

# Topics in Randomized Numerical Linear Algebra

Thesis by

Alex Gittens

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2013

(Defended May 31, 2013)

© 2013

Alex Gittens

All Rights Reserved

# Acknowledgements

I have been privileged to befriend and work with exceptional individuals.

I owe a debt of thank to Drs. Emmanuel Papadakis and Gordon Johnson of the University of Houston for helping me get started on my academic path. I consider myself lucky to have had the guidance of my PhD advisor, Joel Tropp, and remain humbled by his intuitive grasp of the ideas and tools I have wrestled with throughout my graduate career. In addition to Joel, I am grateful to the other collaborators I was privileged to work with: Christos Boutsidis, Michael Mahoney, and Richard Chen. It was a pleasure working with them.

I benefited greatly from interactions with my fellow students; in particular, I thank Catherine Beni, Chia-Chieh Chu, Derek Leong, Jinghao Huang, and Yamuna Phal for being great office mates and friends. My dear friend and amateur herpetologist Yekaterina Pavlova I thank for being her very unique self. Stephen Becker, Michael McCoy, Peter Stobbe, Patrick Sanan, Zhiyi Li, George Chen, and Yaniv Plan may have provided answers to the odd math question, but I appreciate them more for the hours of wide-ranging discussions on everything under the sun.

Finally, I thank my family for their support: my mother Mura; my father Erskine; my sister Lauren; and my grandfather Chesterfield, who was a formative influence on my young mind.

# Abstract

This thesis studies three classes of randomized numerical linear algebra algorithms, namely: (i) randomized matrix sparsification algorithms, (ii) low-rank approximation algorithms that use randomized unitary transformations, and (iii) low-rank approximation algorithms for positive-semidefinite (PSD) matrices.

Randomized matrix sparsification algorithms set randomly chosen entries of the input matrix to zero. When the approximant is substituted for the original matrix in computations, its sparsity allows one to employ faster sparsity-exploiting algorithms. This thesis contributes bounds on the approximation error of nonuniform randomized sparsification schemes, measured in the spectral norm and two NP-hard norms that are of interest in computational graph theory and subset selection applications.

Low-rank approximations based on randomized unitary transformations have several desirable properties: they have low communication costs, are amenable to parallel implementation, and exploit the existence of fast transform algorithms. This thesis investigates the tradeoff between the accuracy and cost of generating such approximations. State-of-the-art spectral and Frobenius-norm error bounds are provided.

The last class of algorithms considered are SPSD “sketching” algorithms. Such sketches can be computed faster than approximations based on projecting onto mixtures of the columns of the matrix. The performance of several such sketching schemes is empirically evaluated using a

suite of canonical matrices drawn from machine learning and data analysis applications, and a framework is developed for establishing theoretical error bounds.

In addition to studying these algorithms, this thesis extends the Matrix Laplace Transform framework to derive Chernoff and Bernstein inequalities that apply to *all* the eigenvalues of certain classes of random matrices. These inequalities are used to investigate the behavior of the singular values of a matrix under random sampling, and to derive convergence rates for each individual eigenvalue of a sample covariance matrix.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1 Introduction and contributions</b>	<b>1</b>
1.1 The sampling approach to matrix approximation . . . . .	3
1.2 The random-projection approach to matrix approximation . . . . .	7
1.3 Nonasymptotic random matrix theory . . . . .	9
1.4 Contributions . . . . .	12
1.4.1 Nonasymptotic random matrix theory . . . . .	12
1.4.2 Matrix sparsification . . . . .	13
1.4.3 Low-rank approximation using fast unitary transformations . . . . .	16
1.4.4 Randomized SPSD sketches . . . . .	17
<b>2 Bounds for all eigenvalues of sums of Hermitian random matrices</b>	<b>20</b>
2.1 Introduction . . . . .	20
2.2 Notation . . . . .	21

2.3	The Courant–Fisher Theorem . . . . .	22
2.4	Tail bounds for interior eigenvalues . . . . .	23
2.5	Chernoff bounds . . . . .	28
2.6	Bennett and Bernstein inequalities . . . . .	32
2.7	An application to column subsampling . . . . .	38
2.8	Covariance estimation . . . . .	41
2.8.1	Proof of Theorem 2.15 . . . . .	47
2.8.2	Extensions of Theorem 2.15 . . . . .	51
2.8.3	Proofs of the supporting lemmas . . . . .	53
<b>3</b>	<b>Randomized sparsification in NP-hard norms</b>	<b>58</b>
3.1	Notation . . . . .	60
3.1.0.1	Graph sparsification . . . . .	61
3.2	Preliminaries . . . . .	63
3.3	The $\infty \rightarrow p$ norm of a random matrix . . . . .	66
3.3.1	The expected $\infty \rightarrow p$ norm . . . . .	66
3.3.2	A tail bound for the $\infty \rightarrow p$ norm . . . . .	68
3.4	Approximation in the $\infty \rightarrow 1$ norm . . . . .	70
3.4.1	The expected $\infty \rightarrow 1$ norm . . . . .	70
3.4.2	Optimality . . . . .	72
3.4.3	An example application . . . . .	74
3.5	Approximation in the $\infty \rightarrow 2$ norm . . . . .	75
3.5.1	Optimality . . . . .	78
3.5.2	An example application . . . . .	79

