

The Nested Periodic Subspaces: Extensions of Ramanujan Sums for Period Estimation

Thesis by
Srikanth Venkata Tenneti

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

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended June 1, 2018

© 2018

Srikanth Venkata Tenneti
ORCID: 0000-0002-5415-3681

All rights reserved except where otherwise noted

*Dedicated to the lessons of selfless love, sacrifice and courage as exemplified by my
mom, dad and cutie pie sister.*

ACKNOWLEDGEMENTS

My sincerest gratitude to the following people for being of invaluable support during my PhD:

Firstly, to my advisor, Prof. P. P. Vaidyanathan for giving me the opportunity to come to Caltech and pursue graduate studies. Over the years, he has not only been a great academic mentor, but was also kind enough to be understanding and guide me whenever I needed personal advice. I would also like to thank my committee members, Prof. Shuki Bruck, Prof. Yaser Abu-Mostafa, Prof. Victoria Kostina and Prof. Venkat Chandrasekaran for their support in helping this thesis reach its current state.

Pursuing PhD studies at one of the most beautiful institutes in the world is definitely a dream-come-true. However, as this phase of my life reaches its conclusion, I have come to realize that the true cost of this PhD has been the time that I have spent away from my family. I would like to thank my mom, dad and sister for their sacrifices and love throughout the years.

Decades from now, when I hear the word ‘Caltech’, the first things that will come to my mind will be the following: the fun times working on countless assignments with the ever so sweet Tahmineh Khazaei, the super kind hearted Kavya Sudhir, my philosophy-over-tea buddy Ishan Tembekhar, and my fellow Mumbaikar Yamuna Phal. It is very difficult to thank them in just a few words. Leaving this school, I sincerely hope that I can continue to keep them as a part of the rest of my life.

I would like to thank my labmates Piya Pal, John Liu, Chun-Lin Liu and Oguzhan Teke for their encouragement and help in my research. In this regard, I am also indebted to all my past teachers and classmates from my previous schools in India. Without their blessings, dedication to teaching and hard work, I would not have been able to attempt the research presented in this thesis. Finally, a big thanks to my friends and coaches at CrossFit Resistance, Pasadena, for making sure that I do not neglect my health and fitness while researching.

Phew... let us end this page with a smiley. :)

ABSTRACT

In the year 1918, the Indian mathematician Srinivasa Ramanujan proposed a set of sequences called Ramanujan Sums as bases to expand arithmetic functions in number theory. Today, exactly a 100 years later, we will show that these sequences re-emerge as exciting tools in a completely different context: For the extraction of periodic patterns in data. Combined with the state-of-the-art techniques of DSP, Ramanujan Sums can be used as the starting point for developing powerful algorithms for periodicity applications.

The primary inspiration for this thesis comes from a recent extension of Ramanujan sums to subspaces known as the Ramanujan subspaces. These subspaces were designed to span any sequence with integer periodicity, and have many interesting properties. Starting with Ramanujan subspaces, this thesis first develops an entire family of such subspace representations for periodic sequences. This family, called Nested Periodic Subspaces due to their unique structure, turns out to be the least redundant sets of subspaces that can span periodic sequences.

Three classes of new algorithms are proposed using the Nested Periodic Subspaces: dictionaries, filter banks, and eigen-space methods based on the auto-correlation matrix of the signal. It will be shown that these methods are especially advantageous to use when the data-length is short, or when the signal is a mixture of multiple hidden periods. The dictionary techniques were inspired by recent advances in sparsity based compressed sensing. Apart from the l_1 norm based convex programs currently used in other applications, our dictionaries can admit l_2 norm formulations that have linear and closed form solutions, even when the systems is under-determined. A new filter bank is also proposed using the Ramanujan sums. This, named the Ramanujan Filter Bank, can accurately track the instantaneous period for signals that exhibit time varying periodic nature. The filters in the Ramanujan Filter Bank have simple integer valued coefficients, and directly tile the period vs time plane, unlike classical STFT (Short Time Fourier Transform) and wavelets, which tile the time-frequency plane. The third family of techniques developed here are a generalization of the classic MUSIC (MULTiple SInal Classification) algorithm for periodic signals. MUSIC is one of the most popular techniques today for line spectral estimation. However, periodic signals are not just any unstructured line spectral signals. There is a nice harmonic spacing between the lines which is not exploited by plain MUSIC. We will show that one can design much more

accurate adaptations of MUSIC using Nested Periodic Subspaces. Compared to prior variants of MUSIC for the periodicity problem, our approach is much faster and yields much more accurate results for signals with integer periods. This work is also the first extension of MUSIC that uses simple integer valued basis vectors instead of using traditional complex-exponentials to span the signal subspace. The advantages of the new methods are demonstrated both on simulations, as well as real world applications such as DNA micro-satellites, protein repeats and absence seizures.

Apart from practical contributions, the theory of Nested Periodic Subspaces offers answers to a number of fundamental questions that were previously unanswered. For example, what is the minimum contiguous data-length needed to be able to identify the period of a signal unambiguously? Notice that the answer we seek is a fundamental identifiability bound independent of any particular period estimation technique. Surprisingly, this basic question has never been answered before. In this thesis, we will derive precise expressions for the minimum necessary and sufficient datalengths for this question. We also extend these bounds to the context of mixtures of periodic signals. Once again, even though mixtures of periodic signals often occur in many applications, aspects such as the unique identifiability of the component periods were never rigorously analyzed before. We will present such an analysis as well.

While the above question deals with the minimum contiguous datalength required for period estimation, one may ask a slightly different question: If we are allowed to pick the samples of a signal in a non-contiguous fashion, how should we pick them so that we can estimate the period using the least number of samples? This question will be shown to be quite difficult to answer in general. In this thesis, we analyze a smaller case in this regard, namely, that of resolving between two periods. It will be shown that the analysis is quite involved even in this case, and the optimal sampling pattern takes an interesting form of sparsely located bunches. This result can also be extended to the case of multi-dimensional periodic signals.

We very briefly address multi-dimensional periodicity in this thesis. Most prior DSP literature on multi-dimensional discrete time periodic signals assumes the period to be parallelepipeds. But as shown by the artist M. C. Escher, one can tile the space using a much more diverse variety of shapes. Is it always possible to account for such other periodic shapes using the traditional notion of parallelepiped periods? An interesting analysis in this regard is presented towards the end of the thesis.

PUBLISHED CONTENT AND CONTRIBUTIONS

Journal Articles

1. S. V. Tenneti and P. P. Vaidyanathan, "Nested Periodic Matrices and Dictionaries: New Signal Representations for Period Estimation", *IEEE Transactions on Signal Processing*, vol.63, no. 14, pp. 3736 - 3750, July, 2015. DOI: 10.1109/TSP.2015.2434318
2. S. V. Tenneti and P. P. Vaidyanathan, "A Unified Theory of Union of Subspaces Models for Period Estimation", *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5217 - 5231, Oct 2016. DOI: 10.1109/TSP.2016.2582473
3. S. V. Tenneti and P. P. Vaidyanathan, "Arbitrarily Shaped Periods in Multi-Dimensional Discrete Time Periodicity", *IEEE Signal Processing Letters*, vol.22, no.10, pp. 1748 - 1751, Oct. 2015. DOI: 10.1109/LSP.2015.2431993
4. S. V. Tenneti and P. P. Vaidyanathan, "Minimum Datalength for Integer Period Estimation", *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2733 - 2745, May 2018. DOI: 10.1109/TSP.2018.2818080

Conference Articles

1. P. P. Vaidyanathan and S. V. Tenneti, "When does periodicity in discrete-time imply that in continuous-time", *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc.*, Canada, April 2018.
2. S. V. Tenneti and P. P. Vaidyanathan, "MUSIC and Ramanujan: MUSIC-like Algorithms for Integer Period Estimation Using Nested Periodic Subspaces", *Proc. Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA, Nov 2017. DOI: 10.1109/ACSSC.2017.8335717
3. S. V. Tenneti and P. P. Vaidyanathan, "Minimum Number of Possibly Non-Contiguous Samples to Distinguish Two Periods", *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc.*, New Orleans, March 2017. DOI: 10.1109/ICASSP.2017.7952872

4. P. P. Vaidyanathan and S. V. Tenneti, "Efficient Multiplier-less structures for Ramanujan Filter Banks", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc., New Orleans, March 2017. DOI: 10.1109/ICASSP.2017.7953400
5. S. V. Tenneti and P. P. Vaidyanathan, "Detection of Protein Repeats Using Ramanujan Filter Bank", Proc. Asilomar Conference on Signals, Systems and Computers, Monterey, CA, 2016. DOI: 10.1109/ACSSC.2016.7869058
6. S. V. Tenneti and P. P. Vaidyanathan, "Critical Data-length for Period Estimation", Proc. IEEE International Symposium on Circuits and Systems, Canada, May. 2016. DOI: 10.1109/ISCAS.2016.7527468
7. S. V. Tenneti and P.P. Vaidyanathan, "Detecting Tandem Repeats in DNA Using Ramanujan Filter Bank", Proc. IEEE International Symposium on Circuits and Systems, Canada, May. 2016. DOI: 10.1109/ISCAS.2016.7527160
8. S. V. Tenneti and P. P. Vaidyanathan, "Minimal Dictionaries for Spanning Periodic Signals", Proc. Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, Nov. 2015. DOI: 10.1109/ACSSC.2015.7421183
9. S. V. Tenneti and P. P. Vaidyanathan, "Period Estimation and Tracking: Filter Bank Design using Truth Tables of Logic", Proc. Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, Nov. 2015. DOI: 10.1109/ACSSC.2015.7421410
10. S. V. Tenneti and P. P. Vaidyanathan, "Ramanujan Filter Banks for Estimation and Tracking of Periodicity", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc., Brisbane, April 2015. DOI: 10.1109/ICASSP.2015.7178692
11. P. P. Vaidyanathan and S. V. Tenneti, "Properties of Ramanujan Filter Banks", Proc. European Signal Processing Conference, France, August 2015. DOI: 10.1109/EUSIPCO.2015.7362898
12. S. V. Tenneti and P. P. Vaidyanathan, "Dictionary Approaches for Identifying Periodicities in Data", Proc. Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, Nov. 2014. DOI: 10.1109/ACSSC.2014.7094814
13. P. P. Vaidyanathan and S. V. Tenneti, "Ramanujan Subspaces and Digital Signal Processing", Proc. Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, Nov. 2014. DOI: 10.1109/ACSSC.2014.7094770

Contributions

All the above papers resulted from research carried out by S. V. T. and Prof. P. P. Vaidyanathan. Hence they are all two-author papers. The papers where S. V. T. is the first author correspond to research where he was the primary contributor. The papers where he is the second author correspond to research which was primarily carried out by Prof. P. P. Vaidyanathan, and where S. V. T. assisted with experimental validations, simulations, demonstrations of theorems etc. Please refer to Chapter 1 for an overview of the content of all the above papers.

TABLE OF CONTENTS

Acknowledgements	iv
Abstract	v
Published Content and Contributions	vii
Table of Contents	x
List of Illustrations	xii
List of Tables	xix
Chapter I: Introduction	1
1.1 A Motivating Example	1
1.2 Outline and Contributions of This Thesis	6
Chapter II: Ramanujan Sums And Subspaces	13
2.1 Ramanujan Sums	13
2.2 Early Use of Ramanujan Sums in DSP	14
2.3 Generalizing Ramanujan Sums to Ramanujan Subspaces	15
2.4 The Ramanujan Periodicity Transform (RPT) Matrix	16
2.5 Estimating the period beyond a particular $\mathcal{V}_{\mathcal{P}}$	17
2.6 Conclusion	18
Chapter III: Nested Periodic Matrices And Dictionaries	20
3.1 Nested Periodicity Matrix: Definition	20
3.2 NPM Properties	21
3.3 Examples of NPMs	23
3.4 Nested Periodic Dictionaries	30
3.5 Real World Examples	40
3.6 Conclusion	43
Chapter IV: A Unified Theory of Union-of-Subspaces Representations of Periodic Signals	46
4.1 Introduction	46
4.2 Review of Existing Subspace Methods	49
4.3 Insights into the Relationships Between Various Techniques	55
4.4 Fundamental Properties of Subspaces That Admit Unique Periodic Decompositions	57
4.5 From Subspaces to Dictionaries	61
4.6 Fundamental Properties of Periodic Dictionaries	66
4.7 Effect of Dictionary Redundancy on Period Estimation: A Numerical Example	70
4.8 The Case of Mixtures of Periodic Signals	73
4.9 Conclusion	74
4.10 Chapter Appendix	75
Chapter V: The Ramanujan Filter Bank and Its Applications	77
5.1 The Ramanujan Filter Bank	77

5.2	Connections to Dictionaries	79
5.3	Simulations	82
5.4	Protein Repeats	85
5.5	Detecting Tandem Repeats in DNA	91
5.6	Adapting the Ramanujan Filter Bank to Detect Tandem Repeats in DNA	93
5.7	Absence Seizure Detection Using Ramanujan Filter Banks	98
5.8	Unique Representation Filter Banks: Removing Redundancy in the RFB Using Truth Tables	104
5.9	Conclusion	111
Chapter VI: iMUSIC: A Family of MUSIC-like Algorithms for Integer Period Estimation		113
6.1	MUSIC and Periodicity: An Overview of Prior Works	116
6.2	The Proposed Methods	119
6.3	Generalizing iMUSIC From Farey Atoms to Other NPS Bases	121
6.4	Experiments	127
6.5	Conclusion	135
6.6	Chapter Appendix	136
Chapter VII: Minimum Datalength for Integer Period Estimation		137
7.1	Introduction	137
7.2	Minimum Datalength for the Single Period Case	140
7.3	Mixtures of Periodic Signals	145
7.4	Minimum Datalength for Estimating The Hidden Periods	150
7.5	Connection to Dictionaries Spanning Periodic Signals	154
7.6	Non-Integer Periodicity and Connections to Caratheodory's Results	157
7.7	Examples of DNA Repeats	160
7.8	Minimal Non-Contiguous Sampling For Period Estimation	162
7.9	Simulations Under Noise	167
7.10	Concluding Remarks	169
7.11	Chapter Appendix	170
Chapter VIII: Arbitrarily Shaped Periods in Multi-Dimensional Periodicity		172
8.1	A Note on Translational Matrices	179
8.2	Conclusion	181
Chapter IX: Concluding Remarks		182

LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
1.1 Examples of signals with periodicity. (a) repeats in Protein Molecules (PDB 1n11), (b) DNA microsatellites used in forensics, (c) EEG waveform during an episode of absence seizure, (d) periodicity in art: (left) Tesellations by M. C. Escher (1943), (right) periodicity in the weave pattern of paintings, commonly used in art forensics [1], (e) The chirp waveform that led to the recent discovery of gravitational waves [2], and (f) periodicity in ECG. . . .	2
1.2 100 samples of a signal that is a mixture of randomly generated period 3, period 7 and period 11 signals.	3
1.3 The spectra of periodic signals have a regular harmonic structure as shown. (a) The ideal spectrum when the datalength is infinitely long. (b) The DFT spectrum when the datalength is finite.	3
1.4 The 100 point DFT magnitude spectrum for the signal shown in Fig. 1.2. . .	4
1.5 Strength vs. Period plots produced for the signal in Fig. 1.2 using Nested Periodic Dictionaries, one of the techniques presented in Chapter 3.	5
1.6 Srinivasa Ramanujan (1887 - 1920).	6
1.7 The Ramanujan Filter Bank can be used to track time varying periodicity. .	8
1.8 Two signals with periods 4 (Top) and period 10 (Bottom) that share the first 11 samples shown in black. What is the minimum datalength needed to identify the period in general?	10
1.9 How many samples do we need to distinguish between periods 8 and 50? (Top) The minimum contiguous data-length we need is 56 samples as derived in Chapter 7. (Bottom) If we allow the samples to be non-contiguous, then just 8 are sufficient! The locations of these samples are shown using impulses in the above plots.	11
2.1 Examples of Ramanujan Sums with one period shown in each case.	14
2.2 Part (a) 256 samples of a period 64 signal. (b) The coefficients α_k 's for the expansion in (2.4). (c) The energies in the various Ramanujan subspaces obtained by solving (2.13). See text for details.	17

2.3	(Top) 100 samples of a noise-less period 10 signal. (Middle) Noise corrupted input (SNR = 5dB). (Bottom) The projection energies plotted vs. the period for the various Ramanujan subspaces. Notice that the LCM of the peak periods 2, 5 and 10 is indeed 10.	19
3.1	Part (a) and (b) – Strength vs period plots for a period 10 signal using 30×30 Natural Basis and Ramanujan Periodicity Transform matrices respectively.	30
3.2	Part (a) and (b) - Strength vs period plots for a period 70 signal using 70×70 Natural Basis and Ramanujan Periodicity Transform matrices respectively. The period 70 signal was generated as a sum of period 7 and period 10 signals.	31
3.3	The Ramanujan Nested Periodic Dictionary	32
3.4	(a) A noisy period 70 signal that was generated as the sum of a period 7 and a period 10 signal (SNR = 5.5dB). Parts (b), (c), (d) and (e) – The strength vs period plots for the solutions of the convex program (3.24) using Ramanujan, DFT, Random and Natural Basis dictionaries respectively.	34
3.5	Strength vs Period plot for the signal in Fig. 3.4(a). Part (a) – using the solution of the convex program (3.24). Part (b) – using the solution for the convex program (3.29).	36
3.6	Part (a) – Strength vs period plots for a period 104 signal generated as a sum of period 8 and period 13 signals using the Ramanujan dictionary. Parts (b), (c) and (d) – Strength vs period plots for shifted versions of the signal, where the shifts are 25, 50 and 75 samples respectively (see text for details).	37
3.7	Part (a) – Strength vs period plots for a period 104 signal generated as a sum of period 8 and period 13 signals using the Natural Basis dictionary. Parts (b), (c) and (d) – Strength vs period plots for shifted versions of the signal, where the shifts are 25, 50 and 75 samples respectively (see text for details).	38
3.8	Periodic decomposition by l_1 methods. Parts (a) and (c) show the strength vs period plots for the solutions of (3.31) for a noiseless period 70 signal that was generated as a sum of period 7 and period 10 signals, using Ramanujan and natural basis dictionaries respectively. Parts (b) and (d) show the strength vs period plots for the same dictionaries using the solutions to the convex program (3.32).	39

3.9	Part (a) – A period 70 signal generated as a sum of period 7 and period 10 signals. Parts (b) and (c) – Triangular and sinusoidal functions were multiplied respectively to the signal in part (a). Parts (d) and (e) show the strength vs period plots for the signals in parts (b) and (c) respectively using the convex program (3.24) and the Ramanujan dictionary.	41
3.10	Parts (a) and (b) – Amino acid sequence of the Antifreeze Protein from <i>Tenebrio Molitor</i> , numerically mapped according to two different metrics (refer text for details). Parts (c) and (d) – Corresponding strength vs period plots using the convex program (24) and the Ramanujan dictionary.	43
3.11	Part (a) - ECG waveform of a 23 year old female patient. Part (b) - DFT coefficients for positive frequencies.	44
3.12	Strength vs Period plot for the ECG signal of Fig. 3.11 using the convex program	44
4.1	The set of frequencies needed to span all periodic signals whose periods lie in the range $1 \leq P \leq 8$. See Sec. 4.5.	48
4.2	An example to illustrate that the projection energies obtained on the different \mathcal{V}_P 's depend on the order of projections. Refer Sec. 4.2 for details.	52
4.3	An example to illustrate that orthogonal subspaces result in unique projections irrespective of the order of decomposition. Refer Sec. 4.2 for details.	52
4.4	An example to illustrate that one can obtain a unique decomposition using parallelogram completion as long as the subspaces are linearly independent. Refer Sec. 4.2 and Sec. 4.4 for details.	57
4.5	Parts (a) and (b): Two different decompositions of a period 8 signal onto $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_8\}$. Part (a) involves $\mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_6$ and \mathcal{V}_8 , while Part (b) involves $\mathcal{V}_2, \mathcal{V}_4$ and \mathcal{V}_8 . Clearly, it is difficult to determine the component periods in the signal using these subspaces. Notice that the LCM property results in an incorrect period estimate in Part (a). Parts (c) and (d) use subspaces that offer unique periodic decompositions: Ramanujan Subspaces in Part (c) and the Natural Basis subspaces in Part (d). Both involve only subspaces with period 8 and its divisors. The LCM property correctly identifies the period as 8 in both these cases. Please see Sec. 4.4 for a discussion.	62

4.6	Effect of Redundancy on Signal Recovery: Part (a) - A randomly generated, noiseless, period 6 signal of length 56 samples. Part (b) - A noisy version of the signal in Part (a) (AWGN with SNR = 0dB). Part (c) - Reconstructed signal using the Ramanujan dictionary, with the threshold factor $T = 0.15$. Parts (d), (e) and (f) - Reconstructed signals using the Identity dictionary, with the threshold factor T being 0.15, 0.525 and 0.9 respectively. Please see Sec. 4.7 for the corresponding reconstruction errors and more details.	71
4.7	Parts (a) and (b) - Strength vs Period plots using the Identity and Ramanujan Dictionaries respectively for the signal shown in Fig. 4.6(b), after solving (3.24). The dotted lines indicate the thresholding parameter T discussed in Sec. 4.7.	72
5.1	Part (a) - Block diagram of the Ramanujan Filter Bank. Part (b) - An example of the impulse response of an RFB filter, $h_{10}(n)$. Part (c) and (d) - Frequency responses of $h_{10}(n)$ for $L = 5$ and $L = 20$ respectively.	78
5.2	The first 50 rows of the Pseudo-inverse matrix in (3.26) for $N = 200$ and $P_{max} = 200$	80
5.3	Parts (a) and (b)- Strength vs period plots for a period 70 signal that was generated as a sum of period 7 and period 10 signals. Part (a) shows the plot obtained by using (3.26), and Part (b) shows the plot obtained by using (5.2).	81
5.4	Parts (a) and (b) - The time vs. period plane for a signal exhibiting localized periodicities using RFB and shifted RFB. Parts (c) and (d) show the time-frequency plane using STFT with window sizes 128 and 32 respectively. Refer text for details.	83
5.5	Part (a) - Sampled inverse chirp signal. Part (b) - The time vs. period plane using shifted RFB. Parts (c) and (d) show the timefrequency plane using STFT with window sizes 256 and 32 respectively. Refer text for details.	84
5.6	HetL: A protein with 40 tandemly repeating pentapeptide repeats. The red arrows indicate two insertion loops that can be predicted using the RFB. See Sec. 5.4 for details. (Image source: [47], [54]–[56])	86
5.7	The period vs location plot produced using the RFB for HetL from Fig. 5.10, indicating the period 5 repeats. See Sec. 5.4 for details.	88

5.8	The output power of the fifth Ramanujan filter for the example of HetL. (Its time-period plane using the RFB is shown in Fig. 5.7.) The region in green above, marks the residues constituting the pentapeptide repeats, as revealed by the crystal structure of this protein [47]. The red regions, L_1 and L_2 , are insertion loops of sizes 6 and 9 residues respectively. These loops can be seen in the crystal structure of this protein shown in Fig. 5.10. Notice that, by observing the valleys in the output of the fifth Ramanujan filter, one can predict the existence of these insertion loops.	90
5.9	The time-period plot produced using an RFB with $K = 5$, for the β -propeller PDB:1hxn shown in Fig. 5.10. See Sec. 5.4 for details.	91
5.10	Protein repeats that were applied as inputs to FTwin [50], the wavelet based method of [53], RADAR [48], REPwin [50], and the RFB. These are abbreviated as FTw., WAV., RAD., REPw., and RFB in the top row. The first column indicates the type of repeat family being considered. The secondary structure of a representative member of each repeat family is also shown. The second column contains the Protein Data Bank (PDB) ID number for each example. For the ANKK1 protein exhibiting ankyrin repeats, its NCBI reference sequence number is shown instead. The wavelet based algorithm in [53] was not available publicly. Hence, we only show their results for two examples from [53]. See text for details. (Image sources: Jmol: http://www.jmol.org/ and JSmol http://wiki.jmol.org/index.php/JSmol . [47], [54], [63]–[68])	92
5.11	The section of the Time vs Period plane containing the tandem repeat of Table 5.3. Notice that the end location does not match with Table 5.5. See Sec. 5.6 for details.	94
5.12	An example of the 3Hz spike and wave discharge pattern in the EEG during an absence seizure.	99
5.13	RFB's test of sensitivity. See Sec. 5.7 for details.	100
5.14	(Top) The entire record of sampled EEG containing Seizure 2 in Patient 1, as measured across the F8-C4 channel. (Bottom) The RFB's time vs period plane.	102
5.15	Comparison of the RFB with STFT. See Sec. 5.7 for details.	103
5.16	Demonstrating a URFB for Table 5.7. See Sec. 5.8 for details.	112
6.1	Part (a) - An arbitrary line spectrum; Part (b) - The harmonic line spectrum of a periodic signal. Can we use the additional structure in the spectrum of a periodic signal to improve MUSIC?	114

6.2	<i>Applications with Integer Periodicity: The protein AnkyrinR (PDB 1n11) that enables red blood cells to resist shear forces. Its period 33 structural repeats can be clearly identified in the plot on the right, produced by the proposed techniques.</i>	114
6.3	<i>Simple Integer Alternatives to Complex-Exponentials: Bases of (a) The Natural Basis Subspaces, and (b) The Ramanujan Subspaces.</i>	116
6.4	Demonstrating the proposed iMUSIC method using a Farey dictionary on a mixture of periods 3, 10 and 13. (a) The noisy periodic signal, (b) conventional MUSIC, and (c) the new iMUSIC method in (6.19).	119
6.5	Demonstrating the NPS based iMUSIC methods on the signal shown in Fig. 6.4 using (a) a Ramanujan dictionary, (b) a natural basis dictionary and (c) a randomly generated NPS dictionary.	122
6.6	The effect of $K < K'$ on the pseudo-spectrum of (a) Ramanujan Subspaces (Farey basis) (b) Ramanujan Subspaces (integer basis) (c) Natural Basis Subspaces and (d) Randomly generated NPSs.	126
6.7	Probability of Estimating both the component periods exactly. Comparison of the proposed Farey-MUSIC with other techniques. See	128
6.8	Probability of Estimating both the component periods exactly. Comparison of the various NPSs. HMUSIC and MUSIC have been included for reference.	130
6.9	A comparison of the CPU Times. See Sec. 6.4 for details.	130
6.10	Pseudospectra of the proposed NPS based techniques for the Ankyrin protein repeats shown in Fig. 6.2.	131
6.11	Top: The protein Ribonuclease Inhibitor (PDB: 1dfj) exhibiting luciferin-rich repeats. The pseudo-spectra obtained from (a) Ramanujan Subspaces (Farey basis) (b) Ramanujan Subspaces (Integer Basis) (c) Natural Basis Subspaces and (d) a randomly generated NPS, are shown. See	133
6.12	Top: An example of DNA microsatellites that are used in DNA fingerprinting. The pseudo-spectra obtained from (a) Ramanujan Subspaces (Farey basis) (b) Ramanujan Subspaces (Integer Basis) (c) Natural Basis Subspaces and (d) a randomly generated NPS, are shown. See	134
7.1	Part (a) - An arbitrary line spectrum; Part (b) - The harmonic line spectrum of a periodic signal. Can we use this additional structure in the spectrum of a periodic signal to reduce the data length required for period estimation?	139

7.2	<i>An example illustrating the necessity of L_{min} (Eq. (7.5)) samples.</i> Part (a) - A period 6 signal; Part (b) - A period 15 signal. The initial 17 samples shown in black are common to both signals.	141
7.3	Error Rate vs Data-Length for a fixed SNR. "Th" refers to the threshold. SNR = 0dB. See text for details.	146
7.4	Demonstration of Theorem 7.2.3 using DNA repeats. (a) Period 4 repeats used in DNA fingerprinting applications (GenBank G08921). (b) A period 10 sequence constructed such that its first 11 samples (shown in black) are the same as in the sequence (a).	160
7.5	DFT spectra for the sequences in Fig. 7.4. The periods ($2\pi/\omega$ values) corresponding to the peak frequencies are shown in parenthesis. See text for details.	161
7.6	Demonstration of Theorem 7.2.1 using DNA repeats. (a) Period 5 repeats used in DNA fingerprinting applications (GenBank M86525). (b) A period 6 sequence constructed such that its first 9 samples (shown in black) are the same as in the sequence (a). See text for details.	162
7.7	Finding the period of $x(n)$ when $(P_1, P_2) = G \neq P_1$. See text for details.	165
7.8	Re-drawing Fig. 7.7 for analysis. See text for details.	166
7.9	Error rate vs. Number of Samples. $P_1 = 9, P_2 = 13, \text{SNR} = 0\text{dB}$. See Sec. 7.9 for details.	168
7.10	Error rate vs. SNR for the Minimum Samples Case. $P_1 = 9, P_2 = 13$, Number of samples = $P_1 = 9$. See Sec. 7.9 for details.	169
8.1	Part (a) - A two dimensional periodic signal according to the definition in (8.1), whose period is represented by the matrix in (8.2). The grid of dots is the set \mathbb{Z}^2 . The numbers shown indicate the value of the signal at those integer points. Part (b) - A convenient way to indicate repetition regions for such periodic signals.	173
8.2	A two dimensional periodic signal with a non-parallelogram repetition region (shaded in black) tiling the plane.	174
8.3	Another example of a two dimensional periodic signal with a non-parallelogram repetition region (shaded in black) tiling the plane.	174
8.4	A parallelepiped repetition region always exists for any signal following Definition 8.0.1. Indicating this for the signal of Fig. 8.2.	180

LIST OF TABLES

<i>Number</i>	<i>Page</i>
3.1 Comparison of CPU times for solving (3.32) and (3.24)	40
5.1 Period Estimation using STFT and the RFB.	84
5.2 Example 1: Human Frataxin Gene	95
5.3 Base pairs 684 - 720 in Example 1	95
5.4 Example 2: Human Genome Sequence AC010136	97
5.5 Ramanujan Filter Bank: Period to Filters Map	105
5.6 Tables Mapping Periods To Unique Sets of Filters. The Table on the Left is Unimplementable.	106
5.7 A URFB for periods 1 to 4.	111
7.1 Comparison between (7.30) and (7.41) for $P_{max} = 20$	156

Chapter 1

INTRODUCTION

One of the most beautiful patterns in nature is periodicity: the ripples on the surface of a pond, the cyclic phases of the moon, chirps of birds, gravitational waves from distant stars - these are just a few among countless examples of periodicity and its variations. Even in our very own human-made world, periodicity is ubiquitous, for example in machine vibrations, communication waves, RADAR and SONAR signals, artistic renditions etc. In fact, one may wonder if the notion of “a pattern” can imply anything but repetitions of some form. Fig. 1.1 shows many other examples.

From the perspective of a DSP researcher, the simplest way to capture periodicity is as follows: A sequence $x(n), n \in \mathbb{Z}$ is said to be periodic if there exists an integer P such that:

$$x(n + P) = x(n) \quad \forall n \in \mathbb{Z} \quad (1.1)$$

Any such P is called a **repetition index** of the signal, and the smallest positive repetition index is known as the **period**. Notice that the period is an integer according to this model. In this thesis, we are primarily interested in estimating the period, given a finite-length, and possibly noisy version of $x(n)$.

1.1 A Motivating Example

Period estimation is a classical problem in DSP. A number of techniques have been proposed for this over many years in various application domains. Why then do we revisit at this problem here? Let us motivate our contribution in this thesis using a simple example. Consider the signal shown in Fig. 1.2. The 100 samples shown were generated by adding randomly-generated signals with periods 3, 7 and 11. Our task is to estimate these three hidden periods from the 100 samples shown.

Most of the popular techniques today for estimating periods are based on the Discrete Fourier Transform or DFT (for e.g. [3]–[7]). The main idea is as follows: in the frequency domain, a periodic signal’s spectrum has a nice harmonic structure as shown in Fig. 1.3 (a). When the data-length is infinitely long, the spectrum consists of lines, spaced at harmonic multiples of a fundamental frequency. Ideally, from such a spectrum, we should be able to identify the fundamental frequency, whose

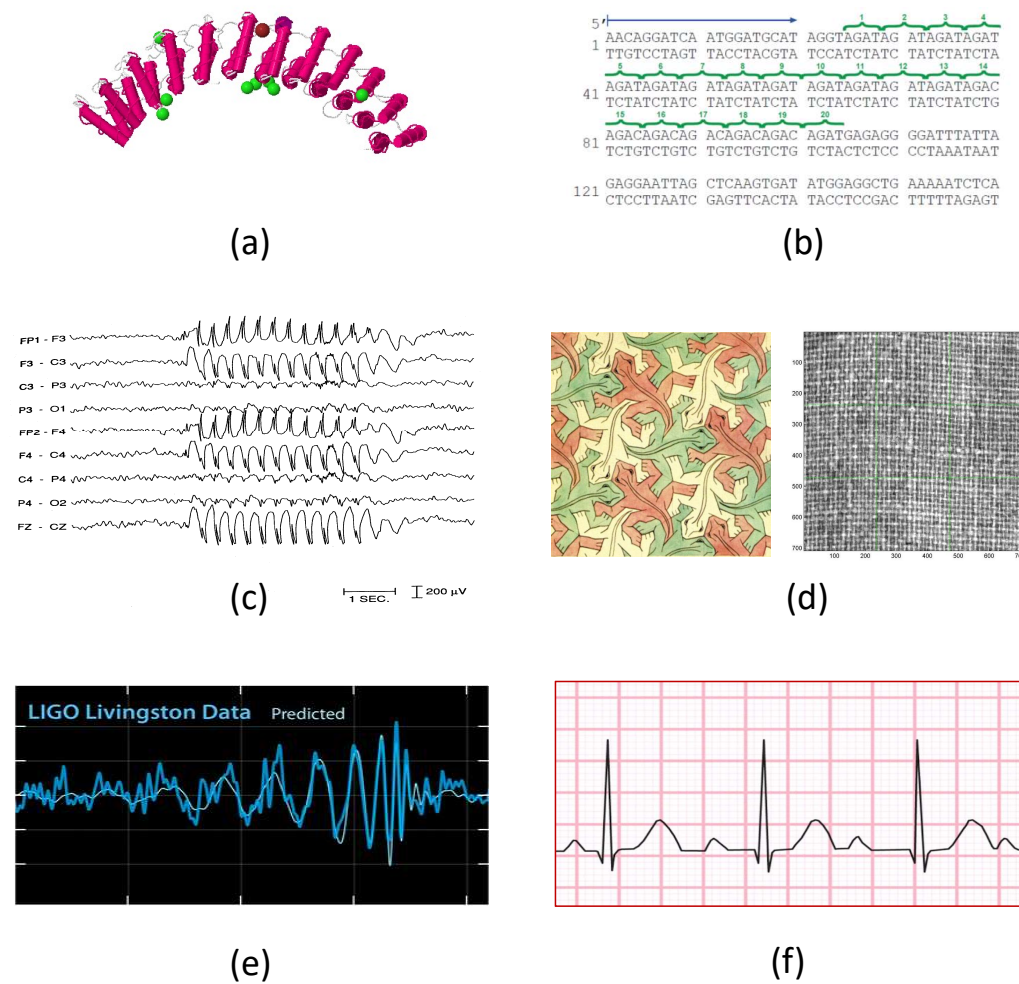


Figure 1.1: Examples of signals with periodicity. (a) repeats in Protein Molecules (PDB 1n11), (b) DNA microsatellites used in forensics, (c) EEG waveform during an episode of absence seizure, (d) periodicity in art: (left) Tesellations by M. C. Escher (1943), (right) periodicity in the weave pattern of paintings, commonly used in art forensics [1], (e) The chirp waveform that led to the recent discovery of gravitational waves [2], and (f) periodicity in ECG.

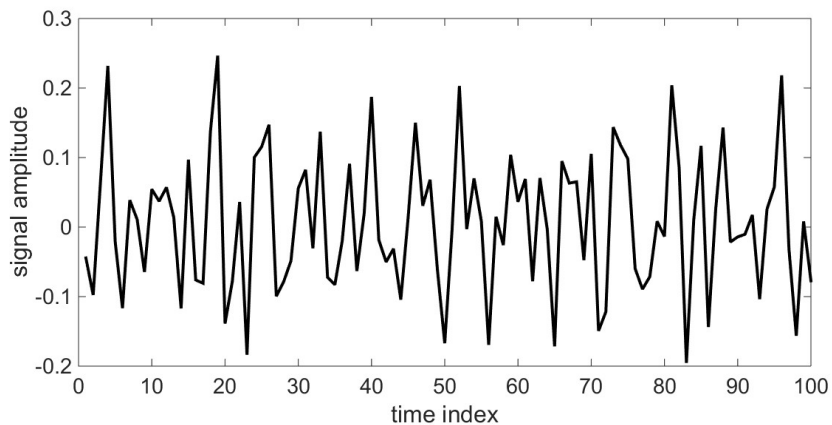


Figure 1.2: 100 samples of a signal that is a mixture of randomly generated period 3, period 7 and period 11 signals.

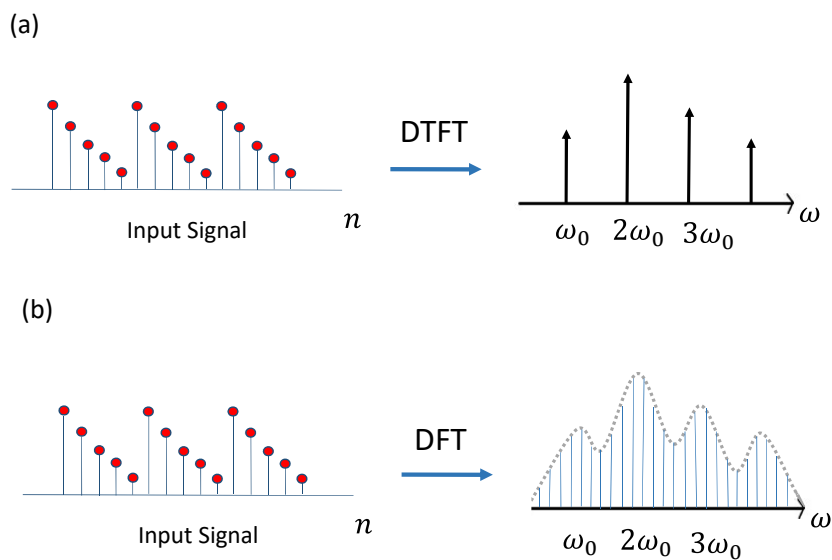


Figure 1.3: The spectra of periodic signals have a regular harmonic structure as shown. (a) The ideal spectrum when the datalength is infinitely long. (b) The DFT spectrum when the datalength is finite.

inverse yields the period:

$$P = \frac{2\pi}{\omega_0} \quad (1.2)$$

Unfortunately, when the available data-length is short, the spectrum computed using the DFT tends to spread out as shown in Fig. 1.3 (b). This becomes problematic,

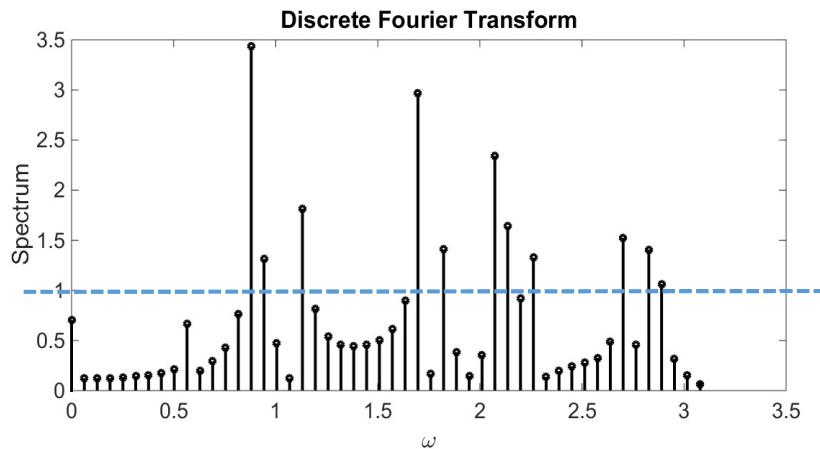


Figure 1.4: The 100 point DFT magnitude spectrum for the signal shown in Fig. 1.2.

especially when the input is a mixture of multiple periodic signals, as is the case here. Fig. 1.4 shows the 100 point DFT magnitude spectrum for the signal in Fig. 1.2. The tallest peaks correspond to the DFT indices:

$$k = 14, 15, 18, 27, 29, 33, 34, 36, 43, 45, 46 \quad (1.3)$$

What periods do these indicate? A peak in the 100 point DFT at $k = k_0$ corresponds to the complex exponential $e^{j2\pi k/100}$. There are two ways to interpret its period:

1. By comparing it with the continuous time signal $e^{j\omega t}$, whose period is period $2\pi/\omega$, we may infer the period of $e^{j2\pi k/100}$ as:

$$P = \frac{100}{k} \quad (1.4)$$

This approach has been commonly used in the literature, but notice that it may not yield an integer estimate for every k . The peaks in Fig. 1.4 according to (1.4) correspond to the following periods:

$$7.14, 6.67, 5.55, 3.70, 3.45, 3.03, 2.94, 2.78, 2.32, 2.22, 2.17 \quad (1.5)$$

Clearly, it is difficult to identify the true periods 3, 7 and 11 from this set.

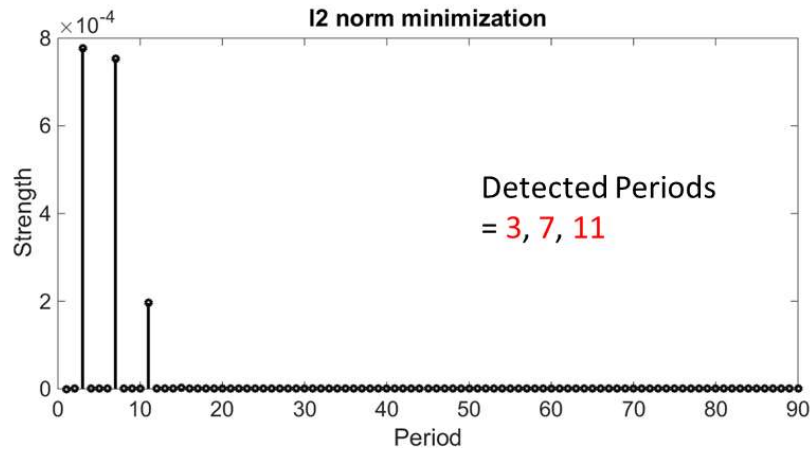


Figure 1.5: Strength vs. Period plots produced for the signal in FIG. 1.2 using Nested Periodic Dictionaries, one of the techniques presented in Chapter 3.

2. If we follow the strict integer period definition of (7.19), then the period of $e^{j2\pi k/100}$ can be shown to be:

$$P = \frac{100}{\text{GCD}(100, k)} \quad (1.6)$$

Unfortunately, the detected periods by this approach can only be divisors of the DFT length, which is 100 in this case. In Fig. 1.4, the peaks correspond to the following periods according to (1.6):

$$100, 50, 25, 20 \quad (1.7)$$

Once again, these do not indicate the true periods 3, 7 and 11. Clearly, neither of the two approaches using DFT works in this example.

As a motivating preview of what we are going to develop in this thesis, Fig. 1.5 shows the results of using Nested Periodicity Dictionaries, one of the techniques that we propose in this thesis. Unlike the DFT spectrum in Fig. 1.4, the x -axis here is directly the period. The y -axis shows the *strength* of the various periods present in the signal. Notice that the correct periods 3, 7 and 11 can be seen much more cleanly in this plot. This plot was in fact produced using an overcomplete dictionary, but using a simple and fast l_2 norm based convex program will be explained in Chapter 3.

While this was just one example, in the following chapters, we will develop a number of techniques that outperform classical methods in applications where the



Figure 1.6: Srinivasa Ramanujan (1887 - 1920).

datalength is short, or when there are multiple hidden periodic components in the signal. The applications we study include DNA microsatellites, protein repeats and absence seizures. Apart from practical contributions, the theory that we develop in this thesis enables us to answer a number of fundamental questions in the context of periodicity that were never addressed before. A summary of our main contributions in this thesis is as follows:

1.2 Outline and Contributions of This Thesis

Chapter 2: Prior Work - Ramanujan Sums and Subspaces - The building blocks for the work presented in this thesis are sequences known as Ramanujan Sums. These were introduced in 1918 by the Indian mathematician Srinivasa Ramanujan (Fig. 1.6) [8]. He showed that several arithmetic functions in number theory such as the divisor function and the Euler totient function can be expanded as a series in Ramanujan sums.

In 2014, it was shown by Vaidyanathan [9] that the Ramanujan sums have several useful properties in a completely different context, namely periodicity analysis. They

were used to construct a set of mutually orthogonal subspaces known as Ramanujan subspaces, which could span any given periodic sequence. It was shown that by projecting a given periodic signal onto the Ramanujan subspaces, one could easily estimate its period. Chapter 2 presents a short overview of these results.

Chapter 3: Nested Periodic Subspaces and Dictionaries - In Chapter 3, we first generalize Ramanujan subspaces to an entire family of such subspaces. These, called Nested Periodic Subspaces, share all the essential properties of Ramanujan subspaces. From a theoretical perspective, the Nested Periodic Subspaces provide interesting connections between Ramanujan subspaces and well known transforms such as DFT and Hadamard transforms. New dictionary based period estimation algorithms are proposed using these subspaces. In addition to the conventional l_1 norm based support recovery [10], [11], the dictionaries constructed using nested periodic subspaces can also admit new l_2 norm based convex programs that have closed form solutions. These result in orders of magnitude faster solutions than the l_1 methods, and give much cleaner period estimates than conventional techniques.

Chapter 4: A General Theory of Subspace Models for Periodic Signals - Are Nested Periodic Subspaces the only possible subspaces that span periodic signals? We will show in Chapter 4 that this is not the case. In fact a few other examples of such subspaces had already been proposed in prior literature [12]–[14]. However, Nested Periodic Subspaces have some very unique properties compared to arbitrary union-of-subspaces models for periodic signals, such as (a) offering unique decompositions, (b) offering the least redundant dictionaries, (c) exhibiting the so-called lcm property, and so on. We will study these in Chapter 4.

So far, most of these other subspace-models for periodic signals were developed independent of each other. There is no unified theory analyzing them from a common perspective. In Chapter 4, all such methods are first unified under one general framework. Further, several fundamental aspects of such subspaces are investigated from a generic perspective, such as the conditions under which a set of subspaces offers unique periodic decompositions, their minimum required dimensions, etc. A number of basic questions in the context of dictionaries spanning periodic signals are also answered. For example, what is the theoretically minimum number of atoms required in any type of dictionary, in order to represent periods $1 \leq P \leq P_{max}$? For each period P , what should be the minimum dimension of the subspace of atoms representing the P th period itself? It will be shown that the Nested Periodic Subspaces turn out to be unique examples of subspaces that satisfy the above

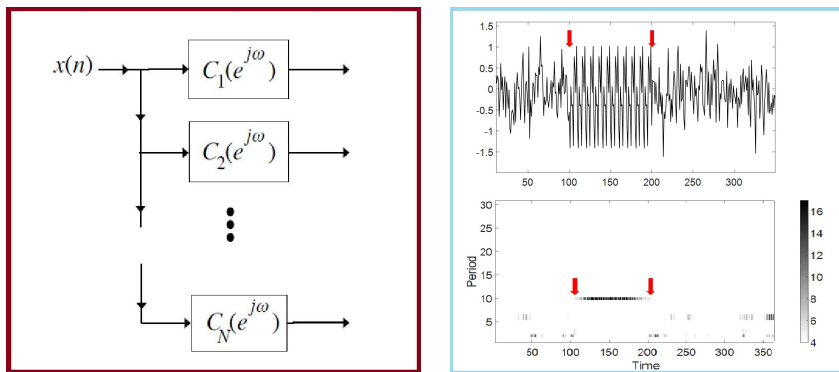


Figure 1.7: The Ramanujan Filter Bank can be used to track time varying periodicity.

minimality criteria. Interestingly, it will be seen that the Euler totient function from number theory plays an important role in providing the answers to all such questions.

Chapter 5: The Ramanujan Filter Bank and Its Applications - In a number of applications, the periodicity in the signal changes across the samples. Examples include chirps and signals with localized periodicity such as DNA microsatellites or epileptic seizures (see Fig. 1.1). One of the ways to process such signals is to use the Nested Periodic Dictionaries in a block-by-block fashion, but this has some limitations. Computational complexity is one. But more importantly, larger periods would require larger block-lengths, while smaller periods may need smaller block-lengths for good localization. The dictionaries on the other hand impose a common block-length across all periods.

In order to efficiently deal with such signals, a new filter bank implementation is developed using Ramanujan sums. This, which we call the Ramanujan Filter Bank, will be derived in Chapter 5 (see Fig. 1.7). It will be shown that if the input has period P , then the lcm of the indices of those filters that have non-zero outputs will be exactly equal to P . Three real world applications will also be shown: namely, DNA microsatellites, Protein repeats and Absence seizures. In each of these applications, the Ramanujan filter bank is compared with the state-of-the-art techniques.

A further extension of the Ramanujan Filter Bank to what are known as Unique Representation Filter Banks is also presented. While the Ramanujan Filter Bank is itself quite efficient computationally, it will be shown that for applications where

the signal is a single periodic signal (as opposed to a mixture), one can design filter banks with far fewer filters.

Chapter 6: iMUSIC - Integer Basis Alternatives to MUSIC for Periodicity

- The MUSIC algorithm [15] is one of the most popular techniques today for line spectral estimation. But if the line spectrum is that of a periodic signal, can we adapt MUSIC to exploit the additional harmonic structure in the spectrum (Fig. 1.3)? Prior work in this direction includes the Harmonic MUSIC (HMUSIC) algorithm [16] and its variations [17]. These involve rather complicated non-convex optimizations to search for harmonically spaced peaks in the conventional MUSIC pseudo-spectrum.

For applications where the period of the discrete signal is an integer (or can be well approximated by an integer), we introduce a new and much simpler class of variants of MUSIC. This new family, called iMUSIC, includes techniques where simple integer valued vectors are used in place of complex exponentials for both representing the signal subspace, and for computing the pseudo-spectrum. It will be shown that the proposed methods not only make the computations much simpler than prior adaptations of MUSIC, but also offer significantly better estimation accuracies for applications with integer periods.

The iMUSIC algorithms are based on Ramanujan Subspaces and Nested Periodic Subspaces. So the resulting signal space bases are non-Vandermonde in structure, unlike MUSIC. Consequently, many aspects of classical MUSIC that were based on the Vandermonde structure of complex-exponentials, such as guarantees for identifiability of the frequencies (periods in our case), are addressed in new ways in this work.

Chapter 7: Minimum Datalength for Integer Period Estimation

- Consider Fig. 1.8. There are two signals shown, with periods 4 and 10. The first 11 samples, shown in black, are common to both signals. Clearly, given these 11 or fewer samples of these signals, it is impossible to estimate what the true underlying period is. More generally, suppose we have a signal $x(n)$ whose period is known to belong to the integer set:

$$\mathbb{P} = \{P_1, P_2, \dots, P_K\} \quad (1.8)$$

For example, this could be the set of all positive integers less than 100. Then, what is the absolute minimum datalength needed to be able to identify the true period?

In spite of period estimation having a rich history, this particular question has

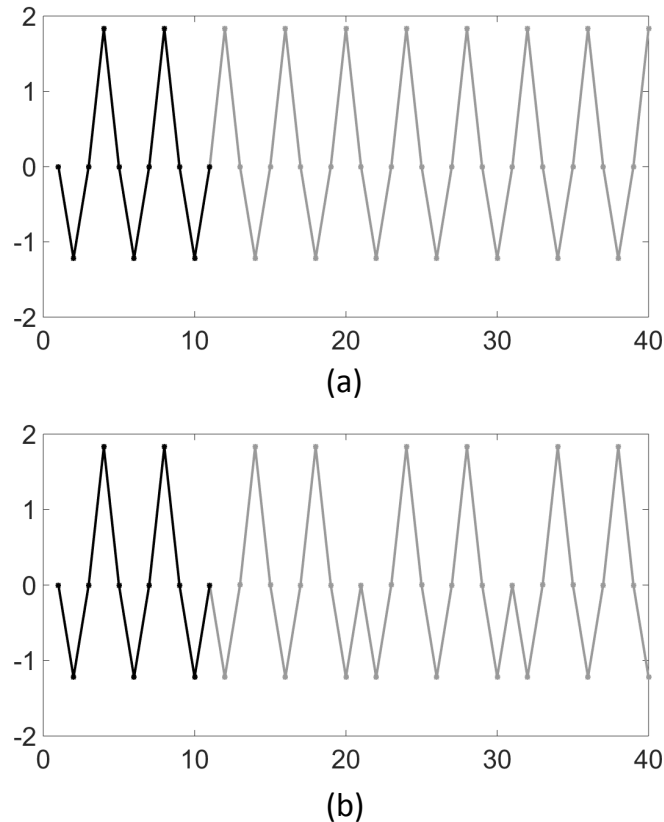


Figure 1.8: Two signals with periods 4 (Top) and period 10 (Bottom) that share the first 11 samples shown in black. What is the minimum datalength needed to identify the period in general?

surprisingly received very little attention from a fundamental perspective. Notice that the answer we seek must be a fundamental bound, i.e., independent of any particular period estimation technique. Common intuition suggests the minimum datalength as twice the largest expected period. However, this is true only under some special contexts. Chapter 7 derives the exact necessary and sufficient bounds to this problem. The above question is also extended to the case of mixtures of periodic signals. First, a careful mathematical formulation discussing the unique identifiability of the component periods (hidden integer periods) is presented. Once again, a rigorous theoretical framework in this regard was missing in the past literature, but is a necessary platform to derive precise bounds on the minimum necessary datalength. The bounds given here are generic, that is, independent of the algorithms used. Specific algorithm-dependent bounds are also presented in the end for the case of dictionary based integer period estimation using Ramanujan subspaces.

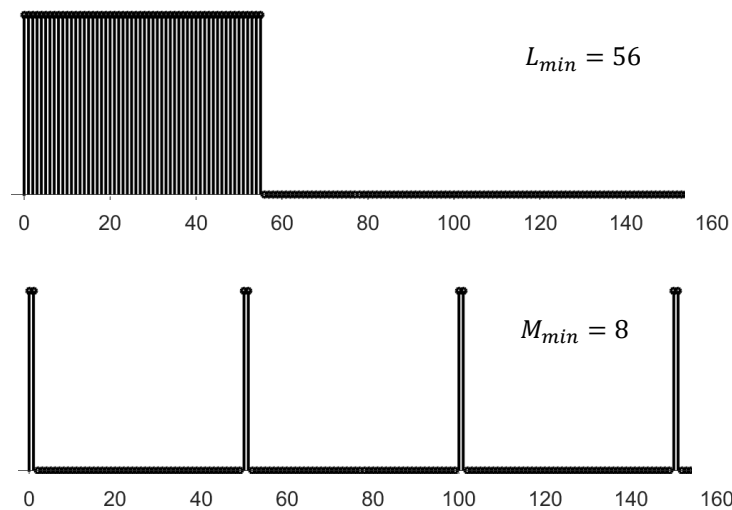


Figure 1.9: How many samples do we need to distinguish between periods 8 and 50? (Top) The minimum contiguous data-length we need is 56 samples as derived in Chapter 7. (Bottom) If we allow the samples to be non-contiguous, then just 8 are sufficient! The locations of these samples are shown using impulses in the above plots.

A further question in this direction is, suppose we are allowed to pick samples in a non-contiguous fashion. Then, how should we select those samples such that we can estimate the period using the least number of samples? This question is quite difficult to answer in general. In this thesis, we address a smaller case in this regard, namely, that of resolving between two periods. It will be shown that the analysis is quite involved even in this case, and the optimal sampling pattern takes the form of sparsely located bunches as shown in Fig. 1.9.

Chapter 7: Multi-Dimensional Discrete Time Periodicity - We address a very basic aspect regarding multidimensional periodicity here: Traditionally, most of the analysis in DSP has been based on defining the period as a parallelepiped. In this chapter, we study whether this framework can incorporate signals that are repetitions of more general shapes than parallelepipeds. For example, the famous Dutch artist M. C. Escher constructed many interesting shapes such as fish, birds and animals, which can tile the continuous 2-D plane¹. Inspired from Escher's tilings, we construct discrete time signals that are repetitions of various kinds of shapes. We look at periodicity in the following way - a given shape repeating itself along fixed directions to tile the entire space. By transcribing this idea into a mathematical

¹See <http://www.tessellations.org/eschergallery1thumbs.shtml> for many interesting examples.

framework, we explore its relationship with the traditional analysis of periodicity based on parallelepipeds.

Notations

Before proceeding, it will be useful to list some notations we use throughout the thesis:

1. $D|P$ denotes that D is a divisor of P .
2. (k, d) denotes the greatest common divisor (gcd) of k and d . The least common multiple of k and d is denoted by $lcm(k, d)$.
3. $\phi(d)$ is the Euler-totient function of d . It is equal to the number of positive integers $\leq d$ and coprime to d .
4. Vectors are denoted by bold lower case font (e.g. \mathbf{x}), matrices by bold upper case font (e.g. \mathbf{A}), and sets by blackboard font (e.g. \mathbb{B}).
5. \mathbf{A}^\dagger denotes the Hermitian transpose (or conjugate transpose) of the matrix \mathbf{A} .
6. \mathbb{Z} denotes the set of all integers. \mathbb{R} denotes the set of all real numbers. And \mathbb{C} denotes the set of all complex numbers.

Chapter 2

RAMANUJAN SUMS AND SUBSPACES

The work that motivated this thesis was the development of Ramanujan subspaces in [9], [18]. These subspaces were proposed as a generalization of number theoretic sequences known as Ramanujan sums [8]. In this chapter, we will summarize some of the important properties of Ramanujan sums and subspaces. These will be important for understanding the results developed in later chapters. Most of the results presented in this chapter were published in [9] and [18].

2.1 Ramanujan Sums

In 1918, the Indian mathematician Srinivasa Ramanujan introduced a set of sequences known as Ramanujan Sums [8]. For every integer $q > 0$, the q^{th} Ramanujan sum was defined as follows:

$$c_q(n) = \sum_{\substack{k=1 \\ (k,q)=1}}^q e^{j2\pi kn/q} \quad (2.1)$$

The notation (k, q) denotes the GCD of k and q . Notice that the sum involves those columns of a $q \times q$ DFT matrix, whose indices are co-prime to q .

Ramanujan showed that these sequences have several interesting properties [8]. The following are the most relevant for us:

1. **Periodicity:** $c_q(n)$ is periodic. Its period is exactly equal to q (and not a proper divisor of it). Notice that adding together arbitrary columns of the $q \times q$ DFT matrix can result in any divisor of q as the period. The Ramanujan sums are special, since the period is exactly q .
2. **Integer Valued:** A surprising property is that $c_q(n)$ is real and integer-valued for $\forall n$. Fig. 2.1 shows some examples.
3. **Orthogonality:** $c_q(n)$ is orthogonal to $c_p(n)$ for any $p \neq q$, when considering segments that are of length a common multiple of p and q .

In his original application, Ramanujan showed that many arithmetic functions in number theory can be expanded as a series using Ramanujan sums. For instance,

$$\begin{aligned}
c_1(n) &= 1 \\
c_2(n) &= 1, -1 \\
c_3(n) &= 2, -1, -1 \\
c_4(n) &= 2, 0, -2, 0 \\
c_5(n) &= 4, -1, -1, -1, -1 \\
c_6(n) &= 2, 1, -1, -2, -1, 1 \\
c_7(n) &= 6, -1, -1, -1, -1, -1, -1 \\
c_8(n) &= 4, 0, 0, 0, -4, 0, 0, 0
\end{aligned}$$

Figure 2.1: Examples of Ramanujan Sums with one period shown in each case.

the divisor function $d(n)$, which denotes the number of divisors of an integer n can be expanded as:

$$-d(n) = \frac{\log 1}{1}c_1(n) + \frac{\log 2}{2}c_2(n) + \frac{\log 3}{3}c_3(n) + \dots \quad (2.2)$$

Similarly, Euler's totient function can be written as:

$$\phi(n) = \frac{6n}{\pi^2} \left(c_1(n) - \frac{c_2(n)}{2^2 - 1} - \frac{c_3(n)}{3^2 - 1} - \frac{c_5(n)}{5^2 - 1} + \dots \right) \quad (2.3)$$

2.2 Early Use of Ramanujan Sums in DSP

From a signal processing perspective, the above expansions of arithmetic functions lead naturally to the following question: Given a signal $x(n)$, are there any merits to an expansion as follows?

$$x(n) = \sum_{q=1}^{\infty} \alpha_q c_q(n) \quad (2.4)$$

The periodicity of $c_q(n)$ led many signal processing researchers to suspect that one may be able to detect hidden periodic components in $x(n)$ using such an expansion ([19]–[22]). For instance if say α_8 is significantly larger than other coefficients for a particular $x(n)$, then this was considered suggestive of a strong period 8 component. This idea was tested on many applications, including cardiology [20], RADAR [22], DNA [21] and so on.

Unfortunately, there is limited mathematical justification for the above idea. For example, suppose $x(n) = c_8(n - 1)$. While this is clearly a period 8 sequence, it can be shown that $c_8(n - 1)$ is orthogonal to $c_q(n)$ for every q , including $q = 8$. So an expansion of the form (2.4) will yield all the α_q 's to be 0.

2.3 Generalizing Ramanujan Sums to Ramanujan Subspaces

Motivated by the above limitation, a generalization of Ramanujan sums to subspaces known as *Ramanujan Subspaces* was proposed in 2014 [9], [18]. Instead of using just one sequence $c_q(n)$ to represent period q as in (2.4), it was proposed to use an entire subspace \mathcal{S}_q instead:

$$\mathcal{S}_q = \text{span}\{e^{j\frac{2\pi k}{q}n} : (k, q) = 1\} \quad (2.5)$$

This was called the q^{th} Ramanujan subspace. Just like the Ramanujan sums, the Ramanujan subspaces have some very interesting properties. The most significant are the following:

1. **Periodicity and Dimension:** \mathcal{S}_q is a subspace where every non-zero signal has period exactly q (and not a proper divisor of q). It is easy to see from (2.5) that its dimension is $\phi(q)$.
2. **Integer Basis:** While (2.5) shows that \mathcal{S}_q can be spanned by a basis of complex exponentials, one can in fact construct much simpler integer valued bases instead. For example, shifted versions of the Ramanujan sums can also be used:

$$\mathcal{S}_q = \text{span}\{c_q(n - m) : 0 \leq m \leq \phi(q) - 1\} \quad (2.6)$$

In Chapter 3, we will discuss many other ways to construct integer bases for \mathcal{S}_q , including connections with some well known transforms such as the Hadamard transform [23].

3. **Orthogonality:** Just like the Ramanujan sums, the Ramanujan subspaces are mutually orthogonal:

$$\mathcal{S}_q \perp \mathcal{S}_p \quad \forall p \neq q \quad (2.7)$$

4. **Direct Sum Property:** The set of all period q signals is not a linear space, since adding two period q signals can reduce their period to a proper divisor of q . However, the set of all signals that repeat after q samples is indeed a linear space:

$$\mathcal{V}_q = \{x(n) : x(n + q) = x(n)\} \quad (2.8)$$

We will keep referring to this space often throughout this thesis. Now, the Ramanujan subspace \mathcal{S}_q by itself cannot span all period q signals. Its dimension is only $\phi(q)$, whereas the number of degrees of freedom in period q signals is q . However, it can be shown that:

$$\bigoplus_{d|q} \mathcal{S}_d = \mathcal{V}_q \quad (2.9)$$

That is, all the Ramanujan subspaces corresponding to divisors of q together span signals in \mathcal{V}_q .

2.4 The Ramanujan Periodicity Transform (RPT) Matrix

To understand how the above properties can be used for period estimation, it is useful to look at the following square matrix. Let P be a positive integer and let d_1, d_2, \dots, d_K be its divisors. Consider the following $P \times P$ matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{C}_{d_1} & \mathbf{C}_{d_2} & \dots & \mathbf{C}_{d_K} \end{bmatrix} \quad (2.10)$$

Here, each \mathbf{C}_{d_i} is a $P \times \phi(d_i)$ matrix whose columns are:

$$\begin{bmatrix} c_{d_i}(0-m) \\ c_{d_i}(1-m) \\ \vdots \\ c_{d_i}(P-1-m) \end{bmatrix}, \quad 0 \leq m \leq \phi(d_i) - 1 \quad (2.11)$$

An example for $P = 8$ is shown below, where the first column is periodic with period 1, the second with period 2, the third and fourth with period 4 and the rest with period 8:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 & 0 & 4 & 0 & 0 & 0 \\ 1 & -1 & 0 & 2 & 0 & 4 & 0 & 0 \\ 1 & 1 & -2 & 0 & 0 & 0 & 4 & 0 \\ 1 & -1 & 0 & -2 & 0 & 0 & 0 & 4 \\ 1 & 1 & 2 & 0 & -4 & 0 & 0 & 0 \\ 1 & -1 & 0 & 2 & 0 & -4 & 0 & 0 \\ 1 & 1 & -2 & 0 & 0 & 0 & -4 & 0 \\ 1 & -1 & 0 & -2 & 0 & 0 & 0 & -4 \end{bmatrix} \quad (2.12)$$

Notice that the columns of \mathbf{C}_{d_i} form a basis for \mathcal{S}_{d_i} using (2.6). Using the orthogonality property of Ramanujan subspaces, it is easy to see that \mathbf{A} has full rank.

Using the direct sum property, it can be seen that the columns of \mathbf{A} , when periodically extended, form a basis for \mathcal{V}_P . Not only this, but one can identify exactly which

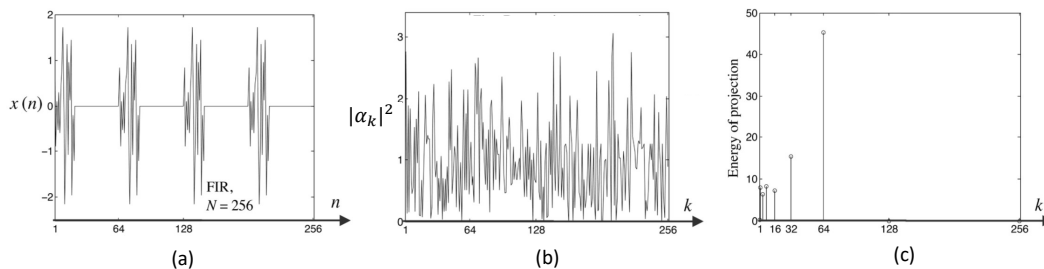


Figure 2.2: Part (a) 256 samples of a period 64 signal. (b) The coefficients α_k 's for the expansion in (2.4). (c) The energies in the various Ramanujan subspaces obtained by solving (2.13). See text for details.

divisor of P the period of the signal is. This follows from the following theorem, proved in [9]:

Theorem 2.4.1. *Let \mathbf{x} be a $P \times 1$ vector of a signal $x(n) \in \mathcal{V}_P$. Suppose the period of $x(n)$ is $q|P$. Let \mathbf{A} be the $P \times P$ RPT matrix. Then, the following system of equations has a unique solution for \mathbf{y} :*

$$\mathbf{x} = \mathbf{A}\mathbf{y} \quad (2.13)$$

Further, if P_1, P_2, \dots, P_M are the periods of those columns of \mathbf{A} that multiply with non-zero entries in \mathbf{y} , then

$$q = \text{LCM}\{P_1, P_2, \dots, P_M\} \quad (2.14)$$

As an example, Fig. 2.2(a) shows 256 samples of a period 64 signal. Fig. 2.2(b) plots the coefficients α_k 's of an expansion based on (2.4). It is clearly difficult to identify the true period 64 from this plot. Fig. 2.2(c) plots the energy in the solution \mathbf{y} for various Ramanujan subspaces. The LCM of the non-zero periods is 64, which is indeed the true period.

Notice that in Theorem 2.4.1, we assumed it to be known a priori that the signal lies in a particular \mathcal{V}_P , and then found its exact period using (2.13). This period can turn out to be any divisor of P . However, in practice, we do not usually know beforehand if a signal belongs to a particular \mathcal{V}_P . How can we use the Ramanujan subspaces in such situations? This is discussed next.

2.5 Estimating the period beyond a particular \mathcal{V}_P

Suppose $x(n)$ is a signal whose period is known to lie in the range $1 \leq P \leq P_{max}$. Let $L = \text{LCM}\{1, 2, 3, \dots, P_{max}\}$. In principle, if we have long enough datalength, we can

construct an $L \times L$ RPT matrix and use it to determine the period by Theorem 2.4.1. However, L could be too large in practice, and the available datalength much smaller. There are two ways to deal with this:

1. **Projections:** Note that, due to the orthogonality of Ramanujan subspaces, one way to solve Eq. 2.13 is to take projections onto the various Ramanujan subspaces. When the data-length is smaller than L , we can still try plotting the projection energies of the signal along the various Ramanujan subspaces. Fig. 2.3 shows an example. We found empirically in all our simulations that the LCM property could still be used to give the correct period estimates, unless the data-length is too small. The precise minimum datalength for this to work is however not known so far.
2. **Dictionaries:** An alternative to projections is to use sparsity based dictionary approaches. These will be discussed in Chapter 3.

2.6 Conclusion

We summarized the most important properties of Ramanujan Subspaces in this chapter. In [9], [18], a number of other interesting properties of these subspaces were derived. In the next chapter, we will show that the Ramanujan subspaces are in fact one example of an entire family of such subspaces called the Nested Periodic Subspaces. We will show that several novel period estimation techniques can be designed using the Ramanujan subspaces and other Nested Periodic Subspaces, which offer unique advantages over the existing methods.

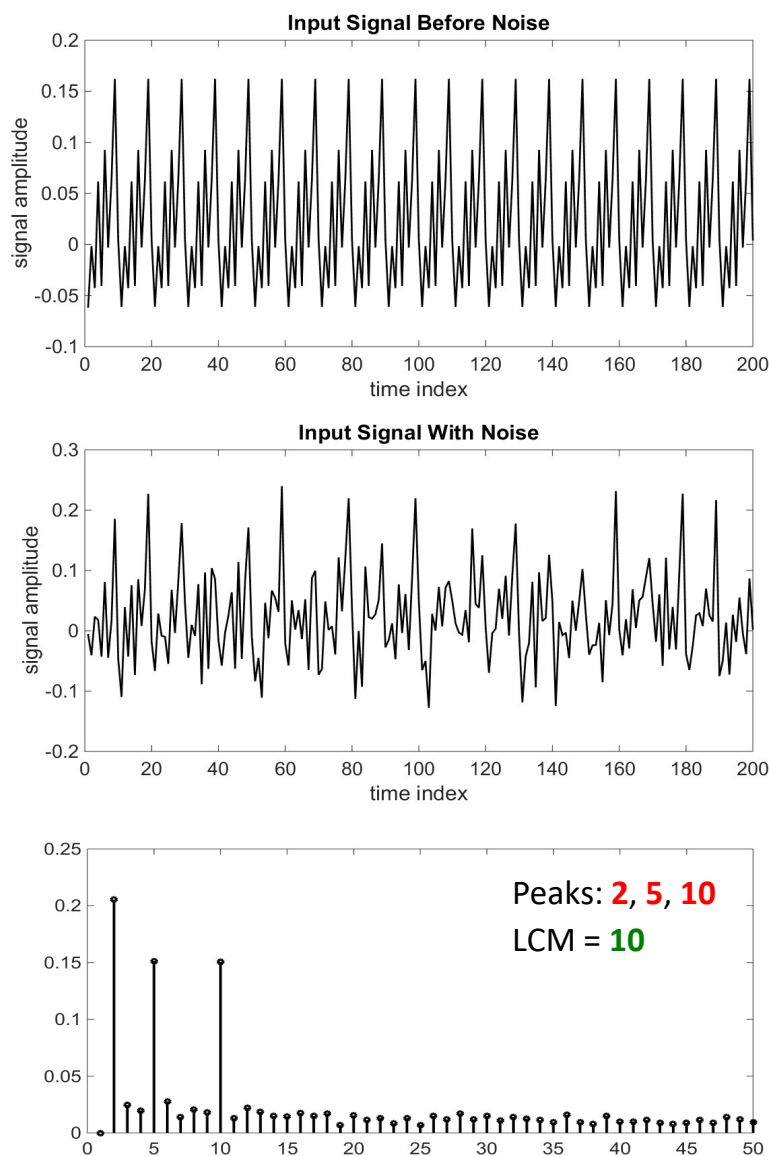


Figure 2.3: (Top) 100 samples of a noise-less period 10 signal. (Middle) Noise corrupted input (SNR = 5dB). (Bottom) The projection energies plotted vs. the period for the various Ramanujan subspaces. Notice that the LCM of the peak periods 2, 5 and 10 is indeed 10.

Chapter 3

NESTED PERIODIC MATRICES AND DICTIONARIES

In this chapter, we will introduce a general family of subspaces that include the Ramanujan subspaces as specific examples, sharing similar properties. These are called Nested Periodic Subspaces [23]. We will start by first generalizing the Ramanujan Periodicity Transform matrix from Sec. 2.4 to what are called the Nested Periodicity Matrices. It will be shown that these matrices can be used in a similar fashion as the RPT in Theorem 2.4.1 for period estimation. Important connections with well-known transforms such as the DFT and the Hadamard transforms will be presented. These matrices will then be used to develop dictionary based period estimation techniques.

3.1 Nested Periodicity Matrix: Definition

Let P be a positive integer. Let d_i , $1 \leq i \leq K$ denote the divisors of P in increasing order, so that $d_1 = 1$ and $d_K = P$. Let $\phi(\cdot)$ denote the Euler totient function. Now consider a matrix of the form:

$$\mathbf{A} = [\mathbf{C}_{d_1} \quad \mathbf{C}_{d_2} \quad \dots \quad \mathbf{C}_{d_K}] \quad (3.1)$$

with the following properties:

1. Each \mathbf{C}_{d_i} is a $P \times \phi(d_i)$ matrix so that the total number of columns in \mathbf{A} is $\sum_{d_i|P} \phi(d_i) = P$ (from [24]). Thus \mathbf{A} is a $P \times P$ matrix.
2. Each column of \mathbf{C}_{d_i} is a length P segment of a sequence with period d_i .
3. \mathbf{A} has full rank .

Since \mathbf{A} has full rank, its columns form a basis for \mathbb{C}^P , so that by periodically extending the columns of \mathbf{A} we can obtain a basis for \mathcal{V}_P (from Eq. 2.8). Such a matrix \mathbf{A} will be referred to as a **Nested Periodic Matrix (NPM)**.

Notice that the Ramanujan Periodicity Transform matrix in Chapter 2 is an example of an NPM. Later in this chapter, we will show several other examples of NPMs.

3.2 NPM Properties

The word ‘‘Nested’’ is used to indicate a very special property of such matrices: In the following lemma, we show that the set of signals obtained by periodically extending the columns of a nested periodic matrix not only form a basis for \mathcal{V}_P , but also contains subsets that are bases for every divisor q of P . Note that this would not be true for an arbitrary $P \times P$ full rank matrix.

Lemma 3.2.1. *For any integer $q|P$, there are precisely q columns of \mathbf{A} with period q or a proper divisor of q . Furthermore, any $P \times 1$ vector with period q can be spanned by these q columns.*

Proof. : Let $q_i, 1 \leq i \leq L$ be the divisors of q . Then \mathbf{A} has $\phi(q_i)$ columns \mathbf{C}_{q_i} with period q_i . Since $\sum_i \phi(q_i) = q$, the submatrix of \mathbf{A} given by

$$\mathbf{C} = [\mathbf{C}_{q_1} \ \mathbf{C}_{q_2} \ \dots \ \mathbf{C}_{q_L}] \quad (3.2)$$

has size $P \times q$. The matrix \mathbf{C} has full column rank because the original matrix \mathbf{A} has full rank. Furthermore it has the form

$$\mathbf{C} = \begin{bmatrix} \mathbf{D} \\ \mathbf{D} \\ \vdots \\ \mathbf{D} \end{bmatrix} \quad (3.3)$$

where \mathbf{D} is $q \times q$. Since \mathbf{C} has full column rank, it is clear that \mathbf{D} has to have full rank. This means that any $q \times 1$ vector can be spanned by the columns of \mathbf{D} . It then follows that any $P \times 1$ vector with q period can indeed be spanned by the columns of \mathbf{C} . We note here that \mathbf{D} itself is a $q \times q$ NPM. ▽ ▽ ▽

In general, when we add two periodic signals with periods P_1 and P_2 , the resulting signal is periodic too, but its period can be any divisor of $\text{LCM}(P_1, P_2)$. But the columns of an NPM have a special property, which forms the backbone of the period estimation techniques developed in later chapters.

Lemma 3.2.2. *Let \mathbf{A} be a $P \times P$ NPM. Consider the linear combination of a subset of L distinct columns.:*

$$\mathbf{v} = \sum_i^L \beta_i \mathbf{a}_i \quad (3.4)$$

Let n_i be the period of \mathbf{a}_i , and assume all $\beta_i \neq 0$ without loss of generality. Then the period of \mathbf{v} is precisely $\text{LCM}\{n_i\}$ (i.e., it cannot be a proper divisor of this lcm).

Proof. Let v be the period of \mathbf{v} . Then

$$v | \text{LCM}\{n_i\} \quad (3.5)$$

And since $n_i | P$, it follows that $v | P$ so that (by Lemma 3.2.1) there is a submatrix \mathbf{V} of \mathbf{A} such that $\mathbf{v} \in \text{span}(\mathbf{V})$. We will now argue that $\mathbf{a}_i \in \text{span}(\mathbf{V})$ for each i .

To see this, let us rewrite (3.4) as:

$$\beta_1 \mathbf{a}_1 = \mathbf{V}\alpha - \sum_{i \neq 1} \beta_i \mathbf{a}_{k_i} \quad (3.6)$$

where $\mathbf{v} = \mathbf{V}\alpha$. So the L.H.S. and the R.H.S. are now expressed in terms of columns of \mathbf{A} . Since \mathbf{A} is a full rank matrix by definition, (3.6) can only happen if \mathbf{a}_1 is a column of \mathbf{V} . So the period of \mathbf{a}_1 , viz. n_1 , has to be a divisor of v . Similarly, it can be shown that $n_i | v$ for each i , so that:

$$\text{LCM}\{n_i\} | v \quad (3.7)$$

Eqs. (3.5) and (3.7) imply $v = \text{LCM}\{n_i\}$, which proves the theorem. $\nabla \nabla \nabla$

Combining the above two Lemmas, we get:

Lemma 3.2.3. *Given a $P \times P$ NPM, any $P \times 1$ vector \mathbf{y} with period $q | P$ can be expressed as:*

$$\mathbf{y} = \mathbf{A}\mathbf{c} \quad (3.8)$$

where the LCM of the periods of those columns of \mathbf{A} that correspond to non-zero entries in \mathbf{c} , is exactly equal to the period of \mathbf{y} .

The above Lemma gives us a way of estimating the periods of signals in \mathcal{V}_P . That is, we can identify which divisor of P the actual period of the signal is. Notice that this cannot be done using arbitrary $P \times P$ matrices instead of an NPM. For example, in a different framework (discussed in Chapter 4), the authors of [13], [25] use the identity matrix to generate a basis for \mathcal{V}_P . To see why this can be problematic, let us consider an example. Suppose we have 6 samples of a period 3 signal. If we expand it in terms of the columns of the 6×6 identity matrix, in general each of the 6 basis vectors would require non-zero coefficients. So it will not be possible to estimate the period of that signal just by looking at its support on the identity matrix. But if we use any of the NPMs for \mathcal{V}_6 , then only the period 3 and period 1 columns of those matrices would be involved in spanning the signal. So we can readily infer that the period is 3.

Another interesting property of NPMs is as follows:

Lemma 3.2.4. *Let \mathbf{z} be a column of \mathbf{A} with period $d|P$. Then it cannot be expressed as a finite sum*

$$\mathbf{z} = \sum_i \mathbf{z}_i \quad (3.9)$$

where \mathbf{z}_i have periods $r_i|P$ with $r_i < d$.

Proof. Suppose that (3.9) were possible. Then, in view of Lemma 3.9, each \mathbf{z}_i is a linear combination of columns of \mathbf{A} that have period *less* than d . That is \mathbf{z} , which is a column of \mathbf{A} with period d , is a linear combination of other columns of \mathbf{A} . This contradicts the fact that \mathbf{A} has full rank. $\nabla \nabla \nabla$

This result can be extended to the following: Suppose \mathbf{z} is a period d column of a $P \times P$ NPM. When \mathbf{z} is periodically extended to a period d signal, then that signal cannot be written as a sum of signals, all of whose periods are strictly smaller than d . This will be studied in detail in Chapter 7.

3.3 Examples of NPMs

There is a lot of freedom in designing NPMs. Here we illustrate five different types of constructions:

1. Natural Basis Matrices
2. DFT Matrices
3. Ramanujan Matrices
4. Hadamard Matrices
5. Random Matrices

Natural Basis Matrices

Consider the identity matrix as a basis for \mathcal{V}_P . Obviously it does not have the nested periodic matrix structure. But we can construct a very simple matrix with the desired structure starting from identity matrices of sizes the divisors of P . For each divisor d of P , we only take the first $\phi(d)$ columns of the identity matrix and periodically extend them to length P . By collecting such columns for all the divisors of P , we obtain a matrix satisfying the conditions of a nested periodic matrix. For example, consider the following 8×8 example. The first column is periodic with

period 1, the second with period 2, the third and fourth with period 4 and the rest with period 8.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (3.10)$$

This construction is very sparse and involves only 0's and 1's. This is quite useful from a computational point of view, for example when solving (3.8). We observed experimentally that the inverses of these matrices have 1's, -1 's and 0's only and are quite sparse. No proof has been possible for this at this time. For example, the inverse of the above 8×8 matrix is shown below.

$$\mathbf{A}^{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \quad (3.11)$$

This too is useful, for example to find the solution of in the context of Lemma 3.2.3. These matrices seem to satisfy another interesting property - their determinant was always 1 or -1 depending on the parity of their size in our experiments. Again, no proof is available at this time.

DFT Matrices

The well known DFT matrices are also examples of NPMs. They are full rank since they are Vandermonde with distinct rows. Additionally, given a sequence W_P^{kn} where $W_P = e^{-j2\pi/P}$ and k is an integer such that $1 \leq k \leq P$, we can always rewrite it as $W_{d_i}^{k_i n}$ where d_i is a divisor of P such that $(d_i, k_i) = 1$ and $1 \leq k_i \leq d_i$. Regarded

as a sequence in n , the signal $W_P^{kn} = W_{d_i}^{k_i n}$ therefore has period $d_i|P$. Since there are $\phi(d_i)$ such values of k_i , it follows that there are $\phi(d_i)$ signals of the form W_P^{kn} with period in exactly equal to d_i . Thus, if we consider the family $x_k(n) = W_P^{kn}$ for all values of k in $1 \leq k \leq P$, we see that for each divisor d_i of P all sequences of the form $W_{d_i}^{k_i n}$ are generated where $(k_i, d_i) = 1$. For each d_i there are obviously $\phi(d_i)$ such sequences because there are $\phi(d_i)$ values of k_i with $(k_i, d_i) = 1$ (in $1 \leq k_i \leq d_i$). The total number of such sequences is

$$\sum_{d_i|P} \phi(d_i) = P \quad (3.12)$$

Summarizing, we have proved this:

Theorem 3.3.1. *The columns of the $P \times P$ DFT matrix can be partitioned into K classes, one for each divisor $d_i|P$. In class i there are $\phi(d_i)$ columns (which can be gathered into a $P \times \phi(d_i)$ matrix \mathbf{C}_{d_i}) and they have the form $W_{d_i}^{k_i n}$ where $(k_i, d_i) = 1$, and all these have period exactly d_i . None of the columns of this DFT matrix can have a period other than a divisor $d_i|P$.*

For example, for the (column permuted) 8×8 DFT matrix shown below, the first column is periodic with period 1, the second with period 2, the third and fourth with period 4 and the rest with period 8.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -j & j & e^{-j\frac{\pi}{4}} & e^{-j\frac{3\pi}{4}} & e^{j\frac{3\pi}{4}} & e^{j\frac{\pi}{4}} \\ 1 & 1 & -1 & -1 & -j & j & -j & j \\ 1 & -1 & j & -j & e^{-j\frac{3\pi}{4}} & e^{-j\frac{\pi}{4}} & e^{j\frac{\pi}{4}} & e^{j\frac{3\pi}{4}} \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -j & j & e^{j\frac{3\pi}{4}} & e^{j\frac{\pi}{4}} & e^{-j\frac{\pi}{4}} & e^{-j\frac{3\pi}{4}} \\ 1 & 1 & -1 & -1 & j & -j & j & -j \\ 1 & -1 & j & -j & e^{j\frac{\pi}{4}} & e^{j\frac{3\pi}{4}} & e^{-j\frac{3\pi}{4}} & e^{-j\frac{\pi}{4}} \end{bmatrix} \quad (3.13)$$

Note that, using the DFT matrix for identifying hidden periods as explained in Lemma 3.2.3 is not the same thing as analyzing the peaks in the DFT of the signal $x(n)$. The latter identifies the exponential W_P^{kn} with the ‘period’ P/k , which need not even be an integer. We instead classify it as a signal with period $P/(P, k)$.

Ramanujan Periodicity Transform

The Ramanujan Periodicity Transform (RPT) matrix was described in Sec. 2.4, and it clearly satisfies the definition of an NPM in Sec. 3.1. We will now show a very interesting property of the RPT matrix here:

Theorem 3.3.2. *Let \mathbf{A} be a $P \times P$ NPM as defined in (3.1) where $d_i | P$, and let the column spaces \mathcal{T}_{d_i} of \mathbf{C}_{d_i} be pairwise orthogonal, that is, $\mathbf{C}_{d_i}^\dagger \mathbf{C}_{d_j} = 0$ for $d_i \neq d_j$. Then \mathcal{T}_{d_i} has to be the d_i^{th} Ramanujan subspace \mathcal{S}_{d_i} for each d_i .*

Proof. We use induction on P to prove the theorem. For $P = 1$, we know $\mathcal{T}_1 = \mathcal{S}_1$ (one dimensional space spanned by $[1 \ 1 \ 1 \ 1 \ \dots]^T$). This is the basis for induction. Now consider the $P \times P$ matrix for an arbitrary P and let the divisors of P be d_1, d_2, \dots, d_K . For induction, assume that the first $K-1$ spaces $\mathcal{T}_{d_1}, \mathcal{T}_{d_2}, \dots, \mathcal{T}_{d_{K-1}}$ are the Ramanujan spaces $\mathcal{S}_{d_1}, \mathcal{S}_{d_2}, \dots, \mathcal{S}_{d_{K-1}}$. We will prove that \mathcal{T}_{d_K} is the Ramanujan space \mathcal{S}_{d_K} . Since the spaces $\mathcal{T}_{d_1}, \mathcal{T}_{d_2}, \dots, \mathcal{T}_{d_{K-1}}$ are orthogonal, their outer sum, call it \mathcal{S} , has dimension given by:

$$n = \phi(d_1) + \phi(d_2) + \dots + \phi(d_{K-1}) \quad (3.14)$$

The space $\mathcal{T}_{d_K} \subset \mathbf{C}^P$ has dimension $\phi(P)$, which is clearly equal to $P - n$. If this space is orthogonal to \mathcal{S} , then it has to be unique. 1 But since the Ramanujan space \mathcal{S}_{d_K} is one such orthogonal space, it follows that $\mathcal{T}_{d_K} = \mathcal{S}_{d_K}$. $\nabla \nabla \nabla$

This theorem shows that if \mathbf{A} is an NPM in which the subspaces with different periods are mutually orthogonal, then these subspaces must necessarily be the Ramanujan subspaces. A corollary is that the columns of the DFT matrix in the previous subsection also span the Ramanujan subspaces. We will next show another well known transform that also spans the Ramanujan subspaces.

Hadamard Transform

A Hadamard matrix is defined to be a square orthogonal matrix whose every entry is either 1 or -1 . It has been postulated that Hadamard matrices of size $P \times P$ exist for some special P , for example multiples of 4 (Hadamard Conjecture). But when P is a power of two, Sylvester showed that such matrices do exist and proposed a simple recursive way to construct them as shown below, where \mathbf{H}_n is of size $2^n \times 2^n$:

$$\mathbf{H}_0 = 1, \quad \mathbf{H}_n = \begin{bmatrix} \mathbf{H}_{n-1} & \mathbf{H}_{n-1} \\ -\mathbf{H}_{n-1} & \mathbf{H}_{n-1} \end{bmatrix} \quad \forall n \in \mathbb{N} \quad (3.15)$$

For the Sylvester construction, it can be shown that the matrices are in fact NPMs:

Theorem 3.3.3. For $n = 0, 1, 2, 3, \dots$, the $2^n \times 2^n$ Hadamard matrix constructed using the Sylvester method is an NPM.

Proof. We shall prove this by induction. For the $n = 0$, the statement of the theorem is obviously true. We assume it to be true for $n = 0, 1, 2, \dots, m - 1$. That is, let the columns of \mathbf{H}_{m-1} be partitionable into $m - 1$ sets $\{\mathbf{C}_k\}_{k=0}^{m-1}$ as in (3.1), where the columns in \mathbf{C}_k are 2^{k-1} in number and are periodic with period 2^k . The matrix \mathbf{H}_m is then constructed as:

$$\begin{bmatrix} \mathbf{H}_{m-1} & \mathbf{H}_{m-1} \\ -\mathbf{H}_{m-1} & \mathbf{H}_{m-1} \end{bmatrix} \quad (3.16)$$

Note that the first 2^{m-1} columns are just repeated versions of the columns of \mathbf{H}_{m-1} :

$$\begin{bmatrix} \mathbf{H}_{m-1} \\ \mathbf{H}_{m-1} \end{bmatrix} \quad (3.17)$$

So these columns have the same periods as the corresponding columns in \mathbf{H}_{m-1} . So they can be partitioned into the sets as $\{\mathbf{C}_k\}_{k=0}^{m-1}$ in (3.1).

It remains to be shown that the last $\phi(2^m) = 2^{m-1}$ columns of \mathbf{H}_m are periodic with period 2^m :

$$\begin{bmatrix} \mathbf{H}_{m-1} \\ -\mathbf{H}_{m-1} \end{bmatrix} \quad (3.18)$$

This is easy to see, since if any of these columns had period less than 2^m , then this period must be a proper divisor of 2^m . This is because, in this section, when we say that a column of a $P \times P$ matrix has period d , we mean that the signal obtained by periodically repeating this $P \times 1$ vector is periodic with period d . For this to be true, d must divide P . So if any column in (3.18) has a proper divisor of 2^m as its period, then the top 2^{m-1} entries of that column must be the same as its bottom 2^{m-1} entries. This means that the corresponding columns of \mathbf{H}_{m-1} , denoted by \mathbf{h}_{m-1} say satisfy:

$$\mathbf{h}_{m-1} = -\mathbf{h}_{m-1} \quad (3.19)$$

so that $\mathbf{h}_{m-1} = 0$. This is a contradiction, since Hadamard matrices are by definition orthogonal and hence must satisfy the full rank property. $\nabla \nabla \nabla$

An example of an 8×8 Hadamard matrix is shown below. Again, the first column is periodic with period one, the second with period 2, the third and fourth with period 4 and the rest with period 8.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} \quad (3.20)$$

The Sylvester construction is not the only class of Hadamard matrices. Two Hadamard matrices of the same size are said to be equivalent and belonging to the same class if one of them can be obtained from the other by negating or interchanging rows or columns. For $N = 1, 2, 4, 8$, and 12 there exists only one class of Hadamard matrices, but for $N = 16$ for example, there exist 5 different classes of Hadamard matrices. We found by inspection that for $N = 16$, one of the five classes does not satisfy nested periodic matrix structure. For $N = 20$, there are three classes of Hadamard matrices and none of them satisfy the nested periodic matrix structure. (Note that a Sylvester construction is not possible for $N = 20$ since 20 is not a power of 2 .)

Note that, if a Hadamard matrix is satisfying the nested periodic matrix structure, then by Theorem 3.3.2, the various periodic subspaces spanned by \mathbf{C}_{d_i} of that Hadamard matrix are same as the corresponding Ramanujan Subspaces.

Hadamard matrices have been used in the past as Hadamard Transforms in various applications, for example in DSP such as in JPEG XR and MPEG-4 AVC, data encryption, quantum computing etc. Their main advantage is the ease of computation compared to the DFT.

Random Periodic Matrices

We observed that it is very easy to construct nested periodic matrices by choosing the entries in a random way. For each divisor d of P , we generate a $d \times 1$ vector with random entries, and repeat it to length P to get the first column of \mathbf{C}_d . The other $\phi(d) - 1$ columns can be obtained by cyclically shifting this first column, or by randomly generating other period d signals. The $P \times P$ matrix so obtained was observed to be full rank with a high probability. Once again, an 8×8 example is shown below:

$$\mathbf{A} = \begin{bmatrix} 6 & 7 & 0 & 2 & 4 & 8 & 4 & 5 \\ 6 & 4 & 2 & 0 & 5 & 4 & 8 & 4 \\ 6 & 7 & 1 & 2 & 4 & 5 & 4 & 8 \\ 6 & 4 & 2 & 1 & 8 & 4 & 5 & 4 \\ 6 & 7 & 0 & 2 & 4 & 8 & 4 & 5 \\ 6 & 4 & 2 & 0 & 5 & 4 & 8 & 4 \\ 6 & 7 & 1 & 2 & 4 & 5 & 4 & 8 \\ 6 & 4 & 2 & 1 & 8 & 4 & 5 & 4 \end{bmatrix} \quad (3.21)$$

Even though this might not be the best way to generate nested periodic matrices, it never-the-less illustrates that such matrices are not ‘very rare’.

Simulations

We illustrate two different examples of period identification using nested periodic matrices.

In our first example, the input data was 30 consecutive samples of a period 10 signal. We solved the equation $\mathbf{y} = \mathbf{A}\mathbf{c}$ in Lemma 3.2.3, taking \mathbf{y} to be the 30×1 data vector and by choosing \mathbf{A} to be the 30×30 Natural Basis and Ramanujan Periodicity Transform matrices in Fig. 3.1(a) and (b) respectively. For each divisor of 30, we have plotted the sum of squares of those entries of \mathbf{c} that correspond to columns of \mathbf{A} that are periodic with period being that divisor. For example, if $G(d)$ represents the function that has been plotted vs the period d , where $d|P$, then $G(\cdot)$ is defined as follows:

$$G(d) = \sum_{k=K+1}^{K+\phi(d)} |c(k)|^2, \quad K = \sum_{\substack{p|P \\ p < d}} \phi(p) \quad (3.22)$$

We will call such plots as *strength vs period plots*. As expected from Lemma 3.2.3, we see that the strength vs period plot is zero except at divisors of 10 in both the plots of Fig. 3.1. This indicates that the input signal had period 10.

The second experiment in Fig. 3.2(a) and (b) show the strength vs period plots for a period 70 signal using the Natural Basis and the Ramanujan Periodicity Transform matrices respectively. In this case, the input signal was generated as a sum of two periodic signals with periods 7 and 10. The matrix \mathbf{A} was chosen to be of

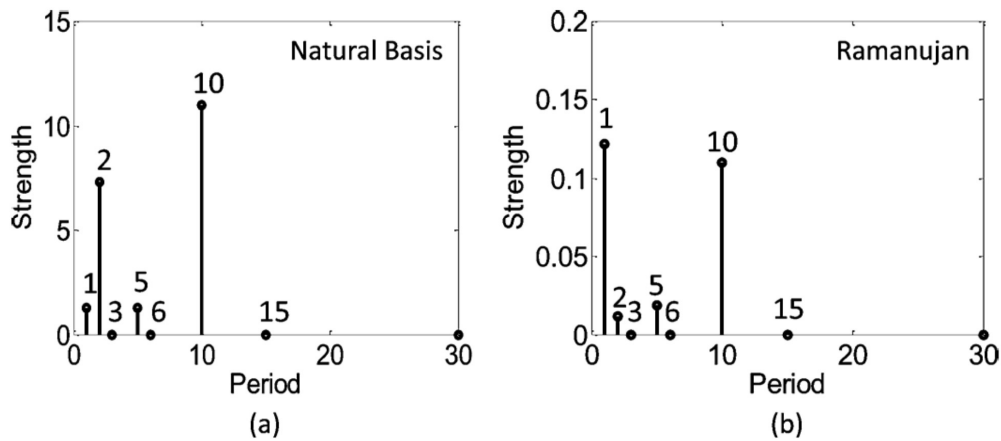


Figure 3.1: Part (a) and (b) – Strength vs period plots for a period 10 signal using 30×30 Natural Basis and Ramanujan Periodicity Transform matrices respectively.

size 70×70 . According to Lemma 3.2.1, the period 7 component can be spanned by columns of \mathbf{A} that have as periods divisors of 7. Similarly for the period 10 component. So we would expect the strength vs period plots for the sum of these two signals to be non-zero only at periods that are divisors of either 7 or 10. This is indeed the case in Fig. 4.

In these examples, there was no noise. But when the input signal was corrupted by noise, we observed that Ramanujan Periodicity Transform and DFT matrices perform the best. We will consider noisy data in the simulations of the next section.

3.4 Nested Periodic Dictionaries

Just like the RPT matrix in Chapter 2, NPMs can only be used to estimate the period within a particular \mathcal{V}_P . That is, in Lemma 3.2.3, we assumed it to be known a priori that the signal lies in a particular \mathcal{V}_P , and then found its exact period using the LCM property. This period can be any divisor of P . However, in practice, we do not usually know beforehand if a signal belongs to a particular \mathcal{V}_P . How can we use the NPMs in such situations?

We approach this problem by modeling the periodic signal in terms of an over-complete dictionary. We formulate the period identification and hidden period

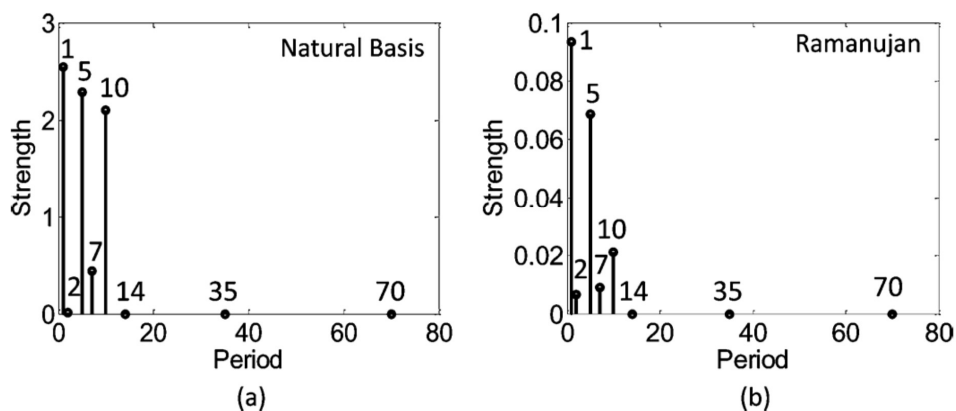


Figure 3.2: Part (a) and (b) - Strength vs period plots for a period 70 signal using 70×70 Natural Basis and Ramanujan Periodicity Transform matrices respectively. The period 70 signal was generated as a sum of period 7 and period 10 signals.

estimation problem as a ‘data-fitting’ problem by re-interpreting the process of finding hidden periods as trying to fit the given data with signals having periods as small possible (it will be shown in Chapter 7 that such a decomposition is guaranteed to yield a unique set of hidden periods). This allows us to formulate a convex program with a closed form expression for its solution. An alternate view point based on sparse representations of periodic signals using the same dictionaries will be discussed in Sec. 3.4.

We first show how a periodic signal can be modeled as a linear combination of the columns of a dictionary.

Dictionary Construction

Suppose that the input data is $x(n), n = 0, 1, 2, \dots, N - 1$. Consider a particular family of NPMs - for example, the Ramanujan construction. For each d going from 1 to P_{max} , where P_{max} is the largest expected period, construct the nested periodic matrix and take only its $\phi(d)$ columns that have period d . Extend these columns periodically to length N , truncating the last period if necessary.

We form a dictionary \mathbf{A} by collecting such columns for each d going from 1 to P_{max} . We will call such a dictionary as a **Nested Periodic Dictionary**. An example is shown in Fig. 3.3 using the Ramanujan construction. The first column corresponds

1	1	2	-1	2	0	4	-1	-1	-1
1	-1	-1	2	0	2	-1	4	-1	-1
1	1	-1	-1	-2	0	-1	-1	4	-1
1	-1	2	2	0	-2	-1	-1	-1	4
1	1	-1	-1	2	0	-1	-1	-1	-1
Period 1	Period 2	Period 3		Period 4		Period 5			
$\phi(1) = 1$	$\phi(2) = 1$	$\phi(3) = 2$		$\phi(4) = 2$		$\phi(5) = 4$			

Figure 3.3: The Ramanujan Nested Periodic Dictionary

to period 1, the second to period 2, the third and fourth to period 3, the fifth and sixth to period 4 and the remaining to period 5.

Note how some of the columns do not end in complete periods. Also, typically P_{max} is of the order of the length of the data, while the sum of Euler totient function from 1 to P_{max} is $O(3P_{max}^2/\pi^2)$ [24]. So \mathbf{A} is generally a fat matrix having many more columns than rows.

If the given signal $x(n)$ is periodic with period less than P_{max} , then it has to be a linear combination of the columns of the dictionary. This is because, if it was periodic with period P , then the columns with periods being divisors of P must be able to span it (follows from Lemma 3.2.1). So the following system of linear equations must have a solution \mathbf{y} , where $\mathbf{x} = [x(0), x(1), \dots, x(N-1)]^T$:

$$\mathbf{x} = \mathbf{A}\mathbf{y} \tag{3.23}$$

But since the dictionary is fat, it is likely that $\mathbf{x} = \mathbf{A}\mathbf{y}$ has multiple solutions. The solution that we are interested in is the one that involves subspaces corresponding to period P and its divisors. We discuss next a way to find this solution.

Formulating a Convex Program

As mentioned earlier, one way to look at the problem of finding hidden periods is that we are trying to fit the given signal with signals having periods as small as possible. If we could relate this problem to those dictionary problems that have closed form solutions, for example finding the least norm solution to an over-determined linear system, then we can compute the solution in a fast way. Of course, there is no reason why we might expect the least norm solution of (3.23) to yield us a decomposition of the signal into its hidden periods. But consider the following optimization problem:

$$\min \|\mathbf{D}\mathbf{y}\|_2 \quad s.t. \quad \mathbf{x} = \mathbf{A}\mathbf{y} \quad (3.24)$$

where \mathbf{D} is a diagonal matrix whose i^{th} diagonal entry is $f(P_i)$, where P_i is the period of the i^{th} column of \mathbf{A} and $f(\cdot)$ is some increasing function. By introducing \mathbf{D} in (24), the columns in \mathbf{A} that have larger periods contribute more towards the objective function than those with smaller periods for similar entries in \mathbf{y} . So in a way, the columns with larger periods are being penalized more and the algorithm will try to use columns of \mathbf{A} with as small periods as possible to fit $x(n)$.

An example using this approach is shown in Fig. 3.4. Part (a) shows two complete periods and a third incomplete period of a period 70 signal contaminated with noise (ENR = 5.5dB). The period 70 signal was generated as a sum of period 7 and period 10 signals. Parts (b), (c), (d) and (e) show the results of solving (3.24) using dictionaries constructed with Ramanujan, DFT, random and natural basis designs respectively. (The entries of the random dictionary were chosen to be integers from 1 to 1000 uniformly). The penalty function was chosen to be $f(P) = P^2$ and P_{max} was chosen to be 90, but we only show the periods 1 to 40 in the plots for clarity. For each period, the plots show the sum of squares of those components of the optimal solution of (3.24) that corresponding to columns of the dictionary with that particular period. These are similar to the strength vs period plots of Sec. 3.3. We deliberately do not show the period 1 component in all such plots, since it is just a DC signal. For example, if $G(P)$ represents the function that has been plotted vs the period P , then $G(\cdot)$ is defined as follows:

$$G(P) = \sum_{k=K+1}^{K+\phi(P)} |c(k)|^2, \quad K = \sum_{q=1}^{P-1} \phi(q) \quad (3.25)$$

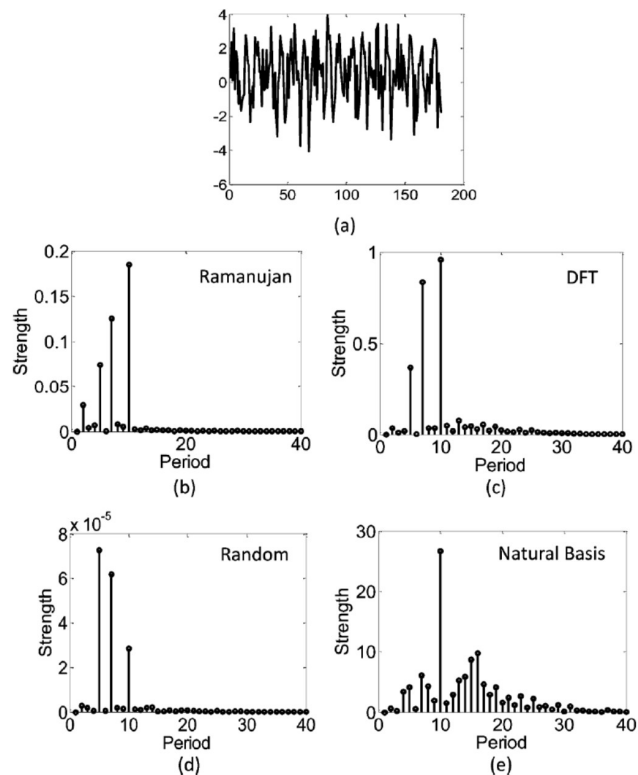


Figure 3.4: (a) A noisy period 70 signal that was generated as the sum of a period 7 and a period 10 signal (SNR = 5.5dB). Parts (b), (c), (d) and (e) – The strength vs period plots for the solutions of the convex program (3.24) using Ramanujan, DFT, Random and Natural Basis dictionaries respectively.

In parts (b), (c) and (d), we see peaks at periods 2, 5, 10 and 7. From Lemma 3.2.3, any linear combination of the columns of with periods 2, 5 and 10 results in a signal that is periodic with period 10. Similarly, there is also a period 7 component in the signal. So using Lemma 3.2.3 again, we can conclude that the given signal was periodic with period 70, but can be decomposed into a period 10 and a period 7 signal.

We would like to make a few remarks at this point:

1. We need to be careful while interpreting the results of Section IV in the context of dictionaries. For instance, when we used Lemma 3.2.3 in the previous paragraph to conclude that the sum of period 7 and period 10 columns of the dictionary \mathbf{A} is a period 70 signal, we considered a 70×70 sub-matrix composed of the columns of \mathbf{A} that have divisors of 70 as their period. We

use Lemma 3.2.3 on this sub-matrix to conclude that the resulting signal was periodic with period 70.

2. Note that the Natural Basis dictionary does not perform well here. The Natural Basis dictionary was in general found to be very sensitive to noise. In this example, when we reduce the noise level further, even the Natural Basis dictionary works well.
3. We would like to remind the reader that finding multiple hidden periods within a signal includes the case when there is only one periodicity in the signal. For example, the given input signal could have just been a period 10 signal, in which case only 2, 5 and 10 would show up in plots similar to Fig. 5.

The advantage of such a method as (3.24) is that we have a closed form expression for the optimal solution \mathbf{y}_\star :

$$\mathbf{y}_\star = \mathbf{D}^{-2} \mathbf{A}^T (\mathbf{A} \mathbf{D}^{-2} \mathbf{A}^T)^{-1} \mathbf{x} \quad (3.26)$$

This is obtained as follows. Under the transformation $\mathbf{z} = \mathbf{D}\mathbf{y}$, the convex program (3.24) changes to:

$$\min \|\mathbf{z}\|_2 \quad s.t. \quad \mathbf{x} = \mathbf{B}\mathbf{z} \quad (3.27)$$

where $\mathbf{B} = \mathbf{A}\mathbf{D}^{-1}$. This has a closed form solution given by:

$$\mathbf{z}_\star = \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{x} \quad (3.28)$$

Substituting $\mathbf{y} = \mathbf{D}^{-1}\mathbf{z}$ and $\mathbf{B} = \mathbf{A}\mathbf{D}^{-1}$ gives us the solution in (3.26). Note that the above expressions are for real valued dictionaries and signals. For the complex case, we can derive similar expressions.

The equivalence of the convex programs (3.24) and (3.27) indicates that using the penalty matrix is similar to rescaling the columns of \mathbf{A} . But does that mean that instead of solving (3.24) with a penalty matrix, we could have just normalized the columns of \mathbf{A} to obtain an $\hat{\mathbf{A}}$ and solved the following problem instead?

$$\min \|\mathbf{y}\|_2 \quad s.t. \quad \mathbf{x} = \hat{\mathbf{A}}\mathbf{y} \quad (3.29)$$

This is similar to taking the i^{th} entry of \mathbf{D} matrix as the norm of the i^{th} column \mathbf{A} of in the convex program (3.24), but not rescaling the solution \mathbf{z}_\star of the corresponding

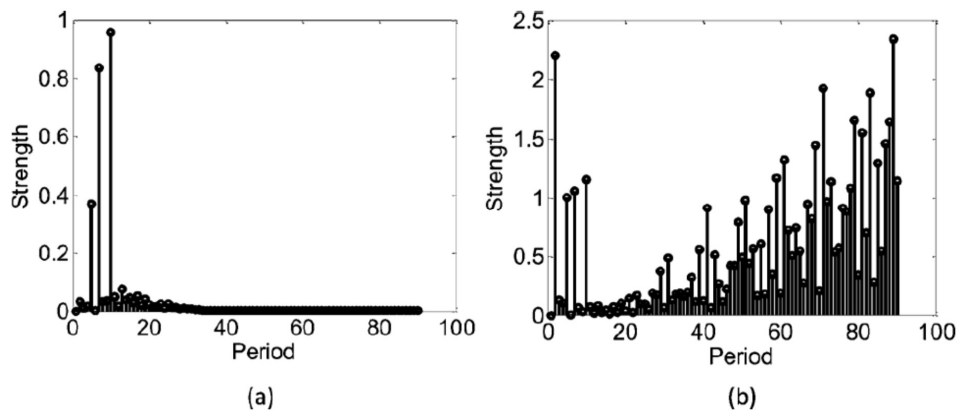


Figure 3.5: Strength vs Period plot for the signal in Fig. 3.4(a). Part (a) – using the solution of the convex program (3.24). Part (b) – using the solution for the convex program (3.29).

equivalent problem in (3.27). Fig. 6(a) and (b) show the strength vs period plots for the solution of (24) and (29) respectively for the signal in Fig. 5(a). We used the DFT subspaces for constructing \mathcal{D} . Notice the spurious peaks at higher periods in part (b). These are due to the absence of rescaling as described before.

Also, we observed that not all designs of periodic subspaces work equally well for this approach in the presence of noise. The Ramanujan and DFT dictionaries were generally much better than the random and natural basis dictionaries at low SNRs. As far as the choice of the penalty function is concerned, we observed that in all our computer generated examples of periodic signals corrupted by noise, many increasing functions work but $f(P) = P^2$ seemed to give good results. Also there is another penalty function which worked very well, especially for real world data, namely $f(P) = \phi(P)$. We show examples in Sec. 3.5. Although $\phi(P)$ is not a monotonically increasing function of P , it is related to the dimension of the subspace spanned by the P -periodic columns in the dictionary.

Shifting the Input Signal

How do the strength vs period plots change if we had a shifted version of the input? Fig. 3.6(a) shows the strength vs period plot using (3.24) and the Ramanujan dictionary for a length 250 signal that was generated as a sum of period 8 and period

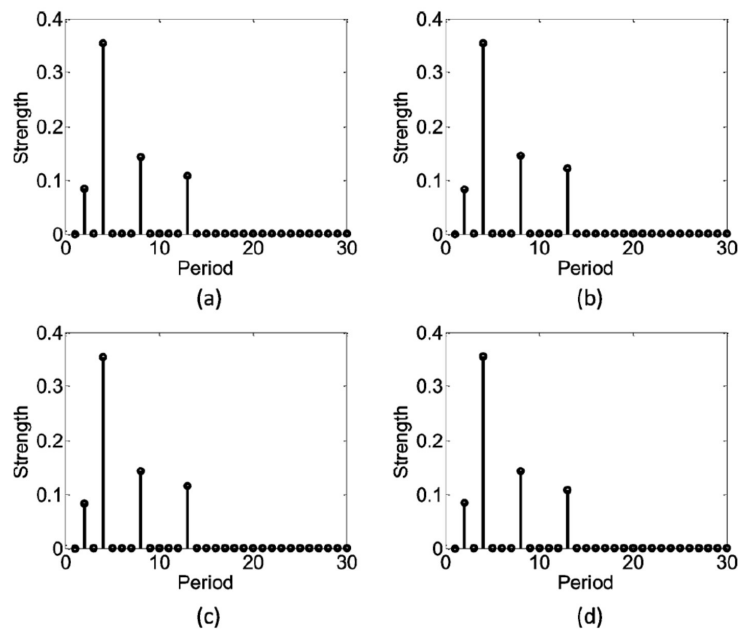


Figure 3.6: Part (a) – Strength vs period plots for a period 104 signal generated as a sum of period 8 and period 13 signals using the Ramanujan dictionary. Parts (b), (c) and (d) – Strength vs period plots for shifted versions of the signal, where the shifts are 25, 50 and 75 samples respectively (see text for details).

13 signals. In parts (b), (c) and (d), we have shown the strength vs period plots for $x(n-25)$, $x(n-50)$ and $x(n-75)$. The data length was fixed at 250 for each of these cases too (that is, n goes from 0 to 249). We did not add noise in this experiment so that we could study the effects of shifting the signal in an exclusive manner. As is evident from the plots, the strength vs period plots do not seem to change with such shifts.

This apparent shift invariance can be related to one of the properties of Ramanujan subspaces. Notice that, if the data length is sufficiently large, the Ramanujan subspaces \mathcal{S}_1 to $\mathcal{S}_{P_{max}}$ are approximately mutually orthogonal. So finding the solution to (3.23) becomes the same thing as decomposing the signal into its projections onto

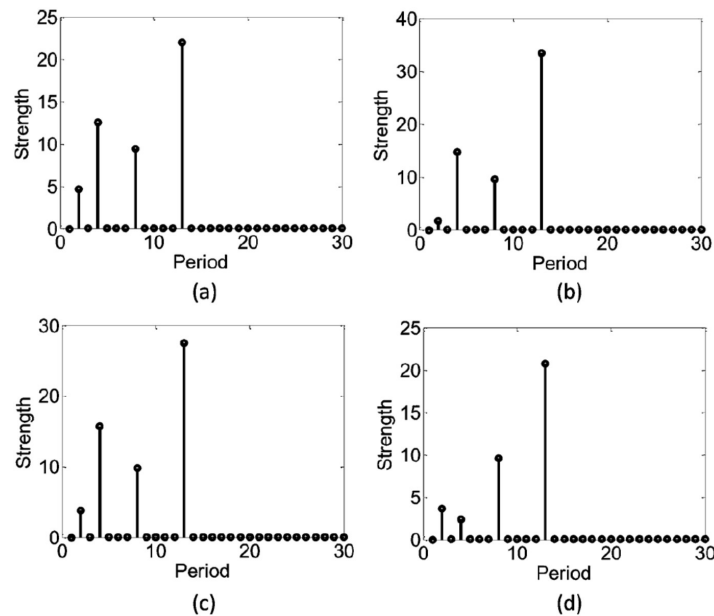


Figure 3.7: Part (a) – Strength vs period plots for a period 104 signal generated as a sum of period 8 and period 13 signals using the Natural Basis dictionary. Parts (b), (c) and (d) – Strength vs period plots for shifted versions of the signal, where the shifts are 25, 50 and 75 samples respectively (see text for details).

the orthogonal Ramanujan subspaces. By a slight extension of Theorem 6 in [9], one can show that if a signal has a particular projection energy on some Ramanujan subspace, then any of its shifted versions also have the same projection energy on that subspace. So if the data length is sufficiently large, it might be reasonable to expect the strength vs period plots to not change too much with such shifts. It must be noted though that the subspaces spanned by arbitrary Nested Periodic Bases in general do not have this shift invariance property. The effect of shifting the signal on the strength vs period plots of the natural basis dictionary is shown in Fig. 3.7.

Minimizing l_1 Norm Instead of l_2 Norm

The only reason we suggested the l_2 norm in (3.24) was to have a closed form expression for the optimal solution. We could have used any other norm instead of l_2 and still applied the penalty matrix idea. Vaidyanathan and Pal in [26] proposed to look at the problem of finding hidden periods as a sparse vector recovery problem instead of the data-fitting approach. That is, in place of (3.24), they formulate the

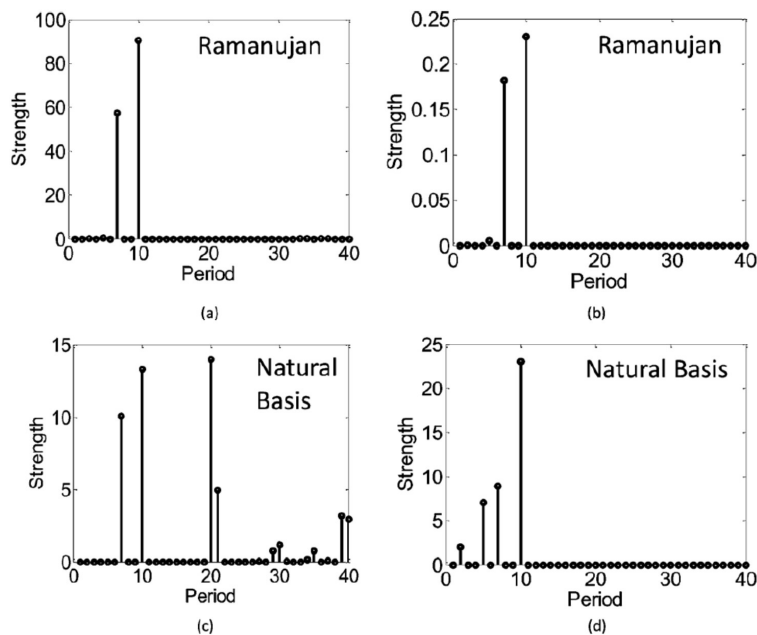


Figure 3.8: Periodic decomposition by l_1 methods. Parts (a) and (c) show the strength vs period plots for the solutions of (3.31) for a noiseless period 70 signal that was generated as a sum of period 7 and period 10 signals, using Ramanujan and natural basis dictionaries respectively. Parts (b) and (d) show the strength vs period plots for the same dictionaries using the solutions to the convex program (3.32).

following:

$$\min \|\mathbf{y}\|_0 \quad s.t. \quad \mathbf{x} = \hat{\mathbf{A}}\mathbf{y} \quad (3.30)$$

where $\hat{\mathbf{A}}$ is a dictionary constructed using the DFT design as in the previous sections, but with the columns normalized to unit norm. They call this the Farey dictionary. As is well known in the sparse vector recovery field, this can be relaxed to the following linear program under certain conditions:

$$\min \|\mathbf{y}\|_1 \quad s.t. \quad \mathbf{x} = \hat{\mathbf{A}}\mathbf{y} \quad (3.31)$$

Unfortunately, this does not have a closed form expression for the optimal solution. Fig. 3.8 shows the strength vs period plots for a noise-free period 70 signal that was generated as a sum of a period 7 and a period 10 signals, using the l_1 method.

Dictionary type	l_1 CPU time (sec)	l_2 CPU time (sec)
Natural Basis	0.90	1.74×10^{-4}
Ramanujan	5.84	2.03×10^{-4}
Random	5.96	1.86×10^{-4}
DFT (Farey)	26.23	5.59×10^{-4}

Table 3.1: Comparison of CPU times for solving (3.32) and (3.24)

The plots on the left, i.e, (a) and (c) are the strength vs period plots for solutions of (3.31) using Ramanujan and Natural Basis dictionaries respectively, each of whose columns were normalized. While there are clear peaks at 10 and 7 for the Ramanujan dictionary, there are many spurious peaks for the Natural Basis dictionary. Instead of normalizing the columns, if we introduce a matrix and solve the following convex program:

$$\min \|\mathbf{D}\mathbf{y}\|_1 \quad s.t. \quad \mathbf{x} = \mathbf{A}\mathbf{y} \quad (3.32)$$

where \mathbf{D} is a penalty matrix as described in the previous section with $f(P) = P^2$, then we obtain the plots on the right side of Fig. 3.8. Now we have clear peaks at period 2, 5, 7 and 10 for all the designs. This suggests that maybe the penalty matrix based data-fitting view point is more appropriate than a sparsity based model. Making the natural dictionary work is of practical significance too as illustrated by the CPU times for solving (3.32) and (3.24) in Table 3.1 (using Matlab 2012b on an Intel Core i7 processor with 2.2 GHz CPU speed): Summarizing, by using the penalty matrix idea, we could make the Natural Basis and the other dictionaries work for the norm based sparsity technique of [26]. These other dictionaries offer at least an order of magnitude faster computations by avoiding the second order cone programming needed for the complex-valued DFT based Farey dictionary.

3.5 Real World Examples

All the examples that we showed till now were signals that were generated as periodically extended random sequences generated by a computer. Here we apply our methods to three examples of more realistic periodic signals.

Periodic Signals With an Envelope

Here we consider periodic signals multiplied by other relatively slowly varying signals. Fig. 3.9 shows two such examples. The original periodic signal, shown in part (a) was generated as a sum of randomly generated period 7 and period 10 signals. We multiply it by a triangular window and by a sinusoidal signal to obtain the signals

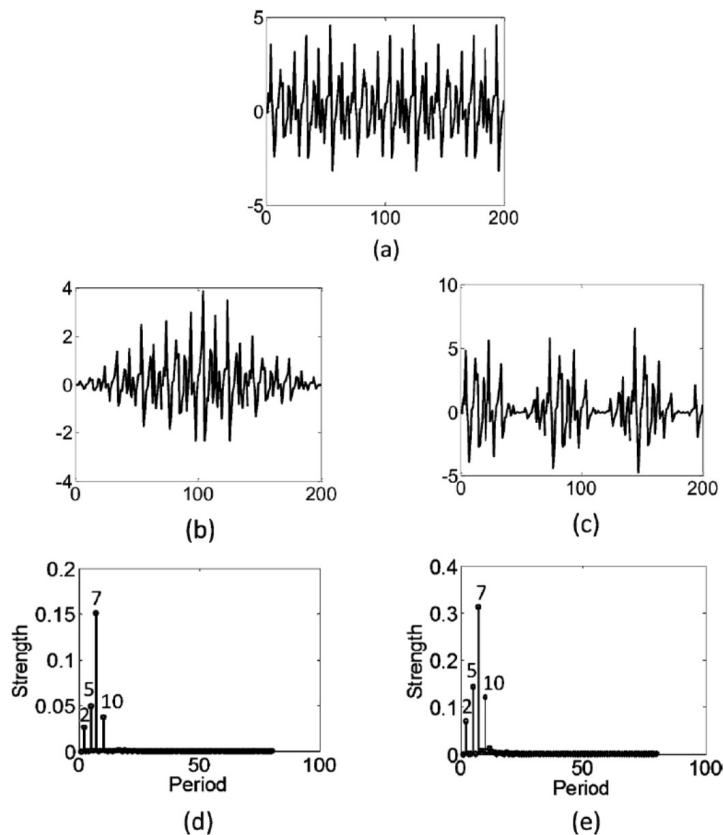


Figure 3.9: Part (a) – A period 70 signal generated as a sum of period 7 and period 10 signals. Parts (b) and (c) – Triangular and sinusoidal functions were multiplied respectively to the signal in part (a). Parts (d) and (e) show the strength vs period plots for the signals in parts (b) and (c) respectively using the convex program (3.24) and the Ramanujan dictionary.

shown in parts (b) and (c) respectively. Such signals might arise for-example in real world applications where the input signal has non-zero rise and fall times, or due to slow modulation. Parts (d) and (e) show the strength vs period plots obtained by solving (3.24) using the Ramanujan dictionary and as the penalty function. We can clearly see peaks at 10, 7 and their divisors, indicating the underlying 10 and 7 periodicities. In our simulations, we observed that as long as the multiplying functions vary slowly as compared to the underlying periodic signal, we were able to recover the hidden periodicities.

Protein Repeats

Proteins are important bio-molecules that are sequences of amino acids. The amino acids that make up proteins belong to a set of 20 as specified by the genetic code. So a protein can be viewed as a discrete-time signal with values chosen from an alphabet of size 20. If we assign numerical values to these amino acids, we can map proteins to discrete ‘time’ signals and process them.

In a significant fraction of proteins (about 14%), the amino acid sequences exhibit repeating periodic patterns. Such patterns are of interest to biologists since they affect the binding and structural properties of the proteins. So detecting and characterizing such repeats has been of interest in various works in the past (see Sec. 5.4).

As an example, we consider a protein called Antifreeze from the beetle *Tenebrio Molitor* (PDB 1EZG). It consists of a sequence of 84 amino acid residues with an underlying periodicity of 12. We apply two different mappings proposed in [27] to convert the amino acid sequence into a sequence of numbers. The first of these is related to the secondary structure while the second is related to the molecular size and volume of the amino acids. Fig. 3.10(a) and (b) show the corresponding discrete time signals. Parts (c) and (d) shows the strength vs period plot for these two mappings using a Ramanujan dictionary and the convex program (3.24) with the penalty function $f(P) = \phi(P)$. The distinct peaks at 12, 6, 4, 3 and 2 clearly indicate the underlying periodicity of 12.

Electrocardiography

An electrocardiogram (ECG) is an electrical recording of the activity of the heart. It is an approximately periodic waveform as show in Fig. 3.11 (a), generated by the cyclic polarization and depolarization of the cardiac tissue. ECG recordings are extremely valuable to cardiologists for monitoring the condition of the heart and also in diagnosing abnormalities.

Fig. 3.11(a) shows a sampled and filtered ECG recording obtained from a 23 year old female patient (source: ECG ID Database, PhysioNet [28]). The original sampling rate was 500 Hz, but we downsampled the signal further by 4 for ease of computations. We manually calculated the average time period as 0.8644 s. Fig. 3.12 shows the strength vs period plot for this signal using the Ramanujan

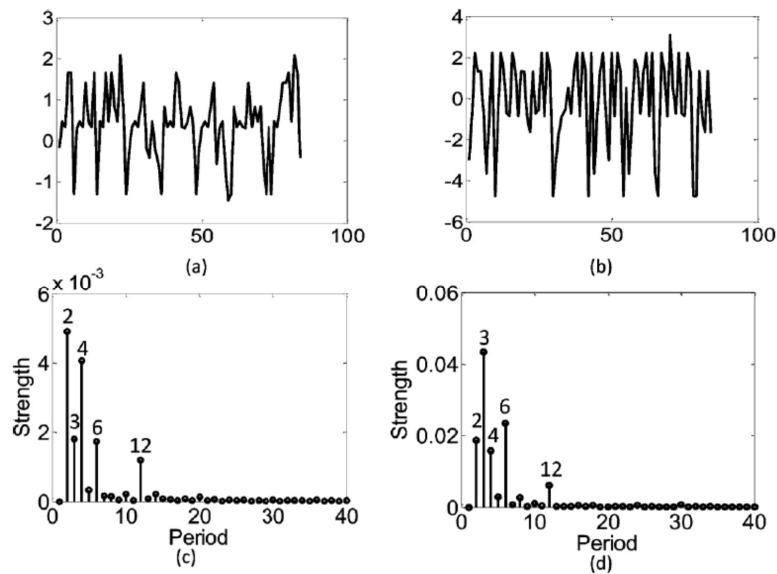


Figure 3.10: Parts (a) and (b) – Amino acid sequence of the Antifreeze Protein from *Tenebrio Molitor*, numerically mapped according to two different metrics (refer text for details). Parts (c) and (d) – Corresponding strength vs period plots using the convex program (24) and the Ramanujan dictionary.

dictionary and the convex program (3.24). The penalty function was $f(P) = \phi(P)$. The peaks at 108, 54, 36, 18 and 12 indicate a periodicity of 108 which corresponds to a period of 0.8640 s. Fig. 3.11(b) shows the DFT of this signal. The fundamental frequency, which is the first peak in the plot, corresponds to a time period of 0.8372 s from this plot. This is an example where our methods out-perform the DFT in terms of a more accurate estimate.

3.6 Conclusion

In this chapter, we presented new methods to identify periodicities in data. Apart from being able to estimate the unknown period of a signal, our methods look for hidden periodicities within the data. We took the idea of the complex-valued Farey dictionary [26] a step further by generalizing it to other much simpler real valued

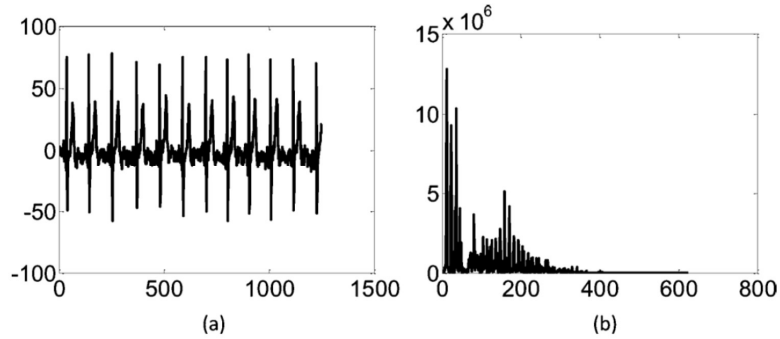


Figure 3.11: Part (a) - ECG waveform of a 23 year old female patient. Part (b) - DFT coefficients for positive frequencies.

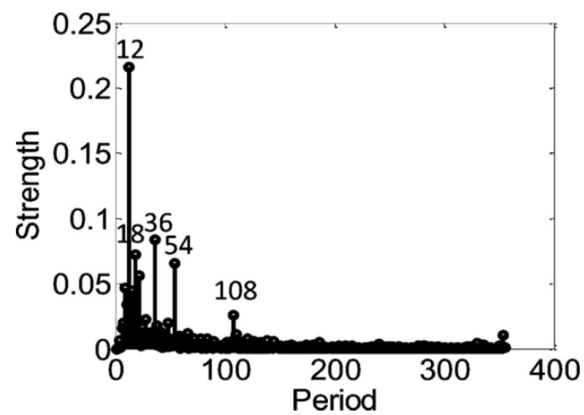


Figure 3.12: Strength vs Period plot for the ECG signal of Fig. 3.11 using the convex program .

dictionaries. Just as the Farey dictionary uses DFT matrices as building blocks, our new designs of dictionaries use a family of alternatives to the DFT matrix called the Nested Periodic Matrices. The theory we developed for this family unifies several important transforms that were relevant to the periodicity problem such as the DFT, Walsh-Hadamard and the Ramanujan Periodicity Transform matrices.

We also proposed a way of replacing the computationally expensive l_1 norm based second order cone programming of the Farey approach of [26] by an l_2 norm based method with closed form solutions. This change resulted in several orders of magnitude faster computations. We also modified the convex programs to incorporate a much larger class of dictionaries than the Farey dictionary. Finally, as a proof of our concept, we applied our methods to two real world applications - analyzing protein repeats and ECG data.

In terms of performance under noise, we observed that dictionaries constructed using either the Ramanujan Periodicity Transform or the DFT matrices offer the best results. This might have to do with the orthogonality property mentioned in Theorem 3.3.2. As is evident from that theorem, the Ramanujan Periodicity Transform matrices provide an integer basis alternative to the same periodic subspaces that are spanned by the columns of the DFT matrix. A more thorough performance analysis and comparisons between the various dictionaries will be a useful direction to pursue in the future. Additionally, an analysis on the choice of the penalty functions in (3.24) and (3.32) will be of interest to us in our future work.

Chapter 4

A UNIFIED THEORY OF UNION-OF-SUBSPACES REPRESENTATIONS OF PERIODIC SIGNALS

4.1 Introduction

The Nested Periodic Matrices and Dictionaries of the previous chapter generalize Ramanujan subspaces to a much bigger family of subspaces. However, as mentioned in Chapter 1, there have been a few more subspace models for periodic signals that have been used for period estimation in the past [12]–[14], [25], [29]. Most of these other works use the column-extended versions of Identity matrices to span period- P signals. This results in a dimension P subspace for each period P , unlike the Nested Periodic Dictionaries which have a dimension $\phi(P)$ subspace for each period P . As we will explain in this chapter, this additional redundancy in the other techniques can cause some fundamental problems.

In spite of many such works that use subspace models for periodicity, there is no unified theory in the literature that studies and compares all such models under one framework. As such, the connection between the elegant Exactly Periodic Subspaces theory of Muresan and Parks [14], the novel periodicity transforms of Sethares and Staley [13], the dictionary approach of Nakashizuka [25], the theory of intrinsic periodic functions of Pei and Lu [30], and the nested periodic dictionaries [23], has not been studied in the literature. Thus, all these above methods remain mostly as isolated pieces of work. Furthermore, there are a number of unanswered questions in the context of dictionaries constructed using such subspaces. For example, (a) what is the provably minimum required dictionary size for the periodicity problem? (b) What are the required dimensions of the various subspaces representing hidden periods in the dictionary? (c) What is a **minimal set** of conditions on the dictionary so that it yields a unique representation for a periodic signal? (d) If a dictionary is based on single-frequency exponentials (Fourier atoms) then how should these frequencies be spaced on the unit circle? If the intuitively appealing uniform-grid is not the best, then what is the best grid to use? Many such questions remain unanswered so far.

Chapter Scope and Outline

The purpose of this chapter is therefore two-fold. First, we unify many of the above methods based on **nested periodic matrices and subspaces**. For this, a brief (but important) review of the relevant past methods is first given in Sec. 4.2. The connection between the methods is then described. For example, we show in Sec. 4.3 that the **exactly periodic subspaces** (EPS) of Muresan and Parks [14] are precisely the Ramanujan subspaces defined in [18] (Theorem 4.3.1). The complex theoretical framework of [14] which arrives at these spaces through an orthogonalization approach can therefore be replaced by the direct methods of [9], [18] and [23]. In the EPS theory, the dimensions of the periodic subspaces are not properly explained or specified. We show, based on the connection with Ramanujan spaces, that this dimension is precisely the Euler totient function $\phi(P)$, where P is the period associated with the subspace. (In a later section, we will show that the Euler totient is even more fundamental than this, please see below.) Third, we also show that the intrinsic integer periodic functions (IIPF), defined from a very different point of view in [30], are in fact identical to the Ramanujan space approach (Theorem 4.3.2).

The second purpose of this chapter is to go beyond this unification, and place the dictionary approaches [23], [25] on a firm theoretical footing. This gives rise to a number of theorems which answer several basic questions about the dictionary approaches, not addressed in any of the earlier papers including [23]. For example, what is the theoretically minimum number of atoms required in any type of dictionary, in order to represent periods $1 \leq P \leq P_{max}$? For each period P , what should be the minimum dimension of the subspace of atoms representing the P th period itself? The answers are found in Theorems 4.6.1 and 4.6.2 (Sec. 4.6). In particular, the answer to the second question is precisely the Euler totient $\phi(P)$. We will also see, rather surprisingly, that a larger-than-minimal dictionary creates difficulties in the process of uniquely identifying periods even in the absence of noise (Sec. 4.6).

Next, what is the set of properties of the subspaces in a dictionary, which allows a signal to be decomposed into periodic components in a **unique** way? We will answer this in Sec. 4.4, and show that among all the past approaches, there are very few methods which allow such a unique solution, and this set includes Ramanujan-space based methods. We answer this in Sec. 4.4. These results, presented as Theorems 4.4.1, 4.4.2 and 4.4.3, are much stronger than earlier results in the sense that, an absolutely minimum set of conditions are imposed on the dictionary subspaces (the LIPS conditions, Sec. 4.4) which make all the previously imposed conditions

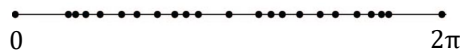


Figure 4.1: The set of frequencies needed to span all periodic signals whose periods lie in the range $1 \leq P \leq 8$. See Sec. 4.5.

[23] such as the Euler structure, the Nested Periodic Property, and so forth, come out as natural consequences of this!

When a signal is expressed as a linear combination of periodic subspace signals then it is sometimes possible to uniquely identify the period of the original signal from a mere knowledge of the indices (= periods) of the participating subspaces. Namely, the period is exactly the LCM of these participating indices. This is a very important practical property, when we have to estimate the period from the dictionary representation. This LCM property is true for the Nested Periodic Dictionaries in Chapter 3 but not for the methods in [12], [13], [25]. What then is the fundamental theoretical condition under which this LCM property holds? Is there a broader class of methods than the Nested Periodic Subspaces methods, with the LCM property? This question is addressed in Sec. 4.6 (Theorem 4.6.3).

Fourier dictionaries and frames are popularly used in a number of signal processing applications such as DOA estimation [31], [32] and [33]. These dictionaries are usually chosen such that the frequency grid is uniformly sampled. In order to avoid the inaccuracies in representation caused by the grid, it is customary to increase the number of atoms in the dictionary to decrease the grid size.¹ We will show that for the case of period estimation, the correct representation departs from this conventional approach in two ways. First, increasing the size of the dictionary is detrimental to the period estimation problem (Sec. 4.6, Sec. 4.7), and second, the best frequency grids are necessarily **nonuniform**, and patterned after the so-called **Farey series** of number theory [24] (please see Fig. 4.1). While such a dictionary was first reported in [26], the fact that this is the **only** dictionary that works if the atoms are Fourier exponentials is new and is proved for the first time in [37], as elaborated in Sec. 4.5.

¹There also exist gridless methods, which take a different approach, e.g., MUSIC [15], or more recent methods [34] and [35], [36].

Finally in Sec. 4.7 we provide some numerical examples to demonstrate the effect of redundant versus minimal dictionaries in the representation of periodic signals in the presence of noise. We also demonstrate how denoised versions of noisy periodic signals can be reconstructed from such representations.

4.2 Review of Existing Subspace Methods

In this section, we analyze several existing period estimation techniques that, either directly or indirectly, utilize union of subspaces models for representing periodic signals. We will start by defining some basic linear algebra on sets of periodic signals (most of the notations are adapted from [13]). We will begin with the formal definition of a periodic signal:

Definition 4.2.1. *A discrete time signal $x(n)$ is said to be periodic if there exists a positive integer P such that:*

$$x(n + P) = x(n) \quad \forall n \in \mathbb{Z} \quad (4.1)$$

Such a P is known as a repetition index of $x(n)$. The smallest repetition index is called the period.

Now, consider the following linear space:

$$\mathcal{V}_P = \{x : x(n + P) = x(n) \quad \forall n \in \mathbb{Z}\} \quad (4.2)$$

\mathcal{V}_P is the set of all signals that have P as a repetition index. Notice that apart from period P signals, \mathcal{V}_P also contains signals that have proper divisors of P as periods. We can further define a set \mathcal{V} as follows:

$$\mathcal{V} = \cup_{P=1}^{\infty} \mathcal{V}_P \quad (4.3)$$

\mathcal{V} can be shown to be an inner product space [13], with the following definition for the inner product:

$$\langle x, y \rangle = \lim_{k \rightarrow \infty} \frac{1}{2k + 1} \sum_{n=-k}^k x(n)y^*(n) \quad (4.4)$$

In practice though, only finite durations of the signals are available to us. Appendix 4.10 discusses techniques to approximate (4.4) in the finite data length situation.

The most important application of defining an inner product over \mathcal{V} is that we can now define notions such as orthogonality, projections, and so on. For example, to

take the projection of a periodic signal $x(n)$ onto an N dimensional subspace \mathcal{G} , we start with an orthogonal basis $\{g_0(n), g_1(n), \dots, g_{N-1}(n)\}$ for \mathcal{G} . The projection of $x(n)$ onto \mathcal{G} , denoted by $\pi_{\mathcal{G}}(x)$, is defined as follows:

$$\pi_{\mathcal{G}}(x) = \sum_{s=0}^{N-1} \frac{\langle x, g_s \rangle}{\|g_s\|^2} g_s \quad (4.5)$$

With this background, we will now start our discussion on some important period estimation and periodic decomposition techniques.

Maximum Likelihood (ML) Period Estimators

An ML framework for period estimation was proposed in [12] in the following way. Suppose we want to find the period P of a signal $s(n)$ from a finite duration, and noise corrupted version of it. Let $\mathbf{s} = [s(1), s(2), \dots, s(K)]^T$ denote the signal vector, and let us assume that it is corrupted by uncorrelated zero mean Gaussian noise $\mathbf{n} \sim N(\mathbf{0}, \sigma \times \mathbf{I}_{K \times K})$, resulting in $\mathbf{r} = \mathbf{s} + \mathbf{n}$. The goal is to estimate P from \mathbf{r} .

If $\mathbf{q} = [q(1), q(2), \dots, q(P)]$ denotes one period of $s(n)$, then the likelihood of \mathbf{r} given $(\mathbf{q}, P, \sigma^2)$ is:

$$P(\mathbf{r}|\mathbf{q}, P, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{r} - \mathbf{s}\|^2\right) \quad (4.6)$$

where,

$$\begin{aligned} \|\mathbf{r} - \mathbf{s}\|^2 = & \sum_{i=0}^{\lfloor \frac{K}{P} \rfloor - 1} \sum_{n=1}^P (r(n+iP) - q(n))^2 \\ & + \sum_{m=1}^{K - \lfloor \frac{K}{P} \rfloor P} \left(r\left(m + \lfloor \frac{K}{P} \rfloor P\right) - q(m) \right)^2 \end{aligned}$$

If $(\hat{\mathbf{q}}, \hat{P}, \hat{\sigma}^2)$ is the global maximizer of (4.6) over all possible triplets $(\mathbf{q}, P, \sigma^2)$, then \hat{P} is defined as the ML estimate of the period. This turns out to be the same as projecting \mathbf{r} onto each of the \mathcal{V}_P 's (4.2) for $1 \leq P \leq P_{max}$, where P_{max} is the largest period expected, and choosing the P for which $\pi_{\mathcal{V}_P}(\mathbf{r})$ (4.5) has the maximum energy.

A major difficulty with the above formulation was that instead of the period itself, a multiple of the period was being estimated. This is because, $\mathcal{V}_P \subset \mathcal{V}_{NP}$ for $N > 1$. To overcome this challenge, a heuristic modification is added to the maximizer of

(4.6) to suppress larger periods compared to the smaller ones. If E_P denotes the energy of $\pi_{\mathcal{V}_P}(r)$, then instead of comparing the E_P 's directly, [12] proposes to compare the following:

$$G_P = E_P - \frac{P}{K} \Phi_{\mathbf{r}}(0) \quad (4.7)$$

where, $\Phi_{\mathbf{r}}$ denotes the autocorrelation function of \mathbf{r} .

This same issue due to $\mathcal{V}_P \subset \mathcal{V}_{NP}$ is encountered in the period estimation algorithms proposed by Sethares and Staley in [13]. We will briefly discuss them next.

Sethares and Staley's Periodicity Transforms

Independent of the ML framework, Sethares and Staley in [13] proposed several methods that involve comparing the projection energies of the signal on \mathcal{V}_P 's. Unlike in [12], where the signal was simultaneously projected onto the different \mathcal{V}_P 's, the authors in [13] proposed several sequential projection schemes to overcome the challenge caused by $\mathcal{V}_P \subset \mathcal{V}_{NP}$. The main idea is to project the signal onto a particular \mathcal{V}_P , and if the projection is of significant energy, then the signal is declared to have a period P component. The resulting residue is then projected onto a different subspace \mathcal{V}_Q to search for a period Q component. This process is continued until all the subspaces $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_{P_{max}}$ are covered. The important question is, in which order do we choose these subspaces for the projections?

The non-orthogonality between the various \mathcal{V}_P 's causes the projection energies obtained in the algorithms of [13] to depend on the order in which we do these projections. For example, consider Fig. 4.2(a), where we have three vectors \mathbf{x} , \mathbf{a} and \mathbf{b} . In Fig. 4.2(b), we project \mathbf{x} on \mathbf{a} first to obtain \mathbf{x}_a and then project the residue $\mathbf{x} - \mathbf{x}_a$ onto \mathbf{b} to give \mathbf{r}_{ab} . Based on the relative sizes of \mathbf{x}_a and \mathbf{r}_{ab} , we might declare that \mathbf{x} is closer to \mathbf{a} than to \mathbf{b} . But on the other hand, if we first project \mathbf{x} on \mathbf{b} and then project the resulting residue on \mathbf{a} , we get the vectors shown in Fig. 4.2(c), which give us the exact opposite conclusion. How can we know a priori which order of projections to choose? The various iterative algorithms presented in [13] can be interpreted as sophisticated ways of selecting an appropriate order of projections.

To avoid this problem, Muresan and Parks in [14] proposed a decomposition of periodic signals onto a set of orthogonal subspaces. Notice that if \mathbf{a} and \mathbf{b} were orthogonal as shown in Fig. 4.3, then any order of projections gives us the same

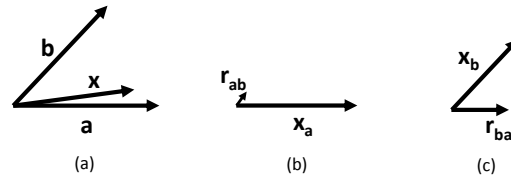


Figure 4.2: An example to illustrate that the projection energies obtained on the different \mathcal{V}_P 's depend on the order of projections. Refer Sec. 4.2 for details.

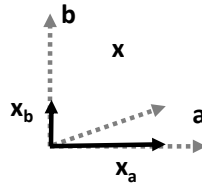


Figure 4.3: An example to illustrate that orthogonal subspaces result in unique projections irrespective of the order of decomposition. Refer Sec. 4.2 for details.

decomposition. In fact, we can project \mathbf{x} on both \mathbf{a} and \mathbf{b} simultaneously to get that decomposition. We will discuss this in more detail next.

The EPS Method of Muresan and Parks

An exactly period- P signal was defined as follows in [14]:

Definition 4.2.2. Exactly Periodic Signals: A signal $x(n) \in \mathcal{V}_P$ is of exactly period P if its projection onto \mathcal{V}_Q is zero for all $Q < P$. \diamond

It was shown in [14] that the collection of all exactly period P signals is in fact a subspace of dimension less than P . Such a subspace was called the Exactly Periodic Subspace of period P (EPS(P)).

Moreover, the following two properties were noted for Exactly Periodic Subspaces:

1. *Direct Sum Property:* $\oplus_{d|P} \text{EPS}(d) = \mathcal{V}_P$.
2. *Orthogonality:* $\text{EPS}(P) \perp \text{EPS}(Q)$ when $P \neq Q$.

These two properties together ensure that if we take projections of a period P signal $x(n)$ onto the subspaces $\text{EPS}(1)$ to $\text{EPS}(P_{max})$, for some $P_{max} \geq P$, then the projections can be non-zero only on those subspaces $\text{EPS}(d_i)$, where $d_i|P$. This can be used to estimate the period of the signal. Also, we do not have to do a sequential process of projecting the signal on a subspace, and then projecting its residue onto the next subspace, like in the algorithms of [13]. Since the subspaces are orthogonal to each other, we can simultaneously project the signal on all of them (Fig. 4.3). In practice though, when the signals are of finite duration, this orthogonality is only approximately true. But if the signals are reasonably long, then the approximation is good, as was demonstrated in [14].

Ramanujan and Nested Periodic Subspaces

The Ramanujan subspaces also satisfy the direct sum and orthogonality properties of the EPSs. But recall from Chapter 2 that they also have the following two properties:

1. *Euler-structure*: For every period P , the Ramanujan subspace \mathcal{S}_P has dimension $\phi(P)$.
2. *LCM Property*: Let $x_{i_1}(n), x_{i_2}(n), \dots, x_{i_K}(n)$ be non-zero signals lying in the Ramanujan subspaces $\mathcal{S}_{i_1}, \mathcal{S}_{i_2}, \dots, \mathcal{S}_{i_K}$ respectively. Then, the sum $x_{i_1}(n) + x_{i_2}(n) + \dots + x_{i_K}(n)$ has its period given by²

$$P = \text{lcm}(i_1, i_2, \dots, i_K) \quad (4.8)$$

The significance of the Euler structure was unknown previously. In Sec. 4.4, we will show that it has a fundamental significance in the context of unique periodic decompositions. The LCM property on the other hand, gives a systematic way of estimating the period from the projections. It tells us that the lcm of the periods of those subspaces that have non-zero projections is exactly equal to P . For example, if a signal $x(n)$ has non-zero projections only on \mathcal{S}_2 and \mathcal{S}_3 , we can infer from the LCM property that its period must be 6.

To bring the Nested Periodic Dictionaries formally into the context of our discussion, we will define Nested Periodic Subspaces based on them in the following way:

²in general, the period of a sum of periodic signals can be either the lcm or a proper divisor of it. Whenever we say lcm, it is a strong statement indicating that the period is not a proper divisor.

Definition 4.2.3. Nested Periodic Subspaces (NPSs): Let $\mathbb{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{P_{max}}\}$ be a set of subspaces such that $\mathcal{T}_i = \text{span}\{\mathbf{C}_i\}$ for some NPM $\mathbf{A} = [\mathbf{C}_1 \ \mathbf{C}_2 \ \mathbf{C}_3 \ \dots]$. Here, as in Sec. 3.1, each \mathbf{C}_i represents the columns in \mathbf{A} with period i . Then, \mathbb{T} is called as a set of Nested Periodic Subspaces.

Notice that in the above definition, we required that the NPM \mathbf{A} contain columns with every possible period in the range $1 \leq P \leq P_{max}$. This is possible by choosing the number of columns in \mathbf{A} to be any multiple of $\text{lcm}(1, 2, \dots, P_{max})$. The Ramanujan Subspaces are special cases of NPSs. If $\mathbb{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{P_{max}}\}$ is a set of Nested Periodic Subspaces such that $\mathcal{T}_i \perp \mathcal{T}_j$ whenever $i \neq j$, then, for every i , \mathcal{T}_i must necessarily be the i^{th} Ramanujan Subspace (Theorem 3.3.2). NPSs can be used for period estimation in the following way:

Theorem 4.2.1. Period Estimation via NPSs: Let $\mathbb{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{P_{max}}\}$ be a set of Nested Periodic Subspaces. For each i , let \mathbb{R}_i denote a basis for \mathcal{T}_i . Then,

1. Given any periodic signal $x(n)$ with period $P \leq P_{max}$, it can be uniquely expressed as a sum of the elements in $\cup_i \mathbb{R}_i$.
2. If signals from $\mathbb{R}_{i_1}, \mathbb{R}_{i_2}, \dots$ are involved in spanning $x(n)$, then the period of $x(n)$ must be equal to $\text{lcm}(i_1, i_2, \dots)$.

The proof follows from the fact that $\{\mathbb{R}_1, \mathbb{R}_2, \mathbb{R}_3, \dots\}$ can be extended to form an NPM, and then using Lemma 3.2.3.

While the NPSs also satisfy the direct sum property like the EPSs and the Ramanujan subspaces, not all sets of NPSs are orthogonal. So how do we interpret the above result in terms of our discussions on Fig. 4.2 and Fig. 4.3? The answer is that period estimation using the NPSs was not based on taking projections along these subspaces. Instead, it uses a different idea based on the fact that the NPSs are linearly independent subspaces. Notice that in Fig. 4.2, if \mathbf{a} and \mathbf{b} are linearly independent, then we can obtain a unique decomposition of \mathbf{x} by completing the parallelogram as shown in Fig. 4.4 instead. By comparing the sizes of the components, we can easily estimate whether \mathbf{x} was closer to \mathbf{a} or to \mathbf{b} . In practice, such a decomposition is done, not using projections, but using dictionaries, as will be explained in Sec. 4.5.

4.3 Insights into the Relationships Between Various Techniques

As mentioned in Sec. 7.1, most of the works discussed in the previous section have so far been studied independent of each other. However, there are some interesting theoretical connections between them that have been largely unexplored. This section aims to fill this gap.

Notice that, as derived in [14], the EPSs are in fact the outcomes of a Gram Schmidt orthogonalization of the subspaces used in [12], [13] and [25], namely the \mathcal{V}_P 's.

Relationship between EPSs and Ramanujan Subspaces

Although $EPS(q)$ and \mathcal{S}_q were formulated in completely different ways, it can be shown that they are in fact the same subspaces. We derive this in the following Theorem.

Theorem 4.3.1. EPS and Ramanujan Subspaces: *EPS(P) is the same subspace as the P^{th} Ramanujan Subspace \mathcal{S}_P .* \diamond

Proof. Let $x(n) \in \mathcal{S}_P$. Let $Q < P$. Using the direct sum property of the Ramanujan subspaces, $\oplus_{d|Q} \mathcal{S}_d = \mathcal{V}_Q$. Since $\mathcal{S}_P \perp \mathcal{S}_d$ for all $d < P$, it follows that $\mathcal{S}_P \perp \mathcal{V}_Q$. Since this is true for all $Q < P$, it follows that $x(n) \in EPS(P)$. This implies that $\mathcal{S}_P \subseteq EPS(P)$.

Conversely, assume that $x(n) \in EPS(P)$. Then, $x(n) \perp \mathcal{S}_d$ for every proper divisor d of P . But $x(n) \in \mathcal{V}_P$ (since $EPS(P) \subseteq \mathcal{V}_P$) and $\oplus_{d|P} \mathcal{S}_d = \mathcal{V}_P$. This implies that $x(n) \in \mathcal{S}_P$. Hence, $EPS(P) \subseteq \mathcal{S}_P$. These two results together show that $\mathcal{S}_P = EPS(P)$. $\nabla \nabla \nabla$

The authors of [9], [18] and [14] were not aware of this connection. Moreover, using the special properties of Ramanujan Sums, many more properties of these subspaces such as the LCM property and the Euler structure were derived in [9] and [18], which were previously unknown for the EPSs. For example, if a signal had non-zero projections on $EPS(6)$ and $EPS(8)$, then previously, it was not possible to say whether the period of the signal was 24, or a proper divisor of 24 (such as 12). This is because, it was not known whether one could apply the LCM property for the EPSs. However, using the equivalence between EPSs and the Ramanujan subspaces, we can conclude that the period must be exactly equal to 24. Similarly, the dimension of $EPS(P)$ was noted to be less than P in [14], but unknown analytically. Using Theorem 4.3.1, we can conclude that the dimension is in fact $\phi(P)$.

Relationship between Ramanujan Subspaces and IIPFs

Apart from the EPSs, the Intrinsic Integer-Periodic Functions (IIPFs) introduced in [30] can also be used to define the Ramanujan subspaces. Any periodic signal $x(n)$ with period P was defined to be a P -IIPF if, for every $d|P$, and for every integer l ,

$$\sum_{k=0}^{\frac{P}{d}-1} x(kd + l) = 0 \quad (4.9)$$

IIPFs are related to the Ramanujan subspaces in the following way:

Theorem 4.3.2. IIPFs and Ramanujan Subspaces: *The set of all P -IIPFs is the P^{th} Ramanujan Subspace.* \diamond

Proof. Let $x(n)$ be a signal in \mathcal{V}_P . Let d be a divisor of P . Then, saying that every downsampled-by- d version of $x(n)$ sums to zero in a period, is equivalent to saying that $\langle x, \delta_d^{(s)} \rangle = 0$ for all $0 \leq s \leq d - 1$, where $\delta_d^{(s)}$ is defined in the following way:

$$\delta_d^{(s)}(n) = \begin{cases} 1 & \text{if } n \bmod d = s \\ 0 & \text{otherwise} \end{cases} \quad (4.10)$$

Since $\{\delta_d^{(s)}(n)\}_{s=0}^{d-1}$ forms an orthogonal basis for \mathcal{V}_d , it follows that $x(n) \perp \mathcal{V}_d$. So, a signal $x(n)$ is a P -IIPF iff $x(n) \perp \mathcal{V}_d$ for every proper divisor d of P .

Now, we will show that if $x(n)$ is a P -IIPF, then $x(n) \in \mathcal{S}_P$. If $x(n) \perp \mathcal{V}_d$ for every proper divisor d of P , then since $\mathcal{S}_d \subseteq \mathcal{V}_P$, it follows that $x(n) \perp \mathcal{S}_d$ for every proper divisor d of P . From the direct sum property of Ramanujan subspaces, it follows that $x(n) \in \mathcal{S}_P$.

To prove the converse, let $y(n)$ be a signal in \mathcal{S}_P . We will prove that $y(n)$ must be a P -IIPF. From the orthogonality property of Ramanujan subspaces, $y(n) \perp \mathcal{S}_d$ for every proper divisor d of P . From the direct sum property of Ramanujan sums, it then follows that $y(n) \perp \mathcal{V}_d$ for every proper divisor d of P . So $y(n)$ must be a P -IIPF. $\nabla \nabla \nabla$

Relationship between Ramanujan Subspaces and NPSs

The Ramanujan Subspaces are special cases of NPSs. If $\mathbb{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{P_{\max}}\}$ is a set of Nested Periodic Subspaces such that $\mathcal{T}_i \perp \mathcal{T}_j$ whenever $i \neq j$, then, for every i , \mathcal{T}_i must necessarily be the i^{th} Ramanujan Subspace. The proof follows from Theorem 3.3.2.

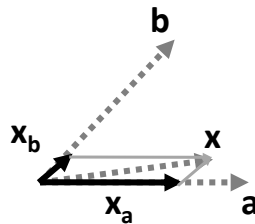


Figure 4.4: An example to illustrate that one can obtain a unique decomposition using parallelogram completion as long as the subspaces are linearly independent. Refer Sec. 4.2 and Sec. 4.4 for details.

This completes the interrelationship between all the previous techniques. Unlike the \mathcal{V}_P 's, the NPSs have a very special property: Any periodic signal can be uniquely decomposed into its components along those subspaces. Are there more general subspaces that have this property? This is explored in the next section.

4.4 Fundamental Properties of Subspaces That Admit Unique Periodic Decompositions

The basic idea behind using NPSs (including EPS/Ramanujan subspaces) for period estimation is as follows: A given periodic signal can be decomposed into its components along a set of linearly independent subspaces, revealing its period (the parallelogram completion idea of Fig. 4.4). These components turn out to be the projections in the case of Ramanujan subspaces and the EPSs, due to their orthogonality. The parallelogram completion idea of Fig. 4.4 would not work when using the \mathcal{V}_P 's since every periodic signal can be expressed in an infinite number of ways as a sum of its components along the \mathcal{V}_P 's. This is due to the linear dependence among the various \mathcal{V}_P 's³ A natural question then is as follows: Beyond the NPSs, are there more general subspaces that have this linear independence property? An analysis of this question reveals some surprising results. We start with the following definition:

Definition 4.4.1. Linearly Independent Periodic Subspaces: Let $\mathbb{G} = \{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{P_{max}}\}$ be a set of subspaces such that the following hold:

1. **Linear Independence:** \mathbb{G} is a linearly independent set of subspaces. That is, if for each i , \mathbb{R}_i is a linearly independent set of signals in \mathcal{G}_i , then $\cup_i \mathbb{R}_i$ is a linearly independent set of signals too.

³An easy way of seeing this linear dependence is using the fact that $\mathcal{V}_P \subset \mathcal{V}_{NP}$, $\forall N > 1$.

2. Direct Sum Condition: $\bigoplus_{d|P} \mathcal{G}_d = \mathcal{V}_P$.

Such a set \mathbb{G} will be called a set of *Linearly Independent Periodic Subspaces (LIPS)*.

◇

Notice that, unlike in the definitions of EPSs, Ramanujan Subspaces or the Nested Periodic Subspaces, we have imposed very few conditions while defining LIPS here. For example, *we did not even insist that \mathcal{G}_i must consist of period i signals*. While the linear independence condition in the above definition is motivated by the need for unique periodic decompositions, the direct sum condition ensures that only those \mathcal{T}_i , with $i|P$, are involved in spanning a period P signal. The latter makes it convenient to be able to estimate the period based on the indices of the subspaces along which the signal has large components. It is easy to verify from the discussion in Sec. 4.2, that the EPSs (equivalently Ramanujan Subspaces) and all other Nested Periodic Subspaces are specific examples of subspaces that satisfy the conditions of being LIPS.

We will now derive some fundamental properties of subspaces that satisfy the LIPS conditions.

Linear Independence Necessitates Euler Structure

First, we prove the following theorem:

Theorem 4.4.1. Origin of the Euler-structure: *Let \mathbb{B}_P be a basis for \mathcal{V}_P such that for every divisor d of P , there exists a subset of \mathbb{B}_P that is a basis for \mathcal{V}_d . Then \mathbb{B}_P must have exactly $\phi(d)$ vectors with period d , for every divisor d of P .* ◇

Proof. Let d be a divisor of P . A basis for \mathcal{V}_d must consist of signals with periods that are divisors of d , since each basis vector must itself belong to \mathcal{V}_d . The rank of the set of all signals in \mathbb{B}_P that have proper divisors of d as periods must be $\leq \sum_{\substack{d_i|d \\ d_i < d}} \phi(d_i)$ (see Appendix 4.10). So for \mathbb{B}_P to contain a basis for the d dimensional space \mathcal{V}_d , there must be at least $d - \sum_{\substack{d_i|d \\ d_i < d}} \phi(d_i) = \phi(d)$ [24] linearly independent signals in \mathbb{B}_P with period d . But the maximum number of vectors in \mathbb{B}_P must be P , and $\sum_{d|P} \phi(d) = P$ [24]. So for each divisor d of P , there are exactly $\phi(d)$ vectors with period d in \mathbb{B}_P . ▽ ▽ ▽

In [23], it was shown that the Euler-structure of the NPMs leads to their nested basis property (Sec. II-E). But the above theorem shows that the converse is also

true. Namely, if \mathbf{A} is a $P \times P$ matrix satisfying the nested basis property, then Theorem 4.4.1 states that \mathbf{A} must have exactly $\phi(d)$ columns with period d for every $d|P$, making it an NPM.

We will now show that every set of LIPS must necessarily have the Euler-structure.

Theorem 4.4.2. The Euler-structure in LIPS: *Let $\mathbb{G} = \{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{P_{max}}\}$ be a set of LIPS. Then for each i , \mathcal{G}_i has to be a $\phi(i)$ dimensional subspace consisting of period i signals only.* \diamond

Proof. For every i , let \mathbb{R}_i denote a basis for \mathcal{G}_i . It follows from the direct sum and the linear independence conditions in Definition 4.4.1 that $\cup_{d|i} \mathbb{R}_d$ is a basis for \mathcal{V}_i .

We will first prove that \mathcal{G}_i consists of period i signals only. It follows from the direct sum condition of LIPS that $\mathbb{G}_i \subseteq \mathcal{V}_i$, and so \mathcal{G}_i must consist of signals that are periodic with period i or a proper divisor of i . To show that \mathcal{G}_i consists of only period i signals, (and not signals with periods that are proper divisors of i) we use a proof by contradiction. Suppose there was a signal $y(n)$ with period j in \mathcal{G}_i such that $j < i$ and $j|i$. In that case, $\cup_{d|j} \mathbb{R}_d$ must be able to span $y(n)$. But $y(n)$ can also be spanned by \mathbb{R}_i , which is a basis of \mathcal{G}_i . This violates the linear independence condition in the definition of LIPS. Hence, \mathcal{G}_i must consist of period i signals only.

To show that the dimension of \mathcal{G}_i is $\phi(i)$, notice that if $d|i$, then $\cup_{d|d} \mathbb{R}_d$, which is a subset of $\cup_{d|i} \mathbb{R}_d$, is a basis for \mathcal{V}_d . So $\cup_{d|i} \mathbb{R}_d$ is a basis for \mathcal{V}_i that satisfies the conditions of Theorem 4.4.1. And since \mathbb{R}_i must consist of period i signals only, it follows that \mathbb{R}_i must consist of exactly $\phi(i)$ period i signals. Hence, the dimension of \mathcal{G}_i must be $\phi(i)$. $\nabla \nabla \nabla$

The above Theorem shows that the Euler-structure is necessary when one tries to design subspaces that offer unique decompositions for periodic signals. The Nested Periodic Subspaces, by definition, have their dimensions given by the Euler totient function (Definition 4.2.3). It follows from the properties of Nested Periodic Matrices that this leads to unique decompositions of periodic signals (see Theorem 4.2.1). However, Theorem 4.4.2 gives us a converse result. Namely, it shows that if one wants unique periodic decompositions, then the dimensions of the subspaces must necessarily be given by the Euler totient function.

This leads us to suspect whether the Nested Periodic Subspaces (Definition 4.2.3) are the *only* examples of subspaces that offer unique periodic decomposition. This is indeed true, as shown by the following theorem:

Theorem 4.4.3. NPSs are the only examples of LIPS: $\mathbb{G} = \{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{P_{max}}\}$ is a set of LIPS iff it is a set of NPS. \diamond

Proof. For every i , let \mathbb{R}_i denote a basis for \mathcal{G}_i . Let $L = lcm\{1, 2, \dots, P_{max}\}$. And let $\mathcal{H}_{P_{max}+1}, \mathcal{H}_{P_{max}+2}, \dots, \mathcal{H}_L$ represent bases for the Ramanujan Subspaces $\mathcal{S}_{P_{max}+1}, \mathcal{S}_{P_{max}+2}, \dots, \mathcal{S}_L$ respectively. Now, consider the set:

$$\mathbb{B} = \{\mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{P_{max}}, \mathbb{H}_{P_{max}+1}, \mathbb{H}_{P_{max}+2}, \dots, \mathbb{H}_L\} \quad (4.11)$$

If we form a matrix \mathbf{B} whose columns are the signals in \mathbb{B} , truncated to their first L elements, then we can show that such a matrix is a Nested Periodic Matrix. The three conditions of an NPM from Sec. 3.1 are met as follows:

- *Conditions 1 and 2:* From Theorem 4.4.2, we know that each \mathbb{R}_i corresponds to an $L \times \phi(i)$ matrix consisting of period i signals. And by construction, the Ramanujan Subspaces have the ϕ -structure.
- *Condition 3:* $\{\mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{P_{max}}\}$ is a linearly independent set because of the linear independence condition in Definition 4.4.1 of LIPS. And the following set:

$$\{\mathbb{H}_{P_{max}+1}, \mathbb{H}_{P_{max}+2}, \dots, \mathbb{H}_L\} \quad (4.12)$$

is also a linearly independent set because of the orthogonality of the Ramanujan Subspaces. Moreover, from Theorem 4.3.1, $\{\mathbb{H}_{P_{max}+1}, \mathbb{H}_{P_{max}+2}, \dots, \mathbb{H}_L\}$ is orthogonal to $\{\mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_{P_{max}}\}$. Hence, \mathbb{B} in (4.11) has full rank L .

So, we have proved that \mathbb{G} is a set of NPS according to Definition 4.2.3. $\nabla \nabla \nabla$

This Theorem shows that, although NPSs were proposed in [23] as basic generalizations of the Ramanujan Subspaces, especially inheriting their ϕ -structure in an ad hoc fashion, they arise as the only subspaces that offer unique decompositions of periodic signals. This result will be used in Sec. 4.6 to prove some important properties of dictionaries that span periodic signals (Theorem 4.6.3 and Corollary 4.6.1).

Importance of the correct choice of subspaces: An example

Let us now illustrate the concept of unique periodic decomposition using an example. Consider a period 8 signal $x(n)$, whose one period is given by the vector [1, 2, 3, 3, 3, 3, 2, 1]. In Fig. 4.5(a) and (b), we have shown two possible decompositions of this signal along the set of subspaces $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_8\}$:

$$x(n) = \sum_{i=1}^8 x_i(n) \quad (4.13)$$

where, for each i , $x_i(n)$ indicates the component along \mathcal{V}_i . Due to the lack of linear independence among these subspaces, we can obtain an infinite number of such decompositions of the same signal. Notice that the LCM property does not apply to the decomposition in Fig. 4.5(a). The LCM property, if blindly applied, estimates the period to be $lcm(2, 3, 6, 8) = 24$, which is incorrect. This is because the \mathcal{V}_P 's are not linearly independent.

Next, let us decompose the same signal along subspaces that offer unique periodic decompositions. We used the Ramanujan subspaces $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_8\}$ in Fig. 4.5(c) and the Natural Basis subspaces $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_8\}$ in Fig. 4.5(d). The Natural Basis subspaces are an example of non-orthogonal NPSs, proposed in [23]. The decompositions shown in Fig. 4.5 are the only possible decompositions of $x(n)$ along these subspaces. In Fig. 4.5(c), the subspaces with non-zero components are \mathcal{S}_4 and \mathcal{S}_8 , while in Fig. 4.5(d), they are \mathcal{T}_2 , \mathcal{T}_4 and \mathcal{S}_8 . Notice that the LCM property correctly predicts the period of the signal as 8 in each case. In these plots, we have ignored the period 1 components, since it is just DC.

4.5 From Subspaces to Dictionaries

Previously, there have been three works that proposed dictionaries for period estimation. These are Nakashizuka's work [25], Vaidyanathan and Pal's Farey dictionary [26], and the Nested Periodic Dictionaries (NPDs) of [23]⁴. While each of these works analyzes a specific type of dictionary, there can be several fundamental questions that apply to any generic dictionary that spans periodic signals. For example, (a) what is the provably minimum required dictionary size to represent periods $1 \leq P \leq P_{max}$? (b) For each period P , what should be the minimum dimension of the subspace of atoms representing the P th period itself? (c) What is a **minimal set** of conditions on the dictionary so that it yields a unique representation for a periodic signal? In order to answer such questions, we need a much more general framework than in [25], [26] and [23]. In this section, we formulate such a framework by proposing a very general definition of a periodic dictionary, that in particular captures [25], [26] and [23] as special cases. We also derive an important

⁴The Farey dictionary is a special case of the Nested Periodic Dictionaries.

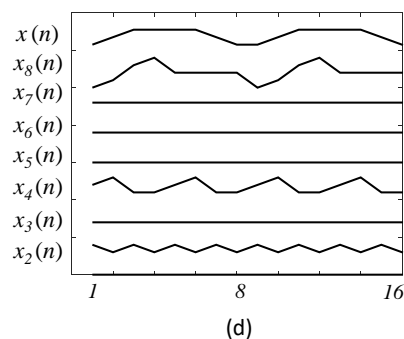
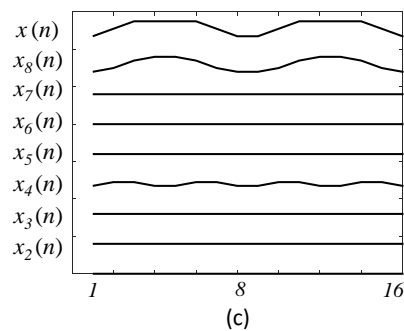
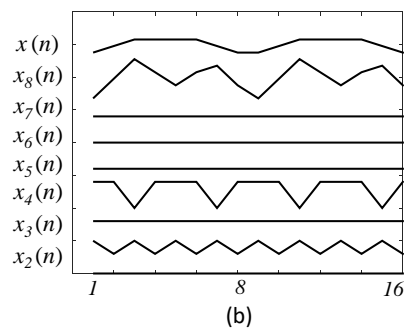
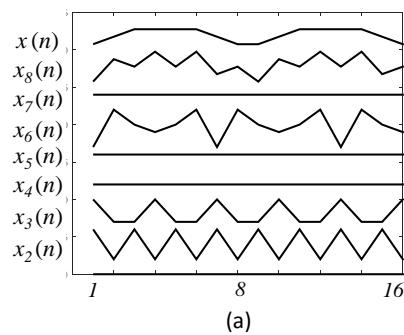


Figure 4.5: Parts (a) and (b): Two different decompositions of a period 8 signal onto $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_8\}$. Part (a) involves $\mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_6$ and \mathcal{V}_8 , while Part (b) involves $\mathcal{V}_2, \mathcal{V}_4$ and \mathcal{V}_8 . Clearly, it is difficult to determine the component periods in the signal using these subspaces. Notice that the LCM property results in an incorrect period estimate in Part (a). Parts (c) and (d) use subspaces that offer unique periodic decompositions: Ramanujan Subspaces in Part (c) and the Natural Basis subspaces in Part (d). Both involve only subspaces with period 8 and its divisors. The LCM property correctly identifies the period as 8 in both these cases. Please see Sec. 4.4 for a discussion.

new property of the Farey dictionary. In the next section, we use this new definition of periodic dictionaries to answer the questions mentioned above.

We will first discuss the connection between dictionaries and the subspace models discussed in the previous sections. Suppose we want to estimate the period of a signal $x(n)$, with the prior knowledge that the period lies in the range $1 \leq P \leq P_{max}$. The main idea in the previous sections is that we can design a set of subspaces $\mathbb{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{P_{max}}\}$ such that $x(n)$ lies in their union. To find the period of $x(n)$, we need to find the exact subset of these subspaces that are involved in spanning $x(n)$.

A popular approach to this problem is to use dictionaries. Let us assume that we have N consecutive samples of $x(n)$ available to us in the form:

$$\mathbf{x} = [x(0), x(1), \dots, x(N-1)]^T \quad (4.14)$$

For every i , let \mathbb{R}_i denote a basis for \mathcal{T}_i . We can form a dictionary \mathbf{A} , whose columns are the signals in $\cup_i \mathbb{R}_i$, truncated to the data length N . That is:

$$\mathbf{A} = [\mathbb{R}_1 \ \mathbb{R}_2 \ \dots \ \mathbb{R}_{P_{max}}] \quad (4.15)$$

Then, the following system of equations will always have a solution for \mathbf{y} :

$$\mathbf{x} = \mathbf{A}\mathbf{y} \quad (4.16)$$

By looking at the locations of the non-zero entries in \mathbf{y} , we can find out those columns of \mathbf{A} (and hence the subspaces in \mathbb{T}) that are involved in spanning $x(n)$. This can be used to estimate the period, as was done for NPDs in Chapter 3.

Typically though, for practical data lengths, \mathbf{A} is a fat matrix. So (4.16) can have multiple solutions for \mathbf{y} . How do we solve for the one that involves subspaces with period P and its divisors? Several techniques have been proposed in the past for this, ranging from the sparsity based techniques from the compressive sensing world ([11], [38]–[41] etc.), to simple l_2 norm convex programs with closed form solutions [23] (Chapter 3).

Periodic Dictionaries - A General Definition

We begin our analysis with the following general definition of a periodic dictionary, that, in particular, captures all the previously proposed dictionaries of [25], [26] and [23] as special examples.

The first column has period 1, the second and third columns have period 2, fourth to sixth have period 3 and so on. We refer to the columns as atoms. It will be shown in the following that this dictionary has a lot of redundancy. The number of atoms in \mathbb{B} is given by:

$$N(P_{max}) = \sum_{P=1}^{P_{max}} P = \frac{P_{max} \times (P_{max} + 1)}{2} \quad (4.18)$$

We will refer to this dictionary as the Identity dictionary.

Notice that, instead of using the identity matrix to generate a basis for \mathcal{V}_P , one may use any full rank $P \times P$ matrix. In [26], the authors chose the $P \times P$ DFT matrix for this purpose. Doing so reduces the number of atoms in the dictionary significantly, since many of the DFT matrices of different sizes give rise to the same columns. It was shown in [26] that removing the repeating copies gives a dictionary of size:

$$N(P_{max}) = \sum_{P=1}^{P_{max}} \phi(P) = \frac{3P_{max}^2}{\pi^2} + O(P_{max} \log P_{max}) \quad (4.19)$$

This dictionary was named as the Farey dictionary in [26]. The difference between (4.18) and (4.19) goes as $O(P_{max}^2)$ [24]. We will now derive an important new property of the Farey dictionary:

Theorem 4.5.1. Uniqueness of The Farey Dictionary: *A set of complex exponentials⁵ \mathbb{B} will be a periodic dictionary of order P_{max} , if and only if, for each period P in $1 \leq P \leq P_{max}$, \mathbb{B} contains the $\phi(P)$ unique complex exponentials that are periodic with period P , namely $\{e^{j\frac{2\pi k}{P}n} : \gcd(k, P) = 1\}$.* \diamond

Proof. Consider a particular P in $1 \leq P \leq P_{max}$. One can show that among the set of all possible complex exponentials, there are exactly $\phi(P)$ complex exponentials with period P . These are in fact the set $\{e^{j\frac{2\pi k}{P}n} : \gcd(k, P) = 1\}$. Suppose \mathbb{B} does not contain one of these $\phi(P)$ complex exponentials. Then clearly, \mathbb{B} does not have a basis for \mathcal{V}_P since none of the other complex exponentials can span this missing complex exponential.

Moreover, if for each P in $1 \leq P \leq P_{max}$, all the $\phi(P)$ P -periodic complex exponentials are present in \mathbb{B} , then the set of all complex exponentials with periods that are divisors of P are in fact the columns of the $P \times P$ DFT matrix. So they will form a basis for \mathcal{V}_P . $\nabla \nabla \nabla$

⁵ A complex exponential is a signal of the form $x(n) = e^{j\omega n}$

The above result says that the Farey dictionary must necessarily be a subset of any periodic dictionary that consists entirely of complex exponentials. This result is of significance when one tries to connect the Farey dictionary with recent works on sinusoidal frequency estimation, such as [42] and [43]. These works target the problem of estimating the frequencies in a mixture of complex exponentials. Dictionaries consisting of complex exponentials as atoms were used for this purpose, with their frequencies lying on a *uniform* grid over 0 to 2π . Since any periodic signal is a mixture of complex exponentials (Fourier series), in principle, one might consider using such techniques for period estimation. However, as shown in Theorem 4.5.1, a complex-exponential dictionary for periodic signals has to have a nonuniform grid as demonstrated in Fig. 4.1.

In [23], the authors showed that the Farey dictionary is an example of a more general set of dictionaries called the Nested Periodic Dictionaries (NPDs). They are based on the NPMs (Sec. 3.1). Consider an NPM \mathbf{A} with number of columns $= \text{lcm}(1, 2, \dots, P_{max})$. Let \mathbb{B} be the set of signals obtained by periodically extending those columns of \mathbf{A} that have periods in the range 1 to P_{max} . It follows from the properties of NPMs that \mathbb{B} is actually a periodic dictionary of order P_{max} . Such dictionaries were called as Nested Periodic Dictionaries (NPDs). NPDs have the same size as a Farey dictionary. In fact the Farey dictionary is also an NPD, since the DFT matrix is itself an NPM [23]. Notice that the NPDs have exactly $\phi(P)$ signals with period P for every P in the range $1 \leq P \leq P_{max}$ by construction. An example, the Natural Basis dictionary for $P_{max} = 4$, is shown below. Note its smaller size compared to (4.17).

$$\mathbb{B} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (4.20)$$

4.6 Fundamental Properties of Periodic Dictionaries

In this section we will show that dictionaries satisfying Definition 4.5.1 automatically satisfy certain fundamental properties. In particular, these properties shall apply to

the dictionaries of [25], [26] and [23].

Minimum Number of Atoms

Let us first look at the number of signals a periodicity dictionary is required to have:

Theorem 4.6.1. Minimum number of atoms needed: *Any periodicity dictionary of order P_{max} must have at least $\sum_{P=1}^{P_{max}} \phi(P)$ number of linearly independent signals.*

◇

Proof. The signals in \mathbb{B} must be able to span the complex exponentials in the Farey dictionary. But the complex exponentials in the Farey dictionary are linearly independent signals. So \mathbb{B} must contain at least as many linearly independent signals as in the Farey dictionary, namely, $\sum_{P=1}^{P_{max}} \phi(P)$ signals. ▽ ▽ ▽

Since the Nested Periodic Dictionaries (NPDs) satisfy Definition 4.5.1, and furthermore have $\sum_{P=1}^{P_{max}} \phi(P)$ atoms, it follows from Theorem 4.6.1 that the NPDs are examples of minimum size dictionaries for periodic signals.

The Basis Inclusive Property

In [25], [26] and [23], the dictionaries have a special feature. For every \mathcal{V}_P , there is a subset of the dictionary that is a basis for \mathcal{V}_P . We shall refer to this as the *basis inclusive property*. Is this a necessary requirement for a periodic dictionary? The answer is no. For example, consider the Farey dictionary. The basis for \mathcal{V}_1 is the signal:

$$\begin{bmatrix} u(n) \end{bmatrix} = \left[\dots \ 1, \ 1, \ 1, \ 1, \ 1, \ 1, \ \dots \right] \quad (4.21)$$

while the basis for \mathcal{V}_2 are the signals:

$$\begin{bmatrix} u(n) \\ v(n) \end{bmatrix} = \begin{bmatrix} \dots \ 1, \ 1, \ 1, \ 1, \ 1, \ 1, \ \dots \\ \dots \ 1, \ -1, \ 1, \ -1, \ 1, \ -1, \ \dots \end{bmatrix} \quad (4.22)$$

As discussed earlier, the Farey dictionary is a linearly independent set. Replacing a vector in a linearly independent set by its sum with another element in the same set doesn't change the linear independence or the span of that set. So, if we replace $u(n)$ by $u(n) + v(n)$ in the Farey dictionary, we will still have a valid periodic dictionary. But $u(n) + v(n)$ is the vector:

$$\left[\dots \ 2, \ 0, \ 2, \ 0, \ 2, \ 0, \ \dots \right] \quad (4.23)$$

And so this new periodic dictionary no longer has any basis for \mathcal{V}_1 . This illustrates that we can have a periodic dictionary that does not contain a basis for every \mathcal{V}_P , in $1 \leq P \leq P_{max}$.

Nevertheless, the basis inclusive property is very useful for support based period estimation. It ensures that all period P signals can be spanned by the same subset of atoms in the dictionary. If the dictionary does not contain a basis for \mathcal{V}_P , then different period P signals might require different supports in the dictionary, making support based period estimation difficult. The following result shows an interesting connection between the basis inclusive property and the Euler structure:

Theorem 4.6.2. Basis inclusivity and the Euler structure: *Let \mathbb{B} be a periodicity dictionary of order P_{max} satisfying the basis inclusive property. Then for each P in $1 \leq P \leq P_{max}$, \mathbb{B} must contain at least $\phi(P)$ linearly independent signals with period P .* \diamond

Proof. Consider a P such that $1 \leq P \leq P_{max}$. A basis for \mathcal{V}_P would consist of signals with periods that are divisors of P , since each basis vector must belong to \mathcal{V}_P . The rank of the set of all signals in \mathbb{B} that have proper divisors of P as periods must be $\leq \sum_{\substack{d|P \\ d < P}} \phi(d)$ (follows from Theorem 4.10.1 in Appendix 4.10). So for \mathbb{B} to contain a basis for the P dimensional space \mathcal{V}_P , there must be at least $P - \sum_{\substack{d|P \\ d < P}} \phi(d) = \phi(P)$ linearly independent signals in \mathbb{B} with period P , since $\sum_{d|P} \phi(d) = P$ for any integer $P > 0$ [24]. $\nabla \nabla \nabla$

The LCM property

Consider the identity dictionary shown in (4.17). Let \mathbf{c}_2 and \mathbf{c}_3 be the second and third columns of this dictionary. Consider the linear combination $\mathbf{x} = \alpha_1 \mathbf{c}_2 + \alpha_2 \mathbf{c}_3$. Clearly, \mathbf{x} will have period 1 if $\alpha_1 = \alpha_2$, and period 2 otherwise. This illustrates that one cannot determine the period of \mathbf{x} just by looking at its support on this dictionary. One needs to know the values of the coefficients α_1 and α_2 as well. Even then, in general, there is no simple mapping that relates the coefficients and the support of a signal to its period.

However, if one uses a Nested Periodic Dictionary (NPD) [23], then there is an easy mapping from the support of a signal to its period. More so, this mapping is independent of the coefficients. For example, if a signal is a linear combination of a set of atoms of an NPD, then its period must be equal to the LCM of the periods of those atoms. This clearly makes support based period estimation very convenient.

Under what conditions does a generic periodic dictionary have the LCM property? Our result in Theorem 4.6.3 shows that any periodic dictionary that satisfies (i) the basis inclusive property and (ii) linear independence of the underlying subspaces corresponding to different periods, will automatically have the LCM property.

Theorem 4.6.3. *Let \mathbb{B} be a periodic dictionary of order P_{max} , satisfying the basis inclusive property. Further, suppose the periodic signals forming the atoms of the dictionary are linearly independent. Then, if $x(n)$ is a linear combination of a subset of atoms in \mathbb{B} , then the period of $x(n)$ has to be the lcm of the periods of those atoms (i.e., it cannot be a proper divisor of this lcm). \diamond*

Proof. From Theorem 4.6.2, we know that \mathbb{B} must contain at least $\phi(P)$ signals with period P for each P in $1 \leq P \leq P_{max}$. But if it contains $> \phi(P)$ signals with period P , then the total number of signals in \mathbb{B} that belong to \mathcal{V}_P would be greater than $\sum_{d|P} \phi(d) = P$. This is against \mathbb{B} being a linearly independent set. Hence, \mathbb{B} will have exactly $\phi(P)$ signals with period P .

Now, let \mathcal{T}_P denote the span of all period P signals in \mathbb{B} . Then, it is easy to verify that $\mathbb{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{P_{max}}\}$ is a set of LIPS (Definition 4.4.1). However, every set of LIPS must necessarily be a set of NPSs (from Theorem 4.4.3). Condition 2 in Theorem 4.2.1 completes the proof. $\nabla \nabla \nabla$

The proof of this result reveals an important corollary. In [23], a Nested Periodic Dictionary was essentially defined as follows: A set of linearly independent signals containing exactly $\phi(P)$ signals with period P for every P . The basis inclusive property and the LCM property for the NPDs were derived as consequences of this definition.

The above proof on the other hand, shows that the basis inclusive property and the linear independence condition make it necessary for the dictionary to have exactly $\phi(P)$ atoms periodic with period P , for every P . This leads to the following corollary.

Corollary 4.6.1. *The NPDs are the only dictionaries that satisfy the basis inclusive property and the linear independence conditions. \diamond*

The Notion of Union of Subspaces

At this point, we would like to explain the phrase “union-of-subspaces” which appears in many of our discussions. A signal with a particular period belongs to a

certain subspace of the column space of a periodic dictionary, defined by a subset of atoms from the dictionary. The problem of identifying the period, or a set of hidden periods can be regarded as the problem of identifying this subspace from the measured data \mathbf{x} (4.16). Before this subspace has been identified, the signal is therefore considered to belong in the union of the various subspaces representing various possible periods. For example, a signal whose period is known to belong to the set $1 \leq P \leq P_{max}$, lies in the union of the subspaces $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_{P_{max}}$. The period identification problem is thus solved by extracting a specific subspace from this union of subspaces to best fit the measured data.

4.7 Effect of Dictionary Redundancy on Period Estimation: A Numerical Example

In this section, we demonstrate the practical importance of less redundancy in periodic dictionaries using an example. We will use periodic dictionaries to estimate the period of a noisy segment of a periodic signal, and also try to recover the original signal. We will use the convex program in Eq. (3.24) for recovering the signal's support. In the following example, we compare two dictionaries:

- The Ramanujan dictionary (an NPD proposed in [23]), which, according to Theorems 4.6.1 and 4.6.2, is a least redundant dictionary, and
- The Identity dictionary shown in (4.17).

A Period Estimation Example

Fig. 4.6(a) shows a randomly generated period 6 signal of length 56 samples. Fig. 4.6(b) shows a noisy version of the same signal (AWGN, SNR = 0dB). We solved (3.24), with \mathbf{x} being the noisy signal vector, $f(P) = P^2$, and $P_{max} = 40$. Let \mathbf{y}_\star denote the optimal solution of (3.24). Fig. 4.7 show the strength vs period plots, which, for every period, show the sum of the squares of the entries in \mathbf{y}_\star that correspond to columns of \mathbf{A} with that particular period. Fig. 4.7(a) shows the results for the Identity dictionary, while Fig. 4.7(b) shows the same for the Ramanujan dictionary. In Fig. 4.7(b), the prominent peaks are at periods 2, 3 and 6. Using the LCM property of the NPDs, we can conclude that the signal has period 6. The

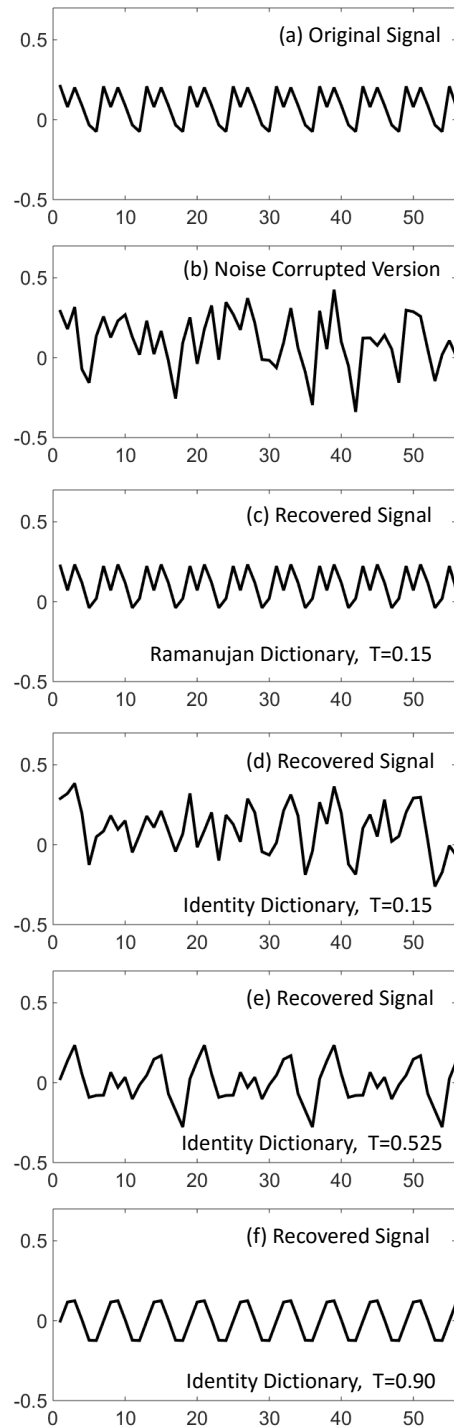


Figure 4.6: **Effect of Redundancy on Signal Recovery:** Part (a) - A randomly generated, noiseless, period 6 signal of length 56 samples. Part (b) - A noisy version of the signal in Part (a) (AWGN with SNR = 0dB). Part (c) - Reconstructed signal using the Ramanujan dictionary, with the threshold factor $T = 0.15$. Parts (d), (e) and (f) - Reconstructed signals using the Identity dictionary, with the threshold factor T being 0.15, 0.525 and 0.9 respectively. Please see Sec. 4.7 for the corresponding reconstruction errors and more details.

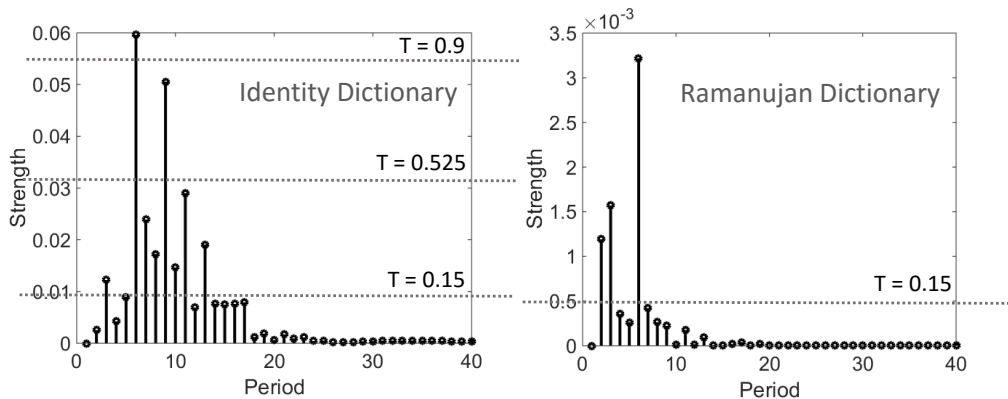


Figure 4.7: Parts (a) and (b) - Strength vs Period plots using the Identity and Ramanujan Dictionaries respectively for the signal shown in Fig. 4.6(b), after solving (3.24). The dotted lines indicate the thresholding parameter T discussed in Sec. 4.7.

Identity dictionary on the other hand has a strong peak at period 6, but there are several spurious peaks, such as at periods 9 and 11, and cannot therefore identify the period correctly. This is due to the high amount of linear dependence between the underlying subspaces corresponding to different periods in this dictionary. For instance, the period 6 and period 9 subspaces in the Identity dictionary, namely \mathcal{V}_6 and \mathcal{V}_9 , have an entire subspace of dimension 3 in common. Even if the Identity dictionary had shown the same periods as the NPDs, one cannot justify the use of the LCM property in this case, as explained in Sec. 4.6.

A Signal Recovery and Denoising Example

A simple way to do signal denoising using a periodic dictionary is to replace the solution \mathbf{y}_\star with a thresholded version, and reconstruct a cleaner version of \mathbf{x} . Let E_0 be the height of the largest peak in the strength vs period plot. Let $\{P_1, P_2, \dots, P_N\}$ be the periods whose strength is more than $T \times E_0$ in the strength vs period plot, where T is a threshold parameter. We may now retain in \mathbf{y}_\star only those entries, which correspond to periods $\{P_1, P_2, \dots, P_N\}$, and set everything else to 0. Let the resulting vector be denoted as \mathbf{y}_{th} . Then, the recovered signal is:

$$\hat{\mathbf{x}} = A \times \mathbf{y}_{th} \quad (4.24)$$

Fig. 4.6(c) shows the recovered signal using the Ramanujan dictionary. Notice that it looks very similar to the original noiseless signal in Fig. 4.6(a). The signal to reconstruction error ratio was found to be 10dB. The threshold T was chosen as 0.15, which, through numerous simulations, was found to perform well for most

examples using this dictionary. This threshold is shown on the strength vs period plot in Fig. 4.7.

Fig. 4.6(d) shows the recovered signal using the Identity dictionary, and $T = 0.15$. Notice from Fig. 4.7 that this threshold chooses a number of incorrect periods due to the several spurious peaks present in the plot. Accordingly, the recovered signal in Fig. 4.6(d) does not resemble the original signal. The signal to reconstruction error ratio in this case was 0.8 dB. Increasing T to around 0.5 does not help much, and results in the recovered signal shown in Fig. 4.6(e). The resulting reconstruction error has an SNR = 0.2 dB. Choosing a threshold of 0.9 would cause only the period 6 component to be chosen in Fig. 4.7. However, even in this case, the recovered signal (shown in Fig. 4.6(f)) does not resemble the original signal well enough (Recovery error SNR = 1.7 dB). This is because, the period 6 component of the original signal must have been spread over several other subspaces such as \mathcal{V}_9 , that are linearly dependent with \mathcal{V}_6 .

This example clearly demonstrates that having linearly independent subspaces, and less redundancy in the union-of-subspaces model, is beneficial for period estimation and signal recovery.

4.8 The Case of Mixtures of Periodic Signals

Before concluding this chapter, we would like to point out that all the results in this chapter also apply to the case of mixtures of periodic signals. For example, suppose we had a mixture of periodic signals, where the period of each component lies in the range $1 \leq P \leq P_{max}$. Such a mixture may not even “look” periodic if the data length is smaller than the lcm of the hidden periods. References [13], [14], [18], [23], [25] show examples of estimating the periods of the components in such mixtures using projection based techniques and dictionaries.

In this context, if one were to ask, what is the minimum number of atoms required in a dictionary to span such periodic signal mixtures, then the answer is again given by Theorem 4.6.1. The basis inclusive property, described in Sec. 4.6, is once again important to be able to infer the component periods in a signal from its support on the dictionary. Without it, different periodic signal mixtures that have the same component periods, might require very diverse supports, making support based period estimation difficult. Imposing the basis inclusive property on the dictionaries results in the Euler structure, as proved in Theorem 4.6.2.

Sec. 4.6 showed that the LCM property of periodic dictionaries is a consequence

of (i) the basis inclusive property, and (ii) linear independence of the underlying periodic subspaces. In the context of estimating the component periods in a mixture of periodic signals, one needs a slight generalization of the LCM property. To see this, let us consider a simple example. If a signal was a mixture of period 3 and period 8 signals, then such a signal would have peaks at periods 1, 2, 3, 4, and 8 in the strength vs period plot, using an NPD [23]. Directly using the LCM property would reveal that the periodic signal mixture has a net period of 24, which is in fact correct. However, the set $\{1, 2, 3, 4, 8\}$ can be divided into two subsets $\{1, 3\}$ and $\{1, 2, 4, 8\}$, and by using the LCM property on each of these subsets, we can infer that the signal is actually a sum of period 3 and period 8 signals. Although this process is intuitive, a careful formulation of the estimation of hidden periods shall be presented in an upcoming work, since it requires a much deeper discussion.

The importance of least redundant dictionaries becomes magnified in the context of mixtures of periodic signals. This is because, in this case, the strength vs period plots inherently have more peaks even when there is no noise (due to a larger number of periodic components). The presence of spurious peaks due to linear dependence between the underlying subspace, such as those in Fig. 4.7(a), would make the identification of hidden periods very complex.

4.9 Conclusion

This chapter derived several fundamental properties of union-of-subspaces models for periodic signals. Previously, several specific examples of such subspaces were proposed in various works, and their properties had been analyzed largely in an independent fashion. This chapter brings together all such subspace models under one unifying framework, deriving not only some important interrelationships between them, but also a number of new properties that apply generically to any union-of-subspaces model for periodic signals. A number of results relating to various aspects of such subspace models, such as unique periodic decompositions, linear independence of the subspaces, minimum dimensionality, the role of the Euler totient function, the origin of the LCM property etc. were derived. In the context of dictionaries, a similar unifying analysis was shown to reveal several interesting properties such as the minimum required size of a dictionary, the minimum number of atoms required with each period, sufficient conditions for period estimation from support information, and so on. Finally, the importance of reducing redundancy in the subspaces was demonstrated using a signal recovery example.

4.10 Chapter Appendix

Computing $\langle x, y \rangle$ in practice

Notice that (4.4) seems to require us to know the values of $x(n)$ and $y(n)$ for all time indices n . However, if the periods of $x(n)$ and $y(n)$ are known a priori, say P_x and P_y respectively, then (4.4) reduces to:

$$\langle x, y \rangle = \frac{1}{K} \sum_{n=0}^{K-1} x(n)y^*(n) \quad (4.25)$$

where $K = lcm(P_x, P_y)$. Generally, in many practical applications where $\langle x, y \rangle$ is to be computed, the period of only one of $x(n)$ or $y(n)$ is known a priori. For example, $x(n)$ could be a signal whose period is to be estimated, while $y(n)$ is a basis vector of some periodic subspace with a known period P_y . Further, $x(n)$ is typically available only over a finite data length, say L samples (where L need not be a common multiple of P_x and P_y). In that case, (4.4) is approximated by the following in [12]–[14]⁶:

$$\langle x, y \rangle = \frac{1}{L'} \sum_{n=0}^{L'-1} x(n)y^*(n) \quad (4.26)$$

where $L' = P_y \lfloor \frac{L}{P_y} \rfloor$. It was observed in [13] and [14] that (4.26) is a good approximation to (4.4) when the data length is sufficiently large.

The dictionary based techniques proposed in [23] and [25] do not need to compute any inner products between periodic signals. So an approximation of (4.4) is not needed in their formulation.

The Nested Basis Property of NPMs

The following theorem was used in proving Theorem 4.4.1, to show that the NPMs are the only square matrices that satisfy the Nested Basis Property. It was also used in Theorem 4.6.2 to show that the Euler structure is fundamental to periodic dictionaries.

Theorem 4.10.1. *Let \mathbb{B} be a set of periodic signals whose periods are proper divisors of P . Then the maximum number of linearly independent signals in \mathbb{B} is $= P - \phi(P)$. \diamond*

Proof. Consider any $P \times P$ Nested Periodic Matrix \mathbf{A} . The columns of \mathbf{A} that have proper divisors of P as periods, when periodically extended, must be able to span the

⁶A slight modification of (4.26), namely $\langle x, y \rangle = \frac{1}{L} \sum_{n=0}^{L-1} x(n)y^*(n)$ is used in [12].

signals in \mathbb{B} . There are $\sum_{\substack{d|P \\ d < P}} \phi(d) = P - \phi(P)$ such (linearly independent) columns of \mathbf{A} . Hence the number of linearly independent signals in \mathbb{B} must be less than or equal to $P - \phi(P)$. $\nabla \nabla \nabla$

Chapter 5

THE RAMANUJAN FILTER BANK AND ITS APPLICATIONS

In the previous chapters, we introduced techniques based on projections and dictionaries using union-of-subspaces models for periodic signals. One of the limitations of such methods is when dealing with time-varying periodic behavior. For example, to use them for signals such as chirps, we will have to break the signal into multiple blocks and apply these methods on each block separately. This can be computationally expensive. Moreover, it would be beneficial to use smaller block lengths for detecting smaller periods and larger block lengths for detecting larger periods to obtain good localization. This cannot be easily done using dictionaries.

In this chapter, we propose an elegant way to address this problem using a new filter bank structure. This filter bank, which we call the Ramanujan Filter Bank (RFB), can be derived in two ways. First, we will derive it in a fundamental manner, starting with the properties of Ramanujan sums. The second approach is a heuristic one, which makes insightful connections with the dictionary based methods of Chapter 3. We will present both of these in the following. The RFB yields a non-uniform tiling of the time vs period plane, similar in spirit to the wavelet tiling of the time frequency plane. But it is fundamentally different from the wavelet tiling as we shall see.

We will illustrate the advantages of the RFB using three different applications: Protein repeats, DNA micro-satellites and Absence seizures.

5.1 The Ramanujan Filter Bank

The RFB is a collection of filters as shown in Fig. 5.1 (a). for every integer $q > 0$, the q^{th} RFB filter has the following impulse response:

$$h_P = \{c_P(0), c_P(1), \dots, c_P(LP - 1)\} \quad (5.1)$$

That is, it is the P^{th} Ramanujan sum $c_P(n)$, truncated to L complete periods for some integer L .

To see how the RFB can be used for period estimation, consider the following theorem:

Theorem 5.1.1. *Any periodic signal $x(n)$ can be expressed as a sum of exponentials*

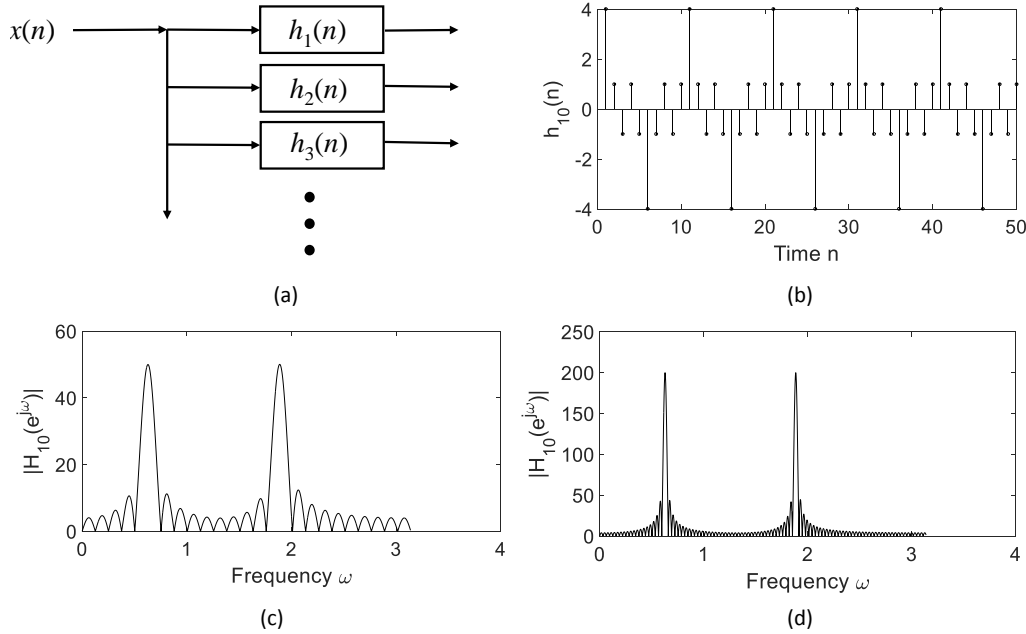


Figure 5.1: Part (a) - Block diagram of the Ramanujan Filter Bank. Part (b) - An example of the impulse response of an RFB filter, $h_{10}(n)$. Part (c) and (d) - Frequency responses of $h_{10}(n)$ for $L = 5$ and $L = 20$ respectively.

in a unique way as

$$x(n) = \sum_{i=1}^K \alpha_i e^{j \frac{2\pi k_i}{q_i} n} \quad \alpha_i \neq 0 \quad (5.2)$$

where k_i and q_i are integers satisfying $(k_i, q_i) = 1 \forall i$. The period of $x(n)$ is exactly equal to $\text{lcm}\{q_i\}$, rather than a proper divisor of it.

Proof. Eq. 5.2 is just the Fourier series expansion of $x(n)$. The essence of the proof is to note that $e^{j \frac{2\pi k_i}{q_i} n}$ belongs to the q_i^{th} Ramanujan subspace when $(k_i, q_i) = 1$, and to then use the LCM property of Ramanujan sums (See Chapter 2 or Theorem 12 of [9]). ▽ ▽ ▽

Suppose $x(n)$ is a signal that is periodic with period P . It can be expressed as a sum of complex exponentials in the form of (5.2) in a unique way through its Fourier series expansion:

$$x(n) = \sum_{k=1}^P \alpha_k e^{j \frac{2\pi k}{P} n}, \quad \alpha_k = \frac{1}{P} \sum_{n=1}^P x(n) e^{-j \frac{2\pi k}{P} n} \quad (5.3)$$

by reducing each k/P to its lowest form. From Theorem 5.1.1 and (5.3), to estimate the period of $x(n)$ we need to find among the set of all frequencies of the form:

$$\left\{ \frac{2\pi k_i}{q_i} : (k_i, q_i) = 1 \right\} \quad (5.4)$$

the ones at which the signal's spectrum has non-zero energy. We can then take the lcm of the periods q_i of those complex exponentials as an estimate for the signal's period.

This is exactly what is happening in the RFB as the length of the filters tends to infinity. The spectrum of the q_i^{th} Ramanujan filter with impulse response $c_{q_i}(n)$ is non-zero only at the frequencies shown in (5.4) (Fig. 5.1 (c) and (d)). So its output will be non-zero if and only if $x(n)$'s decomposition into the form (5.2) has a q_i periodic exponential. So taking the lcm of the indices of those Ramanujan filters that have non-zero outputs (the lcm property) is indeed a valid estimate for the period of the signal.

In practice though, we cannot use infinitely long filters since they would provide no localization for time varying periodicity. Instead, we truncate each Ramanujan filter to L repeats. In the frequency domain, this tends to spread out the spectrum a little as shown in Fig. 5.1 (c) and (d). A large L will yield more accurate period estimates, while a smaller L will result in better time localization. This is a fundamental trade-off that one has to deal with when using FIR Ramanujan filters.

We will now show how to derive the RFB using the dictionaries of Chapter 3.

5.2 Connections to Dictionaries

Among the different choices of periodicity matrices for constructing the dictionary, we specifically consider the Ramanujan dictionary. We consider the l_2 norm based convex program in (3.24). Looking more closely at the form of its leftinverse in (3.26),

$$\mathbf{P} = \mathbf{D}^{-2} \mathbf{A}^T (\mathbf{A} \mathbf{D}^{-2} \mathbf{A}^T)^{-1} \quad (5.5)$$

we noticed that its rows have an interesting pattern. Similar to (the transpose of) the dictionary itself, the rows of the left-inverse seem to be periodic, with exactly $\phi(P)$ rows with period P . For instance, Fig. 5.2 shows a section of the first 50 rows of a left-inverse matrix obtained from a Ramanujan dictionary with parameters $N = 200$ and $P_{max} = 200$. Clearly, the first row has period 1, the second has period 2, the third and fourth have period 3 and so on. Moreover, for many of the periods, the $\phi(P)$

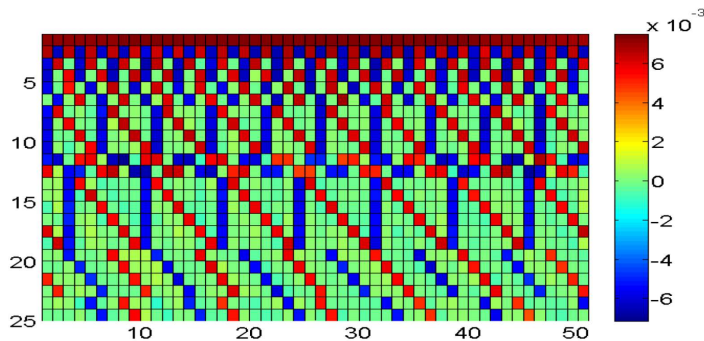


Figure 5.2: The first 50 rows of the Pseudo-inverse matrix in (3.26) for $N = 200$ and $P_{max} = 200$.

rows corresponding to them are approximately shifted versions of each other. Such a strong pattern raises the question if we could directly design suitable ‘leftinverses’ with the same structure without having to formulate it as an optimization problem like (3.24). For instance, we experimentally observed that the following expression instead of (3.26) gives equally good results (see Fig 5.3):

$$\mathbf{y}_\star = \mathbf{D}^{-1} \mathbf{A}^T \mathbf{x} \quad (5.6)$$

Using has an advantage. We can implement it efficiently using a filter bank. To see this, consider the case when the input data is of infinite length (streaming data). One way to process the signal in that case is to apply on successive input blocks of length N , each shifted by one sample. When we use the Ramanujan dictionary, for every period P we are taking the inner product of the input blocks with the following $\phi(P)$ vectors (assuming that N , the data length, is reasonably larger than the signal’s period P):

$$\mathbf{C}_P^{(0)} = \begin{bmatrix} c_P(0) & c_P(1) & \dots & c_P(N-1) \end{bmatrix} \quad (5.7)$$

$$\mathbf{C}_P^{(1)} = \begin{bmatrix} c_P(1) & c_P(2) & \dots & c_P(N) \end{bmatrix} \quad (5.8)$$

and so on till $\mathbf{C}_P^{(\phi(P)-1)}$, where $c_P(n)$ represents the P^{th} Ramanujan sum. After dividing these inner products by the penalty function $f(P)$, we sum their squares to

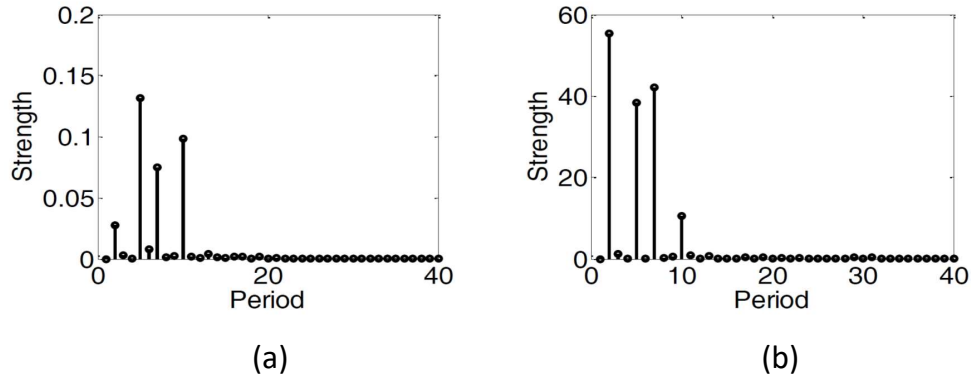


Figure 5.3: Parts (a) and (b)- Strength vs period plots for a period 70 signal that was generated as a sum of period 7 and period 10 signals. Part (a) shows the plot obtained by using (3.26), and Part (b) shows the plot obtained by using (5.2).

define the strength of that period in Fig. 5.2 (b):

$$y_P = \sum_{i=0}^{\phi(P)-1} \left| \frac{\langle \mathbf{x}, \mathbf{C}_P^{(i)} \rangle}{f(P)} \right|^2 \quad (5.9)$$

Notice that the $\mathbf{C}_P^{(i)}$'s are periodic vectors with period P . Moreover, they are nearly shifted versions of the same vector $\mathbf{C}_P^{(0)}$. Based on this observation, we propose a slightly modified implementation for (5.9):

$$y_P(n) = \sum_{i=n-\phi(P)+1}^n \left| \frac{x * h_P(i)}{f(P)} \right|^2 \quad (5.10)$$

where $x * h_P$ denotes convolution and h_P is a filter with impulse response:

$$h_P = \{c_P(0) c_P(-1) \dots c_P(-LP + 1)\} \quad (5.11)$$

for some integer L . That is, h_P consists of L consecutive repeats of the P^{th} Ramanujan sum. Due to symmetry of the Ramanujan sums, h_P can be re-written as:

$$h_P = \{c_P(0) c_P(1) \dots c_P(LP - 1)\} \quad (5.12)$$

Choosing the filter length as LP instead of a fixed N as in (5.2) effectively enables us to detect smaller periods using smaller blocks and larger periods using larger blocks. A collection of such filters for all periods going from 1 to P_{max} , as shown in

Fig. 5.1, is nothing but what we called as the Ramanujan Filter Bank (RFB) in the previous section . A plot of the outputs $y_P(n)$ as a function of P and n will be called as the **RFB time vs period plane**.

5.3 Simulations

We will start with two examples here. Fig. 5.4 (a) shows the time vs period plane for a length 668 signal that has a randomly generated period 3 component between samples 201 and 218 and a sum of randomly generated period 15 and period 11 components from samples 319 to 469. The sum of the period 15 and period 11 signals is actually a signal with period $15 \times 11 = 165$. $L = 15$ and $f(P) = P^2$ were chosen for the RFB. In part (a), the localized period 3 component is detected initially. This is followed by periods 3, 5, 11 and 15 showing up, and using the lcm method, we can conclude that the signal exhibits a periodicity of 165 and is a sum of period 15 and period 11 components. The period 1 DC component is not shown. The outputs of all the filters beyond period 50 were 0 and hence not shown. Note that the q^{th} Ramanujan filter's output is delayed by $qL/2$ due to the causal implementation (5.12). This causes different divisors of 15 to be detected with different delays. To avoid this, part (b) was obtained from part (a) by advancing the output of each Ramanujan filter h_q by $\lfloor qL/2 \rfloor$ so that all the divisors of a particular period are expressed concurrently.

Parts (c) and (d) show the time vs frequency plane using STFT (assuming 1 Hz sampling rate). In part (c), we had to use a rectangular window of size 128 to reasonably identify the period 11 and 15 components. The peaks in the spectrogram correspond to periods 15.06, 11.13, 7.53, 5.56, 5.02, 3.82, 3.66, 3.01, 2.75, 2.51, 2.21 and 2.13. These numbers roughly correspond to 11, 15 and their harmonics. But this window was too wide to detect the period 3 component present between samples 201 and 218. So in part (d), a window of size 32 was chosen. Although the localized period 3 component gets detected well, this window is not sufficient to identify the period 11 and 15 components. We do not have to worry about having different analysis for different periods in the RFB since the length of the each filter was chosen proportional to its period. Moreover, if the periodic signal is a superposition of a number of signals with smaller periods such as the 11 and 15 case, then using the lcm method of the RFB might be more convenient than searching for fundamental frequencies in spectrograms.

In the second example, we consider the inverse chirp signal $x(t) = \sin(1/at)$ in the

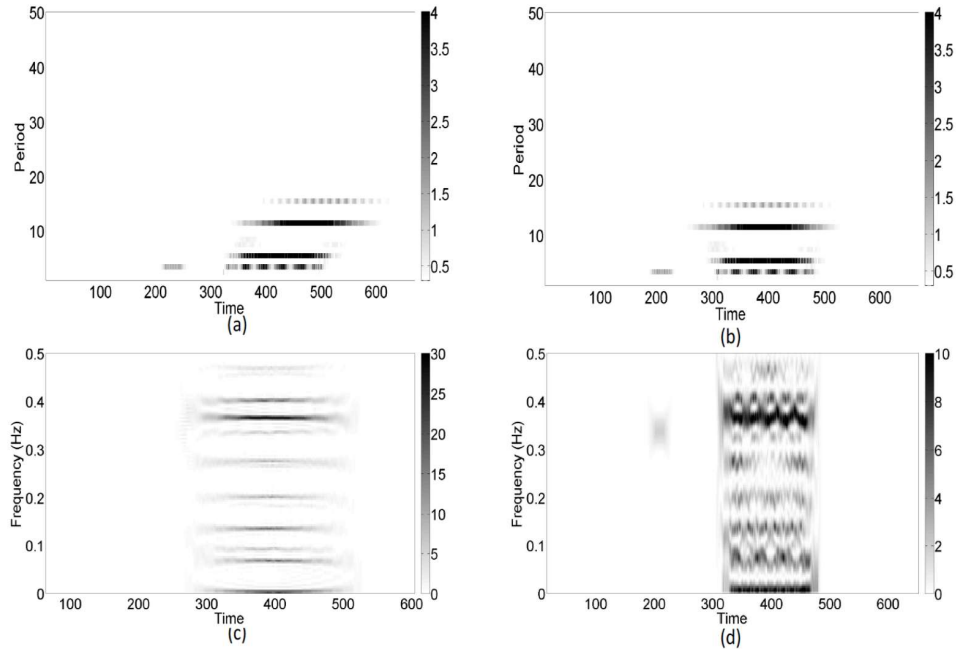


Figure 5.4: Parts (a) and (b) - The time vs. period plane for a signal exhibiting localized periodicities using RFB and shifted RFB. Parts (c) and (d) show the time-frequency plane using STFT with window sizes 128 and 32 respectively. Refer text for details.

interval $t \in [2, 10]$ seconds, with $a = 0.01/2\pi$, sampled every 0.01s (Fig. 5.5 (a)). The instantaneous period of this signal is $2\pi at^2$. This quadratic behavior is evident in the time vs period plane in part (b) ($L = 5$). Part (d) shows the time frequency plane obtained from STFT using a length 32 rectangular window. It captures the small periods well as shown in Table 1. But the larger periods are mis-estimated. When the frequency is very small, the finite frequency resolution of STFT limits the accuracy of the $P = 1/f$ estimate. If we increase the window size to 256 to better estimate the higher periods, as in Fig. 5.5(c), the smaller periods are smeared out in the time frequency plane. The estimate for larger periods is still not very accurate (Table 5.1). The RFB on the other hand offers good estimates for both small and large periods. $f(P)$ was chosen as $(\phi(P))^2$ here to show that a wide choice is available for its selection.

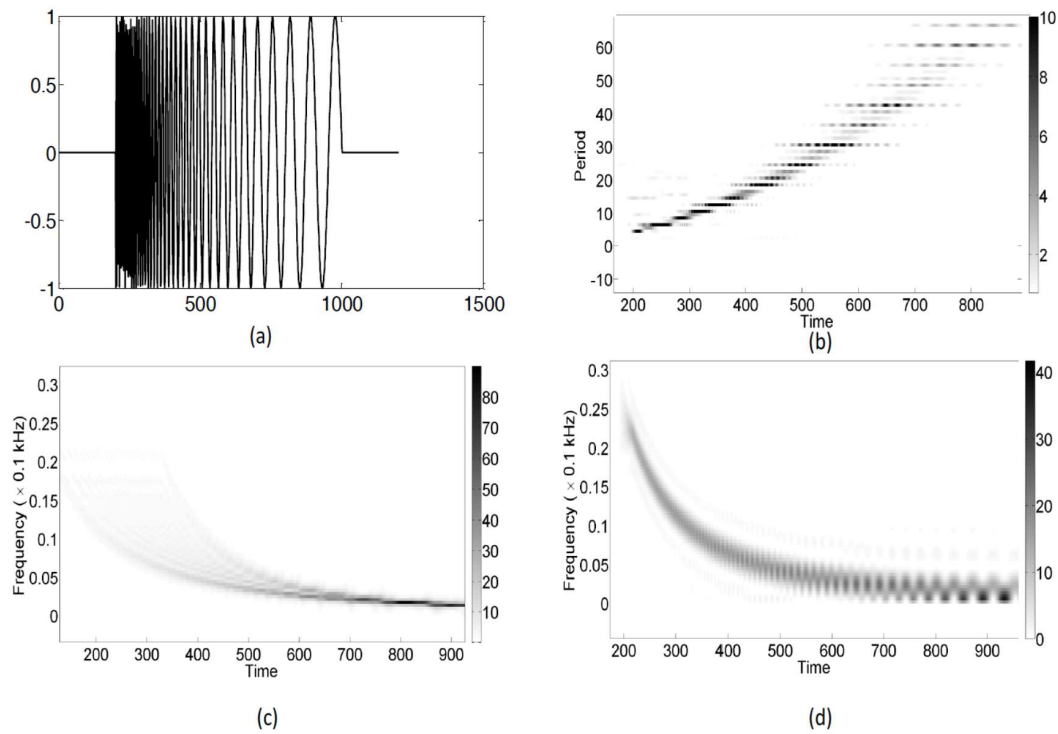


Figure 5.5: Part (a) - Sampled inverse chirp signal. Part (b) - The time vs. period plane using shifted RFB. Parts (c) and (d) show the timefrequency plane using STFT with window sizes 256 and 32 respectively. Refer text for details.

Table 5.1: Period Estimation using STFT and the RFB.

t	Instantaneous Period	STFT (32)	STFT (256)	RFB
2.1 s	44.1 ms	44.9 ms	71.1 ms	40 ms
7.5 s	562 ms	639 ms	512 ms	560 ms

We will now present some real world applications of the RFB.

5.4 Protein Repeats

Proteins are essentially sequences of amino acids. Tandem repeats in proteins are segments within the amino acid sequence that exhibit local periodicity. For example, the sequence $\dots MWACFACFACSY \dots$ has 2 complete, and a partial cycle of the repeat ACF .

Such repeats induce several important structural and binding properties on proteins. For instance, they manifest as characteristic periodic patterns in 3D (for e.g., see Fig. 5.6), offering important cues for the prediction of the protein's structure from its sequence. Functionally, they lead to an enlargement of the binding surface area in the protein. This results in enhanced evolutionary prospects, with some studies even showing a correlation between the complexity of a cell's functions and the number of repeats in its proteins [44]. Tandem repeats have also been known to admit a much higher mutation rate than the background regions. Such mutations have been associated with several diseases. For example, a naturally occurring mutation between glutamine and lysine at position 703 in the ankyrin repeats of ANKK1, causes addictive behaviors to alcohol, nicotine and so on [45]. A different member the same ankyrin family, p16, is known to act as a tumor suppressor [46]. Protein repeats play an important role in several other diverse contexts as well, such as in intracellular transport, DNA repair, cell adhesion, initiating plant responses to UV light, nitrogen fixation and so on [44], [47].

Detecting such repeats from amino acid sequences is not easy in practice, due to (a) high mutation rate in the form of substitutions and insertion-deletions, and (b) small number of repeating copies. Several techniques have been proposed for this problem in the past. They can broadly be divided into the following categories:

- Those based on comparing a sequence with itself using alignment algorithms (trace matrices and dynamic programming). Examples include RADAR [48], TRUST [49], REPwin [50], and so on.
- Hidden Markov Models (HMMs) trained on a set of known repeats. E.g., Pfam [51], SMART [52], etc.
- Classical signal processing techniques such as Short Time Fourier Transform (STFT) [50] and wavelets [53].

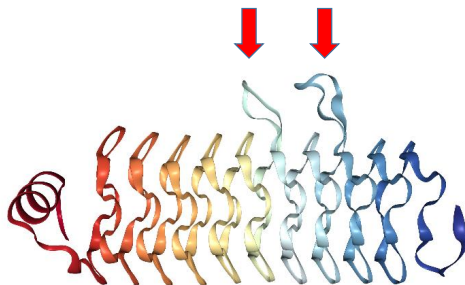


Figure 5.6: HetL: A protein with 40 tandemly repeating pentapeptide repeats. The red arrows indicate two insertion loops that can be predicted using the RFB. See Sec. 5.4 for details. (Image source: [47], [54]–[56])

With the current exponential growth of genomic and proteomic databases¹, algorithmic complexity is increasingly becoming an important parameter to evaluate these techniques. In this regard, although the alignment based methods offer good detection performances, they incur very high computational complexities. For instance, dynamic programming based alignment algorithms such as by Smith and Waterman [57] or Gotoh [58] require $O(N^3)$ and $O(N^2)$ complexities respectively, where N is the length of the protein molecules. The HMM based methods, although performing well for detecting known repeats, are not suited to discover new ones. On the other hand, traditional signal processing techniques such as STFT and wavelets, while offering low computational complexities, have not been as popular. This is due to their relatively poor performance in the presence of mutations and small repeat lengths.

In this section, it will be shown that the RFB performs remarkably well in the context of protein repeats, by comparing its performance with some of the most popular techniques from the literature.

Using the RFB for Detecting Protein Repeats

While the RFB can find periodic segments in sequences of numbers, proteins are sequences of amino acids. In order to apply the RFB to proteins, we need a way to map amino acids to numbers. An arbitrary mapping, in general, may not be useful in this regard. For instance, consider the following sequence:

$$ADLYR \ AILRE \ ADLTG \ AKLVK \quad (5.13)$$

¹RCSB PDB statistics: www.rcsb.org/pdbstatic.do?p=general_information_pdb_statisticsindex.html

Table 1: Kyte Doolittle Scale

A.A.	K.D.	A.A.	K.D.
I	4.5	S	-0.8
V	4.2	Y	-1.3
L	3.8	P	-1.6
F	2.8	H	-3.2
C	2.5	E	-3.5
M	1.9	Q	-3.5
A	1.8	D	-3.5
G	-0.4	N	-3.5
T	-0.7	K	-3.9
W	-0.9	R	-4.5

A.A. = Amino Acid

K.D. = Kyte Doolittle Score

This is a segment from the protein HetL (Fig. 5.6), which is known to contain tandem period 5 repeats [47]. However, looking at the above sequence, it is difficult to see any similarity between adjacent segments of length 5. This example illustrates that we need a notion of ‘distance’ between the amino acids, so that sequences such as the above can be modeled as periodic segments corrupted by noise (which in this case, is substitution and insertion-deletion noise). In prior works that studied the relationships between protein sequences and their structure, several different mappings have been used for this purpose. These include mappings based on the molecular volume of the amino acids [59], their relative composition (namely, the relative weight of carbon to non-carbon atoms) [59], their relative accessible surface area (rASA) to a solvent [60], electron ion interaction potential [61], relative hydrophobicity scores [62] and so on.

In this work, we use the Kyte-Doolittle (KD) hydrophobicity scale, which was proposed in [62] by combining several experimental metrics that measure the average hydrophobicity of an amino acid across several proteins. Each amino acid was qualitatively assigned a number between 4.5 and -4.5 as shown in Table 1, with 4.5 being the most hydrophobic, and -4.5 being the most hydrophilic amino acids. The hydrophobicity of an amino acid is useful in structural analysis of proteins since it

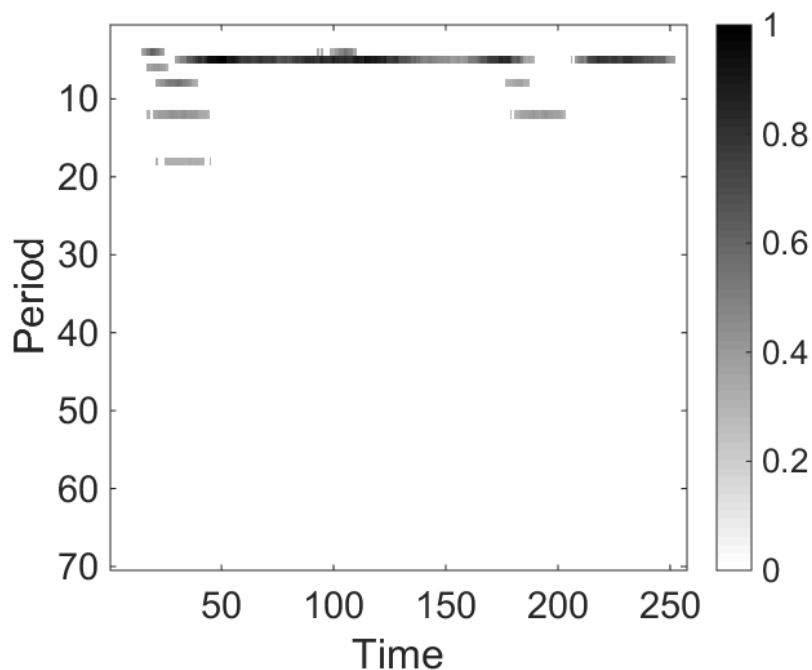


Figure 5.7: The period vs location plot produced using the RFB for HetL from Fig. 5.10, indicating the period 5 repeats. See Sec. 5.4 for details.

plays a significant role in determining the shape into which a protein molecule folds in solvents. From a heuristic perspective as well, we observed in our simulations that the KD scale does result in good location vs period planes from the RFB compared to the other mappings.

Simulation Results

In this section, we will show some examples of protein repeats that were detected using the RFB. Fig. 5.10 shows a set of representative examples from seven well known repeat families. The performance of the RFB is compared with the following four popular techniques mentioned in the previous section: (i) RADAR [48] (ii) REPwin [50] (iii) the STFT based FTwin of [50] and (iv) the wavelet technique of [53]. A red cross indicates that the repeat was not detected by the corresponding algorithm, and a green tick indicates that the repeat was detected and the period was estimated accurately. In all these examples, the RFB could successfully identify the repeats. $L = 5$, and a threshold parameter of 0.3 for the period vs location plane were used for the RFB. That is, all the outputs on the period vs location plane that were less than 30% of the largest peak on that plane were thresholded to 0.

Apart from detecting the repeats, the RFB can provide useful information about the

internal 3D structure in the repeats. This is illustrated by the following example.

Example 1: Pentapeptide repeats in HetL

HetL (PDB:3du1 [47]) is a 237 amino acids protein containing 40 tandem pentapeptide repeats (Fig. 5.10). This protein was known to play a role in the nitrogen fixation process in cyanobacteria, but its exact mechanism in this process was largely unknown. As a first step in understanding this, an X-ray crystallography analysis of the molecule was done in [47]. The crystal structure revealed the following details (Fig. 5.6):

- A 10-residue long α -helix between locations 3 and 12 at the N-terminus,
- Ten periodic coils between locations 15 and 227, consisting of the period 5 pentapeptide repeats,
- A protruding 6-residue long insertion loop between locations 129 and 134,
- A second 9-residue long insertion loop between locations 174 and 182,
- A two-stranded anti-parallel β -sheet between locations 228 and 237 at the C-terminus.

Converting this protein into numbers using the KD scale and using the RFB, we obtained the period vs location plot shown in Fig. 5.7. This plot shows a strong period 5 component, indicative of the pentapeptide repeats.

The two discontinuities in the pentapeptide repeats, in the form of the insertion loops in the crystal structure (Fig. 5.6), can in fact be predicted directly from the RFB's period vs location plot. For instance, Fig. 5.8 shows the output of the fifth RFB filter, which revealed the pentapeptide repeats. The output power of this filter reduces significantly at the locations where the insertion loops occur (see Fig. 5.8 for details).

Example 2: Four bladed β -Propeller in Rabbit Serum Haemopexin

We now consider an example of a repeat with a particularly high amount of noise. The four bladed β -propeller shown in Fig. 5.10 is from the protein rabbit serum haemopexin (PDB: 1hxn [63]). Typically, each blade consists of an α -helix, followed by 4 β -strands. However, due to several insertions of alpha helices and other loops, it is a particularly difficult repeat to detect by most algorithms. The four blades

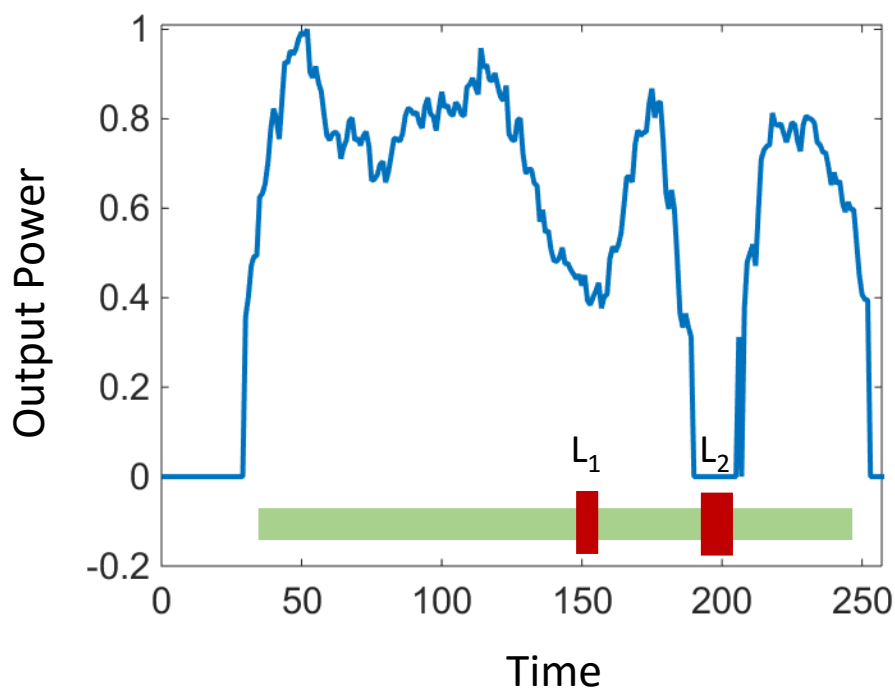


Figure 5.8: The output power of the fifth Ramanujan filter for the example of HetL. (Its time-period plane using the RFB is shown in Fig. 5.7.) The region in green above, marks the residues constituting the pentapeptide repeats, as revealed by the crystal structure of this protein [47]. The red regions, L_1 and L_2 , are insertion loops of sizes 6 and 9 residues respectively. These loops can be seen in the crystal structure of this protein shown in Fig. 5.10. Notice that, by observing the valleys in the output of the fifth Ramanujan filter, one can predict the existence of these insertion loops.

correspond to lengths 44, 48, 52 and 57 residues approximately. RADAR, a popular technique, identifies only the first and third blades, reporting a period of 51. REPwin and FTwin were unable to detect any repeats. The time-period plane of the RFB (shown in Fig. 5.9) not only detects the repeats, but also estimates that the average period of the blades increases from 44 to 54 through the repeats. The other smaller periods that show up in Fig. 5.9 are not errors, but rather show (i) the finer periodic structure within each blade of the propeller, and (ii) the harmonics of the period 44-54 repeats. In all our examples, we observed that the largest periods that show up on the time-period plane of the RFB correspond to the most prominent repeats. The smaller periods correspond to the finer internal structures within each repeat, and the harmonics.

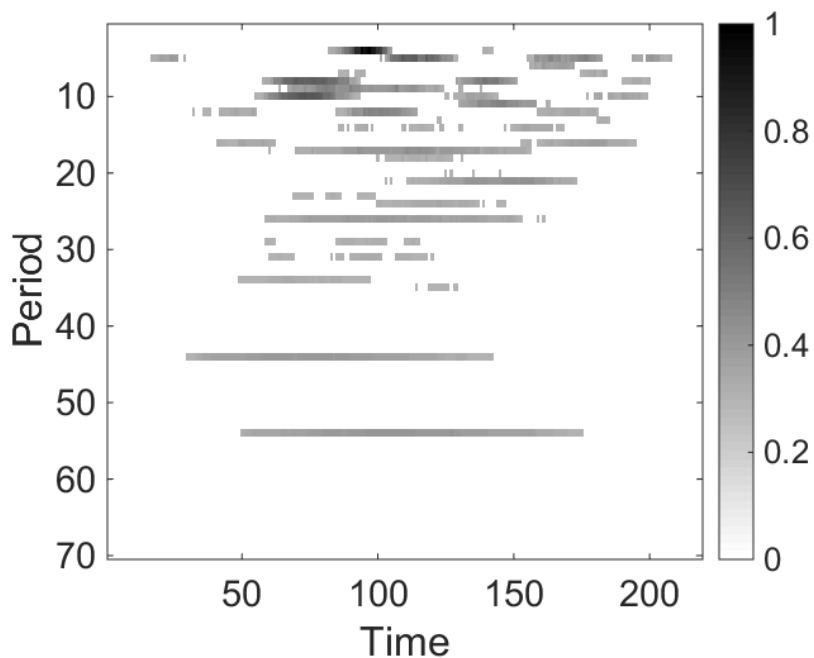


Figure 5.9: The time-period plot produced using an RFB with $K = 5$, for the β -propeller PDB:1hxn shown in Fig. 5.10. See Sec. 5.4 for details.

5.5 Detecting Tandem Repeats in DNA

A tandem repeat is a segment of a DNA sequence that is repeated successively in a periodic manner. For instance, consider the following sequence, where the emphasized segment has 4 copies of the base pairs *TAG* repeating in a tandem fashion:

... *AGCTAGTAGTAGTAGCAAT* ...

Such repeats in the DNA are of significance in a number of contexts. For example, the number of copies of a particular repeat are highly variable in a given population, and hence are used in DNA fingerprinting **str_base**. Moreover, these repeats are inherited from a parent to an offspring, and hence are used in parentage analysis and in studying population evolution [69]. In the field of medicine, it is known that an increase in the number of repeating copies of certain repeats, present in the intronic (i.e., non-coding) regions of genes, can lead to a number of diseases. For example, fragile X syndrome, myotonic dystrophy, Huntington's disease and Friedreich's ataxia [69] are known as triplet expansion diseases, and are caused by copy number expansions of such period 3 repeats. Such mutations in the protein







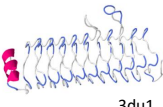
Repeat type	PDB ID	FTw.	WAV.	RAD.	REPw.	RFB
β propeller  1hxn	1hxn	✗	✓	✓	✗	✓
TIM barrel  1tim (chain A)	1tim	✗	✓	✗	✗	✓
LRR  1lrv	1dfj	✗	✓	✓	✓	✓
	1lrv	✗	N.A.	✓	✓	✓
	4cil	✗	N.A.	✓	✓	✓
HEAT  1b3u	1b3u	✗	N.A.	✓	✓	✓
Ankyrin  1n11	1n11	✗	N.A.	✓	✓	✓
	NCBI: NP_848 605.1	✗	N.A.	✓	✓	✓
Armadillo  3wpt	3wpt	✗	N.A.	✓	✓	✓
Pentapeptide  3du1	3du1	✗	N.A.	✗	✓	✓
	2bm4	✗	N.A.	✗	✓	✓
	3n90	✗	N.A.	✗	✓	✓

Figure 5.10: Protein repeats that were applied as inputs to FTwin [50], the wavelet based method of [53], RADAR [48], REPwin [50], and the RFB. These are abbreviated as FTw., WAV., RAD., REPw., and RFB in the top row. The first column indicates the type of repeat family being considered. The secondary structure of a representative member of each repeat family is also shown. The second column contains the Protein Data Bank (PDB) ID number for each example. For the ANKK1 protein exhibiting ankyrin repeats, its NCBI reference sequence number is shown instead. The wavelet based algorithm in [53] was not available publicly. Hence, we only show their results for two examples from [53]. See text for details. (Image sources: Jmol: <http://www.jmol.org/> and JSmol [http://wiki.jmol.org/index.php/JSmol.\[47\], \[54\], \[63\]–\[68\]](http://wiki.jmol.org/index.php/JSmol.[47], [54], [63]–[68]))

coding regions of the genome can affect the corresponding protein actions. For example, the Hand-Foot-Genital Syndrome is caused by length changes in tandemly repeating regions in the polyalanine tracts within the HoxA13 gene [70].

Because of such applications, locating tandem repeats in DNA sequences has been an important problem, and is addressed in several works. The approaches taken in these algorithms are very diverse in nature, and include (sometimes, combinations of) combinatorial searching [71], statistical inferencing techniques [69], spectral estimation techniques such as the Short Time Fourier Transform [72], [73], subspace decomposition based techniques [74], [75] and so on. A good review of these algorithms is presented in [76] and [69].

Here, we demonstrate the applicability of Ramanujan Filter Bank to the DNA repeats problem by proposing a simple algorithm to find DNA repeats. We compare its performance with one of the most popularly used algorithms among the existing ones, namely, the Tandem Repeat Finder (TRF) [69]. In the simplest terms, the TRF algorithm first finds all words of a small size k that repeat in the sequence, and then uses statistical tests to determine if these words are parts of bigger tandem repeats. Our results in the following sections demonstrate that the RFB is able to detect a number of repeats that the TRF algorithm (with its default settings) could not find.

5.6 Adapting the Ramanujan Filter Bank to Detect Tandem Repeats in DNA

To use the RFB for detecting repeats in DNA, we need to add a few new features to it. We explain these in the following subsections.

Mapping Nucleic Acids to Integers

While a DNA sequence is a string of nucleic acids A , T , C and G , the input to the RFB must necessarily be numbers (real or complex). So, to be able to process DNA sequences using the RFB, we must map each nucleic acid to a number. Since the coefficients of the RFB filters are integers, it is convenient (in terms of computations) to map the nucleic acids to integers. We chose the integers 1, 2, 3 and 4 in this work. Notice however that any assignment of these four numbers to the four nucleic acids results in an artificial notion of distance between them. For example, if A , T , C and G are assigned to 1, 2, 3 and 4 respectively, then A has been made ‘closer’ to T than to G , while no such notion of distance exists in reality. This can create biases such as the following: let $TAGTAG$ be an example of a tandem repeat. Consider two mutated versions of it: $TAGTTG$ and $TAGTGG$. If we use the aforementioned assignment of numbers, then $TAGTTG$ will turn out to be closer to $TAGTAG$ (and

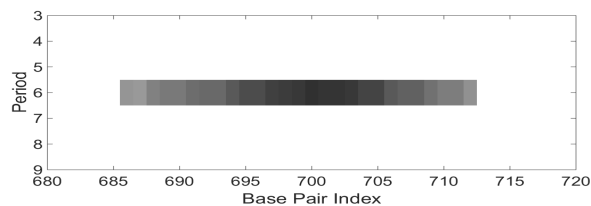


Figure 5.11: The section of the Time vs Period plane containing the tandem repeat of Table 5.3. Notice that the end location does not match with Table 5.5. See Sec. 5.6 for details.

hence closer to being periodic) than *TAGTGG*'s distance to *TAGTAG*.

To avoid such biases, we do the following. Notice that there are $4! = 24$ ways of assigning the integers 1, 2, 3 and 4 to the nucleic acids *A*, *T*, *C* and *G*. For each of those 24 assignments, we produce a time (or location) vs period plane using the RFB. We then take an average of all the 24 time vs period planes to get our final “averaged time vs period plane”.

Choice of Parameters

There are five parameters to select while using the RFB:

1. *Filter Repeat L*: The first is the number of repeats in each filter, namely L in (5.1). A small value for L yields better localization, while choosing a large value for L gives more accurate period estimate. We observed in our experiments that choosing $L = 7$ gave a good trade-off.
2. *Adding an Averaging Filter*: In the time vs period planes so far, we plot the power at the output of each filter versus time. We modify this slightly to plot the average power at the output of each filter. For instance, we average the square of the output of the filter $h_P(n)$ over $K'P$ samples for some integer K' . Choosing $K' = 3$ gave us good results in our experiments.
3. *Number of Filters P_{max}* : P_{max} is the largest period that we want to detect in the DNA. Since most of the applications mentioned in Sec. 7.1 involve repeats with small periods, we choose $P_{max} = 50$.
4. *Threshold T_1* : As mentioned above, we assign numbers to the four nucleic acids in 24 ways. For each assignment, we obtain a time vs period plane. We threshold each of those time vs period planes using the threshold parameter T_1 . That is, we

Table 5.2: Example 1: Human Frataxin Gene

Location	Period	Copy No.	% match	% in-del	RFB	TRF
25-49	4	6	83.3	8.3	√	
684-720	6	6	80.5	5.6	√	
987-1009	6	4	87.5	12.5	√	
2184-2211	3	9.7	100	0	√	√
822 - 854	14	2.4	89	0		√
2183-2211	44	2	90	0		√
2371 - 2401	13	2.8	87	0		√

Table 5.3: Base pairs 684 - 720 in Example 1

A.S.:	gccc-g	gcccag	gcccag	accctc	acccgg	gtcccgc	g
F.S.:	gcccag	gcccag	gcccag	gcccag	gcccag	g-cccag	g

set all the points in the time vs period plane that are less than T_1 to zero. In our experiments, we found that choosing T_1 as 20% of the maximum value on the time vs period plane gave good results. We did not consider the period 2 repeats in computing this threshold, since period 2 repeats tended to give unusually high output powers in the time vs period plane.

5. *Threshold T_2* : We average the 24 (thresholded) time vs period planes corresponding to the 24 different maps to obtain a final ‘averaged time vs period plane’. We threshold this averaged time vs period plane too, using a threshold parameter T_2 . In our experiments, we found that choosing T_2 as 40% of the maximum value on the time vs period plane gave good results. Once again, we did not consider period 2 repeats while computing the threshold.

Example 1: Human Frataxin Gene (Friedreich’s Ataxia), Intron 1

In our first example, we consider the human frataxin gene (Friedreich’s ataxia), intron 1 (Genbank Accession Number U43748). This example was chosen for two reasons: (a) Copy number expansion of the triplet repeat GAA present in the first intron of this gene is responsible for the neurological disorder known as Friedreich’s ataxia, and (b) this same example was also analyzed in the original paper of the TRF algorithm [69].

Table 5.2 summarizes the outputs of both TRF and RFB. In the table, the location column gives the base pair locations between which the repeats are found, copy number is the number of repeating copies of a particular repeat, percentage match indicates the extent of similarity between the copies of the repeat, and percentage

in-del indicates the insertions and deletions in the repeats. A \checkmark in the column corresponding to RFB or TRF indicates that the respective algorithm was able to detect this repeat. For example, consider the period 6 repeats between base pairs 684 and 720 shown in Table 5.3. The first row, labeled A.S. (actual sequence), shows the actual DNA segment. The lower row, labeled F.S. (Fitted Sequence) shows the nearest periodic sequence to the actual sequence. A '-' (dash) in the upper row indicates a deletion error, while a '-' in the lower row indicates an insertion error in the DNA sequence. Notice that there are 7 mismatches and 2 insertions/deletions for 6 copies of the repeat, and so the % match was calculated to be 80.5% and the % in-del was 5.6%.

The section of the time vs period plane that showed this repeat is shown in Fig. 5.11. Notice that while the start location matches with the one mentioned in Table 5.2, namely 684, the end location in the time vs period plane is around 712, whereas, we have reported the end location as 720 in Table 5.2. This is because, we found in our experiments that, looking for 1 or 2 cycles of a repeat on either side of the region reported by the RFB often turned up more cycles of the repeat. This could be because our choice of parameters in Sec. 5.6 makes the RFB a very conservative repeat finder. For instance, notice that most of the mismatch and in-del errors in Table 5.3 are towards the end of the segment relative to the first half of it. Hence, the RFB, although estimating the beginning accurately, was not able to recognize the last cycle. To detect repeats with higher mismatch and indel errors, lowering the thresholds of Sec. 5.6 is an option. It is interesting to note that out of the sets of four repeats found each by the RFB and the TRF, only the period 3 repeat that is relevant to Friedreich's ataxia is common.

Example 2: Human BAC clone AC010136

The GenBank sequence AC010136 is a 118522 base pairs long DNA sequence, found on second chromosome of the human DNA. The period 4 repeats shown in Table 5.4 between base pairs 66975 and 67076 (indicated by \star) are used as markers in DNA fingerprinting and forensics applications² [69], [77], [78]. Both TRF and RFB could identify these repeats. However, notice that there are several repeats that could be identified by only one of either TRF or RFB.

Table 5.2 and Table 5.4 show that there are a number of repeats that the RFB could detect, while TRF couldn't, and vice versa. It seems that RFB is better than TRF at detecting repeats with smaller periods, while TRF seems better with repeats that

²<https://strbase.nist.gov/>

Table 5.4: Example 2: Human Genome Sequence AC010136

Location	Period	Copy No.	% match	% in-del	RFB	TRF
2813-2825	6	2	91.7	8.3	√	
3662-3683	3	6.7	90	10	√	
6506-6547	3	14	88	0	√	
10862-10873	4	3	100	0	√	
11792-11805	3	6.7	95	0	√	
12121-12138	3	6	100	0	√	
12139-12147	3	3	100	0	√	
30684-30707	6	4	88	0	√	
30708-30731	6	4	83	0	√	
30732-30749	6	3	83	0	√	
31023-31058	4	9	86	0	√	
39988-40015	4	7	86	3.6	√	
40207-40226	6	3	83	5.5	√	
42753-42782	11	2.7	93	0	√	
46648-46673	6	4	79	4	√	
48057-48072	4	4	88	0	√	
49458-49476	6	3	89	0	√	
48625-48647	6	4	83	4.2	√	
57291-57314	12	2	88	0	√	
57829-57854	4	5.25	88	0	√	
58508-58536	6	4.7	77	0	√	
60406-60420	3	5	87	0	√	
62794-62809	3	5	87	6.7	√	
62810-62827	3	6	100	5.3	√	
63539-63558	4	5	90	5	√	
75567-75586	4	5	95	10	√	
77116-77138	4	5.75	87	0	√	
82092-82115	3	8	83	0	√	
103964-103985	3	7	95	0	√	
105070-105095	3	8.7	92	0	√	
111827-111846	4	5	100	0	√	
118239-118268	3	10	93	6.7	√	
11470-11530	20	3	95	1.7	√	√
11531-11590	20	3	85	0	√	√
11411-11509	33	3	93	1	√	√
11510-11695	33	5.6	84	0	√	√
31023-31058	4	9	86	0	√	√
42084-42353	24	12	90	6.3	√	√
42153-42274	35	3.5	81	2.4	√	√
42085-42333	46	5.25	85	9.5	√	√
★ 66975-67043	4	17	97	0	√	√
★ 67043-67076	4	7	100	0	√	√
11432-11492	13	5.2	72	22	√	√
11518-11545	13	2.2	100	0	√	√
11669-11709	13	3.2	92	0	√	√
11637-11736	46	2.2	88	0	√	√
17271-17324	26	2.1	92	0	√	√
19445-19480	4	9.8	82	17	√	√
19443-19480	7	5.3	87	12	√	√
19445-19480	10	3.3	81	18	√	√
35208-35271	21	3.0	95	4	√	√
42034-42706	11	59.0	69	11	√	√
42755-42827	33	2.2	90	0	√	√
44812-44853	21	1.9	85	9	√	√
52497-52528	14	2.4	94	5	√	√
53579-53647	35	2.0	97	0	√	√
73366-73405	21	1.9	85	10	√	√
83031-83072	20	2.1	90	0	√	√
83061-83274	10	20.2	65	14	√	√
83067-83125	31	1.9	82	6	√	√
83097-83156	26	2.3	77	5	√	√
83591-83676	39	2.3	80	7	√	√
90266-90315	24	2.1	88	0	√	√
93049-93090	20	2.1	100	0	√	√
93122-93152	14	2.2	100	0	√	√
105387-105421	16	2.2	100	0	√	√
114836-114882	20	2.4	83	16	√	√

have large periods but only a small number of repeating copies (typically 1.5-2.5 copies). An approach that seems encouraging to improve the performance of RFB at larger periods is to use different thresholds for different periods. We will explore this as a part of our future effort. In terms of computations, the RFB requires $O(NP_{max}^2)$ integer multiplications (where the multiplicands are $< P_{max}$) to produce a time vs period plane, where N is the data-length and P_{max} is the largest period we are searching for. Please visit our website³ to access the Matlab codes related to this section.

5.7 Absence Seizure Detection Using Ramanujan Filter Banks

A seizure is defined as a sudden uncontrolled surge in the electrical activity of the brain. It is usually accompanied by physical symptoms such as alterations in behavior, loss of consciousness, uncontrolled muscle spasms and so on. Seizures are generally caused when a large number of neurons get excited simultaneously, leading to a sudden surge in the electrical activity of the brain. While some seizures are hardly noticeable, others can be severely disabling. In both cases, it is important to note the occurrence of a seizure, since the chances of a second seizure is often greater than 50% after the first [79]. Since a seizure is essentially an electrical phenomenon, the most common way to identify them is using electroencephalography (EEG). Seizures produce characteristic EEG patterns, which can be used by doctors for diagnosis.

Absence seizures, the subject of this section, are a type of seizures more commonly occurring in children (ages 4 - 14). They derive their name from a "lack of consciousness" state that occurs during this seizure. Typical symptoms include sudden unresponsiveness, staring, fluttering of the eyelids, mild clonic jerks and possible automatisms. They typically last for 5 to 20 seconds, and are often confused with day dreaming or not paying attention. Currently, the only diagnostic test for these seizures is EEG [80]. Typical absence seizures show a prominent 3 Hz periodic spike and wave discharge pattern in the EEG [81], as shown in Fig. 5.12.

At present, the criterion standard for their diagnosis is EEG video monitoring, which involves monitoring a continuous recording of the EEG, along with a simultaneous video recording of the patient for observing the clinical manifestations of the seizure [82]. Visual inspection of EEG records for seizures is tedious and time consuming, especially when patients are monitored over hours at a stretch. So a number of

³<http://systems.caltech.edu/dsp/students/srikanth/Ramanujan/>

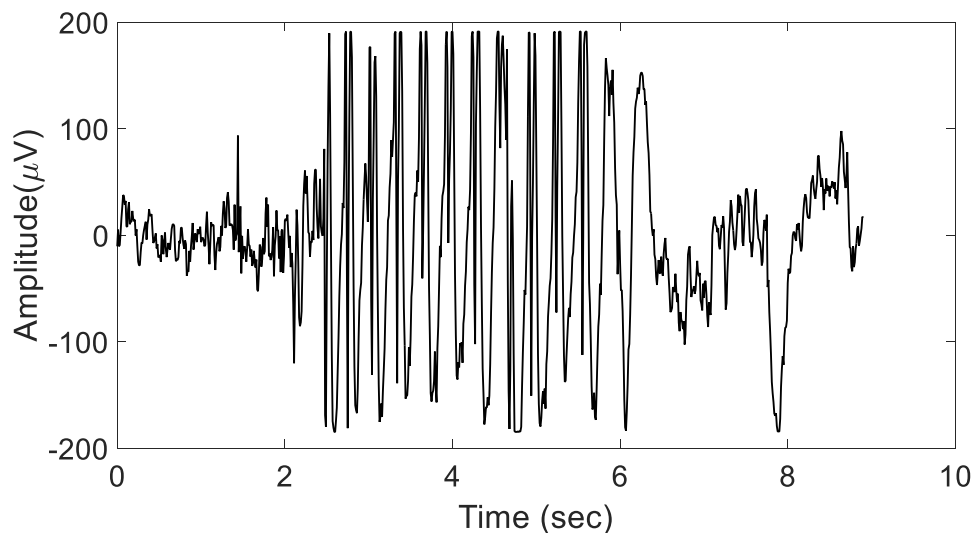


Figure 5.12: An example of the 3Hz spike and wave discharge pattern in the EEG during an absence seizure.

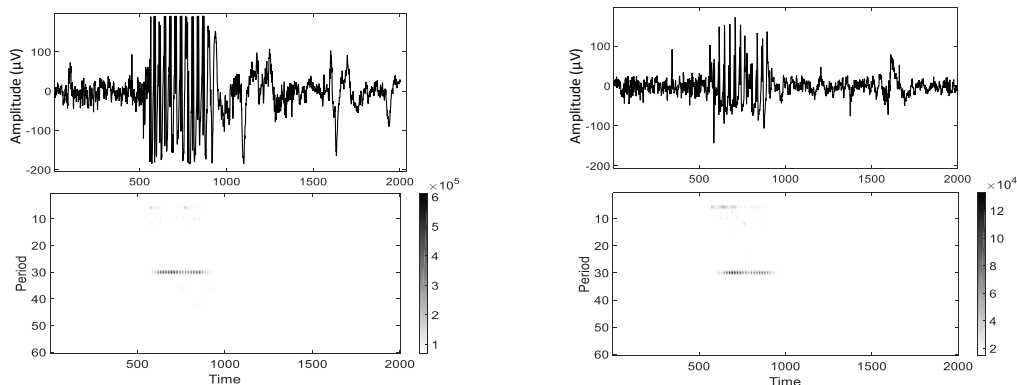
methods have been proposed in the past few years for automatic detection of seizures from EEG signals. An important class of such methods is based on the traditional spectral estimation techniques from Digital Signal Processing (DSP), such as the Short Time Fourier Transform (STFT) [83] and wavelets [84], [85]. These methods were originally designed for estimating the spectrum of a signal. By identifying the fundamental frequency of a periodic signal from among its harmonics in the spectrum, one can estimate the period. Apart from these methods, there are others based on metrics such as the EEG signal's amplitude [86], or more complex statistical features of the signal such as its average entropy and multi-scale variance [87].

Since the EEG waveform during an absence seizure is essentially periodic, the RFB is a suitable technique to be applied here. In this section, we demonstrate that the RFB indeed offers very useful diagnostic information by being able to detect the occurrence of these seizures in the EEG.

Data

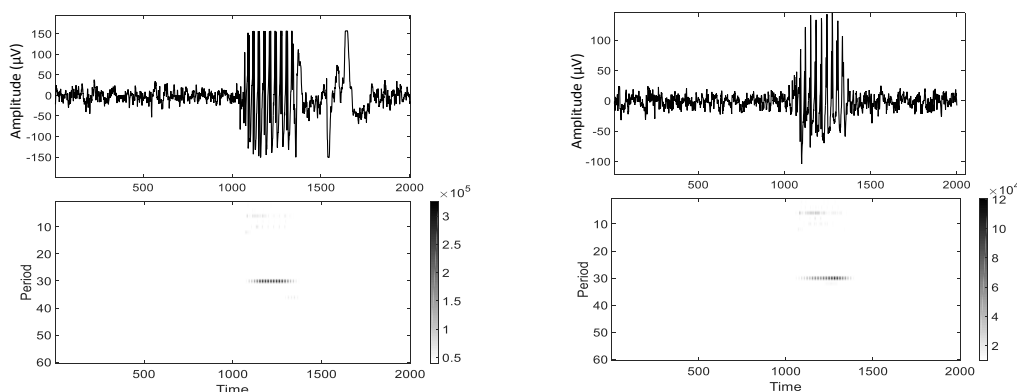
The source for the EEG data in this work is the public FTP site of the Sleep and Sensory Signal Analysis Group from Tampere University of Technology⁴. This data

⁴<ftp://sigftp.cs.tut.fi/pub/eeg-data/>



(a) (Top) Sampled EEG showing Seizure 1 in Patient 1 as measured across the F8-C4 channel. (Bottom) The RFB's time vs period plane.

(b) (Top) Sampled EEG showing Seizure 1 in Patient 1 as measured across the T5-O1 channel. (Bottom) The RFB's time vs period plane.



(c) (Top) Sampled EEG showing Seizure 2 in Patient 1 as measured across the F8-C4 channel. (Bottom) The RFB's time vs period plane.

(d) (Top) Sampled EEG showing Seizure 2 in Patient 1 as measured across the T5-O1 channel. (Bottom) The RFB's time vs period plane.

Figure 5.13: RFB's test of sensitivity. See Sec. 5.7 for details.

has been used in popular seizure detection papers such as [84], but is annotated to a limited extent only. So our verification of the RFB's detection of seizures in the following experiments is based on (a) results from previously published works that used the same data, such as [84], and or or (b) the information provided with the data itself. Encouraged by the RFB's performance on this data set, we are currently trying to acquire completely annotated data from experts. The EEG signals used in the following examples were sampled at 100 Hz (we downsampled the original files by 2 for convenience). This means that the 3 Hz spike and wave discharges associated with absence seizures would appear as periodic patterns with period around 33 samples.

Experiments

In the following, we have divided our examples into three categories. In the following, we test the sensitivity of the RFB in detecting absence seizures. That is, we apply the RFB to multiple instances of absence seizures to see if we are able to detect them in all cases. We also test for the specificity of detection. That is, whether the RFB is able to distinguish absence seizures from normal background EEG in a precise manner. Finally, in we compare the RFB to an STFT based method to illustrate its advantages over the traditional spectral methods. In all the following examples, we chose $K = 5$ for the RFB.

Testing Sensitivity

Fig. 5.13a and Fig. 5.13b show the (sampled) EEG signals of Patient 1, measured between channels F8-C4 and T5-O1 respectively during a seizure. These two channels are located on opposite sides of the scalp, so the seizure waveform appears different in shape in each case. Fig. 5.13c and Fig. 5.13d show EEG signals from the same patient, but during a different episode of absence seizure. In each of these figures (and also in all the other examples in this work), a Time vs Period plane was produced by an RFB with 60 filters. As shown in Fig. 5.13, in each case, the RFB was able to identify a periodic segment with period around 30, which matches with what we would expect during an absence seizure. For clearer plots, the color-bar in these plots was chosen such that all the outputs below 10-15% of the maximum output appear white.

Testing Specificity

The seizures shown in Fig. 5.13 occurred as short segments of much longer data records, which, for most part, contained normal EEG measured during non-seizure intervals. In Fig. 5.14, the red arrow indicates the instance at which the seizure of Fig. 5.13c occurred in its complete data record. As shown in the time vs period plane in Fig. 5.14, the RFB was able to identify this seizure event very precisely.

Comparison with STFT

Fig. 5.15(a) shows an EEG recording from Patient 2, with three epileptic seizures shown by the red arrows. Fig. 5.15(a) shows the time vs period plane obtained using

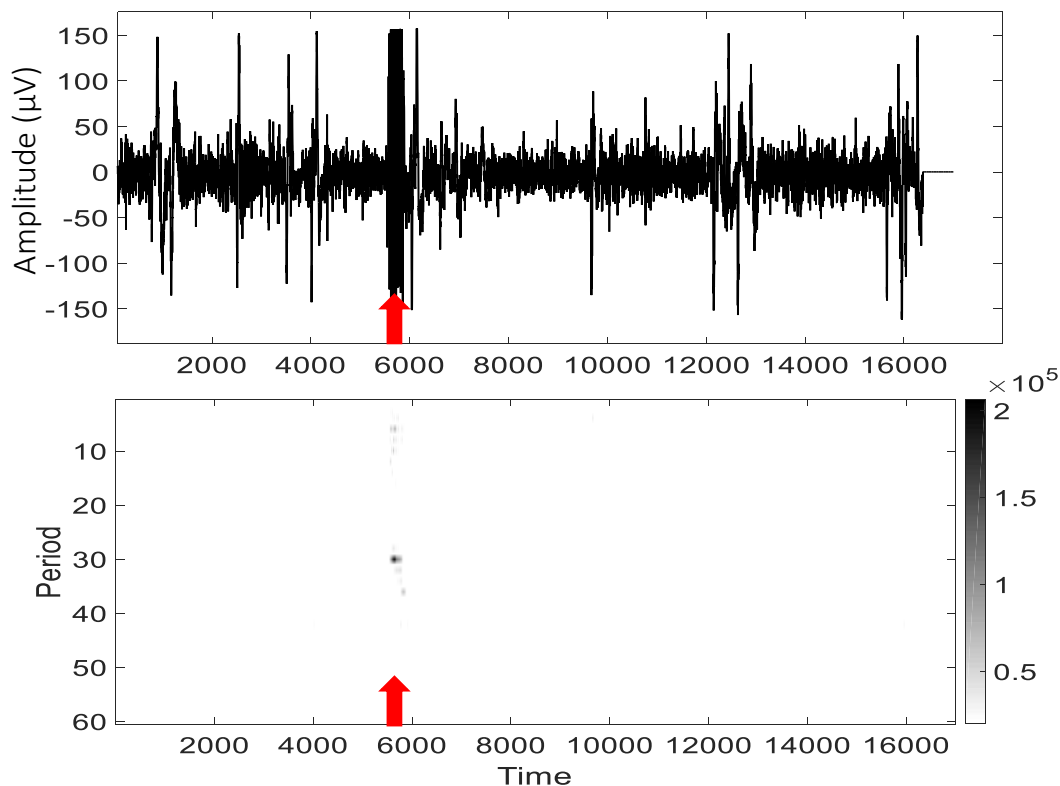


Figure 5.14: (Top) The entire record of sampled EEG containing Seizure 2 in Patient 1, as measured across the F8-C4 channel. (Bottom) The RFB's time vs period plane.

an RFB, and the three seizures can be easily identified. However, notice that the three seizures manifest as a band of periods around period 30. This is because the spike and wave discharges during an absence seizure need not always have a precise periodicity of 3Hz [80]. Fig. 5.15(c) shows the time vs frequency plane obtained by using STFT with a length 128 window. 3Hz in continuous time would correspond to a frequency of around 0.03 on the shown (discrete time normalized) frequency scale. Notice that apart from the seizure features (indicated by the red arrows), there are many other strong features in the time vs frequency plane. This is because the range of frequencies represented by the green band in Fig. 5.15(c) correspond to periods 30 to infinity in the signal. Hence, most of the low frequency noise is being captured in this narrow region of the time vs frequency plane, producing those other features. The only way to increase the resolution in this region of frequencies is to use larger window lengths. Fig. 5.15(d) shows that we only get a slight improvement when we use length 512 STFT. Further increase in the window lengths produce a bad resolution along the time axis. This is an example of a case where the RFB offers clearer results than STFT.

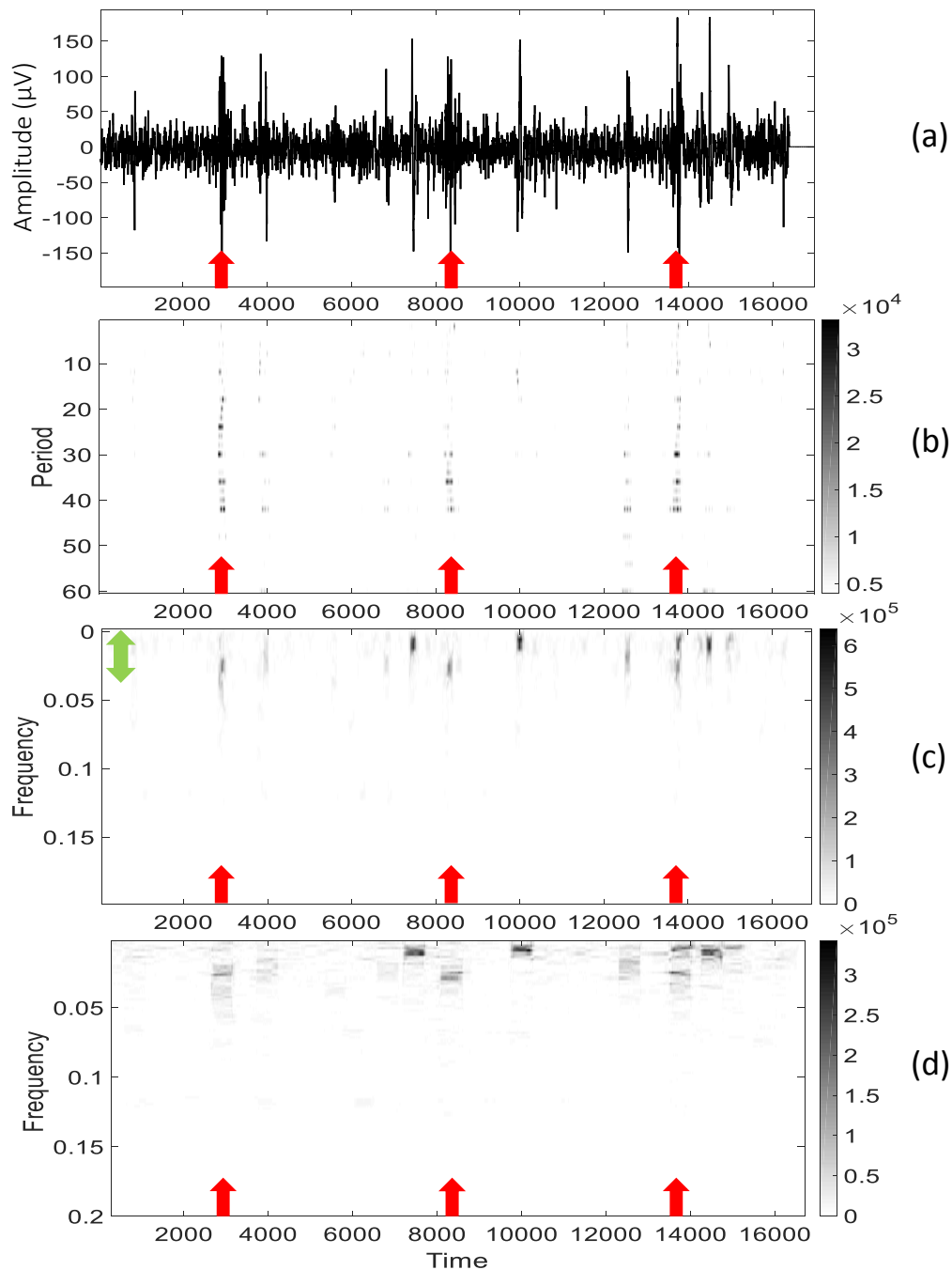


Figure 5.15: Comparison of the RFB with STFT. See Sec. 5.7 for details.

Since the annotation available to us was limited for this data set, we only intend to use these results as a demonstration of RFB's suitability for this application. One of our next goals is to work with an EEG expert on a more formally annotated data set, to demonstrate further the applicability of the RFB.

5.8 Unique Representation Filter Banks: Removing Redundancy in the RFB Using Truth Tables

Notice that, in addition to period estimation, the RFB can determine if a periodic signal is actually a mixture of multiple periodic signals. For instance, adding a period 5 signal to a period 9 signal can give rise to a period 45 signal. The RFB can not only identify the period of the sum, i.e., 45, but also give us the periods of the components, i.e., 5 and 9. However, in a number of applications, one does not care to know whether the given signal can be decomposed into a sum of signals with smaller periods (for example, in epileptic seizures, DNA and protein repeats, ECG etc.). When this is the case, we show in this section that we can have filter banks with far fewer filters than the RFB. For example, to detect periods 1 to 8, while the RFB needs 8 filters, we can do the same with just 3 filters. We call these new filter banks as “Unique Representation Filter Banks”, formally defining them in the next subsection. As we will show, a convenient way to represent and analyze period estimating filter banks is to use truth tables. In the following, we derive sufficient and necessary conditions respectively, for the existence of Unique Representation Filter Banks. The sufficiency proof is constructive, so that we have a technique to construct them once their existence is proved. We will then demonstrate the period estimation capabilities of the new filter banks using simulations.

Unique Representation Filter Banks

In a regular RFB, the period of the input is estimated by taking the LCM of the indices of filters with non-zero outputs (Sec. 5.1). Table 5.5 tabulates this LCM rule for the example of $P_{max} = 8$. In this table, F_1 to F_8 are the RFB filters, and if an entry is 1, it indicates that the output of the corresponding filter is a non-zero signal. For each period, Table 5.5 essentially lists all the possible ways in which the LCM of the filter indices can be equal to that period.

Notice that the RFB does not map a period to a unique set of filters. For example, period 6 can occur in 10 ways. In this section, we want to construct filter banks that map a period to a unique set of filters. The advantage is that we can do the same period estimation task using far fewer filters. For example, if we can construct a filter bank that offers (either of) the mappings shown in Table 5.6, then we can detect periods 1 to 8 using just 3 filters, rather than using 8 RFB filters. We will call such filter banks as *Unique Representation Filter Banks*.

Table 5.5: Ramanujan Filter Bank: Period to Filters Map

Period	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	Hidden Periods
1	1	0	0	0	0	0	0	0	-
2	0	1	0	0	0	0	0	0	-
	1	1	0	0	0	0	0	0	-
3	0	0	1	0	0	0	0	0	-
	1	0	1	0	0	0	0	0	-
4	0	0	0	1	0	0	0	0	-
	1	0	0	1	0	0	0	0	-
	0	1	0	1	0	0	0	0	-
	1	1	0	1	0	0	0	0	-
5	0	0	0	0	1	0	0	0	-
	1	0	0	0	1	0	0	0	-
6	0	0	0	0	0	1	0	0	-
	1	0	0	0	0	1	0	0	-
	0	1	0	0	0	1	0	0	-
	1	1	0	0	0	1	0	0	-
	0	0	1	0	0	1	0	0	-
	1	0	1	0	0	1	0	0	-
	0	1	1	0	0	1	0	0	-
	1	1	1	0	0	1	0	0	-
	0	1	1	0	0	0	0	0	2, 3
1	1	1	0	0	0	0	0	2, 3	
7	0	0	0	0	0	0	1	0	-
	1	0	0	0	0	0	1	0	-
8	0	0	0	1	0	0	0	1	-
	1	0	0	1	0	0	0	1	-
	0	1	0	1	0	0	0	1	-
	1	1	0	1	0	0	0	1	-
	0	0	0	0	0	0	0	1	-
	1	0	0	0	0	0	0	1	-
	0	1	0	0	0	0	0	1	-
	1	1	0	0	0	0	0	1	-

Table 5.6: Tables Mapping Periods To Unique Sets of Filters. The Table on the Left is Unimplementable.

Period	F_1	F_2	F_3
1	0	0	0
2	1	0	0
3	1	1	0
4	0	1	0
5	1	0	1
6	0	0	1
7	1	1	1
8	0	1	1

Period	F_1	F_2	F_3
1	0	0	0
2	1	0	0
3	0	0	1
4	1	1	0
5	0	1	0
6	1	0	1
7	0	1	1
8	1	1	1

It turns out that arbitrary tables that map periods to filters may not have corresponding filter banks that can achieve those tables. For instance, it can be proved that there does not exist any filter bank that can result in the example shown on the left in Table 5.6 (see the discussion after Theorem 5.8.2 in Sec. 5.8), whereas the example shown on the right in Table 5.6 does have a corresponding filter bank.

Which tables result in feasible filters, and how to build those filters when they exist? What are their properties? These are the questions we address in the following sections. Before doing so, we will first introduce a concise notation to represent tables such as those in Table 5.6:

Representing a Table as a Function: Consider the example on the right side of Table 5.6. We can represent it as the following mapping f :

$$\begin{aligned}
 f(1) &= \{\}, & f(2) &= \{1\}, & f(3) &= \{3\}, \\
 f(4) &= \{1, 2\}, & f(5) &= \{2\}, & f(6) &= \{1, 3\}, \\
 f(7) &= \{2, 3\}, & f(8) &= \{1, 2, 3\}
 \end{aligned} \tag{5.14}$$

Notice that f associates, to each period P , the indices of filters that have non-zero outputs when the input is periodic with period P . We shall use this notation throughout the rest of this section to represent tables.

At this point, it is useful to formally define Unique Representation Filter Banks:

Definition 5.8.1. URFB: *Given a filter bank with filters $\{F_1, F_2, \dots, F_N\}$, suppose we define a mapping $\mathbb{G}_{ON}(x)$ for every input signal $x(n)$, that gives the indices of filters that have non-zero output signals when the input is $x(n)$. Then, the filter bank is said to be a Unique Representation Filter Bank (URFB) for periods 1 to P_{max} , if,*

given any pair of periodic signals $x(n)$ and $y(n)$, whose periods lie in $1 \leq P \leq P_{max}$,

$$\mathbb{G}_{ON}(x) = \mathbb{G}_{ON}(y) \quad (5.15)$$

iff both $x(n)$ and $y(n)$ have the same period. \diamond

The above definition essentially says that a filter bank is a URFB if the following property holds: Two periodic inputs result in the same set of filters with non-zero outputs iff their periods are the same. It is obvious that a URFB, if it exists, can be used for period estimation by looking at which set of filters have non-zero outputs. We shall now derive sufficiency conditions for the existence of a URFB.

Sufficiency Conditions for the Existence of URFBs

The following result gives the sufficient conditions for a filter bank to exist given a table f . ($\mathcal{P}(\mathbb{S})$ represents the power set of a set \mathbb{S} , namely, the set of all its subsets).

Theorem 5.8.1. Sufficiency for the existence of a URFB: *Let P_{max} and N be positive integers such that there exists a map $f : \{1, 2, \dots, P_{max}\} \rightarrow \mathcal{P}(\{1, 2, \dots, N\})$ satisfying:*

1. $f(P) = f(Q)$ iff $P = Q$.
2. If $1 \leq d_1 \leq d_2 \leq \dots \leq d_K \leq P \leq P_{max}$ are integers such that $LCM(d_1, d_2, \dots, d_K) = P$, then

$$\cup_i f(d_i) = f(P) \quad (5.16)$$

Then, there exists a URFB with N filters, such that its $\mathbb{G}_{ON}(x) = f(P)$ iff period of $x(n)$ is P . \diamond

Proof. Let the N filters, say F_1, F_2, \dots, F_N , be defined as follows:

$$F_i = \sum_{\substack{1 \leq q \leq P_{max} \\ i \in f(q)}} c_q(n) \quad (5.17)$$

where $c_q(n)$ is the q^{th} Ramanujan Sum. Suppose the input to these filters is a period P signal, say $x(n)$, where $P \leq P_{max}$. Further, let the coprime frequencies in its Fourier series expansion (see (7.6)) have periods d_1, d_2, \dots, d_K . This implies that, if F_i is a filter whose output is non-zero, then it must have a passband at at least one coprime frequency with period d , for some $d \in \{d_1, d_2, \dots, d_K\}$. But from (5.17), filter F_i will have such a passband only if $i \in f(d)$.

Now, by Theorem 5.1.1, we have

$$LCM(d_1, d_2, \dots, d_K) = P \quad (5.18)$$

So if $i \in f(d)$, then $i \in f(P)$ from Condition 2 in the statement of the current Theorem, since $d|P$.

Hence, we have shown that if a filter F_i has a non-zero output, then $i \in f(P)$. It follows that $\mathbb{G}_{ON}(x) \subseteq f(P)$. We will now show that it is in fact equal to $f(P)$. Suppose F_j is a filter such that $j \in f(P)$. Notice that if $j \in f(P)$, then from Condition 2 above, there must exist a $d \in \{d_1, d_2, \dots, d_K\}$ such that $j \in f(d)$. But if $j \in f(d)$, then F_j will have a passband at all the co-prime frequencies with period d (from (5.17)). Now, by our assumption, the Fourier series expansion of the input must have at least one coprime frequency with period d . Hence, the output of F_j cannot be 0.

This shows that every filter in $f(P)$ must have non-zero output when the input is $x(n)$. Hence, $\mathbb{G}_{ON}(x) = f(P)$. This, along with Condition 1, ensure that \mathbb{G}_{ON} satisfies (5.15). Hence, $\{F_1, \dots, F_N\}$ is a URFB. $\nabla \nabla \nabla$

Notice that the example shown on the right in Table 5.6 satisfies the conditions of Theorem 5.8.1. Hence, we can construct a corresponding URFB using (5.17). However, the example shown on the left in Table 5.6 does not satisfy Condition 2.

URFBs: Necessary Properties

We will now show that the two conditions in Theorem 5.8.1 are in fact necessary for any URFB. Further, we will also show that the filters in a URFB must necessarily be of the form (5.17).

Theorem 5.8.2. Necessary Properties of a URFB: *Let $\{F_1, F_2 \dots F_N\}$ be a URFB for periods 1 to P_{max} satisfying Definition 5.8.1. With G_{ON} as in Definition 5.8.1, define a map $f : \{1, 2, \dots, P_{max}\} \rightarrow \mathcal{P}(\{1, 2, \dots, N\})$ as follows:*

$$f(P) = \mathbb{G}_{ON}(x) \quad (5.19)$$

where $x(n)$ is a periodic signal with period P . Then, f must satisfy the following properties:

1. $f(P) = f(Q)$ iff $P = Q$.

2. If $1 \leq d_1 \leq d_2 \leq \dots \leq d_K \leq P \leq P_{max}$ are numbers such that $LCM(d_1, d_2, \dots, d_K) = P$, then

$$\cup_i f(d_i) = f(P) \quad (5.20)$$

◇

Proof. Condition 1 follows directly from Definition 5.8.1. Let us now prove Condition 2. Let $d|P$. We will first show that $f(d) \subseteq f(P)$.

Let $i \in f(d)$. Then, the filter F_i must have non-zero output for all period d inputs. So in particular, it must have a non-zero output when the input is a co-prime frequency with period d . Now, if we add a coprime frequency with period d with another coprime frequency with period P , the resultant signal has period P (from Theorem 7.2.2). If we feed this sum as the input to the filter bank, F_i must continue to have non-zero output. But, by definition, $f(P)$ is the set of filters that have non-zero outputs when the input is periodic with period P . Hence, $i \in f(P)$. This proves that $f(d) \subseteq f(P)$.

It follows that, in Condition 2, since d_1, d_2, \dots, d_K are divisors of P ,

$$\cup_{i=1}^K f(d_i) \subseteq f(P) \quad (5.21)$$

To prove that the above equation holds with an equality, we use proof by contradiction. If possible, let $i \in f(P)$, but $i \notin \cup_{i=1}^K f(d_i)$. Then, $i \notin f(d_j)$ for each $j \in \{1, 2, \dots, K\}$. So F_i must have zero output for every d_j periodic signal. In particular, it must have zero output for coprime frequencies with period d_j . Now, consider an input that is a sum of all coprime frequencies whose periods $\in \{d_1, d_2, \dots, d_K\}$. It must have period P according to Theorem 7.2.2. But from the above discussion, the filter F_i must have a zero output for this particular input. This is a contraction to $i \in f(P)$. Hence,

$$\cup_{i=1}^K f(d_i) = f(P) \quad (5.22)$$

▽ ▽ ▽

Let us now verify Theorem 5.8.2 on the examples shown in Table 5.6. For the example on the left in Table 5.6, $f(2) \cup f(4) \neq f(4)$. So Condition 2 of Theorem 5.8.2 is violated. Hence, no URFB is possible for this case. However, it is easy to check that the example shown on the right in Table 5.6 does satisfy the conditions of Theorem 5.8.2.

We will now prove that the filters must necessarily be of the form (5.17). Consider the following theorem:

Theorem 5.8.3. *Let $\{F_1, F_2 \dots F_N\}$ be a URFB for periods 1 to P_{max} satisfying Definition 1, with the map f defined as in (5.19). Then, for each $1 \leq i \leq N$ and $1 \leq P \leq P_{max}$, filter F_i 's passband must include the coprime frequencies with period P iff $i \in f(P)$. \diamond*

Remark: Notice that the spectrum of any periodic signal is non-zero only at frequencies that are rational multiples of 2π (see Theorem 7.2.2 for instance). So the frequency response of the filters at frequencies other than rational multiples of 2π are irrelevant to period estimation (ideally, in the infinite data length case). The above theorem says that, for any unique representation filter bank, the frequency responses of the filters must match those of the following filters at rational multiples of 2π :

$$F_i = \sum_{\substack{1 \leq q \leq P_{max} \\ i \in f(q)}} \alpha_{i,q} c_q(n) \quad (5.23)$$

where $\alpha_{i,q} \neq 0$. Notice the similarity with (5.17).

Proof of Theorem 5.8.3: Suppose that F_i has a passband that includes the coprime frequencies with period P . Then, if the input is $c_P(n)$, the output of F_i must be non-zero. But if F_i has a non-zero output for a period P signal, then by the definition of a URFB, it must have non-zero outputs for all period P signals. That is, $i \in f(P)$.

Conversely, suppose $i \in f(P)$. Now, if F_i does not have a passband at one of the coprime frequencies with period P , say at the frequency $\frac{2\pi m_0}{P}$, where $(m_0, P) = 1$, then if the input is the signal $e^{\frac{j2\pi m_0}{P}n}$, the output of F_i would be zero. However, this signal has period P , and since $i \in f(P)$, F_i should have given a non-zero output. Hence, we obtain a contradiction. So F_i must have a passband at all the coprime frequencies with period P . $\nabla \nabla \nabla$

A Simulation Example

In this section, we demonstrate a URFB for periods 1 to 4. Notice that Table 5.7's corresponding function f is given by:

$$f(1) = \{\}, \quad f(2) = \{1\}, \quad f(3) = \{2\}, \quad f(4) = \{1, 2\}$$

Table 5.7: A URFB for periods 1 to 4.

Period	F_1	F_2
1	0	0
2	1	0
3	0	1
4	1	1

It is easy to check that f satisfies the conditions of Theorem 4.3.1. Using (5.17), the two filters' impulse responses are given by:

$$h_1(n) = c_2(n) + c_4(n)$$

$$h_2(n) = c_3(n) + c_4(n)$$

For a practical implementation, we truncated these filters to length 20. Fig. 5.16(a) shows a signal whose period increases by 1 from 1 to 4 after every 100 samples. Fig. 5.16(b) shows the output powers of the two filters. Note that they match with Table 5.7.

We did not consider noise in this example, since we only wanted to demonstrate the capability of a URFB to distinguish between different periods. In our experiments, we found that increasing the filter lengths resulted in better noise performance, at the cost of poorer time resolution (due to longer transients. Our current work involves an analysis of how to optimize these filters under noise (for example, by choosing the $\alpha_{i,q}$'s in (5.23) appropriately).

5.9 Conclusion

We introduced a new period estimation technique in this chapter called the Ramanujan Filter Bank (RFB). It was shown that the RFB is especially suited for applications where the periodicity is localized, or is changing with time. Its performance was compared with the state-of-the-art methods for applications such as DNA and protein repeats, and for detecting absence seizures from EEG waveforms. The promising performance of the RFB encourages us to further tailor it for each of these application areas by optimizing its parameters over much larger databases. This could be a promising direction for the future.

By looking at the operation of the RFB in terms of a truth table, we also introduced the so called "Unique Representation Filter Banks" (URFB) which have far fewer

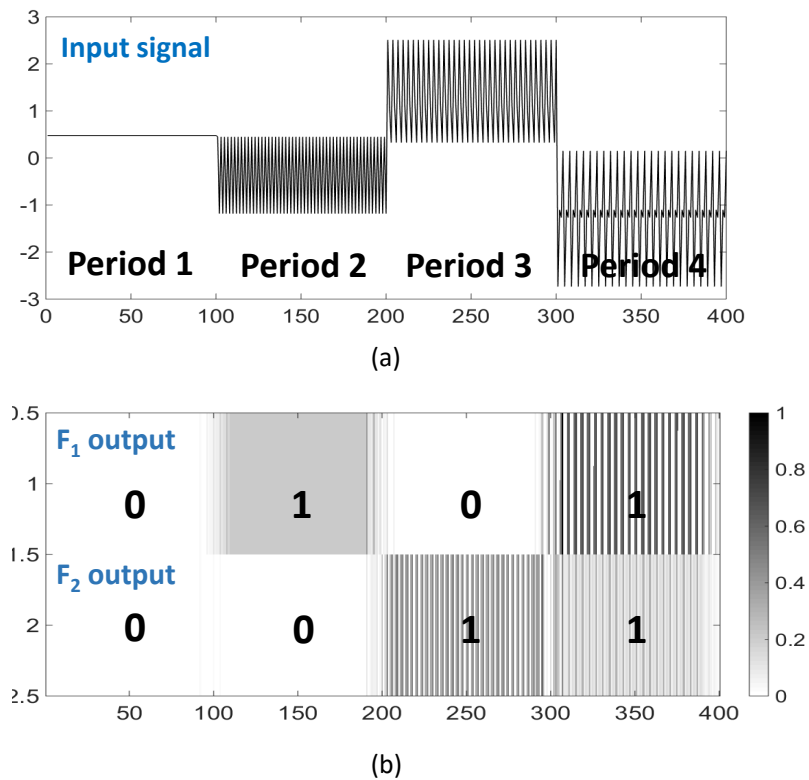


Figure 5.16: Demonstrating a URFB for Table 5.7. See Sec. 5.8 for details.

filters than the RFB. While the RFB can estimate the component periods in a mixture of periodic signals, the URFB are tailored for applications where the input is just a single periodic signal. One of the questions we want to study in this regard is, what is the least number of filters needed for a URFB to detect periods 1 to P_{max} ? One might conjecture that we may need about $\log_2 P_{max}$ filters, but this requires further study.

iMUSIC: A FAMILY OF MUSIC-LIKE ALGORITHMS FOR INTEGER PERIOD ESTIMATION

The MUSIC algorithm (Multiple Signal Classification) [15] is one of the most popular techniques for estimating line spectra in discrete time signals. It has widespread applications, including Direction of Arrival estimation [88]–[90], time delay estimation [91], neuro-imaging [92], [93], and many more. But when the signal of interest is periodic, its spectrum is not just arbitrary lines. There is a nice harmonic structure in the spectrum as shown in Fig. 7.1, which can be modeled mathematically as:

$$x(n) = \sum_{k=0}^{K-1} c_k e^{jk\omega_0 n} \quad (6.1)$$

where $2\pi/\omega_0$ (possibly not an integer) is usually considered as the ‘period’. While MUSIC itself does not exploit this additional harmonic structure, it was shown in an important series of publications [16], [17], [94] that modifying MUSIC’s search over complex exponentials so that we look for harmonically spaced peaks, improves the period estimates significantly. These methods were called Harmonic MUSIC (or HMUSIC). However, they are computationally much more complex than traditional MUSIC, especially when the input is a mixture of multiple periodic signals.

While (6.1) generically applies to several instances of periodicity such as speech, cardiology, EEG analysis and so forth, there is a second class of applications which have more structure than what is captured by (6.1). These are periodic signals whose periods are integers:

$$x(n + P) = x(n) \quad \forall n \in \mathbb{Z} \quad (6.2)$$

for some integer P . Such applications include repeats in protein and DNA sequences (see Chapter 5). In fact, many state of the art methods for conventional periodicity applications such as speech [12], [95], [96] are also based on integer period approximations. This chapter shows that the simplicity of the integer period model (7.19) opens up the possibility for designing a more diverse class of MUSIC-like algorithms than prior works.

More specifically, for such signals with integer periods, this chapter proposes a new formulation of MUSIC called *iMUSIC*, using the Ramanujan Subspaces (Chapter 2)

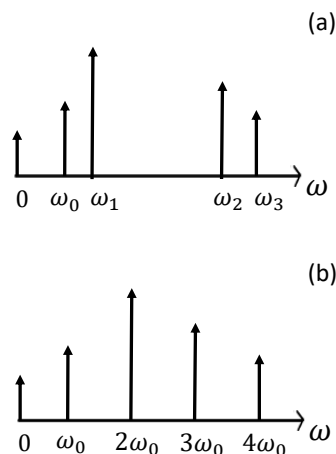


Figure 6.1: Part (a) - An arbitrary line spectrum; Part (b) - The harmonic line spectrum of a periodic signal. Can we use the additional structure in the spectrum of a periodic signal to improve MUSIC?

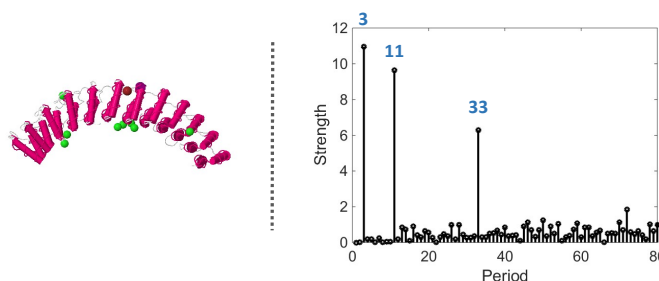


Figure 6.2: *Applications with Integer Periodicity*: The protein AnkyrinR (PDB 1n11) that enables red blood cells to resist shear forces. Its period 33 structural repeats can be clearly identified in the plot on the right, produced by the proposed techniques.

and Nested Periodic Subspaces (NPSs) (Chapter 3). The frequencies corresponding to signals with integer periods can be compactly represented by a non-uniform grid known as the Farey grid (see Sec. 4.5, Fig. 4.1). Using the Farey grid, we propose an alternative to the classical MUSIC pseudospectrum: All the complex exponentials on the Farey grid belonging to a common Ramanujan subspace [9] are grouped together when computing this proposed pseudo-spectrum. It will be shown that the resulting algorithm yields much higher accuracies than classical MUSIC and its prior periodicity variants such as HMUSIC, while keeping the computational complexity very low.

Furthermore, the Ramanujan subspaces can alternatively be spanned by simple

integer valued vectors instead of the Farey grid (Fig. 6.3). In fact, using more general Nested Periodic Subspaces (Chapter 3), one can construct many such examples of simple integer valued vectors that can be used to compute the proposed iMUSIC pseudospectrum instead of complex exponentials. Some of these new bases are very sparse, consisting of only 1's and 0's (Fig. 6.3(a)). These new representations give rise to a rich class of MUSIC-like algorithms that are well suited for integer period applications. Their advantages are demonstrated using examples that include Protein and DNA repeats. To the best of our knowledge, this is the first time MUSIC-like methods have been used on such bio-molecular repeats.

It should be mentioned here that there are other interesting algorithms such as the harmonic matching pursuit (HMP) [97], and expectation-maximization (EM) based algorithms [98] for taking advantage of the harmonicity in line spectra. While our focus in this work is only on MUSIC-like algorithms, we do include HMP and EM in our comparisons.

The mathematical formulation of classical MUSIC benefits greatly from the Vandermonde structure of complex exponentials. For instance, this is used in deriving the conditions for avoiding spurious peaks in the MUSIC pseudospectrum [15], [89]. The absence of a Vandermonde structure in NPSs introduces many new, but interesting challenges. For example, while we cannot guarantee the absence of additional (spurious) peaks in the iMUSIC pseudo-spectrum, *we can still prove that any such peak will not affect the estimated period*. These, and other such deviations from classical MUSIC will be rigorously addressed throughout this chapter.

Chapter Outline

Sec. 6.1 summarizes MUSIC and its prior adaptations to periodic signals (the HMUSIC algorithms). Sec. 6.2 introduces the proposed iMUSIC framework, first using the Farey grid of complex exponentials. This is generalized to iMUSIC using other NPSs in Sec. 6.3, allowing the use of integer valued vectors for spanning the signal subspace. The conditions for identifiability of the true periods using such integer bases are rigorously derived here. Sec. 6.4 contains several simulations and comparisons with other techniques, including examples of protein and DNA repeats.

	G_1	G_2	G_3	G_4
(a)	1	1	1 0	1 0
	1	0	0 1	0 1
	1	1	0 0	0 0
	1	0	1 0	0 0
	1	1	0 1	1 0
	1	0	0 0	0 1
	1	1	1 0	0 0
	S_1	S_2	S_3	S_4
(b)	1	1	2 -1	2 0
	1	-1	-1 2	0 2
	1	1	-1 -1	-2 0
	1	-1	2 -1	0 -2
	1	1	-1 2	2 0
	1	-1	-1 -1	0 2
	1	1	2 -1	-2 0

Figure 6.3: *Simple Integer Alternatives to Complex-Exponentials*: Bases of (a) The Natural Basis Subspaces, and (b) The Ramanujan Subspaces.

6.1 MUSIC and Periodicity: An Overview of Prior Works

In this section, we will briefly outline the MUSIC algorithm [15], [99], and its prior adaptations to periodic signals. Let us begin with the following signal model:

$$x(n) = \sum_{k=0}^{K-1} c_k e^{j\omega_k n} + e(n), \quad (6.3)$$

where ω_k are distinct frequencies in $[-\pi, \pi)$ and $e(n)$ is zero-mean white noise with variance σ_e^2 . Most prior MUSIC-based works model $c_k \in \mathbb{C}$ as random variables [16], [99], [100]. But we will assume them to be constants here, since such is the case in most applications of periodicity.

Assume that there are L samples of $x(n)$, $1 \leq n \leq L$, and define the i th block of data as

$$\mathbf{x}(i) = \left[x(i) \quad x(i+1) \quad \cdots \quad x(i+N-1) \right]^T, \quad (6.4)$$

where $N < L$ is the blocksize. We can call $\mathbf{x}(i)$ the i th “snapshot” but note that successive blocks are not independent (they have an overlap of $N - 1$ samples).

There are

$$M = L - N + 1 \quad (6.5)$$

blocks. Note that we can write the i th block as

$$\mathbf{x}(i) = \begin{bmatrix} \mathbf{a}_0(i) & \mathbf{a}_1(i) & \cdots & \mathbf{a}_{K-1}(i) \end{bmatrix} \underbrace{\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{K-1} \end{bmatrix}}_{\mathbf{c}} + \mathbf{e}(i) \quad (6.6)$$

where $\mathbf{a}_k(i)$ are Vandermonde vectors up to scale:

$$\begin{aligned} \mathbf{a}_k(i) &= \begin{bmatrix} e^{j\omega_k i} & e^{j\omega_k(i+1)} & \cdots & e^{j\omega_k(i+N-1)} \end{bmatrix}^T \\ &= e^{j\omega_k i} \underbrace{\begin{bmatrix} 1 & e^{j\omega_k} & \cdots & e^{j\omega_k(N-1)} \end{bmatrix}}_{\mathbf{w}_k} \end{aligned}$$

Thus

$$\mathbf{x}(i) = \mathbf{A}\mathbf{\Lambda}_\omega(i)\mathbf{c} + \mathbf{e}(i) \quad (6.7)$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\omega_0} & e^{j\omega_1} & \cdots & e^{j\omega_{K-1}} \\ e^{j2\omega_0} & e^{j2\omega_1} & \cdots & e^{j2\omega_{K-1}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(N-1)\omega_0} & e^{j(N-1)\omega_1} & \cdots & e^{j(N-1)\omega_{K-1}} \end{bmatrix} \quad (6.8)$$

is a Vandermonde matrix independent of i , and

$$\mathbf{\Lambda}_\omega(i) = \text{diag} \{ e^{j\omega_0 i}, e^{j\omega_1 i}, \dots, e^{j\omega_{K-1} i} \} \quad (6.9)$$

Define the data matrix to be

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \mathbf{x}(M) \end{bmatrix}. \quad (6.10)$$

Then the sample autocorrelation matrix is

$$\hat{\mathbf{R}} = \frac{1}{M} \mathbf{X}\mathbf{X}^\dagger = \frac{1}{M} \sum_{i=1}^M \mathbf{x}(i)\mathbf{x}^\dagger(i) \quad (6.11)$$

For large M this can be approximated as

$$\widehat{\mathbf{R}} \approx \mathbf{A}\mathbf{\Lambda}_c\mathbf{A}^\dagger + \sigma_e^2\mathbf{I}_N \quad (6.12)$$

where $\mathbf{\Lambda}_c = \text{diag} \{|c_0|^2, |c_1|^2, \dots, |c_{K-1}|^2\}$. (Please see the Appendix for a proof of (6.12)).

Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ be the eigenvalues of $\widehat{\mathbf{R}}$. Since $\text{Rank}(\mathbf{A}\mathbf{\Lambda}_c\mathbf{A}^\dagger) = K$, it can be shown that $\lambda_{K+1} = \lambda_{K+2} = \dots = \lambda_N = \sigma_e^2$. These are commonly referred to as the noise eigenvalues, and their corresponding eigenvectors $\mathbf{U}_e = [\mathbf{u}_{K+1}, \mathbf{u}_{K+2}, \dots, \mathbf{u}_N]$, as the noise eigenvectors. Using (6.12), we obtain

$$\mathbf{A}\mathbf{\Lambda}_c\mathbf{A}^\dagger\mathbf{U}_e = \mathbf{0} \quad (6.13)$$

As long as $N \geq K$, $\mathbf{\Lambda}_c$ will have a full rank and \mathbf{A} will have a full column-rank, because $c_k \neq 0$ and ω_k are distinct in $[-\pi, \pi)$. So (6.13) is equivalent to:

$$\mathbf{A}^\dagger\mathbf{U}_e = \mathbf{0} \quad (6.14)$$

That is, the complex-exponentials in (6.3) turn out to be orthogonal to the noise eigenspace. So one can then use the following to estimate the ω_k :

$$\min_{\omega \in (-\pi, \pi]} \|\mathbf{a}^\dagger(\omega)\mathbf{U}_e\|_2^2 \quad (6.15)$$

where $\mathbf{a}(\omega) = [1, e^{j\omega}, e^{2j\omega}, \dots, e^{(N-1)j\omega}]^T$. It can be proved [15], [89] that as long as $N > K$, the only complex-exponentials that are orthogonal to the noise eigenspace are those in (6.3). Hence, there will be no spurious estimates when solving (6.15).

Now, for applications with periodicity, MUSIC by itself does not exploit the fact that the lines in the spectrum are harmonically spaced (Fig. 7.1). Taking this into account, Christensen et al. [16] proposed to modify (6.15) as

$$\min_{\omega \in (-\pi, \pi]} \min_K \frac{\|\mathbf{B}^\dagger(\omega)\mathbf{U}_e\|_F^2}{KN(N-K)} \quad (6.16)$$

where $\mathbf{B}(\omega) = [\mathbf{a}(0), \mathbf{a}(\omega), \mathbf{a}(2\omega), \dots, \mathbf{a}((K-1)\omega)]$. The factor of $KN(N-K)$ is a normalization term. The resulting algorithm was called the Harmonic MUSIC (HMUSIC) algorithm. It was further generalized to the case of mixtures of periodic signals in [17] as follows:

$$\min_{\{K_l\}_{l=0}^{Q-1}} \min_{\{\omega_l\}_{l=0}^{Q-1}} \sum_{l=0}^{Q-1} \frac{\|\mathbf{B}^\dagger(\omega_l)\mathbf{U}_e\|_F^2}{KN(N-K)} \quad (6.17)$$

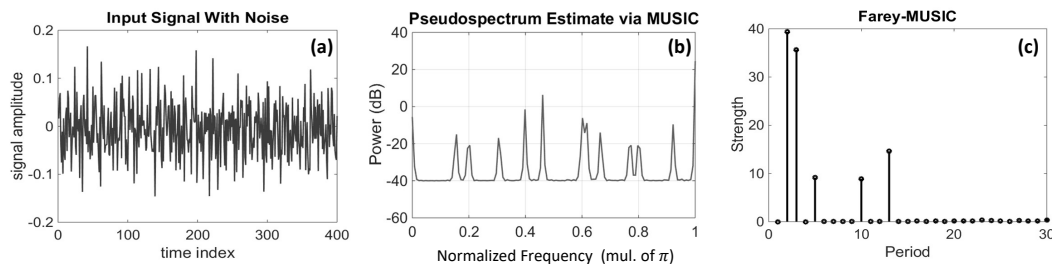


Figure 6.4: Demonstrating the proposed iMUSIC method using a Farey dictionary on a mixture of periods 3, 10 and 13. (a) The noisy periodic signal, (b) conventional MUSIC, and (c) the new iMUSIC method in (6.19).

where Q is the number of component periods in the mixture, and K is the total signal space dimension. These algorithms were shown to offer better estimates than MUSIC in the context of pitch estimation [16], [17]. However, notice that both (6.16) and (6.17) involve computationally intensive discrete optimizations. For signals that can be approximated well by the integer period model of (7.19) (such as DNA and Protein repeats), we can develop much simpler techniques with a significantly higher accuracy as well. We shall present these next.

6.2 The Proposed Methods

As explained above, when the Vandermonde vectors (columns of \mathbf{A}) in Eq. (6.8) have a harmonic structure, it can be exploited to improve the MUSIC algorithm (e.g. HMUSIC [16]). Now, when $x(n)$ has integer period $\leq P_{max}$, the frequencies ω_i in (6.8) can only have the specific form:

$$2\pi\frac{k}{q}, \quad 1 \leq k \leq q, (k, q) = 1 \quad \text{where } 1 \leq q \leq P_{max} \quad (6.18)$$

That is, the Vandermonde vectors are similar to the atoms in the Farey dictionary (Chapter 3). In this case there is a different way to define the MUSIC spectrum which works much better than traditional MUSIC and HMUSIC. We refer to this as **Farey MUSIC**; as we shall see below, this is more than just restricting the computation of traditional MUSIC spectrum to the Farey grid. Since the Farey MUSIC algorithm is specifically designed to find integer periods, we also call it **iMUSIC** (where the i stands for integer period). Also, replacing the Farey atoms with other types of nested periodic bases leads to several generalizations of iMUSIC, as we shall see in Sec. 6.3.

Let us begin by assuming that $x(n)$ in (6.3) is a period- P signal. So the K columns

of \mathbf{A} in (6.8) are a subset of the atoms of the Farey dictionary, with periods being divisors of P . We can follow the derivation in Sec. 6.1 to obtain (6.14). That is, the atoms of the Farey dictionary that span the signal turn out to be orthogonal to the noise subspace. As long as N , the size of the snapshots in (6.4), is larger than K , no other Farey atoms will satisfy (6.14). At this point, we propose the following alternative to the MUSIC (Eq. (6.15)) and HMUSIC pseudo-spectra (Eqs. (6.16) and (6.17)): For every integer P , we compute

$$S_F(P) = \frac{1}{\phi(P)} \sum_{m=1}^{\phi(P)} \frac{1}{\|\mathbf{U}_e^\dagger \mathbf{s}_P^{(m)}\|_2^2} \quad (6.19)$$

where $\{\mathbf{s}_P^{(m)}\}_{m=1}^{\phi(P)}$ are the $\phi(P)$ period- P atoms of the Farey dictionary. The $\phi(P)$ term in the denominator is a normalizing factor. *A plot of (6.19), with the integer P as the x-axis is the discrete iMUSIC pseudospectrum based on the Farey dictionary.*

Note that Eq. (6.19) is not just restricting ordinary music to a special non-uniform grid. It differs from classical MUSIC and HMUSIC in the following ways:

1. Ramanujan Subspaces: Eq. (6.19) consolidates all the Farey atoms in each Ramanujan subspace into one sum. In this way each iMUSIC spectrum line is for one Ramanujan subspace. The LCM property of the Ramanujan subspaces applied to the peaks of (6.19), yields the period. The number of lines is therefore different from the number of lines in ordinary MUSIC.

Notice that HMUSIC in (6.16) and (6.17) also groups together harmonic multiples of a fundamental frequency. But while HMUSIC combines all the harmonics of the fundamental, in Farey-MUSIC, for every period P , we only combine those $e^{j2\pi kn/P}$ for which $(k, P) = 1$.

2. Mixtures of Periodic Signals: Unlike HMUSIC in (6.17), the complexity of iMUSIC does not increase with the number of component signals in a mixture, or with the number of harmonics for each component. The complexity of HMUSIC increases exponentially with the number of hidden periodic components Q in (6.17). This is because the number of ways in which K_l 's in (6.17) can be chosen to add up to the total signal space dimension K increases exponentially with Q . This follows from the theory of partitions [101]. The proposed iMUSIC on the other hand just needs to compute (6.19), irrespective of the number of hidden periodic components.

3. The Period of a Complex-Exponential: There is a subtle distinction in how we

interpret the period of a complex exponential. In prior works, the period of $e^{j2\pi kn/P}$ was interpreted as P/k . However, we follow the strict integer period definition as given in (7.19), so that the period is actually $P/\gcd(P, k)$.

All these differences when put together, result in significantly better accuracies and much simpler algorithmic complexity for integer period estimation, as will be seen in Sec. 6.4. Before proceeding, we will show a simple demonstration of the iMUSIC equation (6.19). Fig. 6.4(a) shows a sum of randomly generated signals with periods 3, 10 and 13 and SNR 5dB. The total signal length (L in Sec. 6.1) was 400. This signal was broken down into successive blocks of length 101 samples (N in (6.4)). K , the number of complex exponentials in (6.3), turns out to be 24 for this choice of periods. In practice, this true value of K is unknown a priori, so we estimate it here using a simple metric: all the eigenvalues of the auto-correlation matrix smaller than 5% of the maximum eigenvalue were considered as noise eigenvalues. Fig. 6.4(b) shows the conventional MUSIC pseudospectrum for reference. The peaks correspond to periods 12.79, 9.85, 6.56, 5.02, 4.34, 3.32, 3.24, 3.01, 2.59, 2.51 and 2.17. Notice that it is quite inconvenient to spot the true periods 3, 10 and 13 from this set. Fig. 6.4(c) shows the iMUSIC pseudospectrum computed using (6.19). It is easy to identify distinct peaks at periods 2, 3, 5, 10 and 13. Using the LCM property, we can deduce that these correspond to periods 3, 10 and 13.

6.3 Generalizing iMUSIC From Farey Atoms to Other NPS Bases

Eq. (6.19) can alternatively be implemented using integer valued basis vectors instead of complex exponentials. This can be done using the Nested Periodic Subspaces (NPSs) [23], [37] described in Chapter 3. The NPSs are generalizations of Ramanujan subspaces, and include several examples of integer bases for representing periodic sequences (Fig. 6.3). In fact, as explained in Chapter 3, the Ramanujan subspaces can themselves be spanned by integer valued vectors instead of the Farey atoms.

Algorithmically, generalizing the iMUSIC spectrum using such NPSs is done as follows: We compute the following for every integer P instead of (6.19):

$$S_N(P) = \frac{1}{\phi(P)} \sum_{m=1}^{\phi(P)} \frac{1}{\|\mathbf{U}_e^\dagger \mathbf{b}_P^{(m)}\|_2^2} \quad (6.20)$$

where $\{\mathbf{b}_P^{(m)}\}_{m=1}^{\phi(P)}$ are the $\phi(P)$ period- P NPS basis vectors. Using the LCM property of NPSs, we can once again determine the period from the peaks of Eq. (6.20).

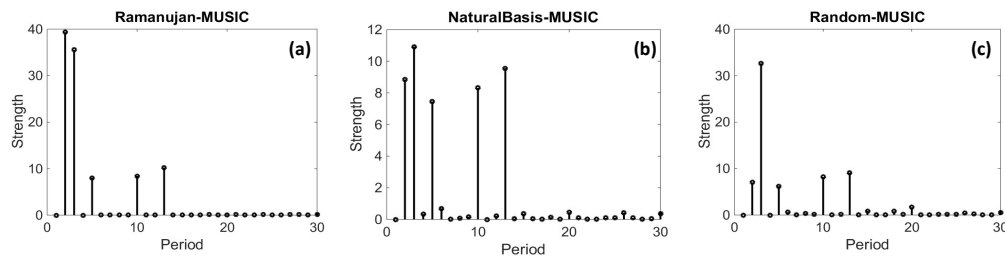


Figure 6.5: Demonstrating the NPS based iMUSIC methods on the signal shown in Fig. 6.4 using (a) a Ramanujan dictionary, (b) a natural basis dictionary and (c) a randomly generated NPS dictionary.

For example, Fig. 6.5 shows plots of Eq. (6.20) vs. P for various integer valued NPS bases, for the signal shown in Fig. 6.4(a). When the atoms $\mathbf{b}_p^{(m)}$ come from a Ramanujan dictionary (Fig. 6.3(b)), we call (6.20) as Ramanujan MUSIC. Natural basis MUSIC and Random NPS MUSIC can be defined similarly using their respective dictionaries [23]. All these plots have clean peaks at periods 2, 3, 5, 10 and 13. Using the LCM property of NPSs, it is easy to see that they represent periods 3, 10 and 13.

Although the above idea is simple, the non-Vandermonde nature of the NPS bases introduces several challenges in the mathematical formulation of (6.20) when compared to classical MUSIC. So in the remainder of this section, we will study (6.20) in a rigorous fashion. To start with, let us assume that $x(n)$ is a period P signal. Following the derivations of Sec. II, we can arrive at (6.13). Now, since $x(n)$ has integer period, the columns of \mathbf{A} themselves have integer periods (atoms of the Farey dictionary), and hence can be spanned by any other set of NPSs (such as say natural basis). We can write this as:

$$\mathbf{A}_{N \times K} = \mathbf{B}_{N \times K'} \mathbf{T}_{K' \times K} \quad (6.21)$$

where the K' columns of \mathbf{B} are the basis vectors of the other NPS. It can be shown [23] that the lcm property applied to the columns of either \mathbf{A} or \mathbf{B} yields the same answer, namely P . It is useful to consider two separate cases at this point, depending on whether K equals K' .

The Case When $K = K'$

In general, K and K' can be different. But to start with, we assume $K = K'$ since it is the most common situation in applications with integer periods. Conceptually, $K = K'$ means that one would require the same number of NPS basis vectors to span

the snapshots, no matter which NPS is chosen. For instance, if $x(n)$ was randomly generated, say by repeating a $P \times 1$ Gaussian random vector \mathbf{x}_P , then it can be shown that $K = K'$ with probability 1. This also applies to mixtures of periodic signals, when each component signal is randomly generated. In applications such as DNA and protein repeats, where the nucleotides or the amino acids are mapped to numbers using scales such as the molecular size, hydrophobicity etc. [102], it is quite natural to expect that $K = K'$. The case of $K \neq K'$ will be addressed later in Sec. 6.3.

We can re-write (6.13) using (6.21) as follows:

$$\mathbf{B}\mathbf{T}\mathbf{\Lambda}_c\mathbf{T}^\dagger\mathbf{B}^\dagger\mathbf{U}_e = \mathbf{0} \quad (6.22)$$

In (6.21), as long as $N > K$, \mathbf{A} will have a full column rank (Vandermonde property). This implies that \mathbf{B} and \mathbf{T} will also have full column ranks $K (= K')$, and hence $\mathbf{B}\mathbf{T}\mathbf{\Lambda}_c\mathbf{T}^\dagger$ will have a full column rank in (6.22). So (6.22) is equivalent to:

$$\mathbf{B}^\dagger\mathbf{U}_e = \mathbf{0} \quad (6.23)$$

Notice that this is similar to (6.14), but involves the columns of an NPS dictionary rather than complex exponentials. So given such an NPS dictionary, plotting (6.20) for every period $\leq P_{max}$ will result in peaks at periods corresponding to the columns of \mathbf{B} . So we may think of using the LCM property on those peaks to estimate the period. But before we can do so, just like in classical MUSIC, we need to address the following question first: Can there be NPS basis vectors other than the columns of \mathbf{B} that are also orthogonal to \mathbf{U}_e ?

We have an interesting deviation from classical MUSIC in this aspect. While we cannot guarantee the absence of such additional (spurious) NPS basis vectors producing peaks in (6.20), we can nevertheless prove that *any such additional peaks will not affect the period estimate*. To see this, we first need the following result which will be proved in Chapter 7:

Theorem 6.3.1. *Let $x(n)$ be a noiseless periodic signal whose period is known to lie in the integer set $\mathbb{P} = \{P_1, P_2, \dots, P_K\}$. To be able to uniquely identify its period using L consecutive samples, it is both necessary and sufficient that:*

$$L \geq L_{min} = \max_{P_i, P_j \in \mathbb{P}} P_i + P_j - (P_i, P_j) \quad (6.24)$$

◇

The above result is a fundamental identifiability result that is independent of which estimation technique is used **myPrep**. We will use it to prove the following:

Theorem 6.3.2. *Suppose the period of $x(n)$ in (6.3) is known a priori to lie in the integer set $\mathbb{P} = \{P_1, P_2, \dots, P_K\}$. If N , the length of the snapshots in (6.4), satisfies:*

$$N \geq L_{min} = \max_{P_i, P_j \in \mathbb{P}} P_i + P_j - (P_i, P_j) \quad (6.25)$$

then the LCM of the periods of all the NPS basis vectors that are orthogonal to \mathbf{U}_e , will be equal to the true period of the signal.

Proof: Let us assume that the input's period is P . As mentioned earlier, the LCM of the periods of the columns in \mathbf{A} in (6.8) and \mathbf{B} in (6.21) will be equal to P . Suppose \mathbf{b} is an NPS basis vector that is not a column of \mathbf{B} , but still satisfies $\mathbf{b}^\dagger \mathbf{U}_e = \mathbf{0}$. There are two possibilities:

Case (i) Period of \mathbf{b} divides P . In this case, even if $\mathbf{b}^\dagger \mathbf{U}_e = \mathbf{0}$, a peak in the psuedospectrum at period of \mathbf{b} will not change the LCM estimate. So such a spurious peak will not lead to a false period estimate.

Case (ii) Period of \mathbf{b} does not divide P . We will show using contradiction that such a \mathbf{b} cannot exist. If there was such a \mathbf{b} , then \mathbf{b} , along with the columns of \mathbf{B} will constitute $K + 1$ vectors in the K dimensional null-space of \mathbf{U}_e^\dagger . When N satisfies (6.25), it follows in particular that $N > P_{max} \geq K$. $N > K$ implies that \mathbf{A} in (6.21) will have full column rank (Vandermonde property), and so \mathbf{B} will also have full column rank K (recall that we assumed $K = K'$ to start with). This would mean that \mathbf{B} is a basis for the null space of \mathbf{U}_e^\dagger , and so:

$$\mathbf{b} = \mathbf{B}\mathbf{v} \quad (6.26)$$

for some vector \mathbf{v} . Notice that the L.H.S. is a length N segment of a signal whose period does not divide P . The R.H.S. is a segment of a signal whose period necessarily divides P , since the columns in \mathbf{B} are NPS basis vectors whose periods are divisors of P . As long as $N \geq L_{min}$ according to Theorem 6.3.1, such an ambiguity in identifying the period is not possible. Hence we arrive at a contradiction to the existence of such a \mathbf{b} .

Remark 1: When the set of possible periods in Theorem 6.3.2 is $\mathbb{P} = \{1, 2, 3, \dots, P_{max}\}$, L_{min} turns out to be $2P_{max} - 2$.

Remark 2: Theorem 6.3.2 can be generalized to mixtures of periodic signals. If $x(n)$ were a mixture of M periodic signals with periods in $\{1, 2, 3, \dots, P_{max}\}$, then the minimum N is approximately:

$$N \geq 2MP_{max} \quad (6.27)$$

For readers already familiar with Chapter 7, (6.27) is in fact an approximation of the following precise lower bound:

$$N \geq N_{min} = \max_{\substack{\mathbb{P}_i, \mathbb{P}_j \subset \mathbb{P} \\ \mathbb{P}_i, \mathbb{P}_j \text{ are} \\ M\text{-sets of size } = N}} \sum_{d \in D.S.(\{\mathbb{P}_i \cup \mathbb{P}_j\})} \phi(d) \quad (6.28)$$

The proof is based on the generalization of Theorem 6.3.1 to mixtures of periodic signals **myPrep**. The details are quite similar to the above proof, so we will skip them here.

Remark 3: Theorem 6.3.2 is tight in the following sense: It is possible to construct examples of NPSs for which spurious peaks will affect the period estimate when N doesn't satisfy (6.25). But for most NPSs, a smaller N may be sufficient. For instance, if we use the Farey atoms, it is easy to show using their Vandermonde structure [89] that we just need:

$$N > \max_{P_i \in \mathbb{P}} P_i \quad (6.29)$$

instead of (6.25) in Theorem 6.3.2. However, deriving the precise necessary and sufficient bounds for other NPSs that do not have a Vandermonde structure is difficult. Theorem 6.3.2 is useful in this regard.

So we have so far shown that as long as $K = K'$, and the snapshot length satisfies (6.25), the period of the signal can be estimated using (6.20). We will now discuss the case of $K \neq K'$.

The case of $K \neq K'$

Let us consider the following two cases separately:

Case A: $K > K'$: This will not happen as long as the snapshot length $N > K$, because then **A** in (6.21) will have full column rank K . So at least K linearly independent columns are needed in **B** in the R.H.S. of (6.21).

Case B: $K < K'$: This can occur in some cases. For instance, if $x(n) = e^{j2\pi n/P}$, then $K = 1$, as only one Farey column is required to span the snapshots of $x(n)$. But

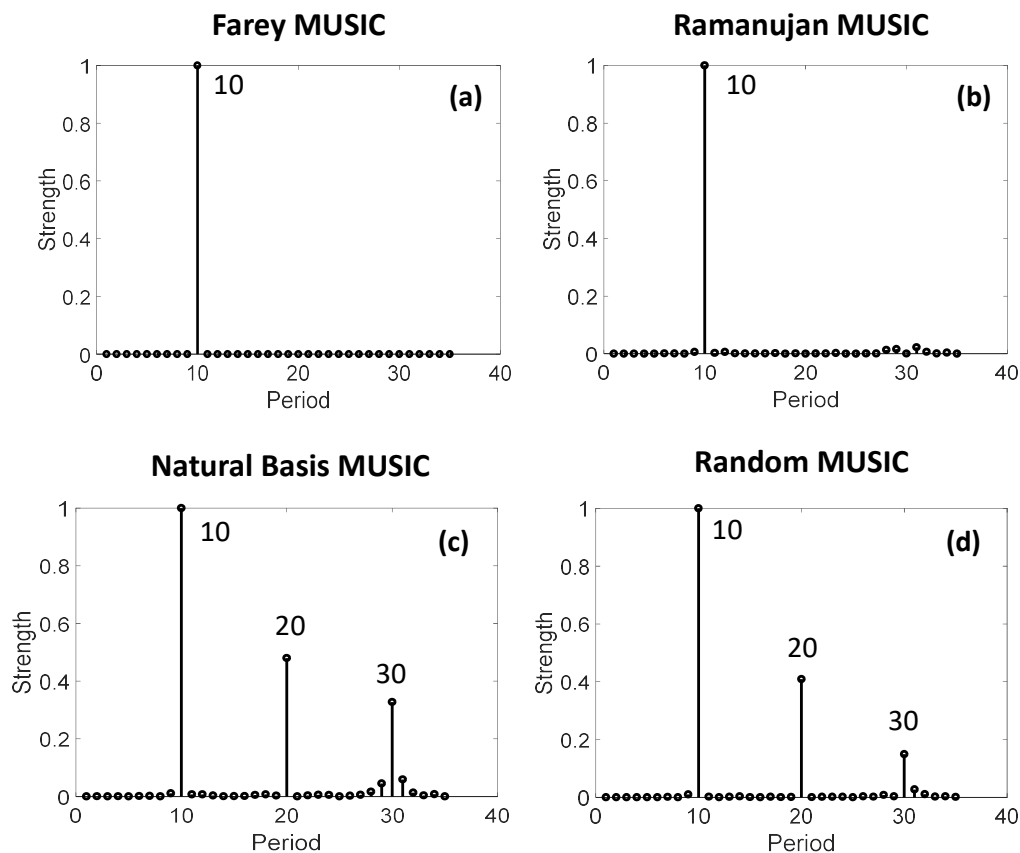


Figure 6.6: The effect of $K < K'$ on the pseudo-spectrum of (a) Ramanujan Subspaces (Farey basis) (b) Ramanujan Subspaces (integer basis) (c) Natural Basis Subspaces and (d) Randomly generated NPSs.

if we use natural basis subspaces (Fig. 7.1), then $K' = P$, as it can be shown that P basis vectors of the natural basis subspaces are needed to span each snapshot of this $x(n)$.

When $K < K'$, \mathbf{T} in (6.21) will not have a full rank. Hence, (6.22) does not imply (6.23). So it is quite possible that some of the columns of \mathbf{B} are not orthogonal to \mathbf{U}_e . So using (6.20) and lcm property is not theoretically guaranteed to give the correct period estimate. This is a fundamental limitation of any non-Farey NPS basis. Nevertheless, it was experimentally observed that:

- For iMUSIC using Ramanujan subspaces, (6.19) gave the correct period estimates even for the non-Vandermonde integer basis vectors.
- For Natural Basis subspaces and randomly generated NPSs, the only spurious peaks observed were smaller peaks at multiples of the true period. So the

period could still be estimated upto a multiple.

As an illustration, let us consider the following signal: $x(n) = e^{j2\pi n/10}$. For this signal, we would need only one Farey basis vector to span its snapshots, while we would need 10 basis vectors from Ramