_1}\max_{\mathbf{u}\in\mathcal{I}_2}\left\{\|\mathbf{w}\|\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w},\mathbf{u})\right\} \geq c\right).$$

Proof. Define two Gaussian processes indexed on the set $\mathcal{I}_1 \times \mathcal{I}_2$:

$$Y_{\mathbf{w},\mathbf{u}} = \mathbf{w}^T\mathbf{G}\mathbf{u} + g\|\mathbf{u}\|\|\mathbf{w}\| \quad \text{and} \quad X_{\mathbf{w},\mathbf{u}} = \|\mathbf{w}\|\mathbf{g}^T\mathbf{u} - \|\mathbf{u}\|\mathbf{h}^T\mathbf{w}.$$

First, we show that the processes defined satisfy the conditions of Gordon's Theorem 3.1.1. Clearly, they are both centered. Furthermore, for all $\mathbf{w},\mathbf{w}' \in \mathcal{I}_1$ and $\mathbf{u},\mathbf{u}' \in \mathcal{I}_2$:

$$\mathbb{E}[X_{\mathbf{w},\mathbf{u}}^2] = \|\mathbf{w}\|^2\|\mathbf{u}\|^2 + \|\mathbf{u}\|^2\|\mathbf{w}\|^2 = \mathbb{E}[Y_{\mathbf{w},\mathbf{u}}^2],$$

and

$$\begin{aligned}\mathbb{E}[X_{\mathbf{w},\mathbf{u}}X_{\mathbf{w}',\mathbf{u}'}] - \mathbb{E}[Y_{\mathbf{w},\mathbf{u}}Y_{\mathbf{w}',\mathbf{u}'}] &= \|\mathbf{w}\|\|\mathbf{w}'\|(\mathbf{u}^T\mathbf{u}') + \|\mathbf{u}\|^2(\mathbf{w}^T\mathbf{w}') \\ &\quad - (\mathbf{w}^T\mathbf{w}')(\mathbf{u}^T\mathbf{u}') - \|\mathbf{u}\|\|\mathbf{u}'\|\|\mathbf{w}\|\|\mathbf{w}'\| \\ &= \left(\underbrace{\|\mathbf{w}\|\|\mathbf{w}'\| - (\mathbf{w}^T\mathbf{w}')}_{\geq 0}\right)\left(\underbrace{(\mathbf{u}^T\mathbf{u}') - \|\mathbf{u}\|\|\mathbf{u}'\|}_{\leq 0}\right),\end{aligned}$$

which is non positive and equal to zero when $\mathbf{w} = \mathbf{w}'$.

Next, for each $(\mathbf{w}, \mathbf{u}) \in \mathcal{I}_1 \times \mathcal{I}_2$, let $\lambda_{\mathbf{w},\mathbf{u}} = -\psi(\mathbf{w}, \mathbf{u}) + c$ and apply Theorem 3.1.1. This completes the proof by observing that

$$\left[\min_{\mathbf{w}\in\mathcal{I}_1}\max_{\mathbf{u}\in\mathcal{I}_2}\{Y_{\mathbf{w},\mathbf{u}} + \psi(\mathbf{w}, \mathbf{u})\} \geq c\right] = \bigcap_{\mathbf{w}\in\mathcal{I}_1}\bigcup_{\mathbf{u}\in\mathcal{I}_2}\left[Y_{\mathbf{w},\mathbf{u}} \geq \lambda_{\mathbf{w},\mathbf{u}}\right],$$

and similar for the process $X_{\mathbf{w},\mathbf{u}}$. □

Proof. (of Theorem 3.2.1) Denote $R_1 := \max_{\mathbf{w}\in\mathcal{S}_\mathbf{w}} \|\mathbf{w}\|$ and $R_2 := \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} \|\mathbf{u}\|$. Fix any $\epsilon > 0$. Since $\psi(\cdot,\cdot)$ is continuous and the sets $\mathcal{S}_\mathbf{w}$,$\mathcal{S}_\mathbf{u}$ are compact, $\psi(\cdot,\cdot)$ is uniformly continuous on $\mathcal{S}_\mathbf{w} \times \mathcal{S}_\mathbf{u}$. Thus, there exists $\delta := \delta(\epsilon) > 0$ such that for every $(\mathbf{w}, \mathbf{u}), (\tilde{\mathbf{w}}, \tilde{\mathbf{u}}) \in \mathcal{S}_\mathbf{w} \times \mathcal{S}_\mathbf{u}$ with $\| \begin{bmatrix}\mathbf{w} & \mathbf{u}\end{bmatrix} - \begin{bmatrix}\tilde{\mathbf{w}} & \tilde{\mathbf{u}}\end{bmatrix} \| \leq \delta$, we have that $|\psi(\mathbf{w}, \mathbf{u}) - \psi(\tilde{\mathbf{w}}, \tilde{\mathbf{u}})| \leq \epsilon$. Let $\mathcal{S}^\delta_\mathbf{w}$,$\mathcal{S}^\delta_\mathbf{u}$ be δ-nets of the sets $\mathcal{S}_\mathbf{w}$ and $\mathcal{S}_\mathbf{u}$, respectively. Then, for any $\mathbf{w} \in \mathcal{S}_\mathbf{w}$, there exists $\mathbf{w}' \in \mathcal{S}^\delta_\mathbf{w}$ such that $\|\mathbf{w} - \mathbf{w}'\| \leq \delta$ and an analogous statement holds for $\mathcal{S}_\mathbf{u}$. In what follows, for any vector $\mathbf{v}$ in a set $\mathcal{S}$, we denote $\mathbf{v}'$ the element in the δ-net of $\mathcal{S}$ that is the closest to $\mathbf{v}$ in the usual ℓ_2-metric. To simplify notation, denote

$$\alpha(\mathbf{w}, \mathbf{u}) := \mathbf{u}^T\mathbf{A}\mathbf{w} + g\|\mathbf{w}\|\|\mathbf{u}\| + \psi(\mathbf{w}, \mathbf{u}) \quad \text{and} \quad \beta(\mathbf{w}, \mathbf{u}) := \|\mathbf{w}\|\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w}, \mathbf{u}).$$

From Lemma A.1.1, we know that for all $c \in \mathbb{R}$:

$$\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}^\delta_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}^\delta_\mathbf{u}} \alpha(\mathbf{w}, \mathbf{u}) \geq c\right) \geq \mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}^\delta_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}^\delta_\mathbf{u}} \beta(\mathbf{w}, \mathbf{u}) \geq c\right). \tag{A.1}$$

In what follows we show that constraining the minimax optimizations over only the δ-nets $\mathcal{S}^\delta_\mathbf{w}, \mathcal{S}^\delta_\mathbf{u}$ instead of the entire sets $\mathcal{S}_\mathbf{w}$,$\mathcal{S}_\mathbf{u}$, changes the achieved optimal values by only a small amount.

First, we calculate an upper bound on

$$\begin{aligned}
\min_{\mathbf{w}\in\mathcal{S}^\delta_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}^\delta_\mathbf{u}} \alpha(\mathbf{w}, \mathbf{u}) - \min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} \alpha(\mathbf{w}, \mathbf{u}) &\leq \min_{\mathbf{w}\in\mathcal{S}^\delta_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}^\delta_\mathbf{u}} \alpha(\mathbf{w}, \mathbf{u}) - \min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}^\delta_\mathbf{u}} \alpha(\mathbf{w}, \mathbf{u}) =: \alpha(\mathbf{w}_1, \mathbf{u}_1) - \alpha(\mathbf{w}_2, \mathbf{u}_2) \\
&\leq \max_{\mathbf{u}\in\mathcal{S}^\delta_\mathbf{u}} \alpha(\mathbf{w}_2', \mathbf{u}) - \alpha(\mathbf{w}_2, \mathbf{u}_2) =: \alpha(\mathbf{w}_2', \mathbf{u}_*) - \alpha(\mathbf{w}_2, \mathbf{u}_2) \\
&\leq \alpha(\mathbf{w}_2', \mathbf{u}_*) - \alpha(\mathbf{w}_2, \mathbf{u}_*) \\
&= \mathbf{u}_*^T\mathbf{A}(\mathbf{w}_2' - \mathbf{w}_2) + g\|\mathbf{u}_*\|(\|\mathbf{w}_2'\| - \|\mathbf{w}_2\|) + (\psi(\mathbf{w}_2', \mathbf{u}_*) - \psi(\mathbf{w}_2, \mathbf{u}_*)) \\
&\leq (\|\mathbf{A}\|_2 + |g|)\underbrace{\|\mathbf{u}_*\|}_{\leq R_2}\underbrace{\|\mathbf{w}_2' - \mathbf{w}_2\|}_{\leq\delta} + \underbrace{|\psi(\mathbf{w}_2', \mathbf{u}_*) - \psi(\mathbf{w}_2, \mathbf{u}_*)|}_{\leq\epsilon} \\
&\leq (\|\mathbf{A}\|_2 + |g|)R_2\delta + \epsilon.
\end{aligned}$$

From this, we have that

$$\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\alpha(\mathbf{w},\mathbf{u})\geq c\right)\geq\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}^\delta}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}^\delta}\alpha(\mathbf{w},\mathbf{u})\geq c+(\|\mathbf{A}\|_2+|g|)R_2\delta+\epsilon\right). \tag{A.2}$$

Using standard concentration results on Gaussians, it is shown in Lemma A.1.0.1 that for all $t>0$,

$$\mathbb{P}(\|\mathbf{A}\|_2+|g|\leq\sqrt{m}+\sqrt{n}+1+t)\geq 1-2\exp(-t^2/4).$$

This, when combined with (A.2) yileds:

$$\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\alpha(\mathbf{w},\mathbf{u})\geq c\right)\geq\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}^\delta}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}^\delta}\alpha(\mathbf{w},\mathbf{u})\geq c+(\sqrt{m}+\sqrt{n}+1+t)R_2\delta+\epsilon\right)-2e^{-t^2/4}. \tag{A.3}$$

Similarly,

$$\begin{aligned}
\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}^\delta}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}^\delta}\beta(\mathbf{w},\mathbf{u})-\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\beta(\mathbf{w},\mathbf{u})&\geq\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}^\delta}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}^\delta}\beta(\mathbf{w},\mathbf{u})-\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}^\delta}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\beta(\mathbf{w},\mathbf{u})=:\beta(\mathbf{w}_1,\mathbf{u}_1)-\beta(\mathbf{w}_2,\mathbf{u}_2)\\
&\geq\beta(\mathbf{w}_1,\mathbf{u}_1)-\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\beta(\mathbf{w}_1,\mathbf{u})=:\beta(\mathbf{w}_1,\mathbf{u}_1)-\beta(\mathbf{w}_1,\mathbf{u}_*)\\
&\geq\beta(\mathbf{w}_1,\mathbf{u}_*')-\beta(\mathbf{w}_1,\mathbf{u}_*)\\
&=\|\mathbf{w}_1\|\mathbf{g}^T(\mathbf{u}_*'-\mathbf{u}_*)+(\|\mathbf{u}_*'\|-\|\mathbf{u}_*\|)\mathbf{h}^T\mathbf{w}_1+(\psi(\mathbf{w}_1,\mathbf{u}_*')-\psi(\mathbf{w}_1,\mathbf{u}_*))\\
&\geq-(\|\mathbf{g}\|+\|\mathbf{h}\|)\underbrace{\|\mathbf{w}_1\|}_{\leq R_1}\underbrace{\|\mathbf{u}_*'-\mathbf{u}_*\|}_{\leq\delta}-\underbrace{|\psi(\mathbf{w}_1,\mathbf{u}_*')-\psi(\mathbf{w}_1,\mathbf{u}_*)|}_{\leq\epsilon}\\
&\geq-(\|\mathbf{g}\|+\|\mathbf{h}\|)R_1\delta-\epsilon.
\end{aligned}$$

Thus,

$$\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\beta(\mathbf{w},\mathbf{u})\geq c+(\|\mathbf{g}\|+\|\mathbf{h}\|)R_1\delta+\epsilon\right)\leq\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}^\delta}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}^\delta}\beta(\mathbf{w},\mathbf{u})\geq c\right),$$

and a further application of Lemma A.1.0.1 shows that for all $t>0$:

$$\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\beta(\mathbf{w},\mathbf{u})\geq c+(\sqrt{m}+\sqrt{n}+t)R_2\delta+\epsilon\right)-2e^{-t^2/4}\leq\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}^\delta}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}^\delta}\beta(\mathbf{w},\mathbf{u})\geq c\right), \tag{A.4}$$

Now, we can apply (A.1) in order to combine (A.3) and (A.4) to yield the following:

$$\begin{aligned}
&\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\alpha(\mathbf{w},\mathbf{u})\geq c\right)\geq\\
&\qquad\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}}\max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}}\beta(\mathbf{w},\mathbf{u})\geq c+(\sqrt{m}+\sqrt{n}+1+t)(R_1+R_2)\delta+2\epsilon\right)-4e^{-t^2/4}.
\end{aligned}$$

This holds for all $\epsilon > 0$ and all $t > 0$. In particular, set $t = \delta^{-\frac{1}{2}}$ and take the limit of the right-hand side as $\epsilon \to 0$. Then, $t \to \infty$ and we can of course choose $\delta \to 0$, which proves that

$$\mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} \alpha(\mathbf{w},\mathbf{u}) \geq c\right) \geq \mathbb{P}\left(\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} \beta(\mathbf{w},\mathbf{u}) > c\right).$$

□

Lemma A.1.0.1. *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$, $g \in \mathbb{R}$, $\mathbf{g} \in \mathbb{R}^m$ *and* $\mathbf{h} \in \mathbb{R}^n$ *have entries i.i.d.* $\mathcal{N}(0,1)$ *and be independent of each other. Then, for all* $t > 0$, *each one of the events*

$$\{\|\mathbf{A}\|_2 + |g| \leq \sqrt{m} + \sqrt{n} + 1 + t\} \quad \textit{and} \quad \{\|\mathbf{h}\|_2 + \|\mathbf{g}\|_2 \leq \sqrt{m} + \sqrt{n} + t\} \tag{A.5}$$

holds with probability at least $1 - 2\exp(-t^2/4)$.

Proof. A well-known non-asymptotic bound on the largest singular value of an $n \times d$ Gaussian matrix shows (e.g. [Ver10a, Corollary 5.35]) that for all $t > 0$:

$$\mathbb{P}\left(\|\mathbf{A}\|_2 > \sqrt{m} + \sqrt{n} + t\right) \leq \exp(-t^2/2).$$

Also, $\|\cdot\|_2$ is an 1-Lipschitz function and for a standard Gaussian vector $\mathbf{v} \in \mathbb{R}^n$: $\mathbb{E}\|\mathbf{v}\|_2 \leq \sqrt{d}$. Applying Proposition 3.1.1 we have that for all $t > 0$ the events $\{|g| > 1 + t\}$, $\{\|\mathbf{g}\|_2 > \sqrt{m} + t\}$ and $\{\|\mathbf{h}\|_2 > \sqrt{n} + t\}$, each one occurs with probability no larger than $\exp(-t^2/2)$. Combining those,

$$\begin{aligned}\mathbb{P}\left(\|\mathbf{A}\|_2 + |g| \leq \sqrt{m} + \sqrt{n} + 1 + t\right) &\geq \mathbb{P}\left(\|\mathbf{A}\|_2 \leq \sqrt{d} + \sqrt{n} + t/2\ ,\ |g| \leq 1 + t/2\right)\\ &\geq 1 - \mathbb{P}\left(\|\mathbf{A}\|_2 > \sqrt{m} + \sqrt{n} + t/2\right) - \mathbb{P}\left(\,|g| > 1 + t/2\right)\\ &\geq 1 - 2\exp(-t^2/4).\end{aligned}$$

The proof of the second statement is identical and is omitted for brevity. □

A.2 Lipschitzness of the (AO)

Lemma A.2.0.2 (Lipschitzness of the AO problem)**.** *Let* $\mathcal{S}_\mathbf{w} \subset \mathbb{R}^n$, $\mathcal{S}_\mathbf{u} \subset \mathbb{R}^m$ *be compact sets and function* $\phi : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$*:*

$$\phi(\mathbf{g},\mathbf{h}) := \min_{\mathbf{w}\in\mathcal{S}_\mathbf{w}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} \|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w},\mathbf{u}).$$

Further let $R_1 = \max_{\mathbf{w}\in\mathcal{S}_\mathbf{w}} \|\mathbf{w}\|_2$ *and* $R_2 = \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} \|\mathbf{u}\|_2$. *Then,* $\phi(\mathbf{g},\mathbf{h})$ *is Lipschitz with constant* $\sqrt{2}R_1R_2$.

Proof. Fix any two pairs $(\mathbf{g}_1, \mathbf{h}_1)$ and $(\mathbf{g}_2, \mathbf{h}_2)$ and let

$$(\mathbf{w}_2, \mathbf{u}_2) = \arg \min_{\mathbf{w} \in \mathcal{S}_\mathbf{w}} \max_{\mathbf{u} \in \mathcal{S}_\mathbf{u}} \|\mathbf{w}\| \mathbf{g}_2^T \mathbf{u} + \|\mathbf{u}\| \mathbf{h}_2^T \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}),$$

and

$$\mathbf{u}_* = \arg \max_{\mathbf{u} \in \mathcal{S}_\mathbf{u}} \|\mathbf{w}_2\| \mathbf{g}_1^T \mathbf{u} + \|\mathbf{u}\| \mathbf{h}_1^T \mathbf{w}_2 + \psi(\mathbf{w}_2, \mathbf{u}).$$

Clearly,

$$\phi(\mathbf{g}_1, \mathbf{h}_1) \leq \|\mathbf{w}_2\| \mathbf{g}_1^T \mathbf{u}_* + \|\mathbf{u}_*\| \mathbf{h}_1^T \mathbf{w}_2 + \psi(\mathbf{w}_2, \mathbf{u}_*),$$

and

$$\phi(\mathbf{g}_2, \mathbf{h}_2) \geq \|\mathbf{w}_2\| \mathbf{g}_2^T \mathbf{u}_* + \|\mathbf{u}_*\| \mathbf{h}_2^T \mathbf{w}_2 + \psi(\mathbf{w}_2, \mathbf{u}_*).$$

Without loss of generality, assume $\phi(\mathbf{g}_1, \mathbf{h}_1) \geq \phi(\mathbf{g}_2, \mathbf{h}_2)$. Then,

$$\begin{aligned}
\phi(\mathbf{g}_1, \mathbf{h}_1) - \phi(\mathbf{g}_2, \mathbf{h}_2) &\leq \|\mathbf{w}_2\| \mathbf{g}_1^T \mathbf{u}_* + \|\mathbf{u}_*\| \mathbf{h}_1^T \mathbf{w}_2 + \psi(\mathbf{w}_2, \mathbf{u}_*) \\
&\qquad - (\|\mathbf{w}_2\| \mathbf{g}_2^T \mathbf{u}_* + \|\mathbf{u}_*\| \mathbf{h}_2^T \mathbf{w}_2 + \psi(\mathbf{w}_2, \mathbf{u}_*)) \\
&\leq \|\mathbf{w}_2\| \mathbf{u}_*^T (\mathbf{g}_1 - \mathbf{g}_2) + \|\mathbf{u}_*\| \mathbf{w}_2^T (\mathbf{h}_1 - \mathbf{h}_2) \\
&\leq \sqrt{\|\mathbf{w}_2\|^2 \|\mathbf{u}_*\|^2 + \|\mathbf{u}_*\|^2 \|\mathbf{w}_2\|^2} \sqrt{\|\mathbf{g}_1 - \mathbf{g}_2\|^2 + \|\mathbf{h}_1 - \mathbf{h}_2\|^2} \\
&\leq R_1 R_2 \sqrt{2} \sqrt{\|\mathbf{g}_1 - \mathbf{g}_2\|^2 + \|\mathbf{h}_1 - \mathbf{h}_2\|^2},
\end{aligned}$$

where the penultimate inequality follows from Cauchy-Schwarz. □

Appendix B

PROOFS FOR CHAPTER 4

B.1 Proof of Theorem 4.2.1

Here, we prove Theorem 4.2.1. The proof consists of several steps and intermediate results that are stated as lemmas. The proofs of the latter are all deferred to Appendix B.2.

Preliminaries

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \lambda f(\mathbf{x}).$$

Recall that $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$. Our goal is to characterize the nontrivial limiting behavior of $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2/\sqrt{n}$. We start with a simple change of variables $\mathbf{w} := (\mathbf{x} - \mathbf{x}_0)/\sqrt{n}$ to directly get a handle on the *error vector* $\mathbf{w}$. Also, we normalize the objective by dividing with n so that the optimal cost is of constant order. Then,

$$\hat{\mathbf{w}} := \arg\min_{\mathbf{w}} \frac{1}{n}\left\{\mathcal{L}\left(\mathbf{z} - \sqrt{n}\mathbf{A}\mathbf{w}\right) + \lambda f\left(\mathbf{x}_0 + \sqrt{n}\mathbf{w}\right)\right\}. \tag{B.1}$$

Instead of the optimization problem above, we will analyze a simpler Auxiliary Optimization (AO) that is tightly related to the Primary Optimization (PO) in (B.1) via the CGMT.

The CGMT for M-estimators

In this section, we show how the CGMT Theorem 3.3.1 can be applied to predict the limiting behavior of the solution $\|\hat{\mathbf{w}}\|_2$ to the minimization in (B.1). The main challenge here is to express (B.1) as a (convex-concave) minimax optimization in which the involved random matrix (here $\mathbf{A}$) appears in a bilinear form, exactly as in (3.11a). Also, some side technical details need to be taken care of. For example, in (3.11a) the optimization constraints are required by Theorem 3.3.1 to be bounded, which is not the case with (B.1). We start with addressing this immediately.

Boundedness of the Error

The constraint set over which $\mathbf{w}$ is optimized in (3.11a) is unbounded. We will introduce "artificial" boundedness constraints that allow the application of Theorem 3.3.1, while they do not affect the optimization itself. For this purpose, recall our goal of proving that $\|\hat{\mathbf{w}}\|_2$ converges to some (finite) α_* defined in Theorem 4.2.1. Define the set $\mathcal{S}_{\mathbf{w}} = \{\mathbf{w} \mid \|\mathbf{w}\|_2 \leq K_\alpha\}$, where

$$K_\alpha := \alpha_* + \zeta \tag{B.2}$$

for a constant $\zeta > 0$, and, consider the "bounded" version of (B.1):

$$\hat{\mathbf{w}}^B := \arg\min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}}} \frac{1}{n}\left\{\mathcal{L}\left(\mathbf{z} - \sqrt{n}\mathbf{A}\mathbf{w}\right) + \lambda f\left(\mathbf{x}_0 + \sqrt{n}\mathbf{w}\right)\right\}. \tag{B.3}$$

We expect that the additional constraint $\mathbf{w} \in \mathcal{S}_{\mathbf{w}}$ in (B.3) will not affect the optimization with high probability when n is large enough. The idea here is that the minimizer of the original unconstrained problem in (B.1) satisfies $\|\hat{\mathbf{w}}\|_2 \approx \alpha_* < K_\alpha$ whp. Of course, this latter statement is yet to be proven! Once this is done, we can return and confirm that our initial expectation is met. Lemma B.1.1 below shows that if $\|\hat{\mathbf{w}}^B\| \xrightarrow{P} \alpha_* < K_\alpha$, then, the same is true for the optimal of (B.1).

Lemma B.1.1. *For the two optimizations in* (B.1) *and* (B.3)*, let* $\hat{\mathbf{w}}$ *and* $\hat{\mathbf{w}}^B$ *be optimal solutions. Also, recall the definition of* K_α *in* (B.2)*. If* $\|\hat{\mathbf{w}}^B\| \xrightarrow{P} \alpha_*$*, then* $\|\hat{\mathbf{w}}\| \xrightarrow{P} \alpha_*$.

Owing to the result of the lemma, henceforth we work with the bounded optimization in (B.3). Using some abuse of notation, we will refer to optimal solution of (B.3) as $\hat{\mathbf{w}}$, rather than $\hat{\mathbf{w}}^B$.

Identifying the (PO)

Here, we bring the minimization in (B.3) it in the form of the (PO) in (3.11a). For this purpose, we will use Lagrange duality. Note that the former can be equivalently expressed as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v}} \frac{1}{n}\left\{\mathcal{L}(\sqrt{n}\mathbf{v}) + \lambda f(\mathbf{x}_0 + \sqrt{n}\mathbf{w})\right\} \quad \text{subject to} \quad \mathbf{v} = \mathbf{z} - \sqrt{n}\mathbf{A}\mathbf{w}.$$

Associating a dual variable $\mathbf{u}$ to the equality constraint above, we write it as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v}} \max_{\mathbf{u}} \frac{1}{\sqrt{n}}\left\{-\mathbf{u}^T(\sqrt{n}\mathbf{A})\mathbf{w} + \mathbf{u}^T\mathbf{z} - \mathbf{u}^T\mathbf{v}\right\} + \frac{1}{n}\left\{\mathcal{L}(\mathbf{v}) + \lambda f(\mathbf{x}_0 + \sqrt{n}\mathbf{w})\right\}. \tag{B.4}$$

It takes not much effort to check that the objective function above is in the desired format of (3.11a): the random matrix $\mathbf{A}$ appears in a bilinear term $\mathbf{u}^T\mathbf{A}\mathbf{w}$ and the rest of the terms form a convex-concave function in $\mathbf{u}, \mathbf{w}$. Furthermore, we can use Assumption 4.2.1(b) to show that the optimal $\mathbf{u}_*$ is bounded, which is a requirement of Theorem 3.3.1. In the same lines as in Section B.1, we henceforth work with the "bounded" version of (B.4), namely,

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}\in\mathcal{S}_\mathbf{w},\mathbf{v}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u}} \frac{1}{\sqrt{n}} \left\{-\mathbf{u}^T(\sqrt{n}\mathbf{A})\mathbf{w} + \mathbf{u}^T\mathbf{z} - \mathbf{u}^T\mathbf{v}\right\} + \frac{1}{n}\left\{\mathcal{L}(\mathbf{v}) + \lambda f(\mathbf{x}_0 + \sqrt{n}\mathbf{w})\right\} \tag{B.5}$$

for $\mathcal{S}_\mathbf{u} := \{\mathbf{u} \mid \|\mathbf{u}\|_2 \le K_\beta\}$ and $K_\beta > 0$ a sufficiently large constant.

Lemma B.1.2. *If Assumption 4.2.1(b) holds, then there exists sufficiently large constant K_β such that the optimization problem in (B.5) is equivalent to that in (B.3), with probability approaching 1 in the limit of $n \to \infty$.*

As a last step, before writing down the corresponding (AO) problem, it will be useful for the analysis of the latter to express f in a variational form through its Fenchel conjugate, which gives,

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}\in\mathcal{S}_\mathbf{w},\mathbf{v}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u},\mathbf{s}} \frac{1}{\sqrt{n}} \left\{-\mathbf{u}^T(\sqrt{n}\mathbf{A})\mathbf{w} + \mathbf{u}^T\mathbf{z} - \mathbf{u}^T\mathbf{v}\right\} + \frac{1}{n}\left\{\mathcal{L}(\mathbf{v}) + \lambda\mathbf{s}^T\mathbf{x}_0 + \lambda\sqrt{n}\mathbf{s}^T\mathbf{w} - \lambda f^*(\mathbf{s})\right\}. \tag{B.6}$$

The (AO)

Having identified (B.6) as the (PO) in our application, it is straightforward to write the corresponding (AO) problem following (3.11b):

$$\min_{\mathbf{w}\in\mathcal{S}_\mathbf{w},\mathbf{v}} \max_{\mathbf{u}\in\mathcal{S}_\mathbf{u},\mathbf{s}} \frac{1}{\sqrt{n}} \left\{\|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u} - \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \mathbf{u}^T\mathbf{z} - \mathbf{u}^T\mathbf{v}\right\} + \frac{1}{n}\left\{\mathcal{L}(\mathbf{v}) + \lambda\mathbf{s}^T\mathbf{x}_0 + \lambda\sqrt{n}\mathbf{s}^T\mathbf{w} - \lambda f^*(\mathbf{s})\right\}. \tag{B.7}$$

Once we have identified the (AO) problem, Corollary 3.3.2 suggests analyzing that one instead of the (PO). Our goal is showing that $\|\hat{\mathbf{w}}\|_2 \xrightarrow{P} \alpha_*$. For this, we wish to apply the corollary to the following set

$$\mathcal{S} = \{\mathbf{w} \mid |\|\mathbf{w}\|_2 - \alpha_*| > \epsilon\},$$

for arbitrary $\epsilon > 0$.

Asymptotic min-max property of the (AO)

It turns out that verifying the conditions of the corollary for the (AO) as it appears in (B.7) is not directly easy. In short, what makes the analysis cumbersome is the fact that the optimization in (B.7) is not convex (e.g. if $\mathbf{g}^T\mathbf{u}$ is negative, then $\|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u}$ is not convex). Thus, flipping the order of min-max operations that would simplify the analysis is not directly justified.

At this point, recall that the (PO) in (B.6) is itself convex. In fact, for it, all conditions of Sion's min-max Theorem [Sio+58] are met, thus, the order of min-max operations can be flipped. According to the CGMT, the (PO) and the (AO) are tightly related in an asymptotic setting. We use this to translate the convexity properties of the (PO) to the (AO). In essence, we show that when dimensions grow, the order of min-max operations in the (AO) can be flipped. Thus, we will instead consider the following problem as the (AO):

$$\phi(\mathbf{g},\mathbf{h}) := \max_{\substack{0\le\beta\le K_\beta\\ \mathbf{s}}} \min_{\substack{\|\mathbf{w}\|_2\le K_\alpha\\ \mathbf{v}}} \max_{\|\mathbf{u}\|_2=\beta} \frac{1}{\sqrt{n}}(\|\mathbf{w}\|_2\mathbf{g}+\mathbf{z}-\mathbf{v})^T\mathbf{u} - \frac{1}{\sqrt{n}}\|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w} + \frac{1}{n}\mathcal{L}(\mathbf{v}) + \frac{\lambda}{n}\mathbf{s}^T\mathbf{x}_0 + \frac{\lambda}{\sqrt{n}}\mathbf{s}^T\mathbf{w} - \frac{\lambda}{n}f^*(\mathbf{s}). \tag{B.8}$$

Observe that the objective function remains the same; it is only the order of min-max operations that is slightly modified compared to (B.7). Since the objective function is not necessarily convex-concave in its arguments, there is no immediate guarantee that the two problems in (B.7) and (B.8) are equivalent for any realizations of $\mathbf{g}$ and $\mathbf{h}$. However, the lemma below essentially shows that such a strong duality holds with high probability over $\mathbf{g}$ and $\mathbf{h}$ in high dimensions. Hence, the problem in (B.8) can be as well used, instead of the one in (B.7), in order to analyze the (PO). For this reason, henceforth, we refer to (B.8) as the (AO) problem.

Lemma B.1.3. *Let $\hat{\mathbf{w}}(\mathbf{A})$ denote an optimal solution of (B.1). Consider the (AO) problem in (B.8). Let α_* be as defined in Theorem 4.2.1. For any $\epsilon > 0$ define the set $\mathcal{S} := \{\mathbf{w} \mid |\|\mathbf{w}\|_2 - \alpha_*| < \epsilon\}$ and let $\phi_{\mathcal{S}^c}(\mathbf{g},\mathbf{h})$ be the optimal cost of the same optimization as in (B.8), only this time the minimization over $\mathbf{w}$ is further constrained such that $\mathbf{w}\notin\mathcal{S}$. Assume that for any $K_\alpha > \alpha_*$ and for any sufficiently large K_β, there exist constants $\overline{\phi} < \overline{\phi}_{\mathcal{S}^c}$ such that for all $\eta > 0$, with probability approaching one in the limit of $n\to\infty$, the following hold:*

(a) $\phi(\mathbf{g},\mathbf{h}) < \overline{\phi} + \eta$,

(b) $\phi_{\mathcal{S}^c}(\mathbf{g},\mathbf{h}) > \overline{\phi}_{\mathcal{S}^c} - \eta$.

Then,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \|\hat{\mathbf{w}}(\mathbf{A})\|_2 - \alpha_* \right| < \epsilon \right) = 1.$$

After Lemma B.1.3, what remains in order to prove Theorem 3.3.1 is satisfying the conditions of the lemma. This involves a thorough analysis of the (AO) problem in (B.8), which is the subject of the next few sections.

Scalarization

Observe that the optimization in (B.8) is over vectors. The purpose of this section is to simplify the (AO) into an optimization involving only scalar variables. Of course, one of these has to play the role of the norm of $\mathbf{w}$, which is the quantity of interest. The main idea behind the "scalarization" step of the (AO) is to perform the optimization over only the direction of the vector variables while keeping their magnitude constant. This is already hinted by the rearrangement of the order of min-max operations going from (B.7) to (B.8). Also, this process is facilitated by the following two facts:

1. The bilinear term $\mathbf{u}^T\mathbf{A}\mathbf{w}$ that appears in the (PO) conveniently "splits" into the two terms $\|\mathbf{w}\|_2\mathbf{g}^T\mathbf{u}$ and $\|\mathbf{u}\|_2\mathbf{h}^T\mathbf{w}$ in the (AO),
2. The term involving the regularizer, i.e. $f(\mathbf{x}_0+\mathbf{w})$ has been expressed in a variational form as $\sup_{\mathbf{s}} \mathbf{s}^T\mathbf{x}_0 + \mathbf{s}^T\mathbf{w} - f^*(\mathbf{s})$.

The details of the reduction step are all summarized in Lemma B.1.4, below which shows that the (AO) reduces to the following *convex* minimax problem on four *scalar* optimization variables:

$$\inf_{\substack{0\le\alpha\le K_\alpha \\ \tau_g>0}} \sup_{\substack{0\le\beta\le K_\beta \\ \tau_h>0}} \frac{\beta\tau_g}{2} + \frac{1}{n}\mathrm{e}_{\mathcal{L}}\left(\alpha\mathbf{g}+\mathbf{z};\frac{\tau_g}{\beta}\right) - \begin{cases} \frac{\alpha\tau_h}{2} + \frac{\beta^2\alpha}{2\tau_h}\frac{\|\mathbf{h}\|^2}{n} - \lambda\cdot\frac{1}{n}\mathrm{e}_f\left(\frac{\beta\alpha}{\tau_h}\mathbf{h}+\mathbf{x}_0;\frac{\alpha\lambda}{\tau_h}\right) & ,\alpha>0 \\ \frac{\lambda}{n}f(\mathbf{x}_0) & ,\alpha=0 \end{cases}, \tag{B.9}$$

where we recall that

$$\mathrm{e}_\omega(\mathbf{u};\tau) := \min_{\mathbf{v}}\{\frac{1}{2\tau}\|\mathbf{u}-\mathbf{v}\|_2^2 + \omega(\mathbf{v})\}$$

denotes the (vector) τ-Moreau envelope of a function $\omega : \mathbb{R}^d \to \mathbb{R}$ evaluated at $\mathbf{u} \in \mathbb{R}^d$.

Lemma B.1.4 (Scalarization of the (AO))**.** *The following statements are true regarding the two minimax optimization problems in* (B.8) *and* (B.9)*:*

(i) They have the same optimal cost.

(ii) The objective function in (B.9) *is continuous on its domain, (jointly) convex in* (α, τ_g) *and (jointly) concave in* (β, τ_h)*.*

(iii) The order of inf-sup in (B.9) *can be flipped without changing the optimization.*

Convergence Analysis

The goal of this section is to show that the (AO) satisfies the conditions of Lemma B.1.3. This requires a convergence analysis of its optimal cost. We work with the scalarized version of the (AO) that was derived in the previous section:

$$\phi(\mathbf{g}, \mathbf{h}, \mathbf{z}, \mathbf{x}_0) = \inf_{\substack{0 \le \alpha \le K_\alpha \\ \tau_g > 0}} \sup_{\substack{0 \le \beta \le K_\beta \\ \tau_h > 0}} \mathcal{R}_n(\alpha, \tau_g, \beta, \tau_h; \mathbf{g}, \mathbf{h}, \mathbf{z}, \mathbf{x}_0), \tag{B.10}$$

$$\mathcal{R}_n = \frac{\beta\tau_g}{2} + \frac{1}{n}\left\{ \mathrm{e}_{\mathcal{L}}\left(\alpha\mathbf{g} + \mathbf{z}; \frac{\tau_g}{\beta}\right) - \mathcal{L}(\mathbf{z}) \right\} - \begin{cases} \frac{\alpha\tau_h}{2} + \frac{\beta^2\alpha}{2\tau_h}\frac{\|\mathbf{h}\|^2}{n} - \frac{\lambda}{n}\left\{ \mathrm{e}_f\left(\frac{\beta\alpha}{\tau_h}\mathbf{h} + \mathbf{x}_0; \frac{\alpha\lambda}{\tau_h}\right) - f(\mathbf{x}_0) \right\} & , \alpha > 0 \\ 0 & , \alpha = 0 \end{cases}.$$

Here, when compared to (B.9), we have subtracted from the objective the terms $\mathcal{L}(\mathbf{z})$ and $f(\mathbf{x}_0)$, which of course does not affect the optimization. The optimization is of course random over the realizations of $\mathbf{g}, \mathbf{h}, \mathbf{z}$ and $\mathbf{x}_0$, and, by the WLLN, it is easy to identify the converging value of the objective function $\mathcal{R}_n$ for fixed parameter values $\alpha, \tau_g, \beta, \tau_h$. Indeed, it converges to the objective function of the (SPO) problem in (4.4). For our goals, we need to show that the minimax of the converging sequence of objectives converges to the minimax of the objective of the (SOP). Convexity of $\mathcal{R}_n$ plays a crucial role here since is being use to conclude local uniform convergence from the pointwise convergence. Uniform convergence is a requirement to conclude the desired.[1]

[1] We remark that the tools used for this part of the proof are similar to those classically used for the study of consistency of M-estimators in the classical regime where n is fixed and m goes to infinity, cf. Arg-min theorems e.g. [LM08, Thm. 7.70], [NM94, Thm. 2.7] .

Lemma B.1.5 (Convergence properties of the (AO))**.** *Let*

$$\mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h) := \mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h;\mathbf{g},\mathbf{h},\mathbf{z},\mathbf{x}_0),$$

be defined as in (B.10)*, and,*

$$\phi_{\mathcal{A}} := \phi_{\mathcal{A}}(\mathbf{g},\mathbf{h},\mathbf{z},\mathbf{x}_0) := \inf_{\substack{\alpha\in\mathcal{A}\\ \tau_g>0}} \sup_{\substack{0\le\beta\le K_\beta\\ \tau_h>0}} \mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h), \tag{B.11}$$

for $\mathcal{A}\subseteq[0,\infty)$*. Further consider the following deterministic convex program*

$$\begin{aligned}\overline{\phi}_{\mathcal{A}} := \inf_{\substack{\alpha\in\mathcal{A}\\ \tau_g>0}} \sup_{\substack{\beta\ge 0\\ \tau_h>0}} \mathcal{D}(\alpha,\tau_g,\beta,\tau_h) := &\begin{cases}\frac{\beta\tau_g}{2}+\delta\cdot L\left(\alpha,\frac{\tau_g}{\beta}\right) & ,\beta>0\\ -\delta\cdot L_0 & ,\beta=0\end{cases}\\ &-\begin{cases}\frac{\alpha\tau_h}{2}+\frac{\alpha\beta^2}{2\tau_h}-\lambda\cdot F\left(\frac{\alpha\beta}{\tau_h},\frac{\alpha\lambda}{\tau_h}\right) & ,\alpha>0\\ 0 & ,\alpha=0\end{cases},\end{aligned} \tag{B.12}$$

where L *and* F *as in Theorem* 4.2.1*. If Assumption* 4.2.1*(a) and* 4.2.2 *hold, then,*

(a) $\mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h)\xrightarrow{P}\mathcal{D}(\alpha,\tau_g,\beta,\tau_h)$*, for all* $(\alpha,\tau_g,\beta,\tau_h)$*, and* $\mathcal{D}(\alpha,\tau_g,\beta,\tau_h)$ *is convex in* (α,τ_g) *and concave in* (β,τ_h)*.*

(b) Assume α_* *is the unique minimizer in* (B.12) *with* $\mathcal{A}:=[0,\infty)$*. For any* $\epsilon>0$*, define* $\mathcal{S}_\epsilon := \{\alpha \mid |\alpha-\alpha_*|<\epsilon\}$*. Then, for any sufficiently large constants* $K_\alpha>\alpha_*$ *and* $K_\beta>0$*, and for all* $\eta>0$*, it holds with probability approaching 1 as* $n\to\infty$*:*

(i) $\phi_{[0,K_\alpha]}<\overline{\phi}_{[0,\infty)}+\eta$*,*

(ii) $\phi_{[0,K_\alpha]\setminus\mathcal{S}_\epsilon}\ge\overline{\phi}_{[0,\infty)\setminus\mathcal{S}_\epsilon}-\eta$*,*

(iii) $\overline{\phi}_{[0,\infty)\setminus\mathcal{S}_\epsilon}>\overline{\phi}_{[0,\infty)}$*.*

Putting all the Pieces Together

We are now ready to conclude the proof of Theorem 4.2.1.

Proof of Theorem 4.2.1. Fix any $\epsilon>0$. Consider the set $\mathcal{S}_\epsilon=\{\mathbf{w}\mid|\|\mathbf{w}\|_2-\alpha_*\|_2<\epsilon$ as in Lemma B.1.3. We use the same notation as in the lemma. Let $K_\alpha>\alpha_*$ and arbitrarily large (but finite) $K_\beta>0$. From Lemma B.1.4(i), $\phi(\mathbf{g},\mathbf{h})$ is equal to the optimal cost of the optimization in (B.9). But, from Lemma B.1.5(b)(i), the latter converges in probability to some constant $\overline{\phi}$ (see Lemma B.1.5 for the exact

value constant). The same line of arguments applies to $\phi_{\mathcal{S}_\epsilon^c}(\mathbf{g}, \mathbf{h})$, showing that it converges to another constant $\overline{\phi}_{\mathcal{S}_\epsilon^c}$. Again from Lemma B.1.5(iii): $\overline{\phi}_{\mathcal{S}_\epsilon^c} > \overline{\phi}$. Thus, the conditions of Lemma B.1.3 are satisfied, and it implies that the magnitude of any optimal minimizer (say) $\hat{\mathbf{w}}^{(PO)}$ of the (PO) problem in (B.6) satisfies $\hat{\mathbf{w}}^{(PO)} \in \mathcal{S}$ in probability, in the limit of $n \to \infty$. □

B.2 Proofs for Section F.1

Proof of Lemma B.1.1

For convenience, denote with $M(\mathbf{w})$ the objective function in (B.1). For some $\epsilon > 0$ such that $\alpha+\epsilon < K_\alpha$ (e.g. $\epsilon = \zeta/2$ in (B.2)), denote $\mathcal{D} := \{\mathbf{w} \mid \alpha-\epsilon \leq \|\mathbf{w}\|_2 \leq \alpha+\epsilon\}$. By assumption, with probability approaching 1 (w.p.a. 1)

$$\hat{\mathbf{w}}^B \in \mathcal{D}. \tag{B.13}$$

For the shake of a contradiction, assume that there exists optimal solution $\hat{\mathbf{w}}$ of (B.1) such that $\hat{\mathbf{w}} \notin \mathcal{D}$ w.p.a. 1. Clearly,

$$M(\hat{\mathbf{w}}) \leq M(\hat{\mathbf{w}}^B). \tag{B.14}$$

Suppose $\hat{\mathbf{w}} \in \mathcal{S}_{\mathbf{w}}$, then $\hat{\mathbf{w}}$ is optimal for (B.3) and satisfies (B.13), which contradicts our assumption. Thus, $\hat{\mathbf{w}} \notin \mathcal{S}_{\mathbf{w}}$. Next, let $\mathbf{w}_\theta := \theta\hat{\mathbf{w}} + (1-\theta)\hat{\mathbf{w}}^B$ for $\theta \in (0,1)$ such that $\mathbf{w}_\theta \notin \mathcal{D}$ and $\mathbf{w}_\theta \in \mathcal{S}_w$ (always possible, by definition of $\mathcal{D}$). By the convexity of F and (B.14), it follows that $M(\hat{\mathbf{w}}_\theta) \leq M(\hat{\mathbf{w}}^B)$. Hence, $\hat{\mathbf{w}}_\theta$ is optimal for (B.3) and satisfies (B.13), which, again, is a contradiction. This completes the proof.

Proof of Lemma B.1.2

It suffices to prove the equivalence of the optimization (B.4) and (B.5). Let $\mathbf{w}_*, \mathbf{v}_*, \mathbf{u}_*$ be optimal in (B.4). To prove the claim, we show that $\mathbf{u}_* \in \mathcal{S}_{\mathbf{u}} \left(\Leftrightarrow \|\mathbf{u}_*\|_2 \leq K_\beta\right)$ w.p.a. 1. From the first order optimality conditions in (B.4), we find that

$$\mathbf{u}_* \in \frac{1}{\sqrt{n}}\partial\mathcal{L}(\mathbf{v}_*), \tag{B.15}$$

$$\mathbf{v}_* = \mathbf{z} - \sqrt{n}\mathbf{A}\mathbf{w}_*. \tag{B.16}$$

Recall Assumption 4.2.1(b) and consider two cases. First, if $\sup_{\mathbf{v}\in\mathbb{R}^m} \sup_{\mathbf{s}\in\partial\mathcal{L}(\mathbf{v})} \|\mathbf{s}\|_2 < \infty$, the claim follows directly by (B.15). Next, assume that w.h.p., $\|\mathbf{z}\|_2 \leq C_1\sqrt{n}$ for constant $C_1 > 0$. Also, a standard high probability bound on the spectral norm of Gaussian matrices gives $\|\mathbf{A}\|_2 \leq C_2$, e.g. [Ver10b]. Using these, boundedness of $\mathbf{w}_*$ and (B.16), we find that $\|\mathbf{v}_*\|_2 \leq C_3\sqrt{n}$ whp. Then, the normalization condition $\frac{1}{\sqrt{n}} \sup_{\mathbf{s}\in\partial\mathcal{L}(\mathbf{v})} \|\mathbf{s}\|_2 \leq C$ for all $\|\mathbf{v}\|_2 \leq c\sqrt{n}$ and all $n \in \mathbb{N}$, yields the desired, i.e. $\|\mathbf{u}_*\|_2 \leq C$ holds with probability approaching 1 as $n \to \infty$.

Proof of Lemma B.1.3

Let $\mathbf{w}_*$ denote an optimal solution of the "bounded" optimization in (B.6). It will suffice to prove that $\mathbf{w}_* \in \mathcal{S}$ in probability. To see this, recall from Lemma B.1.2 that (B.6) is asymptotically equivalent to (B.3). Then, Lemma B.1.1 and the assumption $\alpha_* < K_\alpha$ guarantee that $\hat{\mathbf{w}}(\mathbf{A}) \in \mathcal{S}$ in probability, as desired.

Denote $\Phi := \Phi(\mathbf{A})$ the optimal cost of the minimization in (B.6) and $\Phi_{\mathcal{S}^c} := \Phi_{\mathcal{S}^c}(\mathbf{A})$ the optimal cost of the same problem when the minimization is further restricted to be over the set $\mathbf{w} \in \mathcal{S}^c$. Note that $\mathbf{w}_* \in \mathcal{S}$ iff $\Phi_{\mathcal{S}^c}(\mathbf{A}) > \Phi(\mathbf{A})$; hence, it will suffice to prove that the latter event occurs in probability.

We do so by relating the (PO) in (B.6) to the Auxiliary Optimization (AO) in (B.8) using Theorem 3.3.1. For concreteness, denote the objective function in (B.8) with $A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s})$, and, recall $\mathcal{S}_{\mathbf{w}} := \{\mathbf{w} \mid \|\mathbf{w}\|_2 \le K_\alpha\}$, $\mathcal{S}_{\mathbf{u}} := \{\mathbf{u} \mid \|\mathbf{u}\|_2 \le K_\beta\}$. With these, define

$$\begin{aligned} \phi^P := \phi^P(\mathbf{g}, \mathbf{h}) &:= \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v}} \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}},\mathbf{s}} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}), \\ \text{and } \phi^D := \phi^D(\mathbf{g}, \mathbf{h}) &:= \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}},\mathbf{s}} \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v}} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}). \end{aligned} \tag{B.17}$$

Observe here that the order of min-max in ϕ^P is exactly as in the original formulation of the CGMT, cf. (3.11b); ϕ^D is the dual of it, and ϕ in (B.8) involves yet another change in the order of the optimizations. The reason we prefer to work with the later problem, is that this particular order allows for a number of simplifications performed in Section B.1.

As done before, denote with ${\phi^P}_{\mathcal{S}^c}, {\phi^D}_{\mathcal{S}^c}$ the optimal cost of the optimizations in (B.17) under the additional constraint $\mathbf{w} \in \mathcal{S}^c$. The two problems in (B.17) are related to the one in (B.8) as follows:

$$\begin{aligned} {\phi^P}_{\mathcal{S}^c} &= \min_{\substack{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v} \\ \mathbf{w}\in\mathcal{S}^c}} \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}},\mathbf{s}} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}) = \min_{\substack{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v} \\ \mathbf{w}\in\mathcal{S}^c}} \max_{\beta,\mathbf{s}} \max_{\|\mathbf{u}\|_2=\beta} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}) \\ &\ge \max_{\beta,\mathbf{s}} \min_{\substack{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v} \\ \mathbf{w}\in\mathcal{S}^c}} \max_{\|\mathbf{u}\|_2=\beta} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}) = \phi_{\mathcal{S}^c}, \end{aligned} \tag{B.18}$$

where the inequality follows from the min-max inequality [Roc97, Lem. 36.1]. Similarly,

$$\begin{aligned} \phi^D &= \max_{\mathbf{u}\in\mathcal{S}_{\mathbf{u}},\mathbf{s}} \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v}} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}) = \max_{\beta,\mathbf{s}} \max_{\|\mathbf{u}\|_2=\beta} \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v}} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}) \\ &\le \max_{\beta,\mathbf{s}} \min_{\mathbf{w}\in\mathcal{S}_{\mathbf{w}},\mathbf{v}} \max_{\|\mathbf{u}\|_2=\beta} A(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{s}) = \phi. \end{aligned} \tag{B.19}$$

Furthermore, they are related to the (PO) via the CGMT. From Theorem 3.3.1(i), for all $c \in \mathbb{R}$:

$$\mathbb{P}(\Phi_{\mathcal{S}^c} < c) \leq 2\mathbb{P}(\phi^P{}_{\mathcal{S}^c} \leq c). \tag{B.20}$$

Also, from Theorem 3.3.1(ii)[2]:

$$\mathbb{P}(\Phi > c) \leq 2\mathbb{P}(\phi^D \geq c). \tag{B.21}$$

The remaining of the proof is in the same lines as the proof of 3.3.1(iii), but is included for clarity. Let $\eta := (\phi_{\mathcal{S}^c} - \phi)/3 > 0$. We may apply (B.20) for $c = \overline{\phi}_{\mathcal{S}^c} - \eta$ and combine with (B.18) to find that

$$\mathbb{P}(\Phi_{\mathcal{S}^c} < \overline{\phi}_{\mathcal{S}^c} - \eta) \leq 2\mathbb{P}(\phi^P{}_{\mathcal{S}^c} \leq \overline{\phi}_{\mathcal{S}^c} - \eta) \leq 2\mathbb{P}(\phi_{\mathcal{S}^c} \leq \overline{\phi}_{\mathcal{S}^c} - \eta). \tag{B.22}$$

From assumption (b) the last term above tends to zero as $n \to \infty$. In a similar way, combining (B.21), (B.19) and assumption (a), we find that

$$\mathbb{P}(\Phi > \overline{\phi} + \eta) \leq 2\mathbb{P}(\phi^D \geq \overline{\phi} + \eta) \leq 2\mathbb{P}(\phi \geq \overline{\phi} + \eta), \tag{B.23}$$

goes to zero with $n \to \infty$. Denote the event $\mathcal{E} = \{\Phi_{\mathcal{S}^c} \geq \overline{\phi}_{\mathcal{S}^c} - \eta \text{ and } \Phi \leq \overline{\phi} + \eta\}$. From (B.22) and (B.23) the event occurs with probability approaching 1. Furthermore, in this event, after using assumption (a), we have $\Phi^b{}_{\mathcal{S}^c} \geq \overline{\phi}_{\mathcal{S}^c} - \eta > \overline{\phi} + \eta \geq \Phi^b$; equivalently, the optimal minimizer satisfies $\mathbf{w}_* \in \mathcal{S}$, which completes the proof.

Proof of Lemma B.1.4

(i) We start by showing how the vector optimization in (B.8) can be reduced to the scalar one that appears in (B.9). This requires the following steps.

Optimizing over the direction of $\mathbf{u}$*:* Performing the inner maximization is easy. In particular, using the fact that $\max_{\|\mathbf{u}\|_2=\beta} \mathbf{u}^T\mathbf{t} = \beta\|\mathbf{t}\|_2$ for all $\beta \geq 0$ the problem simplifies to a max-min one:

$$\max_{0\leq\beta\leq K_\beta,\mathbf{s}} \min_{\|\mathbf{w}\|\leq K_\alpha,\mathbf{v}} \frac{\beta}{\sqrt{n}}\| \|\mathbf{w}\|_2\mathbf{g} + \mathbf{z} - \mathbf{v} \|_2 - \frac{\beta}{\sqrt{n}}\mathbf{h}^T\mathbf{w} + \frac{1}{n}\mathcal{L}(\mathbf{v}) + \frac{\lambda}{n}\mathbf{s}^T\mathbf{x}_0 + \frac{\lambda}{\sqrt{n}}\mathbf{s}^T\mathbf{w} - \frac{\lambda}{n}f^*(\mathbf{s}).$$

Optimizing over the direction of $\mathbf{w}$*:* Next, we fix $\|\mathbf{w}\|_2 = \alpha$, and, similar to what was done above, minimize over its direction:

$$\max_{0\leq\beta\leq K_\beta,\mathbf{s}} \min_{0\leq\alpha\leq K_\alpha,\mathbf{v}} \frac{\beta}{\sqrt{n}}\| \alpha\mathbf{g} + \mathbf{z} - \mathbf{v} \|_2 + \frac{1}{n}\mathcal{L}(\mathbf{v}) - \frac{\alpha}{\sqrt{n}}\|\beta\mathbf{h} - \lambda\mathbf{s}\|_2 + \frac{\lambda}{n}\mathbf{s}^T\mathbf{x}_0 - \frac{\lambda}{n}f^*(\mathbf{s}). \tag{B.24}$$

[2] more precisely, please refer to equation (32) in [TOH15].

Changing the orders of min-max: Denote with $M(\alpha, \beta, \mathbf{v}, \mathbf{s})$ the objective function above. It can be checked that M is jointly convex in $(\alpha, \mathbf{v})$ and jointly concave in $(\beta, \mathbf{s})$ (cf. Lemma B.2.4). Thus, $\min_{\mathbf{v}} M$ is convex in α and jointly concave in $(\beta, \mathbf{s})$. Furthermore, the constraint sets are all convex and the one over which minimization over α occurs is bounded. Hence, as in [Sio+58, Cor. 3.3] we can flip the order of $\max_{\beta,\mathbf{s}} \min_{\alpha}$, to conclude with

$$\min_{0\le\alpha\le K_\alpha} \max_{0\le\beta\le K_\beta} \max_{\mathbf{s}} \min_{\mathbf{v}} M(\alpha, \beta, \mathbf{v}, \mathbf{s}).$$

Also, observe that the order of optimization among $\mathbf{v}$ and $\mathbf{s}$ does not affect the outcome.

The square-root trick: We apply the fact that $\sqrt{\chi} = \inf_{\tau>0}\{\frac{\tau}{2} + \frac{\chi}{2\tau}\}$ to both the terms $\frac{1}{\sqrt{n}}\|\alpha\mathbf{g} + \mathbf{z} - \mathbf{v}\|_2$ and $\frac{1}{\sqrt{n}}\|\beta\mathbf{h} - \lambda\mathbf{s}\|_2$:

$$\min_{0\le\alpha\le K_\alpha} \max_{0\le\beta\le K_\beta} \inf_{\tau_g>0} \sup_{\tau_h>0} \frac{\beta\tau_g}{2} + \frac{1}{n}\min_{\mathbf{v}}\left\{\frac{\beta}{2\tau_g}\| \alpha\mathbf{g} + \mathbf{z} - \mathbf{v} \|_2^2 + \mathcal{L}(\mathbf{v})\right\} - \frac{\alpha\tau_h}{2} - \frac{1}{n}\min_{\mathbf{s}}\left\{\frac{\alpha}{2\tau_h}\|\beta\mathbf{h} - \lambda\mathbf{s}\|_2^2 - \lambda\mathbf{s}^T\mathbf{x}_0 + \lambda f^*(\mathbf{s})\right\}. \tag{B.25}$$

Identifying the Moreau envelope: Arguing as before, we can change the order of optimization between β and τ_g. Also, it takes only a few algebra steps and using basic properties of Moreau envelope functions (in particular, Lemma B.2.5(ii)) to rewrite the last summand in (B.25) as below. If $\alpha > 0$, then,

$$\min_{\mathbf{s}}\left\{\frac{\alpha}{2\tau_h}\|\beta\mathbf{h} - \lambda\mathbf{s}\|_2^2 - \lambda\mathbf{s}^T\mathbf{x}_0 + \lambda f^*(\mathbf{s})\right\} = -\frac{\tau_h}{2\alpha}\|\mathbf{x}_0\|_2^2 - \beta\mathbf{h}^T\mathbf{x}_0 + \lambda\cdot \mathrm{e}_{f^*}\left(\frac{\beta}{\lambda}\mathbf{h} + \frac{\tau_h}{\alpha\lambda}\mathbf{x}_0; \frac{\tau_h}{\alpha\lambda}\right) \tag{B.26}$$

$$= \frac{\beta^2\alpha}{2\tau_h}\|\mathbf{h}\|^2 - \lambda\cdot \mathrm{e}_f\left(\frac{\beta\alpha}{\tau_h}\mathbf{h} + \mathbf{x}_0; \frac{\alpha\lambda}{\tau_h}\right). \tag{B.27}$$

Otherwise, if $\alpha = 0$, then the same term equals $-\lambda f(\mathbf{x}_0)$ since $\max_{\mathbf{s}} \mathbf{s}^T\mathbf{x}_0 - f^*(\mathbf{s}) = f(\mathbf{x}_0)$.

(ii) The continuity of the objective function in (B.9) follows directly from the continuity of the Moreau envelope functions, cf. [RW09, Lem. 1.25, 2.26]. In particular, regarding the two branches of the objective: it can be checked, using the continuity of the Moreau envelope, that the limit of the RHS in (B.27) as $\alpha \to 0$ evaluates to

$-\lambda f(\mathbf{x}_0)$. (In fact, this is the unique extension of the upper branch to a continuous finite convex function on the whole $\alpha \geq 0, \tau > 0$, as per [Roc97, Thm. 10.3].)

Convexity of (B.9) can be checked from (B.25). By applying Lemma B.2.4, after minimization over $\mathbf{v}$, the Moreau Envelope remains jointly convex with respect to α and τ_g and concave in β. The same argument (and similar lemma) holds for the last term of (B.25) in which after minimization over $\mathbf{s}$ it remains jointly convex in β and τ_h and concave in α. Then the negative sign before this term makes it jointly concave in β and τ_h and convex over α.

Proof of Lemma B.1.5

(a) By Assumption 4.2.1(a) the normalized Moreau envelope functions in (B.10) converge in probability to L and F, respectively. Also, $\|\mathbf{h}\|_2^2/n \xrightarrow{P} 1$ by the WLLN. This proves the convergence part.

Lemma B.1.4(i) showed $\mathcal{R}_n$ to be convex-concave. Then, the same holds for $\mathcal{D}$ by point-wise convergence and the fact that convexity is preserved by point wise limits.

(b) Call

$$M_n(\alpha) = \sup_{\substack{0\leq\beta\leq K_\beta \\ \tau_h>0}} \inf_{\tau_g>0} \mathcal{R}_n(\alpha,\tau_g,\beta,\tau_h) \quad \text{and} \quad M(\alpha) = \sup_{\substack{0\leq\beta \\ \tau_h>0}} \inf_{\tau_g>0} \mathcal{D}_n(\alpha,\tau_g,\beta,\tau_h). \tag{B.28}$$

The bulk of the proof consists of showing that the following two statements hold:

$$\forall \text{ compact susets } \mathcal{A} \subset (0,\infty) \text{ and sufficiently large } K_\beta := K_\beta(\mathcal{A}) > 0 : \inf_{\alpha\in\mathcal{A}} M_n(\alpha) \xrightarrow{P} \inf_{\alpha\in\mathcal{A}} M(\alpha) \tag{B.29}$$

and

$$\forall \epsilon > 0, \text{w.p.a.1} : \; M_n(0) < M(0) + \epsilon. \tag{B.30}$$

Before proceeding with the proof of these, let us show how the conclusion of the lemma is reached once (B.29) and (B.30) are established.

Using (B.29) and (B.30) to prove the lemma : Fix $K_\alpha > \alpha_*$, any $\delta > 0$ such that $\mathcal{A} := [\alpha_* - 2\delta, \alpha_* + 2\delta] \subset (0, K_\alpha]$ and $K_\beta > 0$ large enough such that (B.29) and (B.30) both hold. Then, for all $\epsilon > 0$, w.p.a.1:

$$\min_{0\leq\alpha\leq K_\alpha} M_n(\alpha) \leq \min_{\alpha\in\mathcal{A}} M_n(\alpha) \leq M_n(\alpha_*) < M(\alpha_*) + \epsilon. \tag{B.31}$$

For the last inequality above: if $\alpha_* = 0$, it follows from (B.30), or otherwise from (B.29).

Next, consider the compact set $\mathcal{A}_l = \{\alpha > 0 \mid \alpha \in [\alpha_* - 2\delta, \alpha - \delta]\}$ and $\mathcal{A}_r = \{\alpha > 0 \mid \alpha \in [\alpha_* + \delta, \alpha + 2\delta]\}$. (Note that if $\alpha_* = 0$, then $\mathcal{A}_l$ is empty.) From (B.29), we know that for all $\epsilon > 0$, w.p.a.1

$$\min_{\alpha\in\mathcal{A}_l} M_n(\alpha) > \min_{\alpha\in\mathcal{A}_l} M(\alpha) - \epsilon \quad \text{and} \quad \min_{\alpha\in\mathcal{A}_u} M_n(\alpha) > \min_{\alpha\in\mathcal{A}_u} M(\alpha) - \epsilon.$$

Let $\mathcal{A}_{lu} = \mathcal{A}_l \cup \mathcal{A}_u$ and combine the above to find

$$\min_{\alpha\mathcal{A}_{lu}} M_n(\alpha) > \min_{\alpha\in\mathcal{A}_{lu}} M(\alpha) - \epsilon. \tag{B.32}$$

By assumption on uniqueness of α_* and on convexity of M, we have

$$M(\alpha_*) < \min_{\alpha\in\mathcal{A}_{lu}} M(\alpha) \tag{B.33}$$

and $M(\alpha_*) = \min_{\alpha\in\mathcal{A}} M(\alpha)$. Thus, Applying (B.31) and (B.32) for $\epsilon = (\min_{\alpha\in\mathcal{A}_{lu}} M(\alpha) - M(\alpha_*))/3$ yields w.p.a.1 :

$$\min_{\alpha\in\mathcal{A}_{lu}} M_n(\alpha) > \min_{\alpha\in\mathcal{A}_{lu}} M(\alpha) - \epsilon > M(\alpha_*) + \epsilon > \min_{\alpha\in[\alpha_*-2\delta,\alpha_*+2\delta]} M_n(\alpha). \tag{B.34}$$

Thus, w.p.a.1,

$$\hat{\alpha}_n := \arg\min_{\alpha\in\mathcal{A}} M_n(\alpha) \in (\alpha_* - \delta, \alpha_* + \delta).$$

In this event, for any $\alpha \notin \mathcal{A}$, there is a convex combination $\alpha_\theta := \theta\hat{\alpha}_n + (1-\theta)\alpha$, $(\theta < 1)$ that equals either $\alpha_* - 2\delta$ or $\alpha_* + 2\delta$. By convexity,

$$M_n(\alpha_\theta) \le \theta M_n(\hat{\alpha}_n) + (1-\theta) M_n(\alpha).$$

Also, from (B.34), $M_n(\hat{\alpha}_n) < M_n(\alpha_\theta)$. Combining those, we find $M_n(\hat{\alpha}_n) < M_n(\alpha)$, implying that $\hat{\alpha}_n$ is the minimizer of M_n over the entire $[0, K_\alpha]$ w.p.a.1. In other words, for all ϵ w.p.a. 1,

$$\min_{\alpha\in[0,K_\alpha]\setminus(\alpha_*-\delta,\alpha_*+\delta)} M_n(\alpha) \ge \min_{\alpha\in\mathcal{A}_{lu}} M_n(\alpha) > \min_{\alpha\in\mathcal{A}_{lu}} M(\alpha) - \epsilon. \tag{B.35}$$

To establish a connection with the three statements (i)-(iii) of the lemma, observe that $\overline{\phi}_{[0,\infty)} = M(\alpha_*)$. Also, $\overline{\phi}_{[0,\infty)\setminus\mathcal{S}_\delta} = \min_{\alpha\in\mathcal{A}_{lu}} M(\alpha)$ (by convexity). With these, (i) corresponds directly to (B.31), (ii) to (B.35), and, (iii) to (B.33).

Proof of (B.29) and (B.30) : From the first statement of the lemma, the objective function $\mathcal{R}_n$ of the (AO) converges point-wise to $\mathcal{D}$. We will use this to show that

the minimax value of $\mathcal{R}_n$ converges to the corresponding minimax of $\mathcal{D}$. The proof is based on a repeated use of Lemma B.2.1 below, about convergence of the infimum of a sequence of convex converging stochastic processes. This fact is essentially a consequence of what is known in the literature as *convexity lemma*, according to which point wise convergence of convex functions implies uniform convergence in compact subsets. Please refer to Section B.2 for the proof.

Lemma B.2.1 (Min-convergence – Open Sets)**.** *Consider a sequence of proper, convex stochastic functions $M_n : (0,\infty) \to \mathbb{R}$, and, a deterministic function $M : (0,\infty) \to \mathbb{R}$, such that:*

(a) $M_n(x) \xrightarrow{P} M(x)$*, for all* $x > 0$*,*

(b) there exists $z > 0$ *such that* $M(x) > \inf_{x>0} M(x)$ *for all* $x \geq z$*.*

Then, $\inf_{x>0} M_n(x) \xrightarrow{P} \inf_{x>0} F(x)$.

1) Fix $\alpha \geq 0, \beta > 0$, and, $\tau_h > 0$. Consider

$$M_n^{\alpha,\beta,\tau_h}(\tau_g) := R_n(\alpha,\tau_g,\beta,\tau_h), \tag{B.36}$$

$$M^{\alpha,\beta,\tau_h}(\tau_g) := \mathcal{D}(\alpha,\tau_g,\beta,\tau_h). \tag{B.37}$$

The functions $\{M_n\}$ are convex. Furthermore, $M_n^{\alpha,\beta,\tau_h}(\tau_g) \xrightarrow{P} M^{\alpha,\beta,\tau_h}(\tau_g)$ point wise in τ_g. Next, we show that M^{α,β,τ_h} is level-bounded, i.e. it satisfies condition (b) of Lemma B.2.1. In view of Lemma B.2.2, it suffices to show that $\lim_{\tau_g\to\infty} M^{\alpha,\beta,\tau_h}(\tau_g) = +\infty$, or $\lim_{\tau_g\to\infty} \left(\frac{\beta}{2} + \delta \cdot \frac{L(\alpha,\tau_g/\beta)}{\tau_g}\right) > 0$. By assumption 4.2.2(c), $\lim_{\tau_g\to\infty} L(\alpha,\tau_g/\beta) = -L_0$. There are two cases to be considered. Either $L_0 < \infty$, or else, Assumption 4.2.2(d) holds. Either way, $\lim_{\tau_g\to\infty} L(\alpha,\tau_g/\beta)/\tau_g = 0$ and we are done. Now, we can apply Lemma B.2.1 to conclude that

$$\inf_{\tau_g>0} M_n^{\alpha,\beta,\tau_h}(\tau_g) \xrightarrow{P} \inf_{\tau_g>0} M^{\alpha,\beta,\tau_h}(\tau_g). \tag{B.38}$$

2) Next, again for fixed $\alpha \geq 0, \tau_h > 0$, consider (we use some abuse of notation here, with the purpose of not overloading notation)

$$M_n^{\alpha,\tau_h}(\beta) := \inf_{\tau_g>0} M_n^{\alpha,\beta,\tau_h}(\tau_g),$$

$$M^{\alpha,\tau_h}(\beta) := \inf_{\tau_g>0} M^{\alpha,\beta,\tau_h}(\tau_g).$$

The functions $\{M_n^{\alpha,\tau_h}\}$ are concave in β, as the point wise minima of concave functions. Furthermore, $M_n^{\alpha,\tau_h}(\beta) \xrightarrow{P} M^{\alpha,\tau_h}(\beta)$ point wise in $\beta > 0$, by (B.38).

$\underline{\alpha > 0:}$ For now and until further notice, restrict attention to the case $\alpha > 0$. Also, consider first $\beta > 0$. We show that M^{α,τ_h} is level-bounded, i.e. it satisfies condition (b) of Lemma B.2.1. In view of Lemma B.2.2, it suffices to show that $\lim_{\beta\to+\infty} M^{\alpha,\tau_h}(\beta) = -\infty$, or $\lim_{\beta\to+\infty} \inf_{\tau_g>0} M^{\alpha,\beta,\tau_h}(\tau_g) = -\infty$. This condition is equivalent to the following

$$(\forall M > 0)(\exists B > 0)\ \left[\beta > B \Rightarrow (\exists\{\tau_g\}_k)\ [D(\alpha,\tau_g,\beta,\tau_h) < -M]\right]. \tag{B.39}$$

First, we show that

$$\lim_{\beta\to+\infty} \frac{\alpha\beta^2}{2\tau_h} - \lambda \cdot F\left(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h}\right) = +\infty. \tag{B.40}$$

This follows by Assumption 4.2.2(a) when applied for $c = \alpha\beta/\tau_h$ and $\tau = \alpha\lambda/\tau_h$ (recall here that $\alpha > 0$).

Next, choose $\{\tau_g\}_k \to 0$. For that choice, $\frac{\beta\tau_g}{2} + L(\alpha, \tau_g/\beta) \to \lim_{\tau\to 0} L(\alpha,\tau) < \infty$, where boundedness follows by Assumption 4.2.2(b). Thus, (B.39) is correct and we may apply Lemma B.2.1 to conclude that

$$\sup_{\beta>0} M_n^{\alpha,\tau_h}(\beta) \xrightarrow{P} \sup_{\beta>0} M^{\alpha,\tau_h}(\beta). \tag{B.41}$$

Now, we investigate the case $\beta = 0$. We have, $M_n^{\alpha,\tau_h}(0) = -\frac{1}{n}\mathcal{L}(\mathbf{z}) - \frac{\alpha\tau_h}{2} + \frac{\lambda}{n}\left(\mathrm{e}_f\left(\mathbf{x}_0; \frac{\alpha\lambda}{\tau_h}\right) - f(\mathbf{x}_0)\right)$ and $M_n^{\alpha,\tau_h}(0) = -\delta L_0 - \frac{\alpha\tau_h}{2} + F(0, \frac{\alpha\lambda}{\tau_h})$.

If $L_0 < \infty$, then by assumption, $M_n^{\alpha,\tau_h}(0) \xrightarrow{P} M^{\alpha,\tau_h}(0)$. Combined with (B.41), we find

$$\sup_{\beta\geq 0} M_n^{\alpha,\tau_h}(\beta) \xrightarrow{P} \sup_{\beta\geq 0} M^{\alpha,\tau_h}(\beta). \tag{B.42}$$

Now, consider the case $L_0 = +\infty$. Clearly, the optimal β for M^{α,τ_h} is not at zero; thus, $\sup_{\beta\geq 0} M^{\alpha,\tau_h}(\beta) = \sup_{\beta>0} M^{\alpha,\tau_h}(\beta)$. Also, by assumption, for all M, $\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\mathcal{L}(\mathbf{z}) > M\right) = 1$. Letting $\epsilon > 0$ and $M := -\sup_{\beta>0} M^{\alpha,\tau_h}(\beta) + \epsilon + \frac{\alpha\tau_h}{2} - F(0, \frac{\alpha\lambda}{\tau_h})$, then w.p.a.1, $M_n^{\alpha,\tau_h}(0) < \sup_{\beta>0} M^{\alpha,\tau_h}(\beta) - \epsilon \leq \sup_{\beta>0} M_n^{\alpha,\tau_h}(\beta)$, where the last inequality follows because of (B.41). Again, this leads to (B.42). To sum up, (B.42) holds for all $\alpha > 0$.

$\underline{\alpha = 0:}$ We show that for all $\epsilon > 0$, the following holds w.p.a.1:

$$\sup_{\beta \geq 0} M_n^{\alpha=0,\tau_h}(\beta) < \sup_{\beta \geq 0} M^{\alpha=0,\tau_h}(\beta) + \epsilon. \tag{B.43}$$

To begin with, note that for all n,

$$\sup_{\beta \geq 0} M_n^{\alpha=0,\tau_h}(\beta) \leq \sup_{\beta > 0} \lim_{\tau_g \to 0} \frac{\beta \tau_g}{2} + \frac{1}{n} \min_{\mathbf{v}} \left\{ \frac{\beta}{2\tau_g} \|\mathbf{z} - \mathbf{v}\|_2^2 + \mathcal{L}(\mathbf{v}) - \mathcal{L}(\mathbf{z}) \right\} = 0, \tag{B.44}$$

where we have used Lemma B.4.1(ix). Next, we show that

$$M^{\alpha=0,\tau_h}(\beta) = 0. \tag{B.45}$$

Using Assumption 4.2.2(c) on the non-negativity of L_0 and Assumption 4.2.2(b) that $\lim_{\tau \to 0} L(c, \tau) = 0$, it follows that $M^{\alpha=0,\tau_h}(\beta) \leq \sup_{\beta>0} \lim_{\tau_g \to 0} \frac{\beta \tau_g}{2} + L(0, \tau_g/\beta) = 0$. Thus, it will suffice for the claim if we prove

$$\lim_{\beta \to \infty} \inf_{\tau_g > 0} \frac{\beta \tau_g}{2} + L(0, \tau_g/\beta) = 0, \tag{B.46}$$

or equivalently,

$$\lim_{\beta \to \infty} \inf_{\kappa > 0} \kappa \left(\frac{\beta^2}{2} + \frac{L(0,\kappa)}{\kappa} \right) = 0.$$

Fix some $\beta > 0$. Note that $\lim_{\kappa \to 0} \frac{\kappa \beta^2}{2} + L(0, \kappa) = 0$, where we have used Assumption 4.2.2(b) that $\lim_{\tau \to 0} L(0, \tau) = 0$. Also, $\lim_{\kappa \to \infty} \kappa \left(\frac{\beta^2}{2} + \frac{L(0,\kappa)}{\kappa} \right) = +\infty$, using Assumption 4.2.2(d) this time. Now, consider only $\beta > \sqrt{-L_{2,+}(0,0)}$ (see Assumption 4.2.2(b)). Then, the function $\frac{\kappa \beta^2}{2} + L(0, \kappa)$ has a positive derivative at $\kappa \to 0^+$. From this and convexity, it follows that for all $\kappa > 0$,

$$\frac{\kappa \beta^2}{2} + L(0, \kappa) \geq \lim_{\kappa \to 0} \frac{\kappa \beta^2}{2} + L(0, \kappa) = 0.$$

This proves (B.46) as desired.

To complete the argument, (B.43) follows by (B.44) and (B.45), and with this we have completed the proof of (B.30).

3) Keep $\alpha > 0$ fixed and consider

$$M_n^{\alpha}(\tau_h) := \sup_{\beta \geq 0} M_n^{\alpha,\tau_h}(\beta),$$

$$M^{\alpha}(\tau_h) := \sup_{\beta \geq 0} M^{\alpha,\tau_h}(\beta).$$

The functions $\{M_n^\alpha\}$ and F are all concave in τ_h, as the point wise maxima of jointly concave functions. Furthermore, $M_n^\alpha(\tau_h) \xrightarrow{P} M^\alpha(\tau_h)$ point wise in τ_h, by (B.42). Next, we show that M^{τ_h} is level-bounded, i.e. it satisfies condition (b) of Lemma B.2.1. In view of Lemma B.2.2, it suffices to show that $\lim_{\tau_h\to\infty} M^\alpha(\tau_h) = +\infty$, or $\lim_{\tau_h\to\infty} \sup_{\beta>0} \inf_{\tau_g>0} \mathcal{D}(\alpha,\tau_g,\beta,\tau_h) = -\infty$. This is equivalent to the following

$$(\forall M > 0)(\exists T > 0)\ \left[\tau_h > T \Rightarrow (\forall\{\beta\}_k)(\exists\{\tau_g\}_k)\ [D(\alpha,\tau_g,\beta,\tau_h) < -M]\ \right]. \tag{B.47}$$

Consider the function

$$\mathcal{H}(\beta,\tau_h) := \frac{\alpha\tau_h}{2} + \frac{\alpha\beta^2}{2\tau_h} - \lambda\cdot F(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h}).$$

We show that

$$\mathcal{H}(\beta,\tau_h) \geq \frac{\alpha\tau_h}{2}.$$

To see this note that $\mathrm{e}_f\,(c\mathbf{h}+\mathbf{x}_0;\tau) \leq \frac{c^2\|\mathbf{h}\|^2}{2\tau} + f(\mathbf{x}_0)$. Thus, $\frac{1}{n}\left\{\mathrm{e}_f\,(c\mathbf{h}+\mathbf{x}_0;\tau) - f(\mathbf{x}_0)\right\} \leq \frac{c^2\|\mathbf{h}\|^2}{2\tau n}$. The LHS converges to $F(c,\tau)$ by Assumption 4.2.1(a) and the RHS converges to $\frac{c^2}{2\tau}$. Therefore, $F(c,\tau) \leq \frac{c^2}{2\tau}$. Applying this for $c = \frac{\alpha\beta}{\tau_h}$ and $\tau = \frac{\alpha\lambda}{\tau_h}$, we have that $\frac{\alpha\beta^2}{2\tau_h} - \lambda\cdot F(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h}) \geq 0$, as desired.

Then,

$$\mathcal{D}(\alpha,\tau_g,\beta,\tau_h) \leq \frac{\beta\tau_g}{2} + \delta\cdot L\left(\alpha, \frac{\tau_g}{\beta}\right) - \frac{\alpha\tau_h}{2}.$$

Also, note that for all $\beta > 0$, we can choose (sequence) of τ_g, such that $\beta\tau_g, \frac{\tau_g}{\beta} \to 0$. Then, $\frac{\beta\tau_g}{2} + \delta\cdot L\left(\alpha, \frac{\tau_g}{\beta}\right) \to \lim_{\tau\to 0} L(\alpha,\tau) =: A < \infty$. It can then be seen that (B.47) holds for (say) $T := T(M) = 4(A+M)/\alpha$.

We can apply Lemma B.2.1 to conclude that

$$\sup_{\tau_h>0} M_n^\alpha(\tau_h) \xrightarrow{P} \sup_{\tau_h>0} M^\alpha(\tau_h). \tag{B.48}$$

4) Finally, consider

$$M_n(\alpha) := \sup_{\tau_h>0} M_n^\alpha(\tau_h),$$

$$M(\alpha) := \sup_{\tau_h>0} M^\alpha(\tau_h). \tag{B.49}$$

The functions $\{M_n\}$ and F are all convex in τ_h, as the point wise maxima of convex functions. Furthermore, $M_n(\alpha) \xrightarrow{P} M(\alpha)$ point wise in α, by (B.48). By assumption of the lemma, F has a unique minimizer α_*, which of course implies level boundedness. Thus, we can apply Lemma B.2.1 to conclude that

$$\inf_{\alpha>0} M_n(\alpha) \xrightarrow{P} \inf_{\alpha>0} M(\alpha). \tag{B.50}$$

Besides, pointwise convergence $M_n(\alpha) \xrightarrow{P} M(\alpha)$ translates to uniform convergence over any compact subset $\mathcal{A} \subset (0, \infty)$ by the Convexity lemma [AG82, Cor.. II.1] ,[LM08, Lem. 7.75]. Hence,

$$\inf_{\alpha\in\mathcal{A}} M_n(\alpha) \xrightarrow{P} \inf_{\alpha\in\mathcal{A}} M(\alpha).$$

This is of course same as the desired in (B.29). Recall, (B.30) was established in (B.43). The only thing remaining is showing that there exists an optimal β_* in $\sup_{\beta\geq 0} M^{\alpha,\tau_h}(\beta)$ that is bounded by some sufficiently large $K_\beta(\mathcal{A})$. This follows from the level-boundedness arguments above as detailed next.

Boundedness of solutions : For a compact subset $\mathcal{A} \subset (0, \infty)$, we argue that there exists *bounded* β_* and sequences $\{\tau_{g_*}\}_k, \{\tau_{h*}\}_k$ such that $(\alpha_*, \{\tau_{g_*}\}_k, \beta_*, \{\tau_{h*}\}_k)$ approaches

$$\min_{\alpha\in\mathcal{A}} \sup_{\substack{\tau_h>0\\ \beta\geq 0}} \inf_{\tau_g>0} \mathcal{D}(\alpha, \tau_g, \beta, \tau_h).$$

This follows from the work above. In particular, at each step in the proof of (B.29) above, we showed level-boundedness of the corresponding functions. For example, (B.47) shows that there exists (sufficiently large) $T_h(\alpha) > 0$ such that $\sup_{\tau_h>0} M^\alpha(\tau_h)$ is equal to $\sup_{T_h(\alpha)\geq\tau_h>0} M^\alpha(\tau_h)$. This holds for all α; so, in particular, is true for $T_h := \max_{\alpha\in\mathcal{A}} T_h(\alpha)$. Next, from (B.40) there exists $K_\beta(\alpha, T_h)$, such that $\sup_{\beta\geq 0} M^{\alpha,\tau_h}(\beta)$ is equal to $\sup_{K_\beta(\alpha,T_h)\geq\beta\geq 0} M^{\alpha,\tau_h}(\beta)$. Again, this holds for all $\alpha \in \mathcal{A}$, thus there exists sufficiently large $K_\beta > 0$ such that (see also Lemma B.2.3)

$$\min_{\alpha\in\mathcal{A}} \sup_{\substack{\tau_h>0\\ \beta\geq 0}} \inf_{\tau_g>0} \mathcal{D}(\alpha, \tau_g, \beta, \tau_h) = \inf_{\alpha\in\mathcal{A}} \sup_{\substack{\tau_h>0\\ K_\beta\geq\beta\geq 0}} \inf_{\tau_g>0} \mathcal{D}(\alpha, \tau_g, \beta, \tau_h).$$

The objective function $\mathcal{D}$ above is convex-concave. Also, the constraint sets over α and β are compact. Furthermore, the optimization of $\mathcal{D}$ over τ_g and τ_h is separable. With these and an application of Sion's minimax theorem, the order of inf–sup between the four optimization variables can be flipped arbitrarily without affecting

the outcome. Thus, for example,

$$\inf_{\alpha\in\mathcal{A}}\sup_{\substack{\tau_h>0\\ \beta\geq 0}}\inf_{\tau_g>0}\mathcal{D}(\alpha,\tau_g,\beta,\tau_h)=\inf_{\substack{\alpha\in\mathcal{A}\\ \tau_g>0}}\sup_{\substack{\tau_h>0\\ \geq\beta\geq 0}}\mathcal{D}(\alpha,\tau_g,\beta,\tau_h).$$

The same is of course true for the corresponding random optimizations (also, Lemma B.1.4(iii)).

Auxiliary Lemmas

Proof of Lemma B.2.1. First, convexity is preserved by point wise limits, so that $F(x)$ is also convex. Using this and level-boundedness condition (b) of the lemma, it is easy to show that $\inf_{x>0} F(x) > -\infty$. Since F is proper and (lower) level-bounded, the only way $\inf_{x>0} F(x) = -\infty$ is if $\lim_{x\to 0} F(x) = -\infty$. But this is not possible as follows: Fix $0 < x_1 < x_2 < x_3$. Then, for any $0 < x < x_1$ and $\theta := \frac{x_3-x_2}{x_3-x}$, convexity gives

$$F(x) \geq \frac{1}{\theta}F(x_2) - \left(1-\frac{1}{\theta}\right)F(x_3) \geq -\frac{x_3-x_1}{x_3-x_2}|F(x_2)| - \frac{x_2-x_1}{x_3-x_2}|F(x_3)|.$$

Next, we show that for sufficiently small $\epsilon > 0$, there exist $x_0 > x_\epsilon > 0$:

$$\inf_{x>0} F(x) \leq F(x_\epsilon) \leq \inf_{x>0} F(x) + \epsilon \quad \text{and} \quad F(x_\epsilon) < F(x_0). \tag{B.51}$$

We show the claim for all $0 < \epsilon < \epsilon_1 := (F(z) - \inf_{x>0} F(x))$. Since $\inf_{x>0} F(x)$ is finite, there exists $x_\epsilon > 0$ such that $F(x_\epsilon) - \epsilon \leq \inf_{x>0} F(x)$. Without loss of generality, $x_\epsilon < z$. Pick any $x_0 > z$. For the sake of contradiction, assume $F(x_0) \leq F(x_\epsilon)$. Then, by convexity, for some $\theta \in (0,1)$

$$F(z) \leq \theta F(x_\epsilon) + (1-\theta)F(x_0) \leq F(x_\epsilon) \leq \inf_{x>0} F(x) + \epsilon < F(z).$$

Thus, $F(x_\epsilon) < F(x_0)$.

In order to establish the desired, it suffices that for all arbitrarily small $\delta > 0$, w.p.a. 1,

$$|\inf_{x>0} F_n(x) - \inf_{x>0} F(x)| < \delta. \tag{B.52}$$

Fix some $0 < \epsilon < \min\{\epsilon_1, \delta\}$ such that (B.51) holds, and, also some

$$0 < \epsilon' < \min\{(F(x_0) - F(x_\epsilon))/4, \delta/4, \delta - \epsilon\}. \tag{B.53}$$

Let $K = [a, b] \subset (0, \infty)$ be compact subset such that $a < x_\epsilon < x_0 \leq b$ and $a = \frac{\delta - 2\epsilon'}{2\delta - \epsilon'} x_\epsilon$. The functions $\{F_n\}$ are convex and they converge point wise to F in the open set $(0, \infty)$. This implies uniform convergence in compact sets by the Convexity lemma [AG82, Cor.. II.1] ,[LM08, Lem. 7.75]. That is, there exists sufficiently large N_1 such that the event

$$\sup_{x \in K} |F_n(x) - F(x)| < \epsilon' \tag{B.54}$$

occurs w.p.a. 1, for all $n > N_1$. In this event,

$$\inf_{x>0} F_n(x) \leq F_n(x_\epsilon) < F(x_\epsilon) + \epsilon' \leq \inf_{x>0} F(x) + \epsilon + \epsilon' \leq \inf_{x>0} F(x) + \delta.$$

It remains to prove the other side of (B.52). In what follows, take $n \geq N_1$ and condition on the high probability event in (B.54).

Let us first show level-boundedness of F_n. Consider the event $\inf_{x>x_0} F_n(x) < \inf_{x \leq x_0} F_n(x)$. If this happens, then, $\inf_{x>x_0} F_n(x) < F_n(x_\epsilon)$, in which case there exists (by continuity of F_n), $x_n > x_0$ such that $F_n(x_n) < F_n(x_\epsilon)$. But then, convexity implies that for some $0 < \theta_n < 1$,

$$F_n(x_0) \leq \theta_n F_n(x_n) + (1 - \theta_n) F_n(x_\epsilon) < F_n(x_\epsilon) \leq F(x_\epsilon) + \epsilon' < F(x_0) - \epsilon', \tag{B.55}$$

where we also used (B.54) and (B.53). Of course, this contradicts (B.54). Thus,

$$\inf_{x>0} F_n(x) = \inf_{x \leq x_0} F_n(x). \tag{B.56}$$

Using (B.56), convexity and properness of $\{F_n\}$, it can be shown that $\inf_{x>0} F_n(x) > -\infty$. The argument is the same as the one used in the beginning of the proof for F, thus is omitted for brevity.

Overall, for all $n > N_1$, conditioned on (B.54), there is some $0 < x_n \leq x_0$ such that

$$\inf_{x>0} F_n(x) \geq F_n(x_n) - \epsilon'. \tag{B.57}$$

If $a \leq x_n \leq b$, then a direct application of (B.54) gives the desired

$$F_n(x_n) \geq F(x_n) - \epsilon' \geq \inf_{x>0} F(x) - \epsilon' \Rightarrow \inf_{x>0} F_n(x) \geq \inf_{x>0} F(x) - 2\epsilon' \geq \inf_{x>0} F(x) - \delta.$$

Next, assume that $0 < x_n < a$. There exists $\theta_n \in (0, 1)$ such that $\theta_n x_n + (1 - \theta_n) x_\epsilon = a$. In fact,

$$\theta_n = \frac{x_\epsilon - a}{x_\epsilon - x_n} \geq (1 - a/x_\epsilon) = \frac{\delta - 2\epsilon'}{2\delta - \epsilon'}. \tag{B.58}$$

Then, by convexity and (B.54), $F_n(a) \leq \theta_n F_n(x_n) + (1-\theta_n)F_n(x_\epsilon)$. Rearranging and using (B.54)

$$\begin{aligned} F_n(x_n) &\geq \frac{1}{\theta_n}F_n(a) - \frac{1-\theta_n}{\theta_n}F_n(x_\epsilon) \\ &\geq \frac{1}{\theta_n}(F(a)-\epsilon') - \frac{1-\theta_n}{\theta_n}(F(x_\epsilon)+\epsilon') \\ &\geq \frac{1}{\theta_n}\left(\inf_{x>0} F(x) - \epsilon\right) - \frac{1-\theta_n}{\theta_n}\left(\inf_{x>0} F(x) + \delta\right). \end{aligned}$$

Combining this with (B.57) and (B.58) yields the desired $\inf_{x>0} F_n(x_n) \geq \inf_{x>0} F(x) - \delta$. □

Lemma B.2.2. *(Level-bounded convex fcns) Let $F : (0,\infty) \to \mathbb{R}$ be convex. Then, the following two statements are equivalent:*

(a) There exists $z > 0$ such that $F(x) > \inf_{x>0} F(x)$ for all $x \geq z$.

(b) $\lim_{x\to\infty} F(x) = +\infty$.

Proof. (a)⇒(b): Clearly, there exists $0 < x_0 < z$, such that $F(z) > F(x_0)$. Then, by convexity, for all $x > z$ it holds

$$F(x) \geq F(z) + \underbrace{\frac{F(z)-F(x_0)}{z-x_0}}_{>0}(x-z).$$

Taking limits of $x \to \infty$ on both sides above proves the claim.

(a)⇐(b): As a proper function, F has a nonempty domain in $(0,\infty)$. Hence, $\inf_{x>0} F(x) < \infty$ and can choose some $M > \inf_{x>0} F(x)$. From (b), there exists $z > 0$ such that $F(x) \geq M$ for all $x \geq z$, as desired. □

Lemma B.2.3 (Saddle-points)**.** *For a convex-concave function $F : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, consider the minimax optimization $\inf_x \sup_y F(x,y)$. Let C, D be compact subsets such that there exists at least one saddle point $(x_*, y_*) \in C \times D$. Then,*

$$\inf_x \sup_y F(x,y) = \inf_{x\in C} \sup_{y\in D} F(x,y).$$

Proof. First observe that

$$\inf_x \sup_y F(x,y) = \inf_{x\in C} \sup_{y} F(x,y).$$

Since F has a saddle-point, the LHS above is equal to $\sup_y \inf_x F(x,y)$ [Roc97, Lem. 36.2]. Also, from Sion's minimax theorem, the RHS is equal to $\sup_y \inf_{x\in C} F(x,y)$. Thus, it suffices to prove that

$$\sup_{y} \inf_{x\in C} F(x,y) = \sup_{y\in D} \inf_{x\in C} F(x,y).$$

Clearly, this holds with a "$\geq$" sign. To prove equality, let (x_*, y_*) be a saddle point. Then,

$$\sup_{y} \inf_{x\in C} F(x,y) = \inf_{x\in C} \sup_{y} f(x,y) \leq \sup_{y} f(x_*,y) \leq f(x_*,y_*) = \sup_{y\in D} \inf_{x\in C} F(x,y).$$

□

Lemma B.2.4. *The function* $h(\alpha,\tau,\mathbf{v}) = \frac{1}{2\tau}\|\alpha\mathbf{x}+\mathbf{z}-\mathbf{v}\|_2^2$ *is jointly convex in its arguments.*

Proof. The function $\|\alpha\mathbf{x}-\mathbf{v}\|_2^2$ is trivially jointly convex in α and $\mathbf{v}$. So its perspective function, which is $\frac{1}{\tau}\|\alpha\mathbf{x}-\mathbf{v}\|_2^2$, is also jointly convex in all its arguments, same as its shifted version which is $h(\alpha,\tau,\mathbf{v})$. □

Lemma B.2.5. *Let* $f:\mathbb{R}^n\to\mathbb{R}$ *be convex. Then,*

(i) $\operatorname{prox}_f(\mathbf{x};\tau) + \tau\cdot\operatorname{prox}_{f^*}\left(\mathbf{x}/\tau;\tau^{-1}\right) = \mathbf{x}$,

(ii) $\mathrm{e}_f(\mathbf{x};\tau) + \mathrm{e}_{f^*}(\mathbf{x}/\tau;1/\tau) = \frac{\|\mathbf{x}\|^2}{2\tau}$.

B.3 Proofs for Separable M-Estimators Satisfying Assumptions 4.2.1 and 4.2.2 Proof of Lemma 4.3.1

Recall,

$$\ell'_+(v) = \max_{s\in\partial\ell(v)} |s|,$$

and that (4.9) gives for all $c\in\mathbb{R}$,

$$\mathbb{E}|\ell'_+(cG+Z)| < \infty \quad \text{and} \quad \mathbb{E}|\ell'_+(cG+Z)|^2 < \infty. \tag{B.59}$$

We make repeated use of Lemma B.4.1 on properties of the Moreau envelope function.

- First, we show that

$$\mathbb{E}\left[\left|\frac{\partial \mathrm{e}_\ell\left(\alpha G+Z;\tau\right)}{\partial \tau}\right|\right]<\infty, \qquad \text{for all } \alpha\in\mathbb{R}, \tau>0. \tag{B.60}$$

From (B.88) $\left|\frac{\partial \mathrm{e}_\ell(\alpha G+Z;\tau)}{\partial \tau}\right| \leq |\ell'_+(\mathrm{prox}_\ell\left(\alpha G+Z;\tau\right))|^2$. Lemma B.4.1(viii) shows that this is no larger than $|\ell'_+(\alpha G+Z)|^2$. Then, (B.60) follows from (B.59).

- It is also useful to prove

$$\mathbb{E}\left[|\ell(\alpha G+Z)-\ell(Z)|\right]<\infty, \qquad \text{for all } \alpha\in\mathbb{R}. \tag{B.61}$$

From convexity of ℓ,

$$|\ell(\alpha G+Z)-\ell(Z)| \leq \max\{|\ell'_+(\alpha G+Z)|, |\ell'_+(Z)|\}\cdot|\alpha G| \leq \left(|\ell'_+(\alpha G+Z)|+|\ell'_+(Z)|\right)\cdot|\alpha G|,$$

and the desired follows by taking expectations and applying (B.59) for $c=\alpha$ and $c=0$.

- Let us now show

$$\mathbb{E}\left[|\mathrm{e}_\ell\left(\alpha G+Z;\tau\right)-\ell(Z)|\right]<\infty, \qquad \text{for all } \alpha\in\mathbb{R}, \tau>0 \tag{B.62}$$

We have, $|\mathrm{e}_\ell\left(\alpha G+Z;\tau\right)-\ell(Z)| \leq |\mathrm{e}_\ell\left(\alpha G+Z;\tau\right)-\ell(\alpha G+Z)|+|\ell(\alpha G+Z)-\ell(Z)|$. In view of (B.61), it suffices for (B.62) to show integrability of the first term. We argue as follows:

$$\begin{aligned}
|\mathrm{e}_\ell\left(\alpha G+Z;\tau\right)-\ell(\alpha G+Z)| &= \lim_{\rho\to 0}|\mathrm{e}_\ell\left(\alpha G+Z;\tau\right)-\mathrm{e}_\ell\left(\alpha G+Z;\rho\right)| \\
&= \lim_{\rho\to 0}\left|\frac{\partial \mathrm{e}_\ell\left(\alpha G+Z;\tau\right)}{\partial\tau}|_{\tau=\xi(\rho)}\right|\cdot|\tau-\rho|.
\end{aligned}$$

It remains to take expectations of both sides and apply the argument below (B.60) to yield (B.62).

- Assumption 4.2.1(a). We have $\frac{1}{m}\{\mathrm{e}_{\mathcal{L}}\left(\alpha\mathbf{g}+\mathbf{z};\tau\right)-\mathcal{L}(\mathbf{z})\} = \frac{1}{m}\sum_{j=1}^m\left(\mathrm{e}_\ell\left(\alpha\mathbf{g}_j+\mathbf{z}_j;\tau\right)-\ell(\mathbf{z}_j)\right)$. Then from the WLLN (e.g. [Dur10, Thm. 2.2.9]) the expression above converges in probability to

$$L\left(\alpha,\tau\right)=\mathbb{E}\left[\mathrm{e}_\ell\left(\alpha G+Z;\tau\right)-\ell(Z)\right], \tag{B.63}$$

where we have also used (B.62) to verify integrability.

• Continuity and convexity of L. The Moreau envelope function is convex in its arguments (see Lemma B.4.1(ii)). Convexity is preserved under affine transformations and nonnegative weighted sums; thus, $L(\alpha, \tau)$ is jointly convex in α, τ. Continuity then follows as a consequence of convexity [Roc97, Thm. 10.1].

• Assumption 4.2.2(c). To compute $\lim_{\tau\to+\infty} \mathbb{E}\left[\mathrm{e}_\ell\left(\alpha G + Z; \tau\right) - \ell(Z)\right]$, we first apply the Dominated Convergence Theorem to pass the limit inside the expectation. This is justified since (B.62) shows integrability, and the limit exists as follows (see Lemma B.4.1(vii)):

$$\lim_{\tau\to+\infty} \mathrm{e}_\ell\left(\alpha G + Z; \tau\right) = \min_v \ell(v) = 0,$$

for all $\alpha, \tau > 0$. Taking expectation of this shows $\lim_{\tau\to+\infty} L(\alpha, \tau) = -L_0$, where $L_0 = \mathbb{E}[\ell(Z)]$ by the WLLN. Also, we need to show that if $\mathbb{E}[\ell(Z)] < \infty$, then $\mathbb{E}\left[\mathrm{e}_\ell\left(\alpha G + Z; \tau\right)\right] \geq 0$. This follows easily since $\mathrm{e}_\ell\left(\alpha G + Z; \tau\right) \geq \min_v \ell(v) = 0$. Finally, the property $L(\alpha, \tau) \geq \lim_{\tau\to\infty} L(\alpha, \tau)$ follows by the non increasing nature of e_ℓ with respect to τ (cf. B.4.1(v)).

• Assumption 4.2.2(d). If $\lim_{\tau\to+\infty} L(\alpha, \tau) < \infty$, the claim is immediate. Otherwise, we apply de l'Hopital rule and (B.68) to get

$$\lim_{\tau\to\infty} \frac{\mathbb{E}[\mathrm{e}_\ell\left(\alpha G + Z; \tau\right) - \ell(Z)]}{\tau} = \lim_{\tau\to\infty} \frac{\partial}{\partial\tau} \mathbb{E}[\mathrm{e}_\ell\left(\alpha G + Z; \tau\right) - \ell(Z)].$$

An application of the Dominated Convergence Theorem in Lemma B.3.1(i) shows that we can interchange the order of differentiation and expectation above. We will prove that

$$\lim_{\tau\to\infty} \frac{\partial}{\partial\tau} (\mathrm{e}_\ell\left(\alpha G + Z; \tau\right) - \ell(Z)) = 0 \tag{B.64}$$

for all G and Z. Then, we can also utilize Dominated Convergence Theorem to pass the limit in the expectation and conclude with the desired.

From standard properties of the Moreau envelopes (cf. (B.88)),

$$\frac{\partial}{\partial\tau} (\mathrm{e}_\ell\left(\alpha G + Z; \tau\right) - \ell(Z)) = \frac{1}{-2\tau^2} (\alpha G + Z - \mathrm{prox}_\ell\left(\alpha G + Z; \tau\right))^2.$$

Thus, it suffices to prove $\lim_{\tau\to\infty} \frac{1}{\tau}(x - \mathrm{prox}_\ell\left(x; \tau\right)) = 0$ for all x. This is shown in Lemma B.4.1(vii).

• Assumption 4.2.2(b). We apply the Dominated Convergence Theorem to compute $\lim_{\tau\to 0^+} \mathbb{E}[\mathrm{e}_\ell\left(\alpha G + Z; \tau\right) - \ell(Z)]$ and exchange limit and expectation. Then,

because $\lim_{\tau\to 0^+} \mathrm{e}_\ell\left(\alpha G + Z;\tau\right) = \ell(\alpha G + Z)$ we have

$$\lim_{\tau\to 0^+} \mathbb{E}\left[\mathrm{e}_\ell\left(\alpha G + Z;\tau\right) - \ell(Z)\right] = \mathbb{E}\left[\lim_{\tau\to 0^+} \mathrm{e}_\ell\left(\alpha G + Z;\tau\right) - \ell(Z)\right] = \mathbb{E}\left[\ell(\alpha G + Z) - \ell(Z)\right] < \infty.$$

Boundedness follows from (B.62). The same argument shows that

$$\lim_{\tau\to 0^+} L(0,\tau) = \lim_{\tau\to 0^+} \mathbb{E}\left[\mathrm{e}_\ell\left(Z;\tau\right) - \ell(Z)\right] = \mathbb{E}\left[\lim_{\tau\to 0^+} \mathrm{e}_\ell\left(Z;\tau\right) - \ell(Z)\right] = \mathbb{E}\left[\ell(Z) - \ell(Z)\right] = 0.$$

Finally, to compute $\lim_{\tau\to 0^+} L_2(0,\tau)$, we apply the Dominated Convergence Theorem twice as was done for the proof of Assumption 4.2.2(d). With this we have,

$$\lim_{\tau\to 0^+} L_2(0,\tau) = \mathbb{E}\left[\lim_{\tau\to 0^+} \frac{\partial}{\partial\tau}\left(\mathrm{e}_\ell\left(\alpha G + Z;\tau\right) - \ell(Z)\right)\Big|_{\alpha=0}\right] = -\frac{1}{2}\mathbb{E}\left[\lim_{\tau\to 0^+} \left(\ell'_{\mathrm{prox}_\ell(Z;\tau),\tau}\right)^2\right] \le 0.$$

The second equality above follows by Lemma B.4.1(iii) (please see (B.86) for the notation $\ell'_{\chi,\tau}$). Besides, due to lemma B.4.1(viii), $(\ell'_{\mathrm{prox}_\ell(Z;\tau),\tau})^2 \le (\ell'_+(Z))^2$ which implies

$$-\mathbb{E}\left[\lim_{\tau\to 0^+} \left(\ell'_{\mathrm{prox}_\ell(Z;\tau),\tau}\right)^2\right] \ge -\mathbb{E}\left[\lim_{\tau\to 0^+} (\ell'_+(Z))^2\right] = -\mathbb{E}\left[(\ell'_+(Z))^2\right] > -\infty.$$

Boundedness follows by (B.59).

Proof of Lemma 4.3.2

- Assumption 4.2.1(a). Assumption 4.2.1(a) is satisfied for $F(c,\tau) = \mathbb{E}\left[\mathrm{e}_f\left(cH + X_0\right);\tau) - f(X_0)\right]$. The proof is exactly the same as in Lemma 4.3.1.

- $\lim_{\tau\to 0^+} F(\tau,\tau) = 0$. This will follow from continuity of the Moreau envelope. In particular, using Lemma B.4.1(ix), we find that for all H, X_0:

$$\lim_{\tau\to 0} \mathrm{e}_f\left(\tau H + X_0;\tau\right) = f(X_0).$$

Then, the desired claim follows from this and an application of the Dominated Convergence Theorem.

- $\lim_{c\to\infty} \frac{c^2}{2\tau} - \mathbb{E}[\mathrm{e}_f(cH + X_0;\tau) - f(X_0)] = \infty$. We have

$$\begin{aligned}
&\frac{c^2}{2\tau} - \mathbb{E}[\mathrm{e}_f(cH + X_0;\tau) - f(X_0)] = \mathbb{E}[\frac{(cH + X_0)^2}{2\tau} - \mathrm{e}_f(cH + X_0;\tau)] + \mathbb{E}[f(X_0) - X_0^2] \\
&= \mathbb{E}[\mathrm{e}_{f^*}((cH + X_0)/\tau; 1/\tau)] + \mathbb{E}[f(X_0) - X_0^2] \\
&= \frac{1}{2}\mathbb{E}[\mathrm{e}_{f^*}((cH + X_0)/\tau; 1/\tau)\big|H > 0] + \frac{1}{2}\mathbb{E}[\mathrm{e}_{f^*}((cH + X_0)/\tau; 1/\tau)\big|H < 0] + \mathbb{E}[f(X_0) - X_0^2] \\
&\geq \frac{1}{2}\mathrm{e}_{f^*}\left(\mathbb{E}[(cH + X_0)/\tau\big|H > 0]; 1/\tau\right) + \frac{1}{2}\mathrm{e}_{f^*}\left(\mathbb{E}[(cH + X_0)/\tau\big|H < 0]; 1/\tau\right) + \mathbb{E}[f(X_0) - X_0^2] \\
&= \frac{1}{2}\mathrm{e}_{f^*}\left(c/\tau\sqrt{\frac{2}{\pi}} + \mathbb{E}[X_0]; 1/\tau\right) + \frac{1}{2}\mathrm{e}_{f^*}\left(-c/\tau\sqrt{\frac{2}{\pi}} + \mathbb{E}[X_0]; 1/\tau\right) + \mathbb{E}[f(X_0) - X_0^2].
\end{aligned} \tag{B.65}$$

The second equality above follows from Lemma B.2.5. For the inequality, $\mathrm{e}_{f^*}(c;\tau)$ is convex in c, thus it follows from Jensen's inequality. From (B.65), observing that $\mathbb{E}[f(X_0) - X_0^2] > -\infty$ and $|\mathbb{E}[X_0]| < \infty$ by boundedness of $\mathbb{E}[X_0^2]$ and non-negativity of f, it suffices to show that

$$\lim_{|c|\to\infty} \mathrm{e}_{f^*}(c/\tau; 1/\tau) = \infty.$$

First, assume that $f(x)$ is defined for some positive value $a > 0$ and $f(a) < \infty$, then

$$\forall M, \quad \forall x > X_M = \frac{f(a)}{a} + \frac{M}{a} : \quad f^*(x) = \max_y xy - f(y) \geq ax - f(a) > M. \tag{B.66}$$

Which means that $\lim_{x\to\infty} f^*(x) = \infty$. Now in order to show that the limit in (B.65) goes to infinity we prove that

$$\begin{aligned}
\forall M \quad \forall x > \tau(X_M + \sqrt{2M/\tau}) \quad \forall v, \quad & \frac{\tau}{2}(x/\tau - v)^2 + f^*(v) > M. \\
\iff \forall u \quad & \frac{\tau}{2}u^2 + f^*(u + x/\tau) > M.
\end{aligned} \tag{B.67}$$

This is easy to show. For the cases that $|u| > \sqrt{2M/\tau}$ we have

$$\frac{\tau}{2}u^2 + f^*(u + x/\tau) > M + f^*(u + x/\tau) \geq M.$$

Note that $f(0) = 0$ implies $f^*(x) \geq 0$ for all x. On the other hand, for the cases that $|u| \leq \sqrt{2M/\tau}$,

$$x/\tau + u \geq x/\tau - |u| \geq x/\tau - \sqrt{2M/\tau} > X_M.$$

Thus due to (B.66),

$$\frac{\tau}{2}u^2 + f^*(u + x/\tau) > \frac{\tau}{2}u^2 + M \geq M,$$

which shows that $\lim_{c\to\infty} \mathrm{e}_{f^*}(c/\tau; 1/\tau) = \infty$.
On the other hand, if $f(x)$ is also defined for some negative value $a < 0$ and $f(a) < \infty$, the same set of arguments proves that $\lim_{c\to-\infty} f^*(c) = \infty$ and also $\lim_{c\to-\infty} \mathrm{e}_{f^*}(c/\tau; 1/\tau) = \infty$.

Strict Convexity of the Expected Moreau Envelope

In this section, we prove Lemmas 4.3.3 and 4.3.4. We have combined the statements in Lemma B.3.1 below.

Lemma B.3.1 (Lemmas 4.3.3 and 4.3.4)**.** *Let $\ell : \mathbb{R} \to \mathbb{R}$ be a proper, closed, convex function, $G \sim \mathcal{N}(0,1)$ and $Z \sim p_z$ such that (4.9) holds. The function $L : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$:*

$$L(\alpha, \tau) := \mathbb{E}_{G,Z}\left[\mathrm{e}_\ell(\alpha G + Z; \tau) - \ell(Z)\right]$$

has the following properties:

(i) It is differentiable with

$$\frac{\partial L}{\partial \alpha} = \mathbb{E}\left[\frac{\partial \mathrm{e}_\ell(\alpha G + Z; \tau)}{\partial \alpha}\right] \quad \text{and} \quad \frac{\partial L}{\partial \tau} = \mathbb{E}\left[\frac{\partial \mathrm{e}_\ell(\alpha G + Z; \tau)}{\partial \tau}\right]. \tag{B.68}$$

(ii) If the conditions (a) and (b) of Lemma 4.3.4 also hold, then it is jointly strictly convex in $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$.

(iii) If $\ell(x) \geq \ell(0) = 0$ and $\ell(x_+) > 0$ for some $x_+ > 0$, then, the function $F(\alpha) := \lim_{\tau\to 0^+} L(\alpha, \tau) = \mathbb{E}\left[\ell(\alpha G + Z) - \ell(Z)\right]$ is strictly convex in $\alpha > 0$.

Proof. We make repeated use of the properties of the Moreau envelope function as listed in Lemma B.4.1. Also, we use the same notation as in that lemma; in particular, recall (B.86), (B.87) and (B.88). For ease of reference we summarize the notation used throughout this section below:

$$\hat{v}_{\chi,\tau} := \mathrm{prox}_\ell(\chi; \tau), \quad \ell'_{\chi,\tau} := \frac{1}{\tau}(\chi - \hat{v}_{\chi,\tau}),$$

$$E_1(\alpha, \tau) := \frac{\partial \mathrm{e}_\ell(\alpha G + Z; \tau)}{\partial \alpha}, \quad E_2(\alpha, \tau) := \frac{\partial \mathrm{e}_\ell(\alpha G + Z; \tau)}{\partial \tau}.$$

(i): The claim follows by the Dominated Convergence Theorem, since the following hold:

- $\mathrm{e}_\ell(\alpha G + Z; \tau)$ is continuously differentiable with respect to both α and τ (cf. Lemma B.4.1(iii)),
- In Section B.3 (see (B.62)) we use (4.9) to show that $\mathbb{E}[|\mathrm{e}_\ell(\alpha G + Z; \tau) - \ell(Z)|] < \infty$ for all α and $\tau > 0$.
- For all $\alpha \in \mathbb{R}$ and $\tau > 0$:

$$\begin{aligned}\mathbb{E}\left[|E_1(\alpha,\tau)|\right] &= \frac{1}{\tau}\mathbb{E}\left[\left|\alpha G + Z - \mathrm{prox}_\ell(\alpha G + Z;\tau)\right| \cdot |G|\right] \\ &\le \frac{1}{\tau}\sqrt{\mathbb{E}\left[\left|\alpha G + Z - \mathrm{prox}_\ell(\alpha G + Z;\tau)\right|^2\right]} = \sqrt{\mathbb{E}\left[|E_2(\alpha,\tau)|\right]},\end{aligned}$$

  where we have used Lemma B.4.1(iii), the Cauchy-Schwarz inequality. In Section B.3 (see (B.60)) we use (4.9) to show that $\mathbb{E}\left[|E_2(\alpha,\tau)|\right] < \infty$.

(ii): For any $\alpha > 0, \tau > 0$, it suffices to show that

$$\Gamma(x,y) := L(\alpha + x, \tau + y) - L(\alpha,\tau) - L_1(\alpha,\tau)x - L_2(\alpha,\tau)y > 0, \quad \text{for all } x \in \mathbb{R}, y > -\tau, \tag{B.69}$$

where we use numerical subscript notation to denote derivation with respect to the corresponding argument, i.e. $L_1 = \partial L/\partial\alpha$ and $L_2 = \partial L/\partial\tau$.

Observe that $\Gamma(x,y)$ defined in (B.69) is differentiable; denote its partial derivatives with respect to x and y as Γ_1 and Γ_2, respectvely. Furthermore, Γ is jointly convex in (x,y) (see Lemma B.4.1(ii)) and $\Gamma(0,0) = 0$. Thus, it suffices for (B.69) to prove strict positivity of the following expression:

$$\begin{aligned}\Gamma_1(x,y)x + \Gamma_2(x,y)y &= (L_1(\alpha + x, \tau + y) - L_1(\alpha,\tau))x + (L_2(\alpha + x, \tau + y) - L_2(\alpha,\tau))y \\ &= \mathbb{E}\left[(E_1(\alpha + x, \tau + y) - E_1(\alpha,\tau))x + (E_2(\alpha + x, \tau + y) - E_2(\alpha,\tau))y\right].\end{aligned}$$

In the last equality above we have interchanged the order of expectation and differentiation. Lemma B.4.1(iv) lower bounds the expression inside the expectation above. To be specific, using (B.89), we find that

$$\Gamma_1(x,y)x + \Gamma_2(x,y)y \ge \left(\tau + \frac{y}{2}\right)\mathbb{E}\left[(\ell'_{\alpha+x,\tau+y} - \ell'_{\alpha,\tau})^2\right].$$

Therefore, it will suffice for our purposes to show that for any fixed x, y,

$$\mathbb{E}\left[\left(\ell'_{\alpha G+Z,\tau} - \ell'_{(\alpha+x)G+Z,\tau+y}\right)^2\right] > 0. \tag{B.70}$$

For this it is enough to prove the existence of (G_*, Z_*) with $p(Z_*) > 0$ such that

$$\ell'_{\alpha G_* + Z_*, \tau} \neq \ell'_{(\alpha+x)G_* + Z_*, \tau+y}. \tag{B.71}$$

Indeed, if this is the case, by continuity of the mapping $(G, Z) \to \alpha G + Z$ and of the prox operator (cf. Lemma B.4.1(i)) there exists an open neighborhood $\mathcal{N}$ around (G_*, Z_*) such that $\ell'_{\alpha G + Z, \tau} \neq \ell'_{(\alpha+x)G + Z, \tau+y}$ for all $(G, Z) \in \mathcal{N}$. Furthermore, there exists subset $\mathcal{J}_1 \times \mathcal{J}_2 \subseteq \mathcal{N}$ of nonzero measure such that: (i) $\mathcal{J}_1$ is a closed interval with $p(G) > 0$ for all $G \in \mathcal{J}_1$, (ii) if Z has a point mass at Z_*, then $\mathcal{J}_2 = Z_*$; otherwise, $\mathcal{J}_2$ is a closed interval with $p(Z) > 0$ for all $Z \in \mathcal{J}_2$. In all cases, $\mathcal{J}_1 \times \mathcal{J}_2$ is a set of nonzero measure, with which we conclude (B.70) as desired. In what follows, we prove (B.71).

Case 1: Assume that there exists an open interval $\mathcal{I}$ on which ℓ is differentiable with strictly increasing derivative:

$$\ell'(v_1) < \ell'(v_2), \qquad \text{for all } v_1 < v_2 \in \mathcal{I}. \tag{B.72}$$

In particular, since ℓ is convex in its entire domain it further holds that

$$v_1 \in \mathcal{I}, v_2 \neq v_1 \Rightarrow \ell'(v_1) \neq \ell'(v_2). \tag{B.73}$$

Consider the set

$$\mathcal{S} := \{(G, Z) \mid \hat{v}_{\alpha G + Z, \tau} \in \mathcal{I}\}. \tag{B.74}$$

Clearly, $\mathcal{S}$ is a nonempty open set (by continuity of the prox operator). Next, we show that there exists $(G_*, Z_*) \in \mathcal{S}$, such that

$$\hat{v}_{\alpha G_* + Z_*, \tau} \neq \hat{v}_{(\alpha+x)G_* + Z_*, \tau+y} \tag{B.75}$$

and $p(Z_*) > 0$. This suffices for proving (B.71), since when combined with $\hat{v}_{\alpha G_* + Z_*, \tau} \in \mathcal{I}$ and (B.73) it implies that $\ell'_{\alpha G_* + Z_*, \tau} \neq \ell'_{(\alpha+x)G_* + Z_*, \tau+y}$.

Choose any two distinct $Z_i, i = 0, 1$ with $p(Z_i) > 0$. This is possible since by assumption $\mathrm{Var}[Z] \neq 0$. Denote, $\mathcal{S}_{Z_i} := \{G \mid (G, Z_i) \in \mathcal{S}\}$. Clearly, $\mathcal{S}_{Z_i}$ are nonempty open sets. If there exists $G_i \in \mathcal{S}_{Z_i}$ such that $(G_*, Z_*) = (G_i, Z_i)$ satisfies (B.75), there is nothing else to prove.

Otherwise, we would have $\hat{v}_{\alpha G + Z_i, \tau} = \hat{v}_{(\alpha+x)G + Z_i, \tau+y} \in \mathcal{I}$ and consequently $\ell'_{\alpha G + Z_i, \tau} = \ell'_{(\alpha+x)G + Z_i, \tau+y}$, for all $G \in \mathcal{S}_{Z_i}$ and $i = 0, 1$. But, Lemma B.3.2 below proves that this cannot happen under our assumptions on the sets $\mathcal{S}_{Z_i}$.

Case 2: Let v_0 be a point where ℓ is not differentiable and consider $\mathcal{I} \subset \partial\ell(v_0)$ a non-empty open subset of the subdifferential of ℓ at v_0. Further consider the nonempty open sets

$$\mathcal{S} := \{(G, Z) \mid \ell'_{\alpha G+Z,\tau} \in \mathcal{I}\} \quad \text{and} \quad \tilde{\mathcal{S}} := \{(G, Z) \mid \ell'_{(\alpha+x)G+Z,\tau+y} \in \mathcal{I}\}. \tag{B.76}$$

Clearly, $\hat{v}_{\alpha G+Z,\tau} = v_0$ for all $(G, Z) \in \mathcal{S}$ and similar for $\tilde{\mathcal{S}}$. Choose any two distinct $Z_i, i = 1, 2$ with $p(Z_i) > 0$. This is possible since by assumption $\mathrm{Var}[Z] \neq 0$. Denote, $\mathcal{S}_{Z_i} := \{G \mid (G, Z_i) \in \mathcal{S}\}$ and $\tilde{\mathcal{S}}_{Z_i} := \{G \mid (G, Z_i) \in \tilde{\mathcal{S}}\}$, which are all nonempty sets. Consider,

$$\mathcal{N}_{Z_i} = \mathcal{S}_{Z_i} \setminus \tilde{\mathcal{S}}_{Z_i}, i = 0, 1.$$

If (say) $\mathcal{N}_{Z_0} \neq \emptyset$, then for any $G_0 \in \mathcal{N}_{Z_0}$, it holds

$$\hat{v}_{(\alpha+x)G_0+Z_0,\tau+y} \neq \hat{v}_{\alpha G_0+Z_0,\tau} \in \mathcal{I} \Rightarrow \ell'_{(\alpha+x)G_0+Z_0,\tau+y} \neq \ell'_{\alpha G_0+Z_0,\tau},$$

where the last implication follows because of monotonicity of the subdifferential. This shows (B.71) as desired.

Otherwise, $\mathcal{N}_i = \emptyset \Rightarrow$ for $i = 0, 1$. In case there exists $G_i \in \mathcal{S}_{Z_i}$ such that $(G_*, Z_*) = (G_i, Z_i)$ satisfies (B.71), there is nothing else to prove. If this was not the case, then we would have $\ell'_{\alpha G_i+Z_i,\tau} = \ell'_{(\alpha+x)G_i+Z_i,\tau+y} \in \mathcal{I}$ and $\hat{v}_{\alpha G_i+Z_i,\tau} = \hat{v}_{(\alpha+x)G_i+Z_i,\tau+y} = v_0$, for all $G \in \mathcal{S}_{Z_i}$. But, Lemma B.3.2 below proves that this cannot happen under our assumptions on the sets $\mathcal{S}_{Z_i}$.

(iii): Suppose that the statement of the lemma is false. Then, there exist $\alpha_1 \neq \alpha_2 > 0$, and, $\alpha_\theta := \theta\alpha_1 + (1-\theta)\alpha_2$ for $\theta \in (0, 1)$ such that $F(\theta\alpha_1 + (1-\theta)\alpha_2) = \theta F(\alpha_1) + (1-\theta)F(\alpha_2)$, or,

$$\mathbb{E}\left[\theta\ell(\alpha_1 G + Z) + (1-\theta)\ell(\alpha_2 G + Z) - \ell(\alpha_\theta G + Z)\right] = 0 \tag{B.77}$$

The convexity of ℓ ensures that, for each $\alpha G + Z$, the argument in the expectation is nonnegative. Therefore, the relation above holds if and only if the argument under the expectation is zero almost surely with respect to the distribution of $\alpha G + Z$. Next, we prove that this leads to a contradiction.

Let x_+ be as in the statement of the lemma, and $x_0 = \max\{x \in [0, x_+] \mid \ell(x) = 0\} < x_+$. For some $\epsilon > 0$ to be specified later in the proof, let $x_1 = x_0 + \epsilon$. Note that $\ell(x_1) > 0$ by definition of x_0 and by convexity. Without loss of generality

assume $\alpha_1 > \alpha_2$. Fix Z_0 such that $p(Z_0) > 0$ and $x_0 \neq Z_0$ (always possible since $\mathrm{Var}[Z] \neq 0$). Consider two cases based on the sign of $x_0 - Z_0$.

$x_0 > Z_0$: Define $G_0 = (x_1 - Z_0)/\alpha_1 > 0$. Note that $\alpha_1 G_0 + Z_0 = x_1$ and call $x_2 := \alpha_2 G_0 + Z_0$. Choose $\epsilon = (\frac{\alpha_1}{\alpha_2} - 1)(x_0 - Z_0)/2 > 0$. Then, it is not hard to check that $x_2 < x_0 < x_1$; thus, for some $\theta \in (0, 1)$, $\alpha_\theta G_0 + Z_0 = x_0$. But $\theta\ell(x_1) + (1 - \theta)\ell(x_2) > 0 = \ell(x_0)$ or

$$\ell(\alpha_\theta G_0 + Z_0) < \theta\ell(\alpha_1 G_0 + Z_0) + (1 - \theta)\ell(\alpha_2 G_0 + Z_0). \tag{B.78}$$

There exists an open ball (of non-zero measure) around $\alpha G_0 + Z_0$, where the same relation as above holds. This contradicts (B.77) and concludes the proof.

$x_0 < Z_0$: Define $G_0 := x_1 - Z_0/\alpha_2 > 0$. Note that $\alpha_2 G_0 + Z_0 = x_1$ and call $x_2 := \alpha_1 G_0 + Z_0$. Choose $\epsilon = (\frac{\alpha_2}{\alpha_1} - 1)(x_0 - Z_0)/2 > 0$. Then, it is not hard to check that $x_2 < x_0 < x_1$ and the same argument as above leads to a contradiction of (B.77).

□

Lemma B.3.2 (Auxiliary)**.** *Suppose $Z_0 \neq Z_1$. For some nonempty set $\mathcal{J} \subset \mathbb{R}$ assume that the sets*

$$\mathcal{G}_{Z_i} := \{G \mid \hat{v}_{\alpha G + Z_i, \tau} \in \mathcal{J}\}, \quad i = 0, 1 \tag{B.79}$$

are non-empty and have at least two elements each. Further suppose that for all $G, G' \in \mathcal{G}_i, i = 0, 1$ the following holds

$$\ell'_{\alpha G + Z_i, \tau} = \ell'_{\alpha G' + Z_i, \tau} \Rightarrow \hat{v}_{\alpha G + Z_i, \tau} = \hat{v}_{\alpha G' + Z_i, \tau}. \tag{B.80}$$

Then, it cannot be true that for all $G \in \mathcal{G}_i$ and $i = 0, 1$:

$$\hat{v}_{\alpha G + Z_i, \tau} = \hat{v}_{(\alpha + x)G + Z_i, \tau + y} \quad \text{and} \quad \ell'_{\alpha G + Z_i, \tau} = \ell'_{(\alpha + x)G + Z_i, \tau + y}. \tag{B.81}$$

Proof. Assume to the contrary of the lemma that the sets $\mathcal{G}_0$ and $\mathcal{G}_1$ satisfy (B.81). When combined with optimality conditions (cf. (B.86)), the properties of the sets give

$$y\ell'_{\alpha G + Z_i, \tau} = xG, \quad \text{for all } G \in \mathcal{G}_i, i = 0, 1. \tag{B.82}$$

Consider separately two cases on the possible values of x and y:

• $x = 0, y \neq 0$: Let $G \neq G'$ both belonging in $\mathcal{G}_0$ (such a pair exists since $\mathcal{G}_0$ is open). Starting from (B.82) and using (B.80), we have:

$$\ell'_{\alpha G+Z_0,\tau} = \ell'_{\alpha G'+Z_0,\tau} = 0 \Rightarrow \hat{v}_{\alpha G+Z_0,\tau} = \hat{v}_{\alpha G'+Z_0,\tau}.$$

These equalities when combined with optimality conditions of the prox (cf. (B.86)) yield a contradiction: $G = G'$.

• $x \neq 0$: Let any $G_0 \in \mathcal{G}_0$, and, consider $G_1 := G_0 + \frac{Z_0 - Z_1}{\alpha} \neq G_0$. Note that $\alpha G_1 + Z_1 = \alpha G_0 + Z_0$. Also, by uniqueness of the prox operator, $\hat{v}_{\alpha G_1+Z_1,\tau} = \hat{v}_{\alpha G_0+Z_0,\tau} \in \mathcal{I}$ and $G_1 \in \mathcal{G}_1$. Furthermore, $\ell'_{\alpha G_0+Z_0,\tau} = \ell'_{\alpha G_1+Z_1,\tau}$. Then, combining with (B.82) we reach the following contradiction:

$$xG_0 = xG_1 \Rightarrow G_0 = G_1.$$

□

Strict convexity $\Longrightarrow$ uniqueness of α_*

Lemma B.3.3. *Suppose all assumptions of Theorem 4.3.1 are satisfied. Then,* (4.4) *has a unique optimal minimizer α_*.*

Proof. During the proof, we borrow notation and results from the proof of Lemma B.1.5 in Section B.2. Under the assumption of the theorem, $L(\alpha, \tau)$ is jointly strictly convex in $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$, by Lemma B.3.1. Also, by assumptions, the set of minimizers of F in (B.49) is bounded. With these, we will show that the set of optima actually consists of a unique point. Consider $M^{\alpha,\beta,\tau_h}(\tau_g)$ as in (B.37). We have shown in Section B.2 that M^{α,β,τ_h} is level bounded. Thus, the minimum is either attained at some τ_{g_*} or is achieved in the limit of $\tau_g \to 0$. Now, consider extending the function at $\tau_g = 0$, by setting $L(\alpha, 0) = \lim_{\tau_g \to 0^+} L(\alpha, \tau_g)$. By assumption, this latter is a strictly convex function of α. Hence, similarly extending M^{α,β,τ_h} at $\tau_g = 0$, the function is jointly strictly convex in (α, τ_g) and the minimum over τ_g is now attained (can be $\tau_{g_*} = 0$). Using those two, Lemma B.3.4 shows that $\inf_{\tau_g > 0} M^{\alpha,\beta,\tau_h}(\tau_g)$ is strictly convex in $\alpha > 0$. Next, consider taking the supremum over $\beta \geq 0$. From the results of Section B.2, the optimal β is attained at some value $\beta_* \geq 0$ (in other words, it does not approach infinity). Suppose $\beta_* = 0$, then the optimal α solves

$$\inf_{\alpha \geq 0} \sup_{\tau_h > 0} -\frac{\alpha \tau_h}{2} + \lambda F(0, \alpha\lambda/\tau_h).$$

In Lemma B.3.6 we show that the set of minimizers of this optimization is unbounded. This contradicts our assumption on the boundedness of α_*. Hence, $\beta_* \neq$

0, and we can apply Lemma B.3.5 to find that $M^{\alpha}(\tau_h) := \sup_{\beta} \inf_{\tau_g>0} M^{\alpha,\beta,\tau_h}(\tau_g)$, remains a strictly convex function of $\alpha > 0$. Lastly, maximizing over τ_h does not affect strict convexity since it is not involved in the term $\frac{\beta\tau_g}{2} + \delta L(\alpha, \tau_g/\beta)$. Overall, $F(\alpha) = \sup_{\beta,\tau_h} \inf_{\tau_g} \mathcal{D}(\alpha, \tau_g, \beta, \tau_h)$ is strictly convex in $\alpha > 0$. Using this it is straightfowrard to show that its minimizer over $\alpha \geq 0$ is unique, thus completing the proof. □

Lemma B.3.4. *Let $\mathcal{X}, \mathcal{Y}$ be convex sets and $F(\mathbf{x}, \mathbf{y}) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be jointly strictly convex. If $F(\mathbf{x}, \cdot)$ attains its minimum value in $\mathcal{Y}$ for all $\mathbf{x} \in \mathcal{X}$, then, $G(\mathbf{x}) := \inf_{\mathbf{y}\in\mathcal{Y}} F(\mathbf{x}, \mathbf{y})$ is strictly convex.*

Proof. For $\theta \in (0, 1)$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$, denote $\mathbf{x}_\theta = (1-\theta)\mathbf{x}_1 + \theta\mathbf{x}_2$, $\mathbf{y}_\theta := \arg\inf_{\mathbf{y}\in\mathcal{Y}} F((1-\theta)\mathbf{x}_1 + \theta\mathbf{x}_2, \mathbf{y})$ and $\mathbf{y}_i := \arg\inf_{\mathbf{y}\in\mathcal{Y}} F(\mathbf{x}_i, \mathbf{y})$, $i = 1, 2$. With these

$$\begin{aligned} G(\mathbf{x}_\theta) = F(\mathbf{x}_\theta, \mathbf{y}_\theta) \leq F(\mathbf{x}_\theta, \theta\mathbf{y}_1 + (1-\theta)\mathbf{y}_2) &< (1-\theta)F(\mathbf{x}_1, \mathbf{y}_1) + \theta F(\mathbf{x}_2, \mathbf{y}_2) \\ &= (1-\theta)G(\mathbf{x}_1) + \theta G(\mathbf{x}_2), \end{aligned}$$

where the first inequality follows from definition of $\mathbf{y}_\theta$, the second from the joint strict convexity of F, and, the third by definition of $\mathbf{y}_1, \mathbf{y}_2$. □

Lemma B.3.5. *Consider $F : \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ and $G(x) := \min_{0\leq y\leq K} F(x, y)$. If F is jointly strictly convex in $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$ and $F(x, 0)$ is also strictly convex, then G is strictly convex.*

Proof. Consider $x_1, x_2 > 0$ and $x_\theta = \theta x_1 + (1-\theta)x_2$ for some $\theta \in (0, 1)$. Let y_1, y_2 and y_θ be defined such that $G(x_1) = F(x_1, y_1)$, $G(x_2) = F(x_2, y_2)$, and $G(x_\theta) = F(x_\theta, y_\theta)$. We distinguish four cases. For each one we prove that $G(x_\theta) < \theta G(x_1) + (1-\theta)G(x_2)$, as desired.

$\underline{y_1, y_2 > 0}$:

$$G(x_\theta) \leq F(x_\theta, \theta y_1 + (1-\theta)y_2) < \theta F(x_1, y_1) + (1-\theta)F(x_2, y_2) = \theta G(x_1) + (1-\theta)G(x_2).$$

The strict inequality follows from the joint convexity of F in $\mathbb{R}_{>0}$.

$\underline{y_1 = 0, y_2 = 0}$:

$$G(x_\theta) \leq F(x_\theta, 0) < \theta F(x_1, 0) + (1-\theta)F(x_2, 0) = \theta G(x_1) + (1-\theta)G(x_2).$$

The strict inequality follows from the joint convexity of $F(\cdot, 0)$.

$y_1 > 0, y_2 = 0, y_\theta \neq \theta y_1$: From the strict convexity of $F(x_\theta, \cdot)$ in $\mathbb{R}_{\geq 0}$ it follows that $G(x_\theta) < F(x_\theta, \theta y_1)$. But, from convexity $F(x_\theta, \theta y_1) \leq \theta G(x_1) + (1-\theta)G(x_2)$.

$y_1 > 0, y_2 = 0, y_\theta = \theta y_1$: Consider the restriction of F on the line segment passing through points (x_1, y_1), (x_θ, y_θ) and $(x_2, 0)$. Call it $H(\rho)$ and let be $H(0) = G(x_1)$ and $H(1) = G(x_2)$. Clearly, $H(1-\theta) = G(x_\theta)$. By strict convexity of F, it follows that $H(\rho)$ is strictly convex for $0 \leq \rho < 1$. Hence,

$$\begin{aligned} H(1-\theta) = H\left(\frac{2(1-\theta)}{2-\theta}\left(1-\frac{\theta}{2}\right)\right) &< \frac{\theta}{2-\theta}H(0) + \frac{2(1-\theta)}{2-\theta}H\left(1-\frac{\theta}{2}\right) \\ &\leq \frac{\theta}{2-\theta}H(0) + \frac{2(1-\theta)}{2-\theta}\left(\frac{\theta}{2}H(0) + \frac{2-\theta}{2}H(1)\right) \\ &= \theta H(0) + (1-\theta)H(1). \end{aligned}$$

The strict inequality follows from strict convexity of H in $(0, 1]$. The last inequality is a consequence of convexity of H in $[0, 1]$.

□

Lemma B.3.6. *Consider the following optimization*

$$\inf_{\alpha \geq 0} \sup_{\tau_h > 0} -\frac{\alpha \tau_h}{2} + \lambda \mathbb{E}\left[\mathrm{e}_f\left(X_0; \alpha\lambda/\tau_h\right) - f(X_0)\right]. \tag{B.83}$$

The set of minimizers over α is unbounded.

Proof. For convenience, denote the objective function as $O(\alpha, \tau_h)$ and its optimal value as O_*. Let us first perform the optimization over τ_h for fixed α. We have

$$\lim_{\tau_h \to 0^+} O(\alpha, \tau_h) = \lambda \mathbb{E}\left[\min_x f(x) - f(X_0)\right],$$

where we have used B.4.1(vi). Also,

$$\lim_{\tau_h \to +\infty} O(\alpha, \tau_h) = -\infty,$$

with an appeal to B.4.1(ix). What we learn from these is that

$$O_* \geq \lambda \mathbb{E}\left[\min_x f(x) - f(X_0)\right] \tag{B.84}$$

and that the optimal τ_h either approaches 0 or is attained. In the latter case, the optimal τ_{h*} satisfies the first-order optimality condition:

$$\frac{1}{\alpha^2}\mathbb{E}\left[\left(X_0 - \mathrm{prox}_f\left(X_0; \alpha\lambda/\tau_{h*}\right)\right)^2\right] = 1.$$

$$e_{\frac{1}{2}(\cdot)^2}(x;\tau) = \frac{x^2}{2(\tau+1)} \qquad e_{|\cdot|}(x;\tau) = \begin{cases} \frac{1}{2\tau}x^2, & \text{if } |x| \le \tau \\ |x| - \frac{\tau}{2}, & \text{otherwise} \end{cases}$$

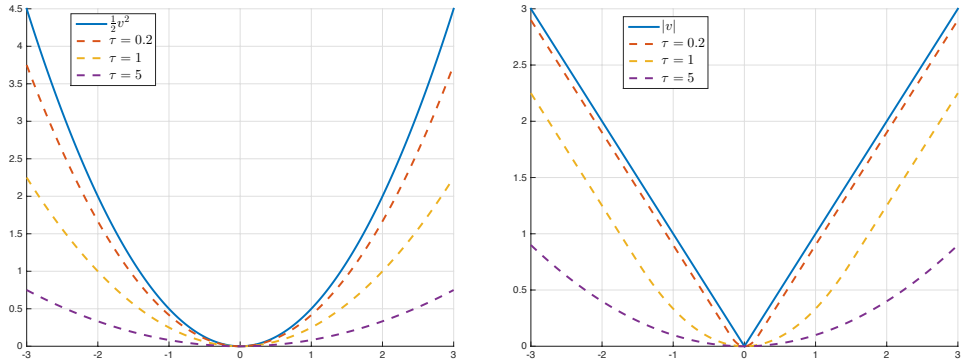

Figure B.1: Graphs of the Moreau envelope functions of a quadratic (left) and of the absolute value (right), for different values of the parameter τ. Moreau envelopes are always *smooth* under-estimators of the original function.

But for any $\tau_{h*} > 0$, by B.4.1(vii), the left hand-side above tends to 0 as $\alpha \to \infty$. Thus, in the limit $\alpha \to \infty$, the optimal τ_h approaches 0, giving

$$\lim_{\alpha \to +\infty} \sup_{\tau_h > 0} O(\alpha, \tau_h) = \lambda \mathbb{E}\left[\min_x f(x) - f(X_0)\right].$$

When combined with (B.84), this completes the proof of the lemma. □

B.4 Useful Properties of Moreau Envelopes

In this section we have gathered some very useful properties of Moreau envelopes of convex functions. We have made heavy use of those results for the proofs in Appendix B.3. Some of the results are standard, while others are more tailored towards our interests.

Lemma B.4.1 (Properties of the Moreau envelope)**.** *Let $\ell : \mathbb{R} \to \mathbb{R}$ be a proper, closed, convex function. For $\tau > 0$, consider its Moreau envelope function and its proximal operator:*

$$\mathrm{e}_\ell(\chi;\tau) := \min_v \frac{1}{2\tau}(\chi - v)^2 + \ell(v), \tag{B.85a}$$

$$\mathrm{prox}_\ell(\chi;\tau) := \arg\min \frac{1}{2\tau}(\chi - v)^2 + \ell(v). \tag{B.85b}$$

The following statements are true:

(i) $\mathrm{prox}_\ell(\chi;\tau)$ *is single valued and continuous. Furthermore,*

$$\ell'_{\chi,\tau} := \frac{1}{\tau}(\chi - \mathrm{prox}_\ell(\chi;\tau)) \in \partial\ell(\mathrm{prox}_\ell(\chi;\tau)). \tag{B.86}$$

(ii) $\mathrm{e}_\ell(\chi;\tau)$ *is jointly convex in* (χ,τ).

(iii) $\mathrm{e}_\ell(\chi;\tau)$ *is continuously differentiable with respect to both* x *and* τ. *The gradients are given by:*

$$E_1(\chi,\tau) := \frac{\partial e_\ell}{\partial \chi} = \frac{1}{\tau}(\chi - \mathrm{prox}_\ell(\chi;\tau)) = \ell'_{\chi,\tau}, \tag{B.87}$$

$$E_2(\chi,\tau) := \frac{\partial e_\ell}{\partial \tau} = -\frac{1}{2\tau^2}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2 = -\frac{1}{2}\left(\ell'_{\chi,\tau}\right)^2. \tag{B.88}$$

(iv) Fix χ *and* $\tau > 0$. *Consider the function* $\Delta : \mathbb{R}\times(-\tau,\infty)\to\mathbb{R}$:

$$\Delta(x,y) := (E_1(\chi+x,\tau+y) - E_1(\chi,\tau))x + (E_2(\chi+x,\tau+y) - E_2(\chi,\tau))y$$

Then,

$$\Delta(x,y) \geq \left(\tau + \frac{y}{2}\right)(\ell'_{\chi+x,\tau+y} - \ell'_{\chi,\tau})^2. \tag{B.89}$$

(v) $\mathrm{e}_\ell(x;\tau)$ *is non-increasing in* τ.

(vi) $\lim_{\tau\to\infty}\mathrm{e}_\ell(x;\tau) = \min_v \ell(v)$.

(vii) $\lim_{\tau\to\infty}\frac{1}{\tau}|x - \mathrm{prox}_\ell(x;\tau)| = 0$.

(viii) If $0 \in \arg\min_v \ell(v)$, *then* $\mathrm{prox}_\ell(x;\tau)\,x \geq 0$, $|\mathrm{prox}_\ell(x;\tau)| \leq |x|$ *and* $|\ell'_{\mathrm{prox}_\ell(x;\tau),\tau}| \leq |\ell'_{x,\tau}|$.

(ix) $\mathrm{e}_\ell(x_n;\tau_n)\to\ell(x)$ *whenever* $x_n\to x$ *while* $\tau_n\to 0^+$ *in such a way that the sequence* $\{|x_n - x|/\tau_n\}_{n\in\mathbb{N}}$ *is bounded.*

Proof. *(i)* From [RW09, Thm. 2.26(a)], $\mathrm{prox}_\ell(\chi;\tau)$ is known to be continuous single valued mapping. Besides, from standard optimality conditions:

$$\frac{1}{\tau}(\chi - \mathrm{prox}_\ell(\chi;\tau)) \in \partial\ell(\mathrm{prox}_\ell(\chi;\tau)).$$

For convenience, we have define $\ell'_{\chi,\tau} := \frac{1}{\tau}(\chi - \mathrm{prox}_\ell(\chi;\tau) \in \partial\ell(\mathrm{prox}_\ell(\chi;\tau))$. Note that if ℓ is differentiable at $\mathrm{prox}_\ell(\chi;\tau)$, then $\ell'_{\chi,\tau}$ is the derivative of ℓ at that point.

(ii) Trivially, $h(\chi,v) := (\chi - v)^2$ is a jointly convex function of v and x. Thus, its perspective function $\tau h(\frac{\chi}{\tau},\frac{v}{\tau}) = \frac{1}{\tau}(\chi - v)^2$ is also jointly convex over τ, x and v and

so after minimization over v, the function remains jointly convex over x and τ (cf. [RW09, Prop. 2.22]).

(iii) See [RW09, Thm. 2.26(b)] for differentiability with respect to x. Next, we mimic the argument to conclude about differentiability with respect to τ. It suffices to show that $h(y) := \mathrm{e}_\ell(\chi;\tau+y) - \mathrm{e}_\ell(\chi;\tau) + \frac{y}{2\tau^2}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2$ is differentiable at $y = 0$ with $\frac{\partial h}{\partial y} = 0$. We know $\mathrm{e}_\ell(\chi;\tau) = \frac{1}{2\tau}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2 + \ell(\mathrm{prox}_\ell(\chi;\tau))$, whereas $\mathrm{e}_\ell(\chi;\tau+y) \le \frac{1}{2(\tau+y)}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2 + \ell(\mathrm{prox}_\ell(\chi;\tau))$. Thus,

$$\begin{aligned} h(y) &\le \frac{1}{2(\tau+y)}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2 - \frac{1}{2\tau}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2 + \frac{y}{2\tau^2}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2 \\ &= \frac{y^2}{2\tau^2(\tau+y)}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2. \end{aligned} \tag{B.90}$$

Besides, because of convexity of $h(y)$, $0 = h(0) \le \frac{1}{2}h(y) + \frac{1}{2}h(-y)$, or equivalently, $h(y) \ge -h(-y)$. Thus, (B.90) gives:

$$h(y) \ge \frac{y^2}{2\tau^2(\tau-y)}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2. \tag{B.91}$$

Combining (B.90) and (B.91) leads to the following:

$$\frac{y^2}{2\tau^2(\tau-y)}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2 \le h(y) \le \frac{y^2}{2\tau^2(\tau+y)}(\chi - \mathrm{prox}_\ell(\chi;\tau))^2. \tag{B.92}$$

Here, $h(y)$ is sandwiched between two continuously differentiable functions at 0 with zero derivatives. This completes the proof.

(iv) From (B.87) and (B.88), we have

$$\begin{aligned} \Delta(x,y) &= (\ell'_{\chi+x,\tau+y} - \ell'_{\chi,\tau})x - \left(\ell'^2(\chi+x,\tau+y) - \ell'^2(\chi,\tau)\right)\frac{y}{2} \\ &= (\ell'_{\chi+x,\tau+y} - \ell'_{\chi,\tau})\left(x - \frac{y}{2}\left(\ell'_{\chi+x,\tau+y} + \ell'_{\chi,\tau}\right)\right). \end{aligned}$$

On the other hand, due to optimality conditions in (B.86),

$$\begin{aligned} \mathrm{prox}_\ell(\chi+x;\tau+y) - \mathrm{prox}_\ell(\chi;\tau) &= x - (\tau+y)\ell'_{\chi+x,\tau+y} + \tau\ell'_{\chi,\tau} \\ &= \left(x - \frac{y}{2}(\ell'_{\chi+x,\tau+y} + \ell'_{\chi,\tau})\right) - (\tau + \frac{y}{2})(\ell'_{\chi+x,\tau+y} - \ell'_{\chi,\tau}). \end{aligned}$$

Finally, from convexity of ℓ, it follows from the monotonicity property of the sub-differential that

$$(\ell'_{\chi+x,\tau+y} - \ell'_{\chi,\tau})(\mathrm{prox}_\ell(\chi+x;\tau+y) - \mathrm{prox}_\ell(\chi;\tau)) \ge 0.$$

Combining the three displays above gives the desired inequality.

(v) This follows directly by non-positivity of the derivative as in (B.88).

(vi) Using the decreasing nature of $\mathrm{e}_\ell(x;\tau)$ w.r.t. τ, we have

$$\lim_{\tau\to\infty} \mathrm{e}_\ell(x;\tau) = \inf_{\tau>0} \min_v \frac{1}{2\tau}(x-v)^2+\ell(v) = \min_v \inf_{\tau>0} \frac{1}{2\tau}(x-v)^2+\ell(v) = \min_v \ell(v).$$

(vii) Fix an $\epsilon > 0$. Since $\lim_{\tau\to\infty} \mathrm{e}_\ell(x;\tau) = \min_v \ell(v)$, there exists T'_ϵ such that for all $\tau \geq T_\epsilon := \max\{2, T'_\epsilon\}$,

$$|\mathrm{e}_\ell(x;\tau) - \min_v \ell(v)| = \frac{1}{2\tau}(x - \mathrm{prox}_\ell(x;\tau))^2 + (\ell(\mathrm{prox}_\ell(x;\tau)) - \min_v \ell(v)) < \epsilon^2.$$

Then $\frac{1}{2\tau}(x - \mathrm{prox}_\ell(x;\tau))^2 < \epsilon^2$, which gives

$$\frac{1}{\tau}|x - \mathrm{prox}_\ell(x;\tau)| < \epsilon\sqrt{\frac{2}{\tau}} \leq \epsilon\sqrt{\frac{2}{T_\epsilon}} \leq \epsilon.$$

Therefore, $\lim_{\tau\to\infty} \frac{1}{\tau}|x - \mathrm{prox}_\ell(x;\tau)| = 0$.

(viii) By (B.86) and the assumption $0 \in \arg\min_v \ell(v)$, we find $\mathrm{prox}_\ell(0;\tau) = 0$. Monotonicity of the prox operator [RW09, Prop. 12.19], gives $(\mathrm{prox}_\ell(x;\tau) - \mathrm{prox}_\ell(0;\tau))x \geq 0$, which then shows $\mathrm{prox}_\ell(x;\tau)\,x \geq 0$. Also, monotonicity of the subdifferential of ℓ gives $\ell'_{x,\tau}x \geq 0$. Those two, when combined with optimality conditions in (B.86) give

$$x - \mathrm{prox}_\ell(x;\tau) = \tau\ell'_{x,\tau} \implies x^2 \geq \mathrm{prox}_\ell(x;\tau)\,x \implies |x| \geq |\mathrm{prox}_\ell(x;\tau)\,x|.$$

It remains to show that $\max_{s\in\partial\ell(\mathrm{prox}_\ell(x;\tau))} |s| \leq \max_{s\in\partial\ell(x)} |s|$. Since $0 \in \arg\min_v \ell(v)$, it follows by convexity that

$$(0 \leq x_1 \leq x_2 \text{ or } x_2 \leq x_1 \leq 0) \implies \max_{s\in\partial\ell(x_1)} |s| \leq \max_{s\in\partial\ell(x_2)} |s|.$$

Observe that the LHS of the implication above is equivalent to ($|x_2| \geq |x_1|$ and $x_1x_2 \geq 0$). Then apply it for $x_1 = \mathrm{prox}_\ell(x;\tau)$ and $x_2 = x$, to conclude.

(ix) Please see [RW09, Thm. 1.25].

□

Appendix C

PROOFS FOR CHAPTER 6

C.1 On Remark 6.3.0.44

Substituting the envelope function of $|\cdot|$ in (6.20) gives:

$$\frac{\beta}{2}+\bar{s}\mathbb{E}\begin{cases}-\frac{\beta(\alpha G+Z)^2}{2\tau^2} & ,|\alpha G+Z|\leq\frac{\tau}{\beta}\\ -\frac{1}{2\beta} & ,\text{otherwise}\end{cases}+(\delta-\bar{s})\mathbb{E}\begin{cases}-\frac{\beta\alpha^2G^2}{2\tau^2} & ,|\alpha G|\leq\frac{\tau}{\beta}\\ -\frac{1}{2\beta} & ,\text{otherwise}\end{cases}\geq 0, \tag{C.1a}$$

$$\bar{s}\mathbb{E}\begin{cases}\frac{\beta G(\alpha G+Z)}{2} & ,|\alpha G+Z|\leq\frac{\tau}{\beta}\\ G\ \text{sign}(\alpha G+Z) & ,\text{otherwise}\end{cases}+(\delta-\bar{s})\mathbb{E}\begin{cases}\frac{\alpha G^2\beta}{\tau} & ,|\alpha G\|\leq\frac{\tau}{\beta}\\ G\ \text{sign}(G) & ,\text{otherwise}\end{cases}-\beta\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}\geq 0, \tag{C.1b}$$

$$\frac{\tau}{2}+\bar{s}\mathbb{E}\begin{cases}\frac{(\alpha G+Z)^2}{2\tau} & ,|\alpha G+Z|\leq\frac{\tau}{\beta}\\ -\frac{\tau}{2} & ,\text{otherwise}\end{cases}+(\delta-\bar{s})\mathbb{E}\begin{cases}\frac{\alpha^2G^2}{2\tau} & ,|\alpha G|\leq\frac{\tau}{\beta}\\ -\frac{\tau}{2} & ,\text{otherwise}\end{cases}-\alpha\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}\leq 0. \tag{C.1c}$$

Define $\kappa := \frac{\tau}{\beta\alpha}$ and $\rho := \frac{\tau}{\beta}$. In order to find a sufficient condition for α to be zero, we assume $\alpha \to 0$, $\tau \to 0$, $\rho \to 0$ and $\kappa \geq 0$ and look for conditions under which the equations in (C.1) are consistent. Under these assumptions, one can check that (C.1c) is satisfied (the argument converges to zero), while, (C.1b) and (C.1a) become

$$2\frac{(\delta-\bar{s})}{\kappa}\int_0^\kappa G^2\phi(G)\mathrm{d}G+(\delta-\bar{s})\int_\kappa^\infty G\phi(G)\mathrm{d}G\geq\beta\sqrt{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}, \tag{C.2a}$$

$$\beta^2\geq\bar{s}+2\frac{\delta-\bar{s}}{\kappa^2}\int_0^\kappa G^2\phi(G)\mathrm{d}G+2(\delta-\bar{s})\int_\kappa^\infty\phi(G)\mathrm{d}G, \tag{C.2b}$$

where $\phi(G)=e^{-G^2/2}/\sqrt{2\pi}$ and we multiplied (C.1a) by β^2 to get (C.2b). Observe that (C.2a) upper bounds β while (C.2b) derives a lower bound on it. Thus, consistency of the set of equations (C.2) is achieved if the following holds:

$$\frac{1}{\overline{\mathbf{D}}_{g,\mathbf{x}_0}}(2\frac{(\delta-\bar{s})}{\kappa}\int_0^\kappa G^2\phi(G)\mathrm{d}G+(\delta-\bar{s})\int_\kappa^\infty G\phi(G)\mathrm{d}G)^2\geq$$
$$\bar{s}+2\frac{\delta-\bar{s}}{\kappa 2}\int_0^\kappa G^2\phi(G)\mathrm{d}G+2(\delta-\bar{s})\int_\kappa^\infty\phi(G)\mathrm{d}G.$$

Or, equivalently,

$$\overline{\mathbf{D}}_{g,\mathbf{x}_0} \leq \frac{(2\frac{(\delta-\bar{s})}{\kappa}\int_0^\kappa G^2\phi(G)\mathrm{d}G + (\delta-\bar{s})\int_\kappa^\infty G\phi(G)\mathrm{d}G)^2}{\bar{s}+2\frac{\delta-\bar{s}}{\kappa^2}\int_0^\kappa G^2\phi(G)\mathrm{d}G + 2(\delta-\bar{s})\int_\kappa^\infty \phi(G)\mathrm{d}G}. \tag{C.3}$$

Thus if the maximum of the right side of (C.3) with respect to κ is greater than $\overline{\mathbf{D}}_{g,\mathbf{x}_0}$, all our variables satisfy (6.20) and the optimal value in (6.19) occurs when $\alpha \to 0, \tau \to 0$ and $\frac{\tau}{\alpha\beta} \to \kappa$ which means $\alpha_* = 0$. We will show that

$$\max_{\kappa>0} \frac{(2\frac{(\delta-\bar{s})}{\kappa}\int_0^\kappa G^2\phi(G)\mathrm{d}G + (\delta-\bar{s})\int_\kappa^\infty G\phi(G)\mathrm{d}G)^2}{\bar{s}+2\frac{\delta-\bar{s}}{\kappa^2}\int_0^\kappa G^2\phi(G)\mathrm{d}G + 2(\delta-\bar{s})\int_\kappa^\infty \phi(G)\mathrm{d}G} \geq$$
$$\delta - \min_{\kappa>0} \bar{s}(1+\kappa^2) + 2(\delta-\bar{s})\int_\kappa^\infty (G-\kappa)^2\phi(G)\mathrm{d}G. \tag{C.4}$$

If both this and (6.24) are true then, there will be a κ for which (C.3) holds and as we discussed, this implies $\alpha_* = 0$.

For convenience, we define $A_\kappa = \int_\kappa^\infty G^2\phi(G)\mathrm{d}G$, $B_\kappa = \int_\kappa^\infty G\phi(G)\mathrm{d}G$ and $C_\kappa = \int_\kappa^\infty \phi(G)\mathrm{d}G$. The optimal κ for the right side of (C.4) satisfies the following due to the first optimality condition

$$2(\delta-\bar{s})\hat{\kappa}B_{\hat{\kappa}} - 2(\delta-\bar{s})\hat{\kappa}^2C_{\hat{\kappa}} = \hat{\kappa}^2\bar{s}. \tag{C.5}$$

For this value of κ, the left side of (C.4) becomes

$$\frac{(2\frac{(\delta-\bar{s})}{\hat{\kappa}}\int_0^{\hat{\kappa}} G^2\phi(G)\mathrm{d}G + (\delta-\bar{s})\int_{\hat{\kappa}}^\infty G\phi(G)\mathrm{d}G)^2}{\bar{s}+2\frac{\delta-\bar{s}}{\hat{\kappa}^2}\int_0^{\hat{\kappa}} G^2\phi(G)\mathrm{d}G + 2(\delta-\bar{s})\int_{\hat{\kappa}}^\infty \phi(G)\mathrm{d}G} = (\delta-\bar{s})(1-2A_{\hat{\kappa}}+2\hat{\kappa}B_{\hat{\kappa}})$$
$$= \delta - \bar{s} - 2(\delta-\bar{s})\hat{\kappa}B_{\hat{\kappa}} + 2(\delta-\bar{s})\hat{\kappa}^2C_{\hat{\kappa}} - 2(\delta-\bar{s})A_{\hat{\kappa}} + 4(\delta-\bar{s})\hat{\kappa}B_{\hat{\kappa}} - 2(\delta-\bar{s})\hat{\kappa}^2C_{\hat{\kappa}}$$
$$= \delta - \bar{s}(1+\hat{\kappa}^2) - 2(\delta-\bar{s})(A_{\hat{\kappa}} - 2B_{\hat{\kappa}} + \hat{\kappa}^2C_{\hat{\kappa}}) = \delta - \bar{s}(1+\hat{\kappa}^2) + 2(\delta-\bar{s})\int_{\hat{\kappa}}^\infty (G-\hat{\kappa})^2\phi(G)\mathrm{d}G,$$

where the first and third equalities follow after substituting $\bar{s}$ using (C.5). This proves (C.4) as desired to conclude the claim of the remark.

C.2 On Section 6.5

Satisfying Assumptions 4.2.1(a) and 4.2.2(b)-(d)

It only takes a few calculations to show that

$$\frac{1}{m}\mathrm{e}_{\sqrt{n}\|\cdot\|_2}(\alpha\mathbf{g}+\mathbf{z};\tau) = \begin{cases} \frac{1}{\sqrt{m\delta}}\|\alpha\mathbf{g}+\mathbf{z}\|_2 - \frac{\tau}{2\delta} & \text{, if } \frac{\sqrt{\delta}\|\alpha\mathbf{g}+\mathbf{z}\|_2}{\sqrt{m}} \geq \tau, \\ \frac{1}{2\tau}\frac{\|\alpha\mathbf{g}+\mathbf{z}\|_2^2}{m} & \text{, otherwise.} \end{cases} \tag{C.6}$$

Assume that $0 < \mathbb{E}\frac{\|\mathbf{z}\|_2^2}{m} =: \sigma^2 < \infty$. From (C.6), it can be seen that $\frac{1}{m}\mathrm{e}_{\sqrt{n}\|\cdot\|_2}(\alpha\mathbf{g}+\mathbf{z};\tau)$ is a Lipschitz convex function of $\frac{\|\alpha\mathbf{g}+\mathbf{z}\|}{\sqrt{m}}$. Also, $\frac{\|\alpha\mathbf{g}+\mathbf{z}\|}{\sqrt{m}}$ converges in probability to $\sqrt{\alpha^2+\sigma^2}$, thus

$$\frac{1}{m}(\mathrm{e}_{\sqrt{n}\|\cdot\|_2}(\alpha\mathbf{g}+\mathbf{z};\tau) - \|\mathbf{z}\|_2\sqrt{n}) \to L(\alpha,\tau) = \begin{cases} \frac{\sqrt{\alpha^2+\sigma^2}-\sigma}{\sqrt{\delta}} - \frac{\tau}{2\delta} & \text{, if } \delta(\alpha^2+\sigma^2) \geq \tau^2, \\ \frac{1}{2\tau}(\alpha^2+\sigma^2) - \frac{\sigma}{\sqrt{\delta}} & \text{, otherwise.} \end{cases}$$

Finally, it remains to show this function satisfies Assumption 4.2.2.

Assumption 4.2.2(b): $\lim_{\tau\to 0} L(\alpha,\tau) = \frac{\sqrt{\alpha^2+\sigma^2}-\sigma}{\sqrt{\delta}}$ and $\lim_{\tau\to 0} L(0,\tau) = 0$. Besides

$$L_{2,+}(\alpha,\tau) = \begin{cases} -\frac{1}{2\delta} & \text{, } \delta(\alpha^2+\sigma^2) \geq \tau^2, \\ -\frac{\alpha^2+\sigma^2}{2\tau^2} & \text{, otherwise.} \end{cases}$$

So, $L_{2,+}(0,0) = -\frac{1}{2\delta}$; thus, condition (b) is satisfied.

Assumption 4.2.2(c): $\frac{1}{m}\mathcal{L}(z) \xrightarrow{P} \frac{\sigma}{\sqrt{\delta}} = -\lim_{\tau\to\infty} L(\alpha,\tau)$. It is also easy to check that $L(\alpha,\tau) \geq -\frac{\sigma}{\sqrt{\delta}}$ for all α and $\tau > 0$ because

$$L(\alpha,\tau) = \begin{cases} \frac{\sqrt{\alpha^2+\sigma^2}-\sigma}{\sqrt{\delta}} - \frac{\tau}{2\delta} & \text{, } \delta(\alpha^2+\sigma^2) \geq \tau^2 \\ \frac{\alpha^2+\sigma^2}{2\tau} - \frac{\sigma}{\sqrt{\delta}} & \text{, otherwise} \end{cases} \geq \begin{cases} \frac{\tau}{2\delta} - \frac{\sigma}{\sqrt{\delta}} & \text{, } \delta(\alpha^2+\sigma^2) \geq \tau^2 \\ \frac{\alpha^2+\sigma^2}{2\tau} - \frac{\sigma}{\sqrt{\delta}} & \text{otherwise} \end{cases} \geq -\frac{\sigma}{\sqrt{\delta}}.$$

Therefore condition (c) is satisfied. Besides, since $L_0 = \frac{\sigma}{\sqrt{\delta}} < \infty$, there is nothing to check regarding Condition 4.2.2(d).

Proving (6.29)⇔(6.31)

It suffices to show that $H(\beta) := -\frac{\alpha\beta^2}{2\tau_h} + F(\frac{\alpha\beta}{\tau_h}, \frac{\alpha\lambda}{\tau_h})$ is a non-increasing function of $\beta > 0$. We prove this for the separable function f satisfying the assumptions of Theorem 4.3.1 where $F(c,\tau) = \mathbb{E}[\mathrm{e}_f(cH + X_0;\tau) - f(X_0)]$. Using Lemma B.3.1(i) and B.4.1(iii), we find

$$\lim_{c\to 0^+} \frac{\partial}{\partial c}\left(\mathbb{E}[\mathrm{e}_f(cH+X_0;\tau) - f(X_0)]\right) = \frac{1}{\tau}\mathbb{E}[H(X_0 - \mathrm{prox}_f(X_0;\tau))].$$

Because of independence of $\mathbf{h}$ and $\mathbf{x}_0$, the RHS above is zero. Thus $\lim_{c\to 0^+}\frac{\partial}{\partial c}F(c,\tau) = 0$ which when combined with concavity of H, it shows that it is non-increasing for $\beta > 0$.

C.3 Proof of Theorem 6.6.1

From (6.37) $\|\hat{\mathbf{x}} - \mathbf{x}_0\|$ converges in probability to the unique minimizer α_* of the following max-min problem:

$$\max_{\substack{0\le\beta\le 1\\ \tau>0}} \min_{\alpha\ge 0} H(\alpha,\beta,\tau) := \beta\sqrt{\delta}\sqrt{\sigma^2+\alpha^2} - \frac{\tau\alpha}{2} + \frac{\tau}{2\alpha} - \frac{\alpha}{2\tau}\mathbb{E}\left[\eta^2\left(\beta H + \frac{\tau}{\alpha}X_0;\lambda\right)\right], \tag{C.7}$$

where the expectation is over $H \sim \mathcal{N}(0,1)$ and $X_0 \sim p_{X_0}$. Here, we prove Theorem 6.6.1 by analyzing the optimality conditions of (C.7). Recall that H is jointly concave in β, p and strongly convex in α.

First Order Optimality Conditions.

We begin with a lemma, which characterizes the first-order optimality conditions of (C.7).

Lemma C.3.1 (Optimality Conditions)**.** *Consider the following pair of equations with respect to β and κ:*

$$\begin{cases} \beta^2\kappa^2\delta = \sigma^2 + \mathbb{E}\left[(\eta(\beta\kappa H + X_0;\kappa\lambda) - X_0)^2\right], & \text{(C.8)}\\ \beta\kappa\delta = \mathbb{E}[(\eta(\beta\kappa H + X_0;\kappa\lambda)\cdot h)]. & \text{(C.9)}\end{cases}$$

Also, define λ_{min} to be the unique non-negative solution to the equation

$$(1+x^2)\int_{-\infty}^{-x} e^{-z^2/2}\mathrm{d}z - xe^{-x^2/2} = \delta\sqrt{\frac{\pi}{2}}.$$

With these, let $(\beta_*,\tau_*,\alpha_*)$ be optimal in (C.7). Then,

$$\alpha_*^2 = \beta_*^2\kappa_*^2\delta - \sigma^2 \quad \text{and} \quad \kappa_* = \frac{\sigma}{\sqrt{\beta_*^2\delta - \tau_*^2}} \tag{C.10}$$

such that,

(i) If $\beta_* = 1$ and $\lambda > \lambda_{min}$, then κ_* is the unique solution to (C.8) for $\beta = 1$,

(ii) If $\beta_* \in (0,1)$,then κ_*, β_* are solutions to the pair of equation (C.8)-(C.9).

Proof. Consider the derivation of the objective function with respect to α, τ and β as follows

$$\frac{\partial}{\partial\alpha}H(\alpha,\tau,\beta) = \frac{\beta\alpha\sqrt{\delta}}{\sqrt{\alpha^2+\sigma^2}} - \frac{\tau}{2} - \frac{1}{2\tau}\mathbb{E}[(\eta(\beta H + \frac{\tau}{\alpha}X_0,\lambda) - \frac{\tau}{\alpha}X_0)^2], \tag{C.11}$$

$$\frac{\partial}{\partial\tau}H(\alpha,\tau,\beta) = -\frac{\alpha}{2} + \frac{\alpha}{2\tau^2}\mathbb{E}[(\eta(\beta H + \frac{\tau}{\alpha}X_0,\lambda) - \frac{\tau}{\alpha}X_0)^2], \tag{C.12}$$

$$\frac{\partial}{\partial\beta}H(\alpha,\tau,\beta) = \sqrt{\delta}\sqrt{\alpha^2+\sigma^2} - \frac{\alpha}{\tau}\mathbb{E}[\eta(\beta H + \frac{\tau}{\alpha}X_0,\lambda)\cdot H]. \tag{C.13}$$

We will prove that the optimal α_*, τ_* and β_* are all strictly positive. First, suppose $\alpha_* = 0$ and $\tau_* > 0$. Then, the first term in (C.11) goes to zero while $-\frac{\tau}{2}$ stays negative and the final term is always non-positive. This shows $\frac{\partial}{\partial\alpha}\big|_{\alpha\to 0} H(\alpha, \tau, \beta) < 0$, which means that $\alpha_* = 0$ cannot be optimal in this case by convexity. Next, assume $\alpha_* = 0$ and the optimal over τ is approached as $\tau \to 0$. In this case, it can be shown that the expected value in (C.11) is strictly positive, thus the derivative remains negative. Thus, $\alpha_* > 0$, as promised. A similar argument shows strict positivity of (C.12) when $\tau \to 0$. Thus, $\tau_* > 0$. Finally, $\frac{\partial}{\partial\beta}\big|_{\beta\to 0} H(\alpha, \tau, \beta) > 0$, by independence of h and X_0, showing $\beta_* > 0$.

The argument above shows that the derivatives are equal to zero at the optimal. For convenience define

$$P\left(\frac{\tau}{\alpha}\right) := \frac{\tau}{2\alpha} - \frac{\alpha\lambda^2}{2\tau}\mathbb{E}\left[\eta^2\left(\frac{\beta}{\lambda}H + \frac{\tau}{\lambda\alpha}X_0; 1\right)\right].$$

Then equating the derivatives in (C.7) with respect to α and τ with zero gives

$$\frac{\beta\sqrt{\delta}\alpha}{\sqrt{\alpha^2+\sigma^2}} - \frac{\tau}{2} - \frac{\tau}{\alpha^2}P'(\frac{\tau}{\alpha}) = 0, \tag{C.14a}$$

$$-\frac{\alpha}{2} + \frac{1}{\alpha}P'(\frac{\tau}{\alpha}) = 0. \tag{C.14b}$$

Here, P' is the derivative of $P(x)$ with respect to x. Any optimal $\beta_*, \tau_*, \alpha_*$ satisfies these. Then, it only takes multiplying (C.14b) by $\frac{\tau}{\alpha}$ and adding the result to (C.14a) to see that

$$\alpha_* = \frac{\tau_*\sigma}{\sqrt{\beta^2\delta - \tau_*^2}}. \tag{C.15}$$

Next, substituting (C.15) in (C.14b) it can be shown that,

$$-\frac{\sigma^2}{2} + \frac{\sigma^2}{2\tau^2}\mathbb{E}[(\eta(\beta H + \frac{\sqrt{\beta^2\delta-\tau^2}}{\sigma}X_0; \lambda) - \frac{\sqrt{\beta^2\delta-\tau^2}}{\sigma}X_0)^2] = 0.$$

To reach this we have also used the following facts: $\eta(x;\lambda)\frac{\partial}{\partial x}\eta(x;\lambda) = \eta(x;\lambda)$, $\lambda\eta(\frac{x}{\lambda}; 1) = \eta(x;\lambda)$ and $\mathbb{E}[X_0^2] = 1$ by assumption (6.33). Multiplying the result with $2\tau^2/\sigma^2$ and defining

$$\kappa := \frac{\sigma}{\sqrt{\beta^2\delta - \tau^2}},$$

we conclude with,

$$\beta^2\delta\kappa^2 - \sigma^2 = \mathbb{E}[(\eta(\beta\kappa H + X_0; \kappa\lambda) - X_0)^2], \tag{C.16}$$

which is same as (C.8). Also, with respect to the optimal κ_* it is easily seen by (C.15) that

$$\alpha_*^2 = \beta_*^2 \kappa_*^2 \delta - \sigma^2. \tag{C.17}$$

The derivative in (C.7) with respect to β gives

$$\begin{aligned}\frac{\partial}{\partial \beta} H(\alpha, \beta, \tau) &= \sqrt{\delta}\sqrt{\sigma^2 + \alpha^2} - \frac{\alpha}{\tau}\mathbb{E}[\eta(\beta H + \frac{\tau}{\alpha} X_0; \lambda) H] \\ &= \beta\delta\kappa - \kappa\mathbb{E}[\eta(\beta H + \frac{X_0}{\kappa}, \lambda) H] = \beta\delta\kappa - \mathbb{E}[\eta(\kappa\beta H + X_0; \lambda\kappa) H],\end{aligned} \tag{C.18}$$

where we have also used (C.17). Note that the above is the same as (C.9) and recall the constraint $0 \le \beta \le 1$ in (C.7) to conclude with the desired.

It only remains to show that the solution with respect to κ of (C.8) (eqv. of (C.16)) is unique when $\beta = 1$ and $\lambda \ge \lambda_{\min}$. For $\beta = 1$, (C.8) is the same as fixed point equation [BM12, Eqn. (1.9)], which in turn was shown to admit a unique solution for all $\lambda > \lambda_{\min}$ in [DMM11] (see [BM12, Prop. 1.3]). □

The Regions of Operation

We build up to the proof of Theorem 6.6.1 through a series of auxiliary lemmas. Through the lemmas, we identify two "regimes of operation" of the LASSO. The first we call $\mathcal{R}_{\text{bad}}$, and it corresponds to values of λ for which the optimal β is in the open set $(0, 1)$. The second regime is such that $\beta = 1$. If $\delta < 1$, we prove in Lemma C.3.5 that there exists a unique critical value λ_{crit} separating the two regimes in the sense that $\mathcal{R}_{\text{bad}}$ extends from 0 to λ_{crit}. If on the other hand $\delta \ge 1$, then there is no $\mathcal{R}_{\text{bad}}$ region (Lemma C.3.6).

First, we need a few useful definitions.

Definition C.3.1. For any $\lambda > 0$, we let $\alpha_*(\lambda)$, $\tau_*(\lambda)$ and $\beta_*(\lambda)$ be optimal solutions in (C.7). Apart from $\alpha_*(\lambda)$, the others are not necessarily unique at this point. Also, $\kappa_*(\lambda)$ is defined as in (C.10).

Definition C.3.2 (Bad Regime)**.** We say that a value $\lambda > 0$ is in the bad regime $\mathcal{R}_{\text{bad}}$, denoted $\lambda \in \mathcal{R}_{\text{bad}}$, if there exists $\beta_*(\lambda) \in (0, 1)$.

Definition C.3.3 (Critical Regime)**.** We say that a value $\lambda_{\text{crit}} > 0$ is in the critical regime $\mathcal{R}_{\text{crit}}$, denoted $\lambda_{\text{crit}} \in \mathcal{R}_{\text{crit}}$ if for some κ_{crit}, the pair $\lambda_{\text{crit}}, \kappa_{\text{crit}}$ solves:

$$\begin{cases} \kappa^2\delta = \sigma^2 + \mathbb{E}\left[(\eta(\kappa H + X_0; \kappa\lambda) - X_0)^2\right], & \text{(C.19)} \\ \kappa\delta = \mathbb{E}[(\eta(\kappa H + X_0; \kappa\lambda) \cdot H)]. & \text{(C.20)} \end{cases}$$

As an immediate consequence of the definition above and the first order optimality conditions in Lemma C.3.1, we have

$$\beta_*(\lambda_{\text{crit}}) = 1, \quad \kappa_*(\lambda_{\text{crit}}) = \kappa_{\text{crit}} \quad \text{and} \quad \alpha_*(\lambda_{\text{crit}}) = \sqrt{\delta\kappa_{\text{crit}}^2 - \sigma^2}. \tag{C.21}$$

Also, the following lemma reveals the importance of λ_{crit}: all $\lambda < \lambda_{\text{crit}}$ are in $\mathcal{R}_{\text{bad}}$ and the squared error is constant in that regime, i.e. $\alpha_*(\lambda) = \alpha_*(\lambda_{\text{crit}})$.

Lemma C.3.2 (Error in $\mathcal{R}_{\text{bad}}$)**.** *Let $\lambda_{crit} \in \mathcal{R}_{crit}$. Then, for all $0 < \lambda' < \lambda_{crit}$, it holds $\lambda' \in \mathcal{R}_{bad}$. Furthermore, $\beta_*(\lambda') = \lambda/\lambda_{crit}$, $\lambda'\kappa_*(\lambda') = \kappa_{crit}\lambda_{crit}$ and $\alpha_*(\lambda') = \alpha_*(\lambda_{crit})$.*

Proof. Fix any $0 < \lambda' < \lambda_{\text{crit}}$. By definition, there exists κ_{crit} such that $\lambda_{\text{crit}}, \kappa_{\text{crit}}$ satisfy (C.19)-(C.20). Define $\beta' := \lambda/\lambda_{\text{crit}}$ and $\kappa' := \kappa_{\text{crit}}/\beta'$. It is then easy to see that β', κ' solve (C.8)-(C.9) (for $\lambda = \lambda'$ therein). Also, $\beta' < 1$ by definition. Thus, $\lambda' \in \mathcal{R}_{\text{bad}}$ and $\beta_*(\lambda') = \lambda/\lambda_{\text{crit}}, \kappa_*(\lambda') = \kappa_{\text{crit}}\lambda_{\text{crit}}/\lambda'$. Also, using (C.10) and (C.21), $\alpha_*(\lambda) = \sqrt{\delta\beta_*^2(\lambda')\kappa_*^2(\lambda) - \sigma^2} = \sqrt{\delta\kappa_*^2(\lambda_{\text{crit}}) - \sigma^2} = \alpha_*(\lambda_{\text{crit}})$. □

It is thus important to identify the critical values of the regularizer parameter, i.e. all $\lambda_{\text{crit}} \in \mathcal{R}_{\text{crit}}$. Values in $\mathcal{R}_{\text{bad}}$ are important towards this aim, since as shown in the next lemma, for any $\lambda \in \mathcal{R}_{\text{bad}}$ there must exist some $\lambda_{\text{crit}} > \lambda$.

Lemma C.3.3 ($\mathcal{R}_{\text{bad}} \to \lambda_{\text{crit}}$)**.** *Let $\lambda_1 \in \mathcal{R}_{bad}$, then there exists $\lambda_2 \in \mathcal{R}_{crit}$ with $\lambda_2 > \lambda_1$.*

Proof. Let $\beta_1, \alpha_1, \kappa_1$ be optimal corresponding to λ_1. Since $\lambda_1 \in \mathcal{R}_{\text{bad}}$, it holds $0 < \beta_1 < 1$. Then, from Lemma C.3.1, κ_1, β_1 solve (C.8)-(C.9). Starting from these and substituting $\lambda_2 := \lambda_1/\beta_1$ and $\kappa_2 := \kappa_1\beta_1$ therein, it is not hard to see that this is equivalent to λ_2, κ_2 satisfying (C.19)-(C.20). Thus, $\lambda_2 \in \mathcal{R}_{\text{crit}}$. Also, clearly $\lambda_2 > \lambda_1$. □

The lemma below is important since it shows that when $\delta < 1$ there exists a *unique* $\lambda_{\text{crit}} \in \mathcal{R}_{\text{crit}}$.

Lemma C.3.4 (Unique λ_{crit})**.** *Suppose $\delta < 1$. The set of equations (C.19)-(C.20) has a* unique *pair of solutions (κ, λ). Thus, there exists unique $\lambda_{crit} \in \mathcal{R}_{crit}$.*

Proof. First, we show that there exists at most one $\lambda_{\text{crit}} \in \mathcal{R}_{\text{crit}}$. For the shake of contradiction assume two different pairs of solutions, say (κ_1, λ_1) and (κ_2, λ_2). By definition, $\lambda_1, \lambda_2 \in \mathcal{R}_{\text{crit}}$. First, note that we cannot have $\lambda_1 = \lambda_2$, since if this was the case then from (C.21) we would also have $\kappa_1 = \kappa_2$. Henceforth, assume w.l.o.g. that $\lambda_1 < \lambda_2$. It follows from Lemma C.3.2 that $\lambda_1 \in \mathcal{R}_{\text{bad}}$ and also $\kappa_*(\lambda_1)\lambda_1 = \kappa_*(\lambda_2)\lambda_2$. Thus,

$$\kappa_*(\lambda_1) < \kappa_*(\lambda_2). \tag{C.22}$$

But also, again from Lemma C.3.2, $\alpha_*(\lambda_1) = \alpha_*(\lambda_2)$. Since, $\lambda_1, \lambda_2 \in \mathcal{R}_{\text{crit}}$, this implies when combined with (C.21) that $\kappa_*(\lambda_1) = \kappa_*(\lambda_2)$, which contradicts (C.22), completing the proof of this part.

Let us now prove that $\mathcal{R}_{\text{crit}}$ is non-empty. To begin with, we show that $\mathcal{R}_{\text{bad}}$ is non-empty in this case. In particular, we show that $\lambda_{\min}$ defined in Lemma C.3.1 is in $\mathcal{R}_{\text{bad}}$. Since, $\delta < 1$, we have $\lambda_{\min} > 0$. Suppose that $(\beta_*(\lambda_{\min}) = 1, \kappa_*(\lambda_{\min}))$ is optimal for some $\kappa_*(\lambda_{\min})$, then, from first-order optimality conditions, $\kappa_*(\lambda_{\min})$, $\lambda_{\min}$ solves (C.8) for $\beta = 1$. But, then as in [BM12, pg. 16] $\kappa_*(\lambda_{\min}) \to \infty$. Also, since $H(\alpha, \tau, \beta)$ is concave in β, the above imply that $\frac{\partial H}{\partial \beta}\big|_{(\beta=0,\kappa\to\infty)} \geq 0$, or equivalently from (C.18),

$$\int_{\lambda_{\min}}^{\infty} h(h - \lambda_{\min}) e^{-h^2/2} \mathrm{d}h \leq \delta \sqrt{\frac{\pi}{2}}.$$

Recalling the definition of $\lambda_{\min}$ in Lemma C.3.1, it can be shown (using standard inequalities on tail functions of Gaussians) that the inequality above is violated for all $0 < \delta < 1$. Hence, it must be $\beta_*(\lambda_{\min}) < 1$. Also, $\beta_*(\lambda_{\min}) > 0$ because of (C.15). Thus, $\lambda_{\min} \in \mathcal{R}_{\text{bad}}$. To complete the proof use Lemma C.3.3 with $\lambda_1 = \lambda_{\min}$ to see that there exists $\lambda_2 \in \mathcal{R}_{\text{crit}}$. □

Lemma C.3.5 ($\delta < 1$)**.** *Suppose $\delta < 1$ and let $\lambda_{crit} \in \mathcal{R}_{crit}$. Furthermore, i) for all $\lambda \leq \lambda_{crit}$, $\alpha_*(\lambda) = \alpha_*(\lambda_{crit})$, and, ii) for all $\lambda > \lambda_{crit}$, $\kappa_*(\lambda)$ is the unique solution to (C.8) for $\beta = 1$.*

Proof. Existence and uniqueness of λ_{crit} is proved in Lemma C.3.4

i) For $\lambda \leq \lambda_{\text{crit}}$, the claim follows directly from Lemma C.3.2.

ii) Next, we show that for $\lambda \geq \lambda_{\text{crit}}$, there exists an optimal solution for which $\beta_*(\lambda) = 1$. This suffices since then $\kappa_*(\lambda)$ solves (C.8) for $\beta = 1$ (by first order optimality conditions), and, also, the solution is unique by [DMM11],[BM12, Prop. 1.3] and the fact that $\lambda_{\min} \leq \lambda_{\text{crit}} \leq \lambda$. To see that $\beta_*(\lambda) = 1$, we argue as follows. First, $\beta_*(\lambda) \notin (0, 1)$. Otherwise, $\lambda \in \mathcal{R}_{\text{bad}}$, thus, by Lemma C.3.3 there

exists $\lambda' > \lambda \geq \lambda_{\text{crit}}$ such that $\lambda' \in \mathcal{R}_{\text{crit}}$, which contradicts the uniqueness of λ_{crit}. Hence, $\beta_*(\lambda) = 1$. □

Lemma C.3.6 ($\delta > 1$)**.** *Suppose $\delta > 1$, then for all $\lambda > 0$, $\kappa_*(\lambda)$ is the unique solution to* (C.8) *for $\beta = 1$.*

Proof. First, let us show that for $\lambda \to 0$, $\beta_*(\lambda) = 1$. Indeed for $\beta = 1$ and $\lambda \to 0$, (C.18) gives

$$\frac{\partial H}{\partial \beta} = \delta - \mathbb{E}[(h + \frac{1}{\kappa} X_0)H] = \delta - 1 > 0.$$

Thus, from concavity of H with respect to β, we find that the unique optimal value for β is

$$\beta_*(\lambda \to 0) = 1. \tag{C.23}$$

Also, as in the proof of Lemma C.3.5, $\beta_*(\lambda \to \infty) = 1$. Thus, again similar to Lemma C.3.5, it suffices to prove that there exists no $\lambda \in \mathcal{R}_{\text{bad}}$. For the sake of contradiction, suppose that there exists $\lambda_1 \in \mathcal{R}_{\text{bad}}$. By Lemma C.3.3, there exists $\lambda_1 < \lambda_{\text{crit}} \in \mathcal{R}_{\text{crit}}$. But, then $\beta_*(\lambda \to 0) \to 0$, which contradicts (C.23). This completes the proof. □

Proof. (of Theorem 6.6.1) The claim of the theorem is now a direct consequence of Lemmas C.3.5 and C.3.6 combined with (C.17). □

Appendix D

CALCULATING THE SUMMARY PARAMETERS

The upper bounds on the NSE of the generalized LASSO presented in Sections 7.4-7.9 are in terms of the summary parameters $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ and $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$. (The same quantities appeared in the study of noiseless problems in Section 2.2.) While the bounds are simple, concise and nicely resemble the corresponding ones in the case of OLS, it may appear to the reader that the formulae are rather abstract, because of the presence of $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$.

However, as discussed here, for a large number of widely-used convex regularizers $f(\cdot)$, one can calculate (tight) upper bounds or even explicit formulae for these quantities. For example, for the estimation of a k-sparse signal $\mathbf{x}_0$ with $f(\cdot) = \|\cdot\|_1$, it has been shown that $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \lesssim 2k(\log\frac{n}{k}+1)$. Substituting this into Theorems 7.5.1 and 7.8.1 results in the "closed-form" upper bounds given in (7.14) and (7.16), i.e. ones expressed only in terms of m, n and k. Analogous results have been derived [Cha+12; OH10; Sto09a; Ame+13; FM14] for other well-known signal models as well, including low rankness and block-sparsity. The first column of Table D.1 summarizes some of the results for $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ found in the literature [Cha+12; Ame+13; FM14]. The second column provides closed form results on $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ when λ is sufficiently large [OTH13b, App. H]. Note that, by setting λ to its lower bound in the second row, one approximately obtains the corresponding result in the first row. This should not be surprising due to (7.35). Also, this value of λ is a good proxy for the optimal regularizer λ_{best} of the ℓ_2-LASSO as was discussed in Sections 7.6 and 7.8.

Table D.1: Closed form upper bounds for $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$.

	$\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$	$\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$
k**-sparse,** $\mathbf{x}_0 \in \mathbb{R}^n$	$2k(\log\frac{n}{k}+1)$	$(\lambda^2+3)k$ for $\lambda \geq \sqrt{2\log\frac{n}{k}}$
Rank r**,** $\mathbf{X}_0 \in \mathbb{R}^{\sqrt{n}\times\sqrt{n}}$	$6\sqrt{n}r$	$\lambda^2 r + 2\sqrt{n}(r+1)$ for $\lambda \geq 2n^{1/4}$
k**-block sparse,** $\mathbf{x}_0 \in \mathbb{R}^{tb}$	$4k(\log\frac{t}{k}+b)$	$(\lambda^2+b+2)k$ for $\lambda \geq \sqrt{b}+\sqrt{2\log\frac{t}{k}}$

We refer the reader to [Cha+12; Ame+13; FM14; OTH13b] for the details and state-of-the-art bounds on $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$. Identifying the subdifferential $\partial f(\mathbf{x}_0)$ and calculating $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ for all $\lambda \geq 0$, are the critical steps. Once those are available, computing $\min_{\lambda\geq 0}\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ provides upper approximation formulae for $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$. This idea was first introduced by Stojnic [Sto09b] and was subsequently refined and generalized in [Cha+12]. Most recently [Ame+13; FM14] proved (7.35), thus showing that the resulting approximation on $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ is in fact highly-accurate. Section 4 of [Ame+13] is an excellent reference for further details and the notation used there is closer to ours.

We should emphasize that examples of regularizers are not limited to the ones discussed here and presented in Table D.1. There are increasingly more signal classes that exhibit low-dimensionality and to which the theorems of Sections 7.4-7.9 would apply. Some of these are as follows.

- Non-negativity constraint: $\mathbf{x}_0$ has non-negative entries, [DT05].
- Low-rank plus sparse matrices: $\mathbf{x}_0$ can be represented as sum of a low-rank and a sparse matrix, [Wri+13].
- Signals with sparse gradient: Rather than $\mathbf{x}_0$ itself, its gradient $\mathbf{d}_{\mathbf{x}_0}(i) = \mathbf{x}_0(i) - \mathbf{x}_0(i-1)$ is sparse, [CX13].
- Low-rank tensors: $\mathbf{x}_0$ is a tensor and its unfoldings are low-rank matrices, [KSV13; GRY11].
- Simultaneously sparse and low-rank matrices: for instance, $\mathbf{x}_0 = \mathbf{s}\mathbf{s}^T$ for a sparse vector $\mathbf{s}$, [Oym+15; RSV12].

Establishing new and tighter analytic bounds for $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ and $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ for more regularizers f is certainly an interesting direction for future research. In the case where such analytic bounds do not already exist in literature or are hard to derive, one can numerically estimate $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ and $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ once there is an available characterization of the subdifferential $\partial f(\mathbf{x}_0)$. Using the concentration property of $\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))$ around $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$, when $\mathbf{h} \sim \mathcal{N}(0, \mathbf{I}_n)$, we can compute $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$, as follows:

1. draw a vector $\mathbf{h} \sim \mathcal{N}(0, \mathbf{I}_n)$,
2. return the solution of the convex program $\min_{\mathbf{s}\in\partial f(\mathbf{x}_0)} \|\mathbf{h} - \lambda\mathbf{s}\|^2$.

Computing $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ can be built on the same recipe by recognizing $\text{dist}^2(\mathbf{h}, \text{cone}(\partial f(\mathbf{x}_0)))$ as $\min_{\lambda\geq 0, \mathbf{s}\in\partial f(\mathbf{x}_0)} \|\mathbf{h} - \lambda\mathbf{s}\|^2$.

To sum up, any bound on $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ and $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ translates, through Theorems 7.5.1–7.9.1, into corresponding upper bounds on the NSE of the generalized LASSO. For purposes of illustration and completeness, we review next the details of computing $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$ and $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ for the celebrated case where $\mathbf{x}_0$ is sparse and the ℓ_1-norm is used as the regularizer.

D.1 Sparse Signals

Suppose $\mathbf{x}_0$ is a k-sparse signal and $f(\cdot) = \|\cdot\|_1$. Denote by S the support set of $\mathbf{x}_0$, and by S^c its complement. The subdifferential at $\mathbf{x}_0$ is [Roc97],

$$\partial f(\mathbf{x}_0) = \{\mathbf{s} \in \mathbb{R}^n \mid \|\mathbf{s}\|_\infty \leq 1 \text{ and } \mathbf{s}_i = \text{sign}(\mathbf{x}_{0,i}), \forall i \in S\}.$$

Let $\mathbf{h} \in \mathbb{R}^n$ have i.i.d $\mathcal{N}(0, 1)$ entries and define

$$\text{shrink}(\chi, \lambda) = \begin{cases} \chi - \lambda & , \chi > \lambda, \\ 0 & , -\lambda \leq \chi \leq \lambda, \\ \chi + \lambda & , \chi < -\lambda. \end{cases}$$

Then, $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ is equal to

$$\begin{aligned} \mathbf{D}(\lambda\partial f(\mathbf{x}_0)) &= \mathbb{E}[\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))] \\ &= \sum_{i\in S} \mathbb{E}[(\mathbf{h}_i - \lambda\text{sign}((\mathbf{x}_0)_i))^2] + \sum_{i\in S^c} \mathbb{E}[\text{shrink}^2(\mathbf{h}_i, \lambda)] = \\ &= k(1+\lambda^2) + (n-k)\sqrt{\frac{2}{\pi}}\left[(1+\lambda^2)\int_\lambda^\infty e^{-t^2/2}\mathrm{d}t - \lambda\exp(-\lambda^2/2)\right]. \end{aligned} \tag{D.1}$$

Note that $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ depends only on n, λ and $k = |S|$, and *not* explicitly on S itself (which is not known). Substituting the expression in (D.1) in place of the $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ in Theorems 7.6.1 and 7.8.2, yields explicit expressions for the corresponding upper bounds in terms of n, m, k and λ.

We can obtain an even simpler upper bound on $\mathbf{D}(\lambda\partial f(\mathbf{x}_0))$ which does not involve error functions as we show below. Denote $Q(t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-\tau^2/2}\mathrm{d}\tau$ the complemen-

tary c.d.f. of a standard normal random variable. Then,

$$\begin{aligned}
\frac{1}{2}\mathbb{E}[\text{shrink}^2(\mathbf{h}_i, \lambda)] &= \int_\lambda^\infty (t-\lambda)^2 \mathrm{d}(-Q(t)) \\
&= -\left[(t-\lambda)^2 Q(t)\right]_\lambda^\infty + 2\int_\lambda^\infty (t-\lambda)Q(t)\mathrm{d}t \\
&\le \int_\lambda^\infty (t-\lambda)e^{-t^2/2}\mathrm{d}t \qquad \text{(D.2)} \\
&\le e^{-\lambda^2/2} - \frac{\lambda^2}{\lambda^2+1}e^{-\lambda^2/2} \qquad \text{(D.3)} \\
&= \frac{1}{\lambda^2+1}e^{-\lambda^2/2}.
\end{aligned}$$

(D.2) and (D.3) follow from standard upper and lower tail bounds on normal random variables, namely $\frac{1}{\sqrt{2\pi}}\frac{t}{t^2+1}e^{-t^2/2} \le Q(t) \le \frac{1}{2}e^{-t^2/2}$. From this, we find that

$$\mathbf{D}(\lambda\partial f(\mathbf{x}_0)) \le k(1+\lambda^2) + (n-k)\frac{2}{\lambda^2+1}e^{-\lambda^2/2}.$$

Letting $\lambda \ge \sqrt{2\log(\frac{n}{k})}$ in the above expression recovers the corresponding entry in Table D.1:

$$\mathbf{D}(\lambda\partial f(\mathbf{x}_0)) \le (\lambda^2+3)k, \text{ when } \lambda \ge \sqrt{2\log(\frac{n}{k})}. \qquad \text{(D.4)}$$

Substituting (D.4) in Theorems 7.6.1 and 7.8.2 recovers the bounds in (7.15) and (7.17), respectively.

Setting $\lambda = \sqrt{2\log(\frac{n}{k})}$ in (D.4) provides an approximation to $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0)))$. In particular, $\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \le 2k(\log(\frac{n}{k}) + 3/2)$. [Cha+12] obtains an even tighter bound ($\mathbf{D}(\text{cone}(\partial f(\mathbf{x}_0))) \le 2k(\log(\frac{n}{k}) + 3/4)$ starting again from (D.1), but using different tail bounds for Gaussians. We refer the reader to Proposition 3.10 in [Cha+12] for the exact details.

Appendix E

PROOFS FOR CHAPTER 8

Here we include a detailed proof of Theorem 8.2.1. In the last section, we provide a short overview of the proof of Theorem 8.2.2 which follows along the same key ideas.

Preliminaries

We rewrite (8.1) in a more convenient format for the purposes of the analysis. In particular, we perform the following operations in the order in which they appear: (i) substitute $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \sigma\mathbf{q}$, (ii) change the decision variable to the quantity of interest, i.e. the normalized error vector $\mathbf{w} := (1/\sigma)(\mathbf{x} - \mathbf{x}_0)$, (iii) move the constraint on $\mathbf{w}$ to the objective function by introducing a Lagrange multiplier λ, and, (iv) rescale by a factor of σ. Then,

$$\hat{\mathbf{w}} := \min_{\mathbf{w}} \max_{\lambda \geq 0} \|\mathbf{A}\mathbf{w} - \mathbf{q}\|_2 + \frac{\lambda}{\sigma}(f(\mathbf{x}_0 + \sigma\mathbf{w}) - f(\mathbf{x}_0)). \tag{E.1}$$

We will derive a precise expression for the limiting (as $n \to \infty$) behavior of $\lim_{\sigma \to 0} \|\hat{\mathbf{w}}\|_2$. Note that after the normalization of $\mathbf{x} - \mathbf{x}_0$ with σ , it is not guaranteed that the optimal minimizer in (E.1) is bounded (think of $\sigma \to 0$). However, we will prove that in the regime of Theorem 8.2.1 this is indeed the case. Many of the arguments that we use in the analysis require boundedness of the constraint sets. To tackle this, we assume that $\hat{\mathbf{w}}$ is bounded by some large constant $K > 0$ (with probability one over $\mathbf{A}, \mathbf{q}$), the value of which to be chosen at the end of the analysis. Recall that at that point we will have a precise characterization of the limiting behavior of $\|\hat{\mathbf{w}}\|_2$, say α_*. If α_* turns out to be independent of the value of K which we started with, then we will assume that this starting value was strictly larger than α_*. Thus, in what follows, we let $K, \Lambda, M, \ldots$ denote such (arbitrarily) large positive quantities. Also, throughout the proof we write $\|\cdot\|$ instead of $\|\cdot\|_2$.

Preparing the grounds for applying the CGMT

Using Lemma 8.3.2, onwards we work with the following (probabilistically) equivalent formulation of (E.1):

$$\hat{\mathbf{w}} := \min_{\|\mathbf{w}\|\leq K} \max_{\lambda\geq 0} \|(\mathbf{G}\mathbf{G}^T)^{-1/2}\mathbf{G}(\mathbf{w}-\mathbf{q})\|_2 + \frac{\lambda}{\sigma}(f(\mathbf{x}_0+\sigma\mathbf{w})-f(\mathbf{x}_0)). \tag{E.2}$$

The goal of this section is to bring this in a format for which CGMT is applicable. We start by using the fact that for any $\mathbf{a}\in\mathbb{R}^m$:

$$\|\mathbf{a}\| = \max_{\|\mathbf{b}\|\leq 1} \mathbf{b}^T\mathbf{a}.$$

In particular, the first term in (E.2) can be expressed as follows; to shorten notation denote $\mathbf{c} := \mathbf{w}-\mathbf{q}$:

$$\begin{aligned}\|(\mathbf{G}\mathbf{G}^T)^{-1/2}\mathbf{G}\mathbf{c}\| &= \max_{\|\mathbf{b}\|\leq 1} \mathbf{b}^T(\mathbf{G}\mathbf{G}^T)^{-1/2}\mathbf{G}\mathbf{c}\\ &= \max_{\|(\mathbf{G}\mathbf{G}^T)^{1/2}\mathbf{b}\|\leq 1} \mathbf{b}^T\mathbf{G}\mathbf{c}\\ &= \max_{\|\mathbf{G}^T\mathbf{b}\|\leq 1} \mathbf{b}^T\mathbf{G}\mathbf{c} \qquad \text{(E.3)}\\ &= \max_{\|\mathbf{b}\|\leq \Lambda} \mathbf{b}^T\mathbf{G}\mathbf{c} - \delta(\mathbf{G}^T\mathbf{b}|\mathcal{B}^{n-1}). \qquad \text{(E.4)}\end{aligned}$$

In the last line above, $\delta(\mathbf{a}|\mathcal{B}^{n-1})$ denotes the indicator function of the unit ball, i.e. takes the value 0 if $\|\mathbf{a}\|\leq 1$ and $+\infty$, otherwise. Also, we are allowed to assume that $\mathbf{b}$ is bounded by some large $0\leq\Lambda\leq\infty$, since the set of optima in (E.3) is a compact set ($\mathbf{G}^T$ has full column rank with probability one). It can be readily checked (also, see [Roc97]) that $\delta(\mathbf{a}|\mathcal{B}^{n-1}) = \sup_{\ell}\mathbf{a}^T\ell - \|\ell\|$, for any $\mathbf{a}\in\mathbb{R}^n$. Thus, continuing from (E.4):

$$\|(\mathbf{G}\mathbf{G}^T)^{-1/2}\mathbf{G}\mathbf{c}\| = \max_{\|\mathbf{b}\|\leq\Lambda}\inf_{\ell}\mathbf{b}^T\mathbf{G}(\mathbf{c}-\ell)+\|\ell\|.$$

As a final step, we will flip the order of max-min above. This is allowed by [Roc97, Cor. 37.3.2] since: (i) the objective function above is continuous, convex in ℓ, and concave in $\mathbf{b}$, (ii) the constraint sets are convex, (iii) the set constraining the maximization is bounded. Thus,

$$\|(\mathbf{G}\mathbf{G}^T)^{-1/2}\mathbf{G}\mathbf{c}\| = \inf_{\ell}\max_{\|\mathbf{b}\|\leq\Lambda}\mathbf{b}^T\mathbf{G}(\mathbf{c}-\ell)+\|\ell\|.$$

We argue that the infimum above is achieved over a bounded set. Indeed, performing the maximization over $\mathbf{b}$ above

$$\inf_{\boldsymbol{\ell}} \max_{\|\mathbf{b}\|\leq\Lambda} \mathbf{b}^T\mathbf{G}(\mathbf{c}-\boldsymbol{\ell}) + \|\boldsymbol{\ell}\| = \inf_{\boldsymbol{\ell}} \Lambda\|\mathbf{G}(\mathbf{c}-\boldsymbol{\ell})\| + \|\boldsymbol{\ell}\|.$$

The sub-level sets of the (continuous) objective function in the minimization on the right-hand side of the equation above are clearly bounded. Hence, by Weierstrass' Theorem [BNO03, Prop. 2.1.1] the set of minimum is nonempty and compact. We may thus assume there exists large but finite N such that constraining the minimization over $\|\boldsymbol{\ell}\| \leq N$ does not increase the optimum. We may now substitute the above in (E.2) to conclude with:

$$\hat{\mathbf{w}} = \min_{\substack{\|\mathbf{w}\|\leq K\\ \|\boldsymbol{\ell}\|\leq N}} \max_{\substack{\lambda\geq 0\\ \|\mathbf{b}\|\leq\Lambda}} \mathbf{b}^T\mathbf{G}(\mathbf{w}-\mathbf{q}-\boldsymbol{\ell}) + \|\boldsymbol{\ell}\| + \frac{\lambda}{\sigma}(f(\mathbf{x}_0+\sigma\mathbf{w}) - f(\mathbf{x}_0)),$$

or, re-defining $\boldsymbol{\ell} := \mathbf{w}-\mathbf{q}-\boldsymbol{\ell}$ and appropriately adjusting N:

$$\hat{\mathbf{w}} = \min_{\substack{\|\mathbf{w}\|\leq K\\ \|\boldsymbol{\ell}\|\leq N}} \max_{\substack{\lambda\geq 0\\ \|\mathbf{b}\|\leq\Lambda}} \mathbf{b}^T\mathbf{G}\boldsymbol{\ell} + \|\mathbf{w}-\mathbf{q}-\boldsymbol{\ell}\| + \frac{\lambda}{\sigma}(f(\mathbf{x}_0+\sigma\mathbf{w}) - f(\mathbf{x}_0)). \tag{E.5}$$

This brings (E.1) in the desired format for the application of the CGMT. In particular, identify $\psi([\boldsymbol{\ell},\mathbf{w}],\mathbf{b}) := \|\mathbf{w}-\mathbf{q}-\boldsymbol{\ell}\| + \max_{\lambda\geq 0}\frac{\lambda}{\sigma}(f(\mathbf{x}_0+\sigma\mathbf{w}) - f(\mathbf{x}_0))$ which is continuous and convex in $[\boldsymbol{\ell},\mathbf{w}]$, as desired. This format is of course the same as in (8.9), modulo the boundedness constraints which were not regarded in the main body of the paper.

The (AO) for arbitrary σ

Let us write the (AO) as it corresponds to (E.5):

$$\tilde{\mathbf{w}}(\mathbf{g},\mathbf{h},\mathbf{q}) = \min_{\substack{\|\mathbf{w}\|\leq K\\ \|\boldsymbol{\ell}\|\leq N}} \max_{\substack{\lambda\geq 0\\ \|\mathbf{b}\|\leq\Lambda}} \|\boldsymbol{\ell}\|\mathbf{g}^T\mathbf{b} - \|\mathbf{b}\|\mathbf{h}^T\boldsymbol{\ell} + \|\mathbf{w}-\mathbf{q}-\boldsymbol{\ell}\| + \frac{\lambda}{\sigma}(f(\mathbf{x}_0+\sigma\mathbf{w}) - f(\mathbf{x}_0)). \tag{E.6}$$

Our goal in the rest of the section is to simplify (E.6). The technique is same as described in Chapter 5, only the details differ. By massaging the objective functions and performing minimizations/maximizations when possible we eventually

reach an equivalent formulation, in which most optimizations are in terms of scalar variables instead of vectors. Two remarks are in place:

(a) We will need to flip the order of min-max several times; except if stated differently we apply [Roc97, Cor. 37.3.2]: here, constraint sets will always be convex and the objective function continuous. We only need to worry about convexity of the objective and boundedness of (at least one of) the constraint sets.

(b) To keep notation short, we will often drop the set constraints over the optimization variables when clear from context. Recall that most of the constraints are just boundedness constraints by constants that can be chosen large.

Maximizing over the direction of b. This is easy to perform, note that $\max_{\|\mathbf{b}\|=\beta} \mathbf{g}^T\mathbf{b} = \beta\|\mathbf{g}\|_2, \beta \geq 0$.

Minimizing over $\boldsymbol{\ell}$. First, let us argue briefly that we can "push" the minimization over $\boldsymbol{\ell}$ on the right of the maximization: (i) it can be seen that after optimizing over the direction of $\mathbf{b}$, the objective function in (E.6) is convex in $\boldsymbol{\ell}$, (ii) it is also concave in λ, β, and (iii) $\boldsymbol{\ell}$ is constrained in a bounded set.

To be able to optimize over $\boldsymbol{\ell}$, we use the following trick. We will express the terms $\|\boldsymbol{\ell}\|$ and $\|\mathbf{w} - \mathbf{q} - \boldsymbol{\ell}\|$ using the fact that:

$$\sqrt{x} = \min_{p \geq 0} \frac{x}{2p} + \frac{p}{2}, \quad \forall x > 0. \tag{E.7}$$

Also, note that the set of minima above is clearly bounded for bounded x. With these,

$$\begin{aligned}
&\min_{\boldsymbol{\ell}} \beta(\|\boldsymbol{\ell}\|\|\mathbf{g}\| - \mathbf{h}^T\boldsymbol{\ell}) + \|\mathbf{w} - \mathbf{q} - \boldsymbol{\ell}\| = \\
&\min_{\substack{0\leq p\leq P \\ 0\leq t\leq T}} \frac{p+t}{2} + \frac{1}{2p}\|\mathbf{q} - \mathbf{w}\|^2 + \\
&\quad \min_{\boldsymbol{\ell}} \frac{1}{p}\left[\frac{t + \beta^2 p\|\mathbf{g}\|^2}{2t}\|\boldsymbol{\ell}\|^2 + (-\beta p\mathbf{h} + \mathbf{q} - \mathbf{w})^T\boldsymbol{\ell}\right],
\end{aligned}$$

and the minimization over $\boldsymbol{\ell}$ contributes the term:

$$-\frac{1}{2p}\frac{t}{t + \beta^2 p\|\mathbf{g}\|^2}\| - \beta p\mathbf{h} + \mathbf{q} - \mathbf{w}\|^2.$$

Linearize f. The function f is continuous and convex, thus, we can express it in terms of its convex conjugate $f^*(\mathbf{u}) = \sup_{\mathbf{x}} \mathbf{u}^T\mathbf{x} - f(\mathbf{x})$. In particular, applying [Roc97, Thm.12.2] we have $f(\mathbf{x}_0 + \sigma\mathbf{w}) = \sup_{\mathbf{u}} \mathbf{x}_0^T\mathbf{u} + \sigma\mathbf{u}^T\mathbf{w} - f^*(\mathbf{u})$. The supremum

here is achieved at $\mathbf{u}_* \in \partial f(\mathbf{x}_0 + \sigma\mathbf{w})$ [Roc97, Thm. 23.5]. Also, from [BNO03, Prop. 4.2.3], $\cup_{\|\mathbf{w}\|\le K}\partial f(\mathbf{x}_0 + \sigma\mathbf{w})$ is bounded. Thus, the set of maximizers $\mathbf{u}_*$ is bounded and for some $0 < M := M(K) < \infty$, $\tilde{\mathbf{w}}$ is given as the solution to

$$\max_{\substack{\lambda\ge 0\\ 0\le\beta\le\Lambda\\ \|\mathbf{u}\|\le M}} \min_{\mathbf{w},p,t} \frac{p+t}{2} + \frac{1}{2p}\|\mathbf{q}-\mathbf{w}\|^2 + \lambda\mathbf{u}^T\mathbf{w} - \frac{1}{2p}\frac{t}{t+\beta^2 p\|\mathbf{g}\|^2}\|-\beta p\mathbf{h}+\mathbf{q}-\mathbf{w}\|^2 + \frac{\lambda}{\sigma}F(\mathbf{u}), \tag{E.8}$$

where we have flipped the orders of min-max for $\mathbf{w}$ and $\mathbf{u}$, and have denoted

$$F(\mathbf{u}) := \mathbf{u}^T\mathbf{x}_0 - f^*(\mathbf{u}) - f(\mathbf{x}_0).$$

Redefine variables. It will be convenient for the calculations to follow to redefine the variables β and t as follows:

$$\beta := \beta p, \quad t := tp \quad \text{and} \quad \lambda := \lambda p.$$

It can be checked that with these changes, the optimization remains convex.

Minimizing over the direction of $\mathbf{w}$. Evaluating the squares in (E.8) and after some algebra, it can be shown that the terms in which $\mathbf{w}$ appears are as follows:

$$\frac{\beta^2\|\mathbf{g}\|^2}{2(\beta^2\|\mathbf{g}\|^2+t)}\|\mathbf{w}\|^2 - (\tilde{\mathbf{f}}-\lambda\mathbf{u})^T\mathbf{w}, \tag{E.9}$$

where

$$\tilde{\mathbf{f}} := \left(-\frac{\beta t}{\beta^2\|\mathbf{g}\|^2+t}\mathbf{h} + \frac{\beta^2\|\mathbf{g}\|^2}{\beta^2\|\mathbf{g}\|^2+t}\mathbf{q}\right), \tag{E.10}$$

which has entries i.i.d. Gaussians of zero mean and standard deviation

$$\sigma_{\tilde{\mathbf{f}}} := \sigma_{\tilde{\mathbf{f}}}(\beta,t) := \frac{\beta\sqrt{t^2+\beta^2\|\mathbf{g}\|^4}}{\beta^2\|\mathbf{g}\|^2+t}. \tag{E.11}$$

Fix the norm of $\|\mathbf{w}\| = \alpha$. Optimizing over the direction of $\mathbf{w}$ the second term in (E.9) gives $-\alpha\|\tilde{\mathbf{f}}-\lambda\mathbf{u}\|$.

Minimize over p. Overall, the min-max problem in (E.6) has reduced itself to:

$$\max_{\substack{\lambda\geq 0\\ 0\leq\beta\leq\Lambda\\ \|\mathbf{u}\|\leq M}} \min_{\alpha,p,t} \Big\{\frac{1}{2p}\Big(t+\|\mathbf{q}\|^2-\frac{t}{\beta^2\|\mathbf{g}\|^2+t}\|\beta\mathbf{h}-\mathbf{q}\|^2+ \frac{\beta^2\|\mathbf{g}\|^2}{\beta^2\|\mathbf{g}\|^2+t}\alpha^2-2\alpha\|\tilde{\mathbf{f}}-\lambda\mathbf{u}\|+2\frac{\lambda}{\sigma}F(\mathbf{u})\Big)+\frac{p}{2}\Big\} =$$
$$\max_{\substack{\lambda\geq 0\\ 0\leq\beta\leq\Lambda\\ \|\mathbf{u}\|\leq M}} \min_{\alpha,t} \Big(t+\|\mathbf{q}\|^2-\frac{t}{\beta^2\|\mathbf{g}\|^2+t}\|\beta\mathbf{h}-\mathbf{q}\|^2+ \frac{\beta^2\|\mathbf{g}\|^2}{\beta^2\|\mathbf{g}\|^2+t}\alpha^2-2\alpha\|\tilde{\mathbf{f}}-\lambda\mathbf{u}\|+2\frac{\lambda}{\sigma}F(\mathbf{u})\Big)^{1/2}. \tag{E.12}$$

In yielding the equality above, we have applied (E.7).

Redifine λ. It is convenient to redefine λ as $\lambda := \lambda/\sigma_{\tilde{\mathbf{f}}}$. Let $\mathbf{f}$ denote standard i.i.d. Gaussian vector, such that $\tilde{\mathbf{f}} \sim \sigma_{\tilde{\mathbf{f}}}\mathbf{f}$. With these, we can express $\tilde{\mathbf{w}}$ as the solution to:

$$\max_{\substack{\lambda\geq 0\\ 0\leq\beta\leq\Lambda\\ \|\mathbf{u}\|\leq M}} \min_{\alpha,t} \Big(t+\|\mathbf{q}\|^2-\frac{t}{\beta^2\|\mathbf{g}\|^2+t}\|\beta\mathbf{h}-\mathbf{q}\|^2+ \frac{\beta^2\|\mathbf{g}\|^2}{\beta^2\|\mathbf{g}\|^2+t}\alpha^2-2\sigma_{\tilde{\mathbf{f}}}(\alpha\|\mathbf{f}-\lambda\mathbf{u}\|-2\frac{\lambda}{\sigma}F(\mathbf{u}))\Big). \tag{E.13}$$

Note that we have essentially considered the square of (E.13). Let us denote the optimal cost of (E.13) above as $\phi(\sigma) := \phi(\sigma;\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f})$.

The (AO) in the limit $\sigma \to 0$

Theorem 3.3.1 relates $\|\tilde{\mathbf{w}}\|$ to $\|\hat{\mathbf{w}}\|$, under appropriate assumptions. Also, recall that we wish to characterize $\lim_{\sigma\to 0}\|\hat{\mathbf{w}}\|$. Thus, in view of (E.13), we wish to analyze the problem

$$\phi_0 := \phi_0(\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f}) := \lim_{\sigma\to 0}\phi(\sigma;\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f}).$$

In (E.13), from Fenchel's inequality:

$$F(\mathbf{u}) = \mathbf{u}^T\mathbf{x}_0 - f^*(\mathbf{u}) - f(\mathbf{x}_0) \leq 0. \tag{E.14}$$

With this observation, we prove in the next lemma that $\phi(\sigma;\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f})$ is non-decreasing in σ.

Lemma E.0.1. *Fix* $\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f}$ *and consider* $\phi(\cdot;\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f}) : (0,\infty) \to \mathbb{R}$ *as defined in* (E.13). $\phi(\sigma;\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f})$ *is non-decreasing in* σ.

Proof. Denote $\mathcal{L}(\sigma, \alpha, t, \beta, \mathbf{u}, \lambda)$ the objective function in (E.13) and consider $0 < \sigma_1 < \sigma_2 < \infty$. Let $\alpha^{(2)}, t^{(2)}$ be an optimal solution to the min-max problem in (E.13) for σ_2. Then, let $(\beta^{(1)}, \mathbf{u}^{(1)}, \lambda^{(1)}) = \arg\max_{\beta,\mathbf{u},\lambda} \mathcal{L}(\sigma_1, \alpha^{(2)}, t^{(2)}, \beta, \mathbf{u}, \lambda)$. Clearly,

$$\phi(\sigma_1) \leq \mathcal{L}(\sigma_1, \alpha^{(2)}, t^{(2)}, \beta^{(1)}, \mathbf{u}^{(1)}, \lambda^{(1)}).$$

Using $1/\sigma_1 > 1/\sigma_2$ and (E.14),

$$\mathcal{L}(\sigma_1, \alpha^{(2)}, t^{(2)}, \beta^{(1)}, \mathbf{u}^{(1)}, \lambda^{(1)}) \leq \mathcal{L}(\sigma_2, \alpha^{(2)}, t^{(2)}, \beta^{(1)}, \mathbf{u}^{(1)}, \lambda^{(1)}).$$

But,

$$\mathcal{L}(\sigma_2, \alpha^{(2)}, t^{(2)}, \beta^{(1)}, \mathbf{u}^{(1)}, \lambda^{(1)}) \leq \phi(\sigma_2).$$

Combine the above chain of inequalities to conclude. □

In particular, when viewed as a function of $\kappa := 1/\sigma$, $\phi(\cdot; \mathbf{g}, \mathbf{h}, \mathbf{q}, \mathbf{f})$ is non-increasing. Thus,

$$\phi_0 = \lim_{\sigma \to 0} \phi(\sigma) = \lim_{\kappa \to \infty} \phi(\kappa) = \inf_{\kappa \geq 0} \phi(\kappa). \tag{E.15}$$

Next, we argue that we can flip the order of min-max. The objective function in (E.13) is continuous, convex in κ, and, concave in $\lambda, \beta, \mathbf{u}$. The constraint set on λ appears to be unbounded, but, it can be checked from (E.13) that the optimal value is in fact bounded. With this and (E.15), we get

$$\max_{\substack{\lambda \geq 0 \\ 0 \leq \beta \leq \Lambda \\ \|\mathbf{u}\| \leq M}} \min_{\alpha, t} \inf_{\kappa \geq 0} \Big(t + \|\mathbf{q}\|^2 - \frac{t}{\beta^2 \|\mathbf{g}\|^2 + t} \|\beta \mathbf{h} - \mathbf{q}\|^2 + \frac{\beta^2 \|\mathbf{g}\|^2}{\beta^2 \|\mathbf{g}\|^2 + t} \alpha^2 - 2\sigma_{\mathbf{f}} \alpha \|\tilde{\mathbf{f}} - \lambda \mathbf{u}\| + \kappa 2 \sigma_{\tilde{\mathbf{f}}} \lambda F(\mathbf{u}) \Big).$$

Recall (E.14) and the fact that equality is achieved iff $\mathbf{u} \in \partial f(\mathbf{x}_0)$ (e.g. [Roc97, Thm. 23.5]). Then, ϕ_0 is given by

$$\max_{\substack{\lambda \geq 0 \\ 0 \leq \beta \leq \Lambda \\ \mathbf{u} \in \partial f(\mathbf{x}_0)}} \min_{\alpha, t} \Big(t + \|\mathbf{q}\|^2 - \frac{t}{\beta^2 \|\mathbf{g}\|^2 + t} \|\beta \mathbf{h} - \mathbf{q}\|^2 + \frac{\beta^2 \|\mathbf{g}\|^2}{\beta^2 \|\mathbf{g}\|^2 + t} \alpha^2 - 2\sigma_{\mathbf{f}} \alpha \|\tilde{\mathbf{f}} - \lambda \mathbf{u}\| \Big),$$

where we have assumed $\infty > M > \max_{\mathbf{s} \in \partial f(\mathbf{x}_0)} \|\mathbf{s}\|$. We can now optimize over $\lambda \mathbf{u}$ (after appropriately flipping the order of min-max): $\min_{\lambda \geq 0, \mathbf{u} \in \partial f(\mathbf{x}_0)} \|\mathbf{f} - \lambda \mathbf{u}\| =$

$\text{dist}(\mathbf{f}, \text{cone}(\partial f(\mathbf{x}_0)))$. Thus, we conclude with "Gordon's optimization" for $\sigma \to 0$ taking the form:

$$\phi_0(\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f}) = \min_{0\le\alpha\le K} \mathcal{L}(\alpha;\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f}), \tag{E.16}$$
$$\mathcal{L}(\alpha;\mathbf{g},\mathbf{h},\mathbf{q},\mathbf{f}) := \min_{0\le t\le T}\max_{0\le\beta\le\Lambda} \Big\{ t + \|\mathbf{q}\|^2 - \frac{t}{\beta^2\|\mathbf{g}\|^2+t}\|\beta\mathbf{h}-\mathbf{q}\|^2 + \frac{\beta^2\|\mathbf{g}\|^2}{\beta^2\|\mathbf{g}\|^2+t}\alpha^2 - 2\sigma_{\tilde{\mathbf{f}}}\alpha\mathbf{d}_{\mathbf{h}}\Big\},$$

where we have denoted $\mathbf{d}_{\mathbf{h}} := \text{dist}(\mathbf{f}, \text{cone}(\partial f(\mathbf{x}_0)))$.

It is now easy to optimize (E.18) over α. We summarize the result in the following lemma.

Lemma E.0.2. *In* (E.18)*, fix* $\mathbf{g},\mathbf{h},\mathbf{q}$ *and let* $\tilde{\mathbf{w}} := \tilde{\mathbf{w}}(\mathbf{g},\mathbf{h},\mathbf{q})$ *be optimal. Denote,*

$$\tilde{\mathbf{f}} := \tilde{\mathbf{f}}(\beta,t) := -\frac{\beta t}{\beta^2\|\mathbf{g}\|^2+t}\mathbf{h} + \frac{\beta^2\|\mathbf{g}\|^2}{\beta^2\|\mathbf{g}\|^2+t}\mathbf{q},$$

and $\upsilon(\beta,t) := \textit{dist}(\tilde{\mathbf{f}}, \text{cone}(\lambda\partial f(\mathbf{x}_0)))$*. Then,*

$$\|\tilde{\mathbf{w}}\| = \frac{\beta^2\|\mathbf{g}\|^2+t}{\beta\|\mathbf{g}\|^2}\upsilon(\beta,t), \tag{E.17}$$

where β, t *are optimal solutions to the following optimization:*

$$\max_{\Lambda\ge\beta\ge 0}\min_{T\ge t\ge 0}\Big(t + \|\mathbf{q}\|^2 - \frac{t}{\beta^2\|\mathbf{g}\|^2+t}\|\beta\mathbf{h}-\mathbf{q}\|^2 + - \frac{t^2+\beta^2\|\mathbf{g}\|^4}{\|\mathbf{g}\|^2(\beta^2\|\mathbf{g}\|^2+t)}\upsilon(\beta,t)\Big). \tag{E.18}$$

Note that $\tilde{\mathbf{f}} \sim \sigma_{\tilde{\mathbf{f}}}\mathbf{f}$ where $\mathbf{f}$ is standard i.i.d. Gaussian and

$$\sigma_{\tilde{\mathbf{f}}} := \sigma_{\tilde{\mathbf{f}}}(\beta,t) := \beta\sqrt{t^2+\beta^2\|\mathbf{g}\|^4}/(\beta^2\|\mathbf{g}\|^2+t).$$

Probabilistic Analysis

Lemma E.0.2 derives an expression for $\|\tilde{\mathbf{w}}\|$, for fixed $\mathbf{g},\mathbf{h},\mathbf{q}$. Here, we evaluate the limiting behavior of this expression. Recall that $\mathbf{g},\mathbf{h},\mathbf{q}$ are all i.i.d. standard Gaussian vectors and assume the large-system limit linear regime as in the statement of Theorem 8.2.1. Recall the use of the following notation: let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and $\{c_n\}$ a deterministic sequence, then $X_n \xrightarrow{P} c_n$ iff for all $\epsilon > 0$, the event $|X_n - c_n| \le \epsilon c_n$ occurs w.p. 1 in the limit $n \to \infty$. For the purpose of this section, convergence is to be understood in the aforementioned meaning.

From standard concentration results on Gaussian r.v.s.: $\|\mathbf{g}\|^2 \xrightarrow{P} m$, $\|\mathbf{h}\|^2 \xrightarrow{P} n$, $\|\mathbf{q}\|^2 \xrightarrow{P} n$, $\|\beta\mathbf{h} - \sigma\mathbf{q}\|^2 \xrightarrow{P} (\beta^2 + \sigma^2)n$, and $\upsilon_{f,\mathbf{x}_0}(\beta, t; \mathbf{g}, \mathbf{h}, \mathbf{v}) \xrightarrow{P} \sigma_{\tilde{\mathbf{f}}}\omega_{f,\mathbf{x}_0}$. For the last relation, we have used the property of the Gaussian width as in[Ame+13, Prop. 10.1]. Hence, for any fixed β, t, the objective function in (E.18) converges to

$$d(\beta, t) = t + \frac{\beta^2(m-t)}{\beta^2 m + t}n - \frac{t^2 + \beta^2 m^2}{m(\beta^2 m + t)}\overline{\omega}^2_{f,\mathbf{x}_0}. \tag{E.19}$$

It can be checked that the objective function in (E.18) is convex in t and concave in β. Also, the constraint sets are compact. Thus, it follows from [AG82, Cor. II.1] ("point-wise convergence in probability of concave functions implies uniform convergence in compact spaces") that the convergence in (E.19) is uniform over β and t. As will be shown next, provided that the constants determining the constraint sets are large enough, then there exist unique β_*^2 and t_* that are optimal in (E.19). Hence, as in [NM94, Thm. 2.7], the optimal solutions of (E.18) indeed converge to the deterministic solutions of (E.19), which we calculate below. Let the constant bounds on the variables β, t, namely Λ, T, to be specified later. Denote β_*, t_* optimal solutions in

$$\max_{0\le\beta\le B}\min_{0\le t\le T} d(\beta, t).$$

Let us write $\omega := \omega_{f,\mathbf{x}_0}$. We differentiate the objective with respect to both β and t to find:

$$\frac{\partial d(\beta_*, t_*)}{\partial\lambda} = 1 - \frac{\beta_*^2 m(\beta_*^2 - 1)}{(t_* + \beta_*^2 m)}n - \frac{t_*^2 + 2t_*\beta_*^2 m - m^2\beta_*^2}{(\beta_*^2 m + t_*)^2}\frac{\omega^2}{m}, \tag{E.20a}$$

$$\frac{\partial d(\beta_*, t_*)}{\partial\beta} = \frac{2\beta_* t_*(m - t_*)}{(\beta_*^2 m + t_*)^2}(n - \omega^2). \tag{E.20b}$$

Setting them to zero, from (E.20b) we have $\beta_* = 0$, $t_* = 0$ or $t_* = \sigma m$. We consider each case separately. Assume $\beta_* = 0$, then $t_* = \arg\min d(\lambda, \beta_*) = \arg\min \lambda - \lambda\frac{\omega^2}{m} = 0$ and $d(t_*, \beta_*) = 0$. Next, suppose $t_* = m$. Substituting this in (E.20a) we find

$$(n-m)\beta_*^4 + ((n-m) - (m-\omega^2))\beta_*^2 - (m - \omega^2) = 0.$$

Solving this yields $\beta_*^2 = \frac{m-\omega^2}{n-m}$ and $d(\beta_*, t_*) = m - \omega^2 > 0$. Choose, Λ, T such that β_*, t_* are feasible. From convexity, first-order optimality conditions are sufficient.

What is left is to substitute those limit values β_* and t_* in (E.17) in Lemma E.0.2, to conclude with

$$\frac{\|\tilde{\mathbf{w}}\|^2}{\sigma^2} \xrightarrow{P} \frac{\overline{\omega}^2_{f,\mathbf{x}_0}(n - \overline{\omega}^2_{f,\mathbf{x}_0})}{m - \overline{\omega}^2_{f,\mathbf{x}_0}}.$$

Proof Outline of Theorem 8.2.2

In the next few lines we outline only the main checkpoints involved in the proof of Theorem 8.2.1. The analysis follows along the same lines as in the proof of Theorem 8.2.1. In fact, things here are less involved since we are only interested in lower bounding the optimal cost of a min-max problem. Hence, a single application of the GMT Theorem 3.2.1 (and not the CGMT 3.3.1) suffices.

Denote, $C := \mathcal{T}_f(\mathbf{x}_0) \cap \mathcal{S}^{n-1}$. We write the mCSV of $\mathbf{A}$ as

$$\sigma_{\min}(\mathbf{A}; \mathcal{T}_f(\mathbf{x}_0)) = \min_{\mathbf{x}\in C} \max_{\|\mathbf{y}\|\leq 1} \mathbf{y}^T \mathbf{A}\mathbf{x}. \tag{E.21}$$

We prove a high-probability lower bound on the optimal cost of this min-max optimization. We do so by applying Gordon's GMT, just as is done in the Gaussian case. But first, we need to bring (E.21) in a format where GMT is applicable After applying the GMT, it can be shown that it suffices to lower bound the optimal cost of the following "Gordon's optimization" problem instead:

$$\min_{\mathbf{x}\in C, \boldsymbol{\ell}} \max_{B\geq\beta\geq 0} \|\mathbf{x}-\boldsymbol{\ell}\| + \beta(\|\boldsymbol{\ell}\|\|\mathbf{g}\| - \mathbf{h}^T\boldsymbol{\ell}). \tag{E.22}$$

Next, we perform a deterministic (fixed $\mathbf{g}, \mathbf{h}$) analysis of this to simplify it as possible into a scalar optimization problem. Caution should be taken here that the constraint set C on $\mathbf{x}$ is non-convex, thus we are not allowed to flip min-max operations "carelessly". It can be shown that (E.23) has optimal cost $\sqrt{F}$, where $F := F(\mathbf{g}, \mathbf{h})$ is the optimal cost of the following optimization:

$$\min_{\mathbf{x}\in C, T\geq t\geq 0} \max_{B\geq\beta\geq 0} \frac{\|\mathbf{g}\|^2 - t\beta^2\|\mathbf{h}\|^2 - 2t\beta\mathbf{h}^T\mathbf{x} + t\beta^2\|\mathbf{g}\|^2 + t^2\beta^2}{\|\mathbf{g}\|^2 + t}. \tag{E.23}$$

Now, it is easy to optimize over $\mathbf{x}$ by choosing it to maximize $\mathbf{h}^T\mathbf{x}$ in C and F is the optimal cost to only a scalar optimization problem involving the r.v.s. $\|\mathbf{g}\|$, $\|\mathbf{h}\|$ and $\max_{\mathbf{x}\in C} \mathbf{h}^T\mathbf{x}$. All three are 1-Lipschitz functions thus they concentrate (thus, converge in the proportional regime) to their mean values $\sqrt{m}$, $\sqrt{n}$ and $\omega_{f,\mathbf{x}_0}$ respectively. Also, the problem is convex in β, t, thus we can yield the expression of Theorem 8.2.2 (with the correspondence $\beta \leftrightarrow \chi$, $t \leftrightarrow \rho$), by first-order optimality conditions.

Appendix F

PROOFS FOR CHAPTER 10

F.1 Proof of Theorem 10.1.1

The proof of the theorem is of course based on the CGMT framework and is almost identical to the proof of Theorem 9.4.1. In particular, from Section 9.5, it suffices to prove that the BER of the corresponding (AO) problem converges to the desired quantity $Q(1/\tau_*)$. Hence, we only include the part of the proof that involves the analysis of the (AO).

For simplicity, we write $\|\cdot\|$ for the ℓ_2-norm.

The error vector. As usual, it is convenient to re-write (10.1a) by changing the variable to the error vector $\mathbf{w} := \mathbf{x} - \mathbf{x}_0$:

$$\hat{\mathbf{w}} := \min_{-2 \le \mathbf{w}_i \le 0} \|\mathbf{z} - \mathbf{A}\mathbf{w}\|. \tag{F.1}$$

Without loss of generality we assume for the analysis that $\mathbf{x}_0 = \mathbf{1}_n = (1, 1, \ldots, 1)$. Then, we can write (10.2a) in terms of the error vector $\mathbf{w}$ as: $BER = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\hat{\mathbf{w}}_i \le -1\}}$.

Identifying the (PO) and the (AO). Using the CGMT for the analysis of the BER, requires as a first step expressing the optimization in (10.1a) in the form of a (PO) as it appears in (3.11a). It is easy to see that (F.1) is equivalent to

$$\min_{-2 \le \mathbf{w}_i \le 0} \max_{\|\mathbf{u}\| \le 1} \mathbf{u}^T \mathbf{A}\mathbf{w} - \mathbf{u}^T \mathbf{z}. \tag{F.2}$$

Observe that the constraint sets above are both convex and compact; also, the objective function is convex in $\mathbf{w}$ and concave in $\mathbf{u}$. Hence, according to the CGMT we can perform the analysis of the BER for the corresponding (AO) problem instead, which becomes (note the normalization to account for the variance of the entries of $\mathbf{A}$)

$$\frac{1}{\sqrt{n}} \min_{-2 \le \mathbf{w}_i \le 0} \max_{\|\mathbf{u}\| \le 1} (\|\mathbf{w}\|\mathbf{g} - \sqrt{n}\mathbf{z})^T \mathbf{u} - \|\mathbf{u}\| \mathbf{h}^T \mathbf{w}. \tag{F.3}$$

We refer to the optimization in (F.3) as the (AO) problem.

Computing the BER via the (AO). Call $\tilde{\mathbf{w}}$ the optimal solution of the (AO). Fix any $\epsilon > 0$ and consider the set

$$\mathcal{S} = \{\mathbf{v} : \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{v}_i \leq -1\}} - Q(1/\tau_*)\right| < \epsilon\}, \tag{F.4}$$

where τ_* is defined in the statement of Theorem 10.1.1. We will apply Theorem 3.3.1 for the above set $\mathcal{S}$. In particular, we show that (i) the (AO) in (F.3) converges in probability (after proper normalization with n), and, (ii) $\tilde{\mathbf{w}} \in \mathcal{S}$ with probability one. These will suffice to conclude that $\hat{\mathbf{w}} \in \mathcal{S}$ with probability one, which would complete the proof of Theorem 10.1.1.

Simplifying the (AO). We begin by simplifying the (AO) problem as it appears in (F.3). First, since both $\mathbf{g}$ and $\mathbf{z}$ have entries iid Gaussian, then, $\|\mathbf{w}\|\mathbf{g} - \sqrt{n}\mathbf{z}$ has entries iid $\mathcal{N}(0, \sqrt{\|\mathbf{w}\|^2 + n\sigma^2})$. Hence, for our purposes and using some abuse of notation so that $\mathbf{g}$ continues to denote a vector with iid standard normal entries, the first term in (F.3) can be treated as $\sqrt{\|\mathbf{w}\|^2 + n\sigma^2}\mathbf{g}^T\mathbf{u}$, instead. As a next step, fix the norm of $\mathbf{u}$ to say $\|\mathbf{u}\| = \beta$. Optimizing over its direction is now straightforward, and gives $\min_{-2\leq \mathbf{w}_i \leq 0} \max_{0\leq\beta\leq 1} \frac{\beta}{\sqrt{n}}\left(\sqrt{\|\mathbf{w}\|^2 + n\sigma^2}\|\mathbf{g}\| - \mathbf{h}^T\mathbf{w}\right)$. In fact, it is easy to now optimize over β as well; its optimal value is 1 if the term in the parenthesis is non-negative and is 0 otherwise. With this, the (AO) simplifies to the following:

$$(\min_{-2\leq \mathbf{w}_i \leq 0} \sqrt{\frac{\|\mathbf{w}\|^2}{n} + \sigma^2}\|\mathbf{g}\| - \frac{1}{\sqrt{n}}\mathbf{h}^T\mathbf{w})_+, \tag{F.5}$$

where we defined $(\chi)_+ := \max\{\chi, 0\}$. To facilitate the optimization over $\mathbf{w}$, we express the term in the square-root in a variational form, using $\sqrt{\chi} = \inf_{\tau>0} \frac{\tau}{2} + \frac{\chi}{2\tau}$. With this trick, the minimization over the entries of $\mathbf{w}$ becomes separable:

$$\min_{\tau\geq 0} \frac{\tau\|\mathbf{g}\|}{2} + \frac{\sigma^2\|\mathbf{g}\|}{2\tau} + \sum_{i=1}^{n} \min_{-2\leq \mathbf{w}_i \leq 0} \frac{\|\mathbf{g}\|}{2\tau n}\mathbf{w}_i^2 - \frac{\mathbf{h}_i}{\sqrt{n}}\mathbf{w}_i.$$

Then, the optimal $\tilde{\mathbf{w}}_i$ satisfies

$$\tilde{\mathbf{w}}_i = \begin{cases} 0 & , \text{if } \mathbf{h}_i \geq 0, \\ \frac{\tau\sqrt{n}}{\|\mathbf{g}\|}\mathbf{h}_i & , \text{if } -\frac{2\|\mathbf{g}\|}{\tau\sqrt{n}} \leq \mathbf{h}_i < 0, \\ -2 & , \text{if } \mathbf{h}_i < -\frac{2\|\mathbf{g}\|}{\tau\sqrt{n}}. \end{cases} \tag{F.6}$$

where τ is the solution to the following:

$$\left(\min_{\tau>0} \frac{\tau\|\mathbf{g}\|}{2} + \frac{\sigma^2\|\mathbf{g}\|}{2\tau} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \upsilon(\tau; \mathbf{h}_i, \|\mathbf{g}\|)\right)_+, \tag{F.7}$$

$$\upsilon(\tau;\mathbf{h}_i,\|\mathbf{g}\|):=\begin{cases}0 & ,\text{if } \mathbf{h}_i\geq 0,\\ -\frac{\tau\sqrt{n}}{2\|\mathbf{g}\|}\mathbf{h}_i^2 & ,\text{if } -\frac{2\|\mathbf{g}\|}{\tau\sqrt{n}}\leq \mathbf{h}_i<0,\\ 2\frac{\|\mathbf{g}\|}{\tau\sqrt{n}}+2\mathbf{h}_i & ,\text{if } \mathbf{h}_i\leq -\frac{2\|\mathbf{g}\|}{\tau\sqrt{n}}.\end{cases}$$

Convergence of the (AO). Now that the (AO) is simplified as in (F.7), we can get a handle on the limiting behavior of the optimization itself as well as of the optimal $\tilde{\mathbf{w}}$. But first, we need to properly normalize the (AO) by dividing the objective in (F.7) by $\sqrt{n}$. Also, for convenience, redefine $\tau:=\frac{\tau}{\sqrt{\delta}}$. By the WLLN, we have $\frac{\|\mathbf{g}\|}{\sqrt{n}}\xrightarrow{P}\sqrt{\delta}$, and, for all $\tau>0$, $\frac{1}{n}\sum_{i=1}^n\upsilon(\tau;\mathbf{h}_i,\|\mathbf{g}\|)\xrightarrow{P}Y(\tau):=-\frac{\tau}{2}\int_0^{\frac{2}{\tau}}h^2p(h)\mathrm{d}h+\frac{2}{\tau}Q\left(\frac{2}{\tau}\right)-2\int_{\frac{2}{\tau}}^{\infty}hp(h)\mathrm{d}h$. With these we can evaluate the point-wise (in τ) limit of the objective function in (F.7). Next, we use the fact that the objective is convex in τ and Lemma [AG82, Cor.. II.1], to conclude that the convergence is indeed uniform in τ. Hence, the random optimization in (F.7) converges to the following deterministic optimization $\min_{\tau>0}\frac{\tau\delta}{2}+\frac{\sigma^2}{2\tau}+Y(\tau)$; some algebra shows that the latter is the same as (10.3). If $\delta>1/2$, then, it can be shown via differentiation that the objective function of it is strictly convex. Also, it is nonnegative; thus, the entire expression in (F.7), which is nothing but the (AO) problem we started with, converges in probability to (10.3). What is more, using [NM94, Thm. 2.7] it can be shown that the optimal $\tau_*(\mathbf{g},\mathbf{h})$ of the (AO) converges in probability to the unique optimal solution τ_* of (10.3). This is crucial for the final step of the proof.

Proving $\tilde{\mathbf{w}}\in\mathcal{S}$. Recall the definition in (F.4). We prove that $\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\{\tilde{\mathbf{w}}_i\leq -1\}}\xrightarrow{P}Q(1/\tau_*)$. From (F.6), $\mathbb{1}_{\{\tilde{\mathbf{w}}_i\leq -1\}}=\mathbb{1}_{\{\mathbf{h}_i\leq -\frac{\|\mathbf{g}\|}{\sqrt{n}\sqrt{\delta}\tau}\}}$. Recall, $\|\mathbf{g}\|/\sqrt{n}\xrightarrow{P}\sqrt{\delta}$ and $\tau\xrightarrow{P}\tau_*$. Conditioning on those high probability events it can be shown that $\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\{\mathbf{h}_i\leq -\frac{\|\mathbf{g}\|}{\sqrt{n}\sqrt{\delta}\tau}\}}\xrightarrow{P}\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\{\mathbf{h}_i\leq -\frac{1}{\tau_*}\}}\xrightarrow{P}Q(\frac{1}{\tau_*})$.

F.2 Proof of Theorem 10.3.1

The proof of the theorem is an almost straightforward extension of the proof of Theorem 10.1.1. We omit most of the details for brevity.

It can easily be verified that the corresponding (AO) becomes (cf. Eqn. (F.5)):

$$\min_{\tau\geq 0}\frac{\tau\|\mathbf{g}\|}{2}+\frac{\sigma^2\|\mathbf{g}\|}{2\tau}+\sum_{i=1}^n\min_{\ell_i\leq\mathbf{w}_i\leq u_i}\frac{\|\mathbf{g}\|}{2\tau n}\mathbf{w}_i^2-\frac{\mathbf{h}_i}{\sqrt{n}}\mathbf{w}_i,$$

where $\ell_i=-(M-1)-\mathbf{x}_{0,i}$ and $u_i=(M-1)-\mathbf{x}_{0,i}$. For simplicity in notation,

denote $A = \frac{\|\mathbf{g}\|}{\tau\sqrt{n}}$. Then, the optimal $\tilde{\mathbf{w}}_i$ satisfies

$$\tilde{\mathbf{w}}_i = \begin{cases} \ell_i & , \text{if } \mathbf{h}_i < A\ell_i, \\ \frac{1}{A}\mathbf{h}_i & , \text{if } A\ell_i \le \mathbf{h}_i \le Au_i, \\ u_i & , \text{if } \mathbf{h}_i > Au_i. \end{cases} \tag{F.8}$$

where, τ is the solution to the following:

$$\left(\min_{\tau>0} \frac{\tau\|\mathbf{g}\|}{2} + \frac{\sigma^2\|\mathbf{g}\|}{2\tau} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \upsilon(\tau; \mathbf{h}_i, \|\mathbf{g}\|) \right)_+, \tag{F.9}$$

$$\upsilon(\tau; \mathbf{h}_i, \|\mathbf{g}\|) := \begin{cases} \frac{A}{2}\ell_i^2 - \mathbf{h}_i\ell_i & , \text{if } \mathbf{h}_i < A\ell_i, \\ -\frac{1}{2A}\mathbf{h}_i^2 & , \text{if } A\ell_i \le \mathbf{h}_i \le Au_i, \\ \frac{A}{2}u_i^2 - \mathbf{h}_i u_i & , \text{if } \mathbf{h}_i > Au_i. \end{cases}$$

Proceeding exactly as in the case of BPSK signal recovery, we need to properly normalize the (AO) by dividing the objective in (F.9) by $\sqrt{n}$. Also, for convenience, redefine $\tau := \frac{\tau}{\sqrt{\delta}}$. By the WLLN, we have $\frac{\|\mathbf{g}\|}{\sqrt{n}} \xrightarrow{P} \sqrt{\delta}$, and using the assumption on uniform distribution of the entries of $\mathbf{x}_0$ over the constellation, we further have that for all $\tau > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} \upsilon(\tau; \mathbf{h}_i, \|\mathbf{g}\|) \xrightarrow{P} Y(\tau),$$

where

$$Y(\tau) := \frac{1}{M} \sum_{i=1,3,\ldots,M-1} \int_{\frac{-u_i}{\tau}}^{\frac{\ell_i}{\tau}} \tau\frac{h^2}{2} p(h)\mathrm{d}h + \int_{-\infty}^{\frac{-u_i}{\tau}} (\frac{1}{2\tau}u_i^2 + hu_i)p(h)\mathrm{d}h + \int_{\frac{\ell_i}{\tau}}^{\infty} (\frac{1}{2\tau}\ell_i^2 - h\ell_i)p(h)\mathrm{d}h$$
$$+ \frac{1}{M} \sum_{i=1,3,\ldots,M-1} \int_{\frac{-\ell_i}{\tau}}^{\frac{u_i}{\tau}} \tau\frac{h^2}{2} p(h)\mathrm{d}h + \int_{-\infty}^{\frac{-\ell_i}{\tau}} (\frac{1}{2\tau}\ell_i^2 + h\ell_i)p(h)\mathrm{d}h + \int_{\frac{u_i}{\tau}}^{\infty} (\frac{1}{2\tau}u_i^2 - hu_i)p(h)\mathrm{d}h,$$

and (using some abuse of notation) $\ell_i := (M-1) - i$ and $u_i = (M-1) + i$. With some algebra, this can be simplified to

$$Y(\tau) := \frac{1}{M} \sum_{i=1,3,\ldots,M-3} \left(-\tau + \tau \int_{\frac{\ell_i}{\tau}}^{\infty} (h - \frac{\ell_i}{\tau})^2 p(h)\mathrm{d}h + \tau \int_{\frac{u_i}{\tau}}^{\infty} (h - \frac{u_i}{\tau})^2 p(h)\mathrm{d}h. \right)$$
$$+ \frac{1}{M} \left(-\frac{\tau}{2} + \tau \int_{\frac{2(M-1)}{\tau}}^{\infty} (h - \frac{2(M-1)}{\tau})^2 p(h)\mathrm{d}h. \right) \tag{F.10}$$

Hence, the random optimization in (F.9) converges to the following deterministic optimization $\min_{\tau>0} \frac{\tau\delta}{2} + \frac{\sigma^2}{2\tau} + Y(\tau)$; some algebra shows that the latter is the same as

(10.7). Here, we also used the fact that for $\mathbf{x}_0$ sampled uniformly from an M-PAM constellation, it holds $\mathbb{E}[\mathbf{x}_{0,i}^2] = \frac{2}{M}\sum_{i=1,3\ldots,M-1} i^2 = (M^2-1)/3$, hence

$$\sigma^2 = \left(\frac{M^2-1}{3}\right)\frac{1}{\mathrm{SNR}}.$$

Finally, it can be readily verified that the optimal cost of the minimization in (10.7) is nonnegative if $\delta > 1 - 1/M$.

Using this fact, and with arguments same as in the proof of Theorem 10.1.1, it can be shown that the optimal $\tau_*(\mathbf{g}, \mathbf{h})$ of the (AO) converges in probability to the unique optimal solution τ_* of (10.7).

To complete the proof, we need to show that $\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\mathbf{x}_i^* \neq \mathbf{x}_{0,i}\}} \xrightarrow{P} 2(1-1/M)Q(1/\tau_*)$. It is easily seen that if $\mathbf{x}_{0,i} \in \{\pm 1, \pm 3, \ldots, \pm(M-3)\}$, then there is an error when $|\tilde{\mathbf{w}}_i| > 1$. From (F.8) this event corresponds to $|\mathbf{h}_i| > A$. On the other hand, if $\mathbf{x}_{0,i} = M-1$ (or $\mathbf{x}_{0,i} = -(M-1)$), then the error event corresponds to $\tilde{\mathbf{w}}_i < -1$ (or $\tilde{\mathbf{w}}_i > 1$, reps.), which in view of (F.8) translates to $\mathbf{h}_i < -A$ (or $\mathbf{h}_i > A$, resp.). Putting these together and conditioning on the high-probability events $\|\mathbf{g}\|/\sqrt{n} \xrightarrow{P} \sqrt{\delta}$ and $\tau \xrightarrow{P} \tau_*$, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\arg\min_{s\in C} |\mathbf{x}_{0,i}+\tilde{\mathbf{w}}_i - s| \neq \mathbf{x}_{0,i}\}} &\xrightarrow{P} \frac{1}{M}\left(2(M-2)Q(1/\tau_*) + 2Q(1/\tau_*)\right) \\ &= 2(1-1/M)Q(1/\tau_*).\end{aligned}$$

Appendix G

PROOFS FOR CHAPTER 11

G.1 Proof of Theorem 11.2.1

In Chapter 11 we presented Corollary 11.2.1 as a consequence of Theorem 11.2.1. To prove the result, we follow the reverse direction, i.e. we first prove Corollary 11.2.1 and subsequently show how Theorem 11.2.1 follows from that. The proof is based on the CGMT framework and it follows largely the same steps outlined in Chapter 5. Nevertheless, a few crucial modifications are required to account for the non-linear nature of the measurements.

Assume a sequence of problem instances as described in Section 11.2. To keep notation simple, we simply use $\|\mathbf{v}\|$ (rather than $\|\mathbf{v}\|_2$) for the Euclidean norm of $\mathbf{v}$ and we shall also drop the superscript (n) when referring to elements of the sequence. Thus, we write

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{\sqrt{n}}\|\vec{g}(\mathbf{A}\mathbf{x}_0) - \mathbf{A}\mathbf{x}\| + \frac{\lambda}{\sqrt{n}} f(\mathbf{x}), \tag{G.1}$$

but it is to be understood that the above actually produces a sequence of solutions $\hat{\mathbf{x}}^{(n)}$ indexed by n. Our goal is to characterize the nontrivial limiting behavior of $\|\hat{\mathbf{x}} - \mu\mathbf{x}_0\|$.

We start with a simple but useful change of variables $\mathbf{w} := \mathbf{x} - \mu\mathbf{x}_0$, to directly get a handle on the error vector $\mathbf{w}$. Then, (G.1) becomes:

$$\begin{aligned}\hat{\mathbf{w}} &:= \arg\min_{\mathbf{w}} \frac{1}{\sqrt{n}}\|\vec{g}(\mathbf{A}\mathbf{x}_0) - \mu\mathbf{A}\mathbf{x}_0 - \mathbf{A}\mathbf{w}\| + \frac{\lambda}{\sqrt{n}} f(\mu\mathbf{x}_0 + \mathbf{w}) \\ &= \arg\min_{\mathbf{w}} \max_{\|\mathbf{u}\|\le 1} \frac{1}{\sqrt{n}}(-\mathbf{u}^T\mathbf{A}\mathbf{w} + \mathbf{u}^T(\vec{g}(\mathbf{A}\mathbf{x}_0) - \mu\mathbf{A}\mathbf{x}_0)) + \frac{\lambda}{\sqrt{n}} f(\mu\mathbf{x}_0 + \mathbf{w}),\end{aligned} \tag{G.2}$$

where the second line follows after using the fact $\|\mathbf{v}\| = \max_{\|\mathbf{u}\|\le 1} \mathbf{u}^T\mathbf{v}$.

A Key Decomposition

The first key step in the proof is a trick adapted from the proofs of [Ver10b, Lem. 4.3] and [PVY14, Thm. 1.3]. Until further notice, we condition on $\mathbf{x}_0$. Also, we repeatedly make use of the assumption that $\|\mathbf{x}_0\| = 1$ without direct reference. The trick amounts to decomposing each measurement vector $\mathbf{a}_i$ in its projection on the direction of $\mathbf{x}_0$ and its orthogonal complement. Denoting $\mathbf{P}^{\perp} = (\mathbf{I} - \mathbf{x}_0\mathbf{x}_0^T)$ for the

projector onto the orthogonal complement of the span of $\mathbf{x}_0$ (recall $\|\mathbf{x}_0\|_2 = 1$), we have $\mathbf{a}_i^T = (\mathbf{a}_i^T\mathbf{x}_0)\mathbf{x}_0^T + \mathbf{a}_i^T\mathbf{P}^\perp$, or, in matrix form:

$$\mathbf{A} = (\mathbf{A}\mathbf{x}_0)\mathbf{x}_0^T + \mathbf{A}\mathbf{P}^\perp.$$

Then, (G.2) becomes:

$$\min_{\mathbf{w}} \max_{\|\mathbf{u}\|\le 1} \frac{1}{\sqrt{n}} \left\{-\mathbf{u}^T\mathbf{A}\mathbf{P}^\perp\mathbf{w} + \mathbf{u}^T(\vec{g}(\mathbf{A}\mathbf{x}_0) - \mu\mathbf{A}\mathbf{x}_0 - (\mathbf{A}\mathbf{x}_0)\mathbf{x}_0^T\mathbf{w})\right\} + \frac{\lambda}{\sqrt{n}} f(\mu\mathbf{x}_0 + \mathbf{w}). \tag{G.3}$$

Using the Gaussianity assumption on the entries of $\mathbf{A}$ it is straightforward to show that $\mathbf{P}^\perp\mathbf{a}_i$ is *independent* of $\mathbf{a}_i^T\mathbf{x}_0$, for all $i = 1, \dots, m$. Also, conditioned on $\mathbf{a}_i^T\mathbf{x}_0$, $\mathbf{P}^\perp\mathbf{a}_i$ is independent of $(\vec{g}_i(\mathbf{a}_i^T\mathbf{x}_0) - \mu\mathbf{a}_i^T\mathbf{x}_0)$ since the latter only depends on $\mathbf{a}_i$ through $\mathbf{a}_i^T\mathbf{x}_0$. Combining those, it follows that $\mathbf{P}^\perp\mathbf{a}_i$ is also *independent* of $(\vec{g}_i(\mathbf{a}_i^T\mathbf{x}_0) - \mu\mathbf{a}_i^T\mathbf{x}_0)$ [Ver10b, pg. 13]. Overall, $\mathbf{A}\mathbf{P}^\perp\mathbf{w}$ is independent of the rest of the terms in in (G.3). This shows that the objective function of (G.3) is distributed identically even after replacing the $\mathbf{A}\mathbf{P}^\perp\mathbf{w}$ with $\mathbf{G}\mathbf{P}^\perp\mathbf{w}$, where $\mathbf{G}$ is an independent copy of $\mathbf{A}$. After all these, (G.3) is identically distributed with the following:

$$\min_{\mathbf{w}} \max_{\|\mathbf{u}\|\le 1} \frac{1}{\sqrt{n}} \{-\mathbf{u}^T\mathbf{G}\mathbf{P}^\perp\mathbf{w} + \mathbf{u}^T(\mathbf{z}_\mathbf{e} - (\mathbf{x}_0^T\mathbf{w})\mathbf{e})\} + \frac{\lambda}{\sqrt{n}} f(\mu\mathbf{x}_0 + \mathbf{w}), \tag{G.4}$$

where $\mathbf{G}$ and $\mathbf{e} := \mathbf{A}\mathbf{x}_0$ have entries i.i.d. standard normal and are independent of each other. Also, $\mathbf{z}_\mathbf{e} := \vec{g}(\mathbf{e}) - \mu\mathbf{e}$ for convenience.

Applying the CGMT

After the decomposition step in the previous section, we have transformed the initial problem to that of analyzing the (probabilistically) equivalent one in (G.4). In particular, we wish to evaluate the limiting behavior of $\|\hat{\mathbf{w}}\|$, i.e. the norm of the minimizer of the optimization in (G.4). The analysis is possible thanks to the CGMT framework.

In (G.4) identify the bilinear term $\mathbf{u}^T\mathbf{G}\mathbf{P}^\perp\mathbf{w}$ and note that the rest of the objective function is convex in $\mathbf{w}$ (recall that f is convex), and, linear (thus, concave) in $\mathbf{u}$. Overall, this is in the appropriate format of a (PO) problem as in (3.11a) modulo the extra factor $\mathbf{P}^\perp$ in the bilinear term. It is straightforward to show that this extra factor only requires a natural change of the corresponding terms in the (AO) problem as follows: $\|\mathbf{P}^\perp\mathbf{w}\|\mathbf{g}^T\mathbf{u} + \|\mathbf{u}\|_2\mathbf{h}^T\mathbf{P}^\perp\mathbf{w}$. With this minor modification, the CGMT continues to hold. A remaining technical caveat is that the minimization

over $\mathbf{w}$ in it appears unconstrained. For this, we assume that the minimizer of (G.4) satisfies $\|\hat{\mathbf{w}}\| \leq K_{\mathbf{w}}$ for sufficiently large constant $K_{\mathbf{w}} > 0$ *independent of* n. If our assumption is valid, then by the end of the proof we will have identified a quantity $\alpha_* > 0$ to which $\|\hat{\mathbf{w}}\|$ converges; if α_* turns out to be independent of the choice of $K_{\mathbf{w}}$, then we may explicitly choose $K_{\mathbf{w}} = 2\alpha_*$ (say) and α_* is the true limit; on the other hand, if α_* turns out to depend on $K_{\mathbf{w}}$, this means that we could have chosen $K_{\mathbf{w}}$ arbitrarily large in the first place, and so the true limit diverges. Thus, assuming that $\|\hat{\mathbf{w}}\|$ the minimization in (G.4) is not affected by imposing the constraint $\|\mathbf{w}\| \leq K_{\mathbf{w}}$. With these, we can write the corresponding (AO) problem as

$$\tilde{\mathbf{w}} = \arg \min_{\|\mathbf{w}\| \leq K_{\mathbf{w}}} \max_{\|\mathbf{u}\| \leq 1} \frac{1}{\sqrt{n}} \{ \|\mathbf{P}^{\perp}\mathbf{w}\| \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\| \mathbf{h}^T \mathbf{P}^{\perp}\mathbf{w} + \mathbf{u}^T (\mathbf{z}_{\mathbf{e}} - (\mathbf{x}_0^T \mathbf{w})\mathbf{e}) \} + \frac{\lambda}{\sqrt{n}} f(\mu \mathbf{x}_0 + \mathbf{w}). \tag{G.5}$$

We will see that analyzing this problem, with the goal of determining the converging value of the magnitude of its minimizer $\tilde{\mathbf{w}}$, is simpler than analyzing the (PO) (and certainly more so than the one we started with in (G.2)). The CGMT essentially shows that $\|\tilde{\mathbf{w}}\|$ converges to the same value as $\|\hat{\mathbf{w}}\|$. Recall $\hat{\mathbf{w}}$ being the minimizer of the (PO) and the goal of Theorem 11.2.1 being to evaluate the converging value of its magnitude.

Analysis of the Auxiliary Optimization

The goal of this section is that of analyzing the (AO) problem in (G.5). In particular, we will prove that (i) the optimal cost of the (AO) problem converges to the optimal cost of the deterministic optimization in (11.9), which involves three scalar optimization variables α, β, τ, (ii) the min-max problem in (11.9) is *strictly* convex in α and jointly concave in β, τ, (iii) $\|\tilde{\mathbf{w}}\|$ converges to the unique optima α_* in (11.9). With these, the claim of the Theorem follows by Theorem 3.3.1(iii) (see also Chapter 5).

As described in Chapter 5 the analysis requires several steps here as well. The randomness in (G.5) is over $\mathbf{e}$, $\mathbf{g}$, $\mathbf{h}$, $\mathbf{x}_0$ and possibly the link function g; at each step we condition on all but a subset of these and identify convergence of the objective function of the (AO) with respect to the remaining. Pointwise convergence (with respect to the involved optimization variables) needs to be turned into uniform convergence to guarantee that not only the objective function, but also the min/max value and the optimizer converge appropriately. (Strict) convexity of the objective will turn out to be crucial for the latter.

Introducing the Frenchel conjugate. To begin with, let us rewrite the (AO) problem above by expressing f in terms of its Frenchel conjugate, i.e.

$$f(\mathbf{x}) = \sup_{\bar{\mathbf{v}}} \bar{\mathbf{v}}^T \mathbf{x} - f^*(\bar{\mathbf{v}}) = \sup_{\mathbf{v}} \sqrt{n}\mathbf{v}^T \mathbf{x} - f^*(\sqrt{n}\mathbf{v}). \tag{G.6}$$

Translating to our problem and after rescaling this gives,

$$n^{-1/2} f(\mu \mathbf{x}_0 + \mathbf{w}) = \sup_{\mathbf{v}} \mathbf{v}^T (\mu \mathbf{x}_0 + \mathbf{w}) - n^{-1/2} f^*(\sqrt{n}\mathbf{v}). \tag{G.7}$$

Now, from standard optimality conditions of (G.6), the optimal $\bar{\mathbf{v}}_*$ satisfies $\bar{\mathbf{v}}_* \in \partial f(\mathbf{x})$. Then, using condition (b) of Section 11.2, $\|\mathbf{v}_*\| = O\left(\sqrt{n}\right)$ for all $\mathbf{x}$ such that $\|\mathbf{x}\| = O(1)$. From this, and $\|\mathbf{w} + \mu \mathbf{x}_0\| = O(1)$ we conclude that the optimal $\mathbf{v}_*$ in (G.7) satisfies $\|\mathbf{v}_*\| \leq K_{\mathbf{v}} < 0$ for sufficiently large constant $K_{\mathbf{v}}$ independent of n. Putting everything together, (G.5) is equivalent to

$$\min_{\|\mathbf{w}\| \leq K_{\mathbf{w}}} \max_{\substack{\|\mathbf{u}\| \leq 1 \\ 0 \leq \|\mathbf{v}\| \leq K_{\mathbf{v}}}} \frac{1}{\sqrt{n}} \mathbf{u}^T (\mathbf{z}_{\mathbf{e}} - (\mathbf{x}_0^T \mathbf{w})\mathbf{e} - \|\mathbf{P}^{\perp}\mathbf{w}\|\mathbf{g}) - \|\mathbf{u}\| \bar{\mathbf{h}}^T \mathbf{P}^{\perp} \mathbf{w} + \lambda \mathbf{v}^T (\mu \mathbf{x}_0 + \mathbf{w}) - \lambda \bar{f}^*(\mathbf{v}), \tag{G.8}$$

where we have also denoted $\bar{\mathbf{h}} := n^{-1/2}\mathbf{h}$ and $\bar{f}^*(\mathbf{v}) = n^{-1/2} f^*(\sqrt{n}\mathbf{v})$. Observe again that by condition (b) of Section 11.2, $\bar{f}^*(\mathbf{v}) = \max_{\mathbf{x}} \mathbf{x}^T \mathbf{v} - n^{-1/2} f(\mathbf{x}) = O(1)$ since $\mathbf{v} = O(1)$.

In order to somewhat simplify the exposition, we often omit explicitly carrying over the constraints $\|\mathbf{w}\| \leq K_{\mathbf{w}}$, $\|\mathbf{v}\| \leq K_{\mathbf{v}}$ until the very last step, but we often recall and actually make use of them.

Optimizing over the directions of u and w. Observe that the maximization over the direction of $\mathbf{u}$ is easy in (G.8), which then becomes:

$$\min_{\mathbf{w}} \max_{\substack{0 \leq \beta \leq 1 \\ \mathbf{v}}} \frac{1}{\sqrt{n}} \beta \| \mathbf{z}_{\mathbf{e}} - (\mathbf{x}_0^T \mathbf{w})\mathbf{e} - \|\mathbf{P}^{\perp}\mathbf{w}\|\mathbf{g} \| - \beta \bar{\mathbf{h}}^T \mathbf{P}^{\perp} \mathbf{w} + \lambda \mathbf{v}^T (\mu \mathbf{x}_0 + \mathbf{w}) - \lambda \bar{f}^*(\mathbf{v}). \tag{G.9}$$

At this point, the form of the objective function suggests that it is possible to do the same trick over $\mathbf{w}$, i.e. fix its magnitude and optimize over only its direction. The caveat is that the minimization over $\mathbf{w}$ in (G.9) is done only after the maximization over β and $\mathbf{v}$. What is more, the objective function is not be convex in $\mathbf{w}$; thus, flipping the order of min-max operations that would resolve the issue is not directly justified by (say) Sion's minimax theorem [Sio+58].

The very same issue was encountered in the analysis under noisy linear measurements in Section 5.2 (specifically, going from (5.8) to (5.9)). In essence, it was shown that the flipping under question is indeed possible when dimensions are large. Hence, we have:

$$\max_{\substack{0\le\beta\le 1\\ \mathbf{v}}} \min_{\mathbf{w}} \frac{\beta}{\sqrt{n}}\| \mathbf{z_e} + (\mathbf{x}_0^T\mathbf{w})\mathbf{e} - \|\mathbf{P}^\perp\mathbf{w}\|\mathbf{g} \| - \beta\bar{\mathbf{h}}^T\mathbf{P}^\perp\mathbf{w} + \lambda\mathbf{v}^T(\mu\mathbf{x}_0+\mathbf{w}) - \lambda\bar{f}^*(\mathbf{v})$$

$$= \max_{\substack{0\le\beta\le 1\\ \mathbf{v}}} \min_{\alpha_1,\alpha_2\ge 0} \frac{\beta}{\sqrt{n}}\| \mathbf{z_e} + \alpha_2\mathbf{e} - \alpha_1\mathbf{g} \| - \max_{\substack{\|\mathbf{P}^\perp\mathbf{w}\|=\alpha_1\\ \mathbf{x}_0^T\mathbf{w}=\alpha_2}} \left\{\beta\bar{\mathbf{h}}^T\mathbf{P}^\perp\mathbf{w} - \lambda\mathbf{v}^T(\mu\mathbf{x}_0+\mathbf{w}) + \lambda\bar{f}^*(\mathbf{v})\right\}.$$

By decomposing $\mathbf{w}$ as $\mathbf{P}^\perp\mathbf{w} + (\mathbf{x}_0^T\mathbf{w})\mathbf{x}_0$, it is not hard to perform the maximization over $\mathbf{w}$ to equivalently write the last display above as:

$$\max_{\substack{0\le\beta\le 1\\ \mathbf{v}}} \min_{\alpha_1,\alpha_2\ge 0} \frac{\beta}{\sqrt{n}}\| \mathbf{z_e} + \alpha_2\mathbf{e} - \alpha_1\mathbf{g} \| - \alpha_1\|\beta\mathbf{P}^\perp\bar{\mathbf{h}} - \lambda\mathbf{P}^\perp\mathbf{v}\| + \lambda\mu\mathbf{v}^T\mathbf{x}_0 + \alpha_2\lambda(\mathbf{v}^T\mathbf{x}_0) - \lambda\bar{f}^*(\mathbf{v}). \tag{G.10}$$

The randomness of e, g and g. Until further notice condition on $\bar{\mathbf{h}}$ and $\mathbf{x}_0$. All randomness in (G.10) is now on the first term.

Consider $\beta, \mathbf{v}$ fixed for now. For any pair α_1, α_2 by the WLLN, $m^{-1}\|\mathbf{z_e} + \alpha_2\mathbf{e} - \alpha_1\mathbf{g}\|^2 \xrightarrow{P} \mathbb{E}[(g(\gamma) - \mu\gamma + \alpha_2\gamma - \alpha_1\gamma')^2]$, where $\gamma, \gamma' \sim \mathcal{N}(0,1)$ and independent. Recall, $\mathbb{E}[(g(\gamma)-\mu\gamma)^2] = \sigma^2$, $\mathbb{E}[(g(\gamma)-\mu\gamma)\gamma] = \mu - \mu = 0$ and $m/n = \delta$, to conclude that $n^{-1/2}\|\mathbf{z_e} + \alpha_2\mathbf{e} - \alpha_1\mathbf{g}\| \xrightarrow{P} \sqrt{\delta}\sqrt{\sigma^2 + \alpha_1^2 + \alpha_2^2}$, where convergence is point-wise in α_1, α_2. Also, the objective function in (G.10) is jointly convex in $[\alpha_1, \alpha_2]$. Thus, point-wise convergence translates to uniform as in [AG82, Cor.. II.1] , from which, it follows that (for any $\beta, \mathbf{v}$) the minimum over α_1, α_2 in (G.10) converges to

$$\min_{\alpha_1,\alpha_2\ge 0} \beta\sqrt{\delta}\sqrt{\sigma^2 + \alpha_1^2 + \alpha_2^2} - \alpha_1\|\beta\mathbf{P}^\perp\bar{\mathbf{h}} - \lambda\mathbf{P}^\perp\mathbf{v}\| + \lambda\mu\mathbf{v}^T\mathbf{x}_0 + \alpha_2\lambda(\mathbf{v}^T\mathbf{x}_0) - \lambda\bar{f}^*(\mathbf{v}). \tag{G.11}$$

Furthermore, the function $\sqrt{\sigma^2 + \alpha_1^2 + \alpha_2^2}$ is (by direct differentiation) jointly strongly convex over $[\alpha_1, \alpha_2]$; thus (G.11) has a unique minimizer. Then, we can apply the Argmin theorem [NM94, Thm. 2.7] to conclude that the optimal α_1, α_2 of (G.10) converge to the corresponding (unique) optima of (G.11).

Up to now, $\beta, \mathbf{v}$ were assumed fixed and the convergence from (G.10) to (G.11) holds point-wise with respect to $\beta, \mathbf{v}$. The point-wise minimum of concave functions is still concave, thus, uniform convergence is indeed true by [NM94, Thm. 2.7]

. Hence, (G.10) converges to

$$\max_{\substack{0\le\beta\le 1\\ \mathbf{v}}} \min_{\alpha_1,\alpha_2\ge 0} \beta\sqrt{\delta}\sqrt{\sigma^2+\alpha_1^2+\alpha_2^2} - \alpha_1\|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{P}^{\perp}\mathbf{v}\| + \lambda\mu\mathbf{v}^T\mathbf{x}_0 + \alpha_2\lambda(\mathbf{v}^T\mathbf{x}_0) - \lambda\bar{f}^*(\mathbf{v}), \tag{G.12}$$

and the optimal α_1, α_2 of the former converge to the corresponding optima of the latter.

Merging α_1 and α_2. It is important to note that $\alpha_1^2+\alpha_2^2$ in (G.12) correspond exactly to the squared norm of the error. Here, we simplify (G.12) by introducing the quantity $\alpha_1^2+\alpha_2^2$ as the minimization variable rather than sperately α_1 and α_2. By first order optimality conditions in (G.12) we find

$$\alpha_1\beta\sqrt{\delta} = \|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{P}^{\perp}\mathbf{v}\|\sqrt{\alpha_1^2+\alpha_2^2+\sigma^2} \quad \text{and} \quad -\alpha_2\beta\sqrt{\delta} = \lambda\mathbf{v}^T\mathbf{x}_0\sqrt{\alpha_1^2+\alpha_2^2+\sigma^2}. \tag{G.13}$$

Substituting this in (G.12), the objective becomes (ignoring the terms that do not involve α_1 or α_2):

$$\beta\sqrt{\delta}\sqrt{\sigma^2+\alpha_1^2+\alpha_2^2} - \frac{\sqrt{\sigma^2+\alpha_1^2+\alpha_2^2}}{\beta\sqrt{\delta}}\left(\|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{P}^{\perp}\mathbf{v}\|^2 + (\lambda\mathbf{v}^T\mathbf{x}_0)^2\right).$$

But, from (G.13) we find $\sqrt{\sigma^2+\alpha_1^2+\alpha_2^2}\sqrt{\|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{P}^{\perp}\mathbf{v}\|^2 + (\lambda\mathbf{v}^T\mathbf{x}_0)^2} = \beta\sqrt{\delta}\sqrt{\alpha_1^2+\alpha_2^2}$. Combining, we conclude that (G.12) can be written as

$$\max_{\substack{0\le\beta\le 1\\ \mathbf{v}}} \min_{\alpha\ge 0} \beta\sqrt{\delta}\sqrt{\sigma^2+\alpha^2} - \alpha\|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{v}\| + \lambda\mu\mathbf{v}^T\mathbf{x}_0 - \lambda\bar{f}^*(\mathbf{v}), \tag{G.14}$$

where the new optimization variable α plays the role of $\sqrt{\alpha_1^2+\alpha_2^2}$, thus it represents the norm of the error vector $\|\mathbf{w}\|$. We have also identified $\|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{P}^{\perp}\mathbf{v}\|^2 + (\lambda\mathbf{v}^T\mathbf{x}_0)^2 = \|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{v}\|^2$.

Introducing a new optimization variable. To get a better handle on it, we square the norm term in (G.14) at the expense of introducing a new scalar optimization variable. This is based on the following trick:

$$\sqrt{x} = \min_{\tau>0} \frac{\tau}{2} + \frac{x}{2\tau}, \tag{G.15}$$

for any $x \ge 0$. Thus, (G.14) becomes

$$\max_{\substack{0\le\beta\le 1\\ \mathbf{v},\tau>0}} \min_{\alpha\ge 0} \beta\sqrt{\delta}\sqrt{\sigma^2+\alpha^2} - \frac{\alpha\tau}{2} - \frac{\alpha}{2\tau}\|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{v}\|^2 + \lambda\mu\mathbf{v}^T\mathbf{x}_0 - \lambda\bar{f}^*(\mathbf{v}), \tag{G.16}$$

where we have also flipped the order of min-max between α and τ. We could do this as in [Roc97, Cor. 37.3.2] since the objective is convex in α and concave in τ, the constraint sets are both convex and both of them are bounded. To argue the boundedness, recall that $\alpha \le K_{\mathbf{w}}$; for τ it suffices to combine optimality conditions of (G.15) and boundedness of $\mathbf{v}$, $\|\mathbf{v}\|_2 \le K_{\mathbf{v}}$.

Optimizing over v. Note that the objective in (G.16) is concave in $\mathbf{v}$, convex in α and the constraint sets are convex compact. Thus, as might be expected by now, we use [Roc97, Cor. 37.3.2] to flip the corresponding order of max-min. Also, after some simple algebra while using $\mathbf{P}^{\perp}\mathbf{x}_0 = 0$ and $\|\mathbf{x}_0\| = 1$, it can be shown that

$$\|\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} - \lambda\mathbf{v}\|^2 - 2\frac{\tau}{\alpha}\lambda\mu\mathbf{v}^T\mathbf{x}_0 = \|\lambda\mathbf{v} - (\beta\mathbf{P}^{\perp}\bar{\mathbf{h}} + \frac{\tau}{\alpha}\mu\mathbf{x}_0)\|^2 - \mu^2\frac{\tau^2}{\alpha^2}.$$

Combining, we conclude with

$$\begin{aligned}(\text{G.16}) &= \max_{\substack{0\le\beta\le 1\\ \tau>0}} \min_{\alpha\ge 0} \beta\sqrt{\delta}\sqrt{\sigma^2+\alpha^2} - \frac{\alpha\tau}{2} + \mu^2\frac{\tau}{2\alpha} \\ &\qquad - \frac{\alpha\lambda^2}{\tau}\min_{\mathbf{v}}\left\{\frac{1}{2}\|\mathbf{v} - (\frac{\beta}{\lambda}\mathbf{P}^{\perp}\bar{\mathbf{h}} + \frac{\tau}{\alpha\lambda}\mu\mathbf{x}_0)\|^2 + \frac{\tau}{\lambda\alpha}\bar{f}^*(\mathbf{v})\right\}, \quad \text{(G.17)} \\ &= \max_{\substack{0\le\beta\le 1\\ \tau>0}} \min_{\alpha\ge 0} G(\alpha,\beta,\tau).\end{aligned}$$

Here, $G(\alpha,\beta,\tau)$ is convex in α (see (G.14)) and jointly concave in β,τ. To see the latter it suffices to show that $\frac{\alpha\lambda^2}{\tau}\|\mathbf{v} - \frac{\beta}{\lambda}(\mathbf{P}^{\perp}\bar{\mathbf{h}} + \frac{\mu\tau}{\lambda\alpha}\mathbf{x}_0)\|^2$ is jointly convex over $\beta,\tau,\mathbf{v}$ (minimization over $\mathbf{v}$ does not change the joint convexity over τ and β.). Norm is separable over its entries, so we equivalently show that for scalars τ,β,v, the function $\frac{1}{\tau}(v - c_1\beta - c_2\tau)^2$ is jointly convex over $\tau > 0,\beta$; this is true as the perspective function of $(v - c_1\beta - c_2)^2$. One more remark is in place here regarding the form of $G(\alpha,\beta,\tau)$: even-though α appears in the denominator in (G.17), the limit of $\alpha \to 0$ of the expression is finite using the continuity of the Moreau envelope[RW09]. Another way to see this is by noting that the objective in (G.17) is equivalent to that in (G.14). Hence, evaluating G at $\alpha = 0$ in the minimization in (G.17) subsumes $G(0,\beta,\tau) = \lim_{\alpha\to 0} G(\alpha,\beta,\tau)$.

The randomness of $\bar{\mathbf{h}}$ and $\mathbf{x}_0$. Fix β,τ,α, denote $c_1 = \frac{\beta}{\lambda}, c_2 = \frac{\tau\mu}{\alpha\lambda}, c_3 = \frac{\tau}{\alpha\lambda}$, and, consider

$$R(\bar{\mathbf{h}},\mathbf{x}_0) := R(\alpha,\beta,\tau;\bar{\mathbf{h}},\mathbf{x}_0) := -\frac{c_2}{2} + \frac{1}{c_2}\min_{\mathbf{v}}\left\{\frac{1}{2}\|\mathbf{v} - c_1\mathbf{P}^{\perp}\bar{\mathbf{h}} - c_2\mathbf{x}_0\|^2 + c_3\bar{f}^*(\mathbf{v})\right\}.$$

Recall from Assumption 1, and, from the modeling condition (a) in Section 11.2 that

$$A(\bar{\mathbf{h}}, \mathbf{x}_0) := R(\alpha, \beta, \tau; \bar{\mathbf{h}}, \mathbf{x}_0) := -\frac{c_2}{2}\frac{\|\bar{\mathbf{x}}_0\|^2}{n} + \frac{1}{c_2}\min_{\mathbf{v}}\left\{\frac{1}{2}\|\mathbf{v} - c_1\bar{\mathbf{h}} - c_2\frac{\bar{\mathbf{x}}_0}{\sqrt{n}}\|^2 + c_3\bar{f}^*(\mathbf{v})\right\} \tag{G.18}$$

converges to $\{-\frac{c_2}{2} + \frac{\mu}{c_2}F(c_1, c_2, c_3)\}$ in probability. Also, recall $\bar{\mathbf{x}}_0 = \mathbf{x}_0\|\bar{\mathbf{x}}_0\|$. Next, we show that for all constant $\zeta > 0$

$$|R(\bar{\mathbf{h}}, \mathbf{x}_0) - A(\bar{\mathbf{h}}, \mathbf{x}_0)| \leq \zeta \tag{G.19}$$

with probability approaching one in the limit of $n \to \infty$. Combining this with Assumption 1, will prove that $R(\bar{\mathbf{h}}, \mathbf{x}_0)$ converges in $\{-\frac{c_2}{2} + \frac{\mu}{c_2}F(c_1, c_2, c_3)\}$ in probability, as well.

Proof of (G.19): Fix any $\epsilon > 0$. We condition on the following events:

$$\begin{cases} |\bar{\mathbf{h}}^T\mathbf{x}_0| \leq \epsilon, \\ 1 - \epsilon \leq n^{-1/2}\|\bar{\mathbf{x}}_0\| \leq 1 + \epsilon. \end{cases} \tag{G.20}$$

Each one of the events occurs with probability approaching one as $n \to \infty$; the first follows since $\bar{\mathbf{h}} \sim \mathcal{N}(0, \frac{1}{n}\mathbf{I}_n)$ and $\|\mathbf{x}_0\| = 1$ and from standard tail bounds on Gaussians; the second is due to condition (a) of Section 11.2. Without loss of generality assume $R(\bar{\mathbf{h}}, \mathbf{x}_0) \geq A(\bar{\mathbf{h}}, \mathbf{x}_0)$, and let $\mathbf{v}_*$ be optimal in (G.18), then

$$\begin{aligned}
|R(\bar{\mathbf{h}}, \mathbf{x}_0) - A(\bar{\mathbf{h}}, \mathbf{x}_0)| &\leq \frac{c_2}{2}\left(\frac{\|\bar{\mathbf{x}}_0\|^2}{n} - 1\right) + \frac{1}{2c_2}\|\mathbf{v}_* - c_1\mathbf{P}^{\perp}\bar{\mathbf{h}} - c_2\mathbf{x}_0\|^2 - \frac{1}{2c_2}\|\mathbf{v}_* - c_1\bar{\mathbf{h}} - c_2\frac{\bar{\mathbf{x}}_0}{\sqrt{n}}\|^2 \\
&= \frac{c_2}{2}\left(\frac{\|\bar{\mathbf{x}}_0\|^2}{n} - 1\right) \\
&\quad + \left(\frac{c_1}{c_2}(\mathbf{x}_0^T\bar{\mathbf{h}})\mathbf{x}_0 + \bar{\mathbf{x}}_0\left(\frac{1}{\sqrt{n}} - \frac{1}{\|\bar{\mathbf{x}}_0\|}\right)\right)^T\left(\mathbf{v}_* - c_1\bar{\mathbf{h}} - \frac{1}{2}c_2\bar{\mathbf{x}}_0\left(\frac{1}{\sqrt{n}} + \frac{1}{\|\bar{\mathbf{x}}_0\|}\right) + \frac{1}{2}c_1(\mathbf{x}_0^T\bar{\mathbf{h}})\mathbf{x}_0\right) \\
&= -\frac{1}{2}\frac{c_1^2}{c_2}(\mathbf{x}_0^T\bar{\mathbf{h}})^2 + \frac{c_1}{c_2}(\mathbf{x}_0^T\mathbf{h})(\mathbf{x}_0^T\mathbf{v}_*) - c_1(\mathbf{x}_0^T\bar{\mathbf{h}})\frac{\|\bar{\mathbf{x}}_0\|}{\sqrt{n}} + (\mathbf{x}_0^T\mathbf{v}_*)\left(\frac{\|\bar{\mathbf{x}}_0\|}{\sqrt{n}} - 1\right) \\
&\leq \frac{1}{2}\frac{c_1^2}{c_2}\epsilon^2 + \frac{c_1}{c_2}\|\mathbf{v}_*\|\epsilon + c_1\epsilon(1+\epsilon) + \|\mathbf{v}_*\|\epsilon
\end{aligned} \tag{G.21}$$

where the last line follows after bounding the absolute values of the summands using (G.20). Recall now that $\|\mathbf{v}_*\| \leq K_{\mathbf{v}} < \infty$. Also, note that c_1 and $\frac{c_1}{c_2}$ are also bounded constants. Then, for all $\zeta > 0$ in (G.19) we can find sufficiently small $\epsilon > 0$ such that the value of the last expression in the panel above is no larger than ζ, thus completing the proof of (G.19).

Thus, we have shown that $G(\alpha, \beta, \tau)$ in (G.17) converges pointwise to

$$\beta\sqrt{\delta}\sqrt{\alpha^2+\sigma^2} - \frac{\alpha\tau}{2} - \frac{\alpha\beta^2}{2\tau} + \lambda \cdot F\left(\frac{\alpha\beta}{\tau}, \frac{\alpha\lambda}{\tau}\right)$$

in the limit of $n \to \infty$. Above, we have applied Lemma B.2.5(b) and have further made use of Assumption 11.2.1. Note that H is strongly convex in α and jointly concave in $\beta, \mathbf{v}$ since taking limits does not affect convexity properties (recall that G is convex-concave). With these, it follows as per [NM94, Thm. 2.7] that (i)

$$\max_{0\le\beta\le 1,\tau>0} \min_{\alpha\ge 0} G(\alpha,\beta,\tau) \xrightarrow{P} \max_{0\le\beta\le 1,\tau>0} \min_{\alpha\ge 0} H(\alpha,\beta,\tau), \tag{G.22}$$

and (ii) $\alpha_*(\mathbf{h}, \mathbf{x}_0) \xrightarrow{P} \alpha_*$, where α_* the unique minimizer of the second optimization in (G.22). This completes the proof of the corollary.

Theorem 11.2.1

As mentioned, Theorem 11.2.1 is a direct consequence of what we have already shown. In particular, we showed that the value α_* to which the error converges only depends on g through the parameters μ and σ^2. Those are the same (by definition) for the non-linear and the linear case considered. Therefore, the errors are the same.

G.2 Proofs for Section 11.3

The LM Algorithm

The Lloyd-Max algorithm is an algorithm for finding the quantization threshold t_i and the representation points ℓ_i. Given real values $x \in \mathbb{R}$ sampled from some probability density $\phi(x)$ it looks for optimal sets $\hat{\mathbf{t}}, \hat{\boldsymbol{\ell}}$ that minimize the mean-square-error (MSE) between x and their corresponding quantized values $Q_q(x; \boldsymbol{\ell}, \mathbf{t})$, i.e.

$$(\hat{\boldsymbol{\ell}}, \hat{\mathbf{t}}) := \arg\min_{\boldsymbol{\ell},\mathbf{t}} \mathbb{E}_{x\sim\phi}[(x - Q_q(x;\boldsymbol{\ell},\mathbf{t}))^2]. \tag{G.23}$$

The algorithm simply alternates between i) optimizing the threshold $\mathbf{t}_i$ for a given set of $\boldsymbol{\ell}$, and then ii) optimizing the levels $\boldsymbol{\ell}_i$ for the new thresholds. It is well known that the converging points $\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM}$ of the algorithm satisfy

$$\mathbf{t}_i^{LM} = \frac{\boldsymbol{\ell}_i^{LM} + \boldsymbol{\ell}_{i+1}^{LM}}{2} \qquad i = 1, \ldots, L-1, \tag{G.24a}$$

$$\boldsymbol{\ell}_i^{LM} = \left(\int_{\mathbf{t}_{i-1}^{LM}}^{\mathbf{t}_i^{LM}} \phi(x)\mathrm{d}x\right)^{-1} \left(\int_{\mathbf{t}_{i-1}^{LM}}^{\mathbf{t}_i^{LM}} x\phi(x)\mathrm{d}x\right) \qquad i = 1, \ldots, L. \tag{G.24b}$$

Furthermore, they are stationary points of the objective function in (G.23).

Gaussian Case

Assume that the values x are sampled from a standard Gaussian distribution, i.e. $x \sim \mathcal{N}(0,1)$ and $\phi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. Also, recall the definition of the parameters μ, σ^2 in (11.3); setting $g = Q_q$ therein, we find (also, to compare with (11.14))

$$\mu := \mu(\boldsymbol{\ell}, \mathbf{t}) = 2\sum_{i=1}^{L} \ell_i \int_{\mathbf{t}_{i-1}}^{\mathbf{t}_i} x\phi(x)\mathrm{d}x, \tag{G.25a}$$

$$\tau^2 := \tau^2(\boldsymbol{\ell}, \mathbf{t}) = 2\sum_{i=1}^{L} \ell_i^2 \int_{\mathbf{t}_{i-1}}^{\mathbf{t}_i} \phi(x)\mathrm{d}x. \tag{G.25b}$$

In this notation, the objective in (G.23) can be writthen as $\tau^2 - 2\mu + 1$. Thus, $\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM}$ satisfy

$$(\tau^2)'\big|_{(\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM})} = 2\mu'\big|_{(\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM})}. \tag{G.26}$$

Here and onwards we use $(\tau^2)'$, μ' to denote the gradient of τ^2 and μ with respect to the vector $[\boldsymbol{\ell}^T, \mathbf{t}^T]$. The gradients are evaluated at the point $(\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM})$ in (G.26).

q-Bit Compressive Sensing

We prove that the LM algorithm is an efficient algorithm when the objective is minimizing the LASSO reconstruction error of a signal $\mathbf{x}_0$ to which we have access through q-bit quantized linear measuments $Q_q(\mathbf{a}_i^T\mathbf{x}; \boldsymbol{\ell}, \mathbf{t})$. It was shown in Section 11.3 that the problem can be posed as that of finding $\boldsymbol{\ell}_*, \mathbf{t}_*$ such that

$$(\mathbf{t}_*, \boldsymbol{\ell}_*) = \arg\min_{\mathbf{t},\boldsymbol{\ell}} \frac{\sigma^2(\mathbf{t}, \boldsymbol{\ell})}{\mu^2(\mathbf{t}, \boldsymbol{\ell})} = \arg\min_{\mathbf{t},\boldsymbol{\ell}} \frac{\tau^2(\mathbf{t}, \boldsymbol{\ell})}{\mu^2(\mathbf{t}, \boldsymbol{\ell})}. \tag{G.27}$$

The following Lemma proves the claim made in Section 11.3, i.e. the converging points of the LM algorithm are stationary points of the objective function in (G.27).

Lemma G.2.1. *The converging points of the LM algorithm, say* $(\mathbf{t}^{LM}, \boldsymbol{\ell}^{LM})$ *satisfy*

$$\begin{aligned} \frac{\partial}{\partial \ell_i}\left(\frac{\tau^2(\boldsymbol{\ell}, \mathbf{t})}{\mu^2(\boldsymbol{\ell}, \mathbf{t})}\right)\Big|_{(\boldsymbol{\ell}, \mathbf{t}) = (\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM})} &= 0 \qquad , i = 1, \ldots, L, \\ \frac{\partial}{\partial t_i}\left(\frac{\tau^2(\boldsymbol{\ell}, \mathbf{t})}{\mu^2(\boldsymbol{\ell}, \mathbf{t})}\right)\Big|_{(\boldsymbol{\ell}, \mathbf{t}) = (\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM})} &= 0 \qquad , i = 0, \ldots, L-1. \end{aligned} \tag{G.28}$$

Proof. Call $R(\mathbf{t}, \boldsymbol{\ell}) = \frac{\tau^2(\mathbf{t}, \boldsymbol{\ell})}{\mu^2(\mathbf{t}, \boldsymbol{\ell})}$. We denote $R' := R'(\mathbf{t}, \boldsymbol{\ell})$ for its gradient with respect to the vector $[\mathbf{t}^T, \boldsymbol{\ell}^T]$. It suffices to prove that $R'\big|_{(\boldsymbol{\ell}^{LM}, \mathbf{t}^{LM})} = 0$, or equivalently, that

at the point $(\mathbf{t}, \ell) = (\mathbf{t}^{LM}, \ell^{LM})$ the following holds:

$$(\tau^2)'\mu^2 = 2\tau^2\mu\mu'. \tag{G.29}$$

To see that this is the case, note that

$$\tau^2(\mathbf{t}^{LM}, \ell^{LM}) = \mu(\mathbf{t}^{LM}, \ell^{LM}). \tag{G.30}$$

This follows by direct substitution of (G.24) in (G.25). Then, (G.29) follows from (G.30) and (G.26). □

Appendix H

A NOTE ON SIMPLE DENOISING

The problem of simple denoising was introduced in Chapter 12. It refers to the recovery of a structured signal $\mathbf{x}_0 \in \mathbb{R}^n$ from uncompressed observations $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$. For the estimation, we use (12.1) and ask what is the resulting squared error $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2$. The purpose of this short note is to showcase that some of the mechanics of the CGMT framework, which are presented in this thesis, might still be applicable towards answering that question. The details go beyond our main purpose; instead, our intention is to motivate a potentially interesting research direction. Hence, we have decided to keep the presentation short and be somewhat informal with the focus being on conveying the main idea.

H.1 Regularized Least-squares

We will show that an appropriate application of (subset of) the mechanics of the CGMT framework prescribed in Chapter 5 results in a precise characterization of the squared error of regularized least-squares in the simple denosing setting. To begin, let $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$ with $\mathbf{z} \sim p_{\mathbf{z}}$ and $\mathbf{x}_0 \sim p_{\mathbf{x}_0}$, and consider

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda f(\mathbf{x}), \tag{H.1}$$

for some convex f and $\lambda > 0$. We characterize the squared error $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2$ as a function of f, λ, $p_{\mathbf{z}}$, and $p_{\mathbf{x}_0}$.

As in Chapter 5.2, the first step is to introduce the error vector $\mathbf{w} := \mathbf{x} - \mathbf{x}_0$. With this, (H.2) is expressed as

$$\hat{\mathbf{w}} := \arg\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{z} - \mathbf{w}\|_2^2 + \lambda f(\mathbf{x}_0 + \mathbf{w}), \tag{H.2}$$

and we wish to find the converging limit of $\|\hat{\mathbf{w}}\|_2^2/n$. Our strategy is simple as follows: we write (H.2) in a min-max form, treat that as an (AO) [1], and apply Steps 2–4 of Chapter 5.

First, note that we can rewrite (H.2) as

$$\hat{\mathbf{w}} := \arg\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|_2^2 - \mathbf{z}^T\mathbf{w} + \lambda f(\mathbf{x}_0 + \mathbf{w}).$$

[1] Strictly speaking, things here are much simpler; there is no random Gaussian matrix $\mathbf{A}$ involved in (H.2), hence there is no (obvious) place or need to apply any comparison theorem! However, it will turn out that writing (H.2) as a min-max is still useful.

Of course, the minimization above is equivalent to

$$\min_{\mathbf{w},\mathbf{s}} \frac{1}{2}\|\mathbf{w}\|_2^2 - \mathbf{z}^T\mathbf{w} + \lambda f(\mathbf{s}), \qquad \text{s.t.} \quad \mathbf{x}_0 + \mathbf{w} = s$$

or,

$$\min_{\mathbf{w},\mathbf{s}} \max_{\mathbf{u}} \frac{1}{2}\|\mathbf{w}\|_2^2 - \mathbf{z}^T\mathbf{w} + \lambda f(\mathbf{s}) + \mathbf{u}^T\mathbf{x}_0 + \mathbf{u}^T\mathbf{w} - \mathbf{u}^T\mathbf{s}. \tag{H.3}$$

Following the same tricks as in the CGMT framework, we reduce the optimization in (H.3) into one that only involves scalar optimization variables. This corresponds to the "scalarization" step of Chapter 5. We start by flipping the order of min-max in (H.3). The objective function is appropriately convex-concave, but thecompactness of the constraint sets needs to be taken into account. This can be done rigorously in a similar manner as in Section 5.2. Details are omitted here since they don't serve our main purpose. Once flipped, we can optimize over the direction of $\mathbf{w}$ while keeping its norm fixed to (say) α, i.e. (H.3) becomes as follows:

$$\min_{\alpha\geq 0,\mathbf{s}} \max_{\mathbf{u}} \frac{1}{2}\alpha^2 - \alpha\|\mathbf{z}-\mathbf{u}\|_2 + \lambda f(\mathbf{s}) + \mathbf{u}^T\mathbf{x}_0 - \mathbf{u}^T\mathbf{s}, \tag{H.4}$$

and the minimizer α_* satisfies $\|\hat{\mathbf{w}}\|_2 = \alpha_*$. Next, we write the term $\|\mathbf{z}-\mathbf{u}\|_2$ in its variational form $\min_{\tau>0} \frac{\tau}{2} + \frac{\|\mathbf{z}-\mathbf{u}\|_2^2}{2\tau}$, which allows maximizing over $\mathbf{u}$. With these, (H.4) reduces to

$$\min_{\alpha\geq 0,\mathbf{s}} \max_{\tau} \frac{1}{2}\alpha^2 - \frac{\alpha\tau}{2} + \frac{\tau}{2\alpha}\|\mathbf{x}_0-\mathbf{s}\|_2^2 + \mathbf{z}^T(\mathbf{x}_0-\mathbf{s}) + \lambda f(\mathbf{s}).$$

It only takes completing the squares above to reach the more convenient form:

$$\min_{\alpha\geq 0} \max_{\tau} \frac{1}{2}\alpha^2 - \frac{\alpha\tau}{2} + \min_{\mathbf{s}} \left\{ \frac{\tau}{2\alpha}\|\mathbf{x}_0 + \frac{\alpha}{\tau}\mathbf{z} - \mathbf{s}\|_2^2 + \lambda f(\mathbf{s}) \right\} - \frac{\alpha}{2\tau}\|\mathbf{z}\|_2^2,$$

where we can clearly identify the Moreau envelope of f. To conclude, we have reduced (H.2) to the following optimization that only involves scalar variables:

$$\min_{\alpha\geq 0} \max_{\tau} \frac{1}{2}\alpha^2 - \frac{\alpha\tau}{2} - \frac{\alpha}{2\tau}\|\mathbf{z}\|_2^2 + \lambda \cdot e_f\left(\mathbf{x}_0 + \frac{\alpha}{\tau}\mathbf{z}; \frac{\alpha}{\lambda\tau}\right). \tag{H.5}$$

Compare this to (5.10) of Section 5.2. As it was the case with the latter, under appropriate conditions on f, $p_{\mathbf{x}_0}$, and $p_{\mathbf{z}}$, it is easy to find the converging limit of the objective function in (H.5) when the optimization variables α and τ are fixed. For a mere illustration, assume here that f is separable, $\mathbf{z}_i \overset{iid}{\sim} p_Z$, $\mathbf{x}_{0,i} \overset{iid}{\sim} p_{X_0}$, and $\mathbb{E}_{Z\sim p_Z} Z^2 = \sigma^2 < \infty$. Then, for fixed α and τ, the objective function in (H.5) (after normalized by n) converges by the WLLN to the following:

$$\frac{1}{2}\alpha^2 - \frac{\alpha\tau}{2} - \frac{\alpha}{2\tau}\sigma^2 + \lambda \cdot F(\frac{\alpha}{\tau}, \frac{\alpha}{\lambda\tau}), \tag{H.6}$$

where F is the following expected Moreau envelope function:

$$F(c,\tau) = \mathbb{E}_{\substack{Z\sim p_Z \\ X_0\sim p_{X_0}}} e_f\left(X_0 + cZ;\tau\right).$$

It now takes replicating the technical work involved in the "convergence analysis" step of the CGMT framework to prove that the random optimization in (H.5) converges to the min-max of the deterministic function in (H.6), which then becomes the Scalar Performance Optimization (SPO). Finally, again with arguments that are same as in the CGMT framework, it is shown that the random minimizer $\alpha_*(\mathbf{x}_0, \mathbf{z})$ of (H.6), which corresponds to the quantity of interest $\|\hat{\mathbf{w}}\|_2$, converges to the minimizer of the SPO. The details go beyond the scope of this note.

To conclude, we have prescribed a machinery (basically an adaptation of the CGMT framework of Chapter 5) that yields a precise (asymptotic) characterization of the squared error of regularized least-squares in the simple denoising setting.This result extends recent results of Chatterjee [Cha+14] and of Oymak and Hassibi [OH15]. Oymak and Hassibi characterize the squared error only when noise is Gaussian with vanishing variance (high-SNR regime). Instead, the machinery presented here concludes about general noise distributions. Chatterjee achieves to characterize the error of constrained least-squares. We have extended the result to the more often encountered in practice regularized version. (We remark however that the analysis in both [OH15] and [Cha+14] is non-asymptotic).

It is interesting to explore the extent to which these new ideas are still applicable to characterize the error under different choices of the loss function in (12.1) other than the least-squares discussed here. This would yield a theorem as general as Theorem 4.2.1 applied to the simple denoising setting.