3.6	A spectral error bound . . . . .	80
3.6.1	Comparison with previous results . . . . .	82
3.6.1.1	A matrix quantization scheme . . . . .	82
3.6.1.2	A nonuniform sparsification scheme . . . . .	83
3.6.1.3	A scheme which simultaneously sparsifies and quantizes . . . . .	85
3.7	Comparison with later bounds . . . . .	86
<b>4</b>	<b>Preliminaries for the investigation of low-rank approximation algorithms</b>	<b>89</b>
4.1	Probabilistic tools . . . . .	89
4.1.1	Concentration of convex functions of Rademacher variables . . . . .	89
4.1.2	Chernoff bounds for sampling without replacement . . . . .	90
4.1.3	Frobenius-norm error bounds for matrix multiplication . . . . .	91
4.2	Linear Algebra notation and results . . . . .	96
4.2.1	Column-based low-rank approximation . . . . .	98
4.2.1.1	Matrix Pythagoras and generalized least-squares regression . . . . .	99
4.2.1.2	Low-rank approximations restricted to subspaces . . . . .	100
4.2.2	Structural results for low-rank approximation . . . . .	101
4.2.2.1	A geometric interpretation of the sampling interaction matrix . . . . .	102
<b>5</b>	<b>Low-rank approximation with subsampled unitary transformations</b>	<b>105</b>
5.1	Introduction . . . . .	105
5.2	Low-rank matrix approximation using SRHTs . . . . .	110
5.2.1	Detailed comparison with prior work . . . . .	111
5.3	Matrix computations with SRHT matrices . . . . .	118
5.3.1	SRHTs applied to orthonormal matrices . . . . .	120



5.3.2	SRHTs applied to general matrices . . . . .	123
5.4	Proof of the quality of approximation guarantees . . . . .	132
5.5	Experiments . . . . .	142
5.5.1	The test matrices . . . . .	143
5.5.2	Empirical comparison of the SRHT and Gaussian algorithms . . . . .	146
5.5.3	Empirical evaluation of our error bounds . . . . .	147
<b>6</b>	<b>Theoretical and empirical aspects of SPSD sketches</b>	<b>153</b>
6.1	Introduction . . . . .	153
6.1.1	Outline . . . . .	156
6.2	Deterministic bounds on the errors of SPSD sketches . . . . .	157
6.3	Comparison with prior work . . . . .	162
6.4	Proof of the deterministic error bounds . . . . .	164
6.4.1	Spectral-norm bounds . . . . .	165
6.4.2	Frobenius-norm bounds . . . . .	168
6.4.3	Trace-norm bounds . . . . .	173
6.5	Error bounds for Nyström extensions . . . . .	175
6.6	Error bounds for random mixture-based SPSD sketches . . . . .	181
6.6.1	Sampling with leverage-based importance sampling probabilities . . . . .	182
6.6.2	Random projections with subsampled randomized Fourier transforms . . . . .	185
6.6.3	Random projections with i.i.d. Gaussian random matrices . . . . .	188
6.7	Stable algorithms for computing regularized SPSD sketches . . . . .	192
6.8	Computational investigations of the spectral-norm bound for Nyström extensions	197
6.8.1	Optimality . . . . .	197

6.8.2	Dependence on coherence . . . . .	198
6.9	Empirical aspects of SPSP low-rank approximation . . . . .	201
6.9.1	Test matrices . . . . .	202
6.9.2	A comparison of empirical errors with the theoretical error bounds . . . . .	207
6.9.3	Reconstruction accuracy of sampling and projection-based sketches . . . . .	209
6.9.3.1	Graph Laplacians . . . . .	209
6.9.3.2	Linear kernels . . . . .	216
6.9.3.3	Dense and sparse RBF kernels . . . . .	219
6.9.3.4	Summary of comparison of sampling and mixture-based SPSP Sketches . . . . .	225
6.10	A comparison with projection-based low-rank approximations . . . . .	226
	<b>Bibliography</b>	<b>231</b>

# List of Figures

2.1	Spectrum of a random submatrix of a unitary DFT matrix. . . . .	41
5.1	Residual errors of low-rank approximation algorithms . . . . .	145
5.2	Forward errors of low-rank approximation algorithms . . . . .	148
5.3	The number of column samples required for relative error Frobenius-norm approximations . . . . .	149
5.4	Empirical versus predicted spectral-norm residual errors of low-rank approximations	152
6.1	Empirical demonstration of the optimality of Theorem 6.9. . . . .	198
6.2	Spectral-norm errors of regularized Nyström extensions as coherence varies . . . . .	200
6.3	Spectral-norm error of regularized Nyström extensions as regularization parameter varies . . . . .	202
6.4	Relative errors of non-rank-restricted SPSD sketches of the GR and HEP Laplacian matrices . . . . .	210
6.5	Relative errors of non-rank-restricted SPSD sketches of the Enron and Gnutella Laplacian matrices . . . . .	211
6.6	Relative errors of rank-restricted SPSD sketches of the GR and HEP Laplacian matrices	212
6.7	Relative errors of rank-restricted SPSD sketches of the Enron and Gnutella Laplacian matrices . . . . .	213

6.8	Relative errors of non-rank-restricted SPSD sketches of the linear kernel matrices .	217
6.9	Relative errors of rank-restricted SPSD sketches of the linear kernel matrices . . .	218
6.10	Relative errors of non-rank-restricted SPSD sketches of the dense RBFK matrices .	220
6.11	Relative errors of rank-restricted SPSD sketches of the dense RBFK matrices . . . .	221
6.12	Relative errors of non-rank-restricted SPSD sketches of the sparse RBFK matrices .	222
6.13	Relative errors of rank-restricted SPSD sketches of the sparse RBFK matrices . . .	223
6.14	Comparison of projection-based low-rank approximations with one-pass SPSD sketches . . . . .	230

## List of Tables

6.1	Asymptotic comparison of our bounds on SPSP sketches with prior work . . . . .	163
6.2	Information on the SPSP matrices used in our empirical evaluations . . . . .	203
6.3	Statistics of our test matrices . . . . .	205
6.4	Comparison of empirical errors of SPSP sketches with predicted errors . . . . .	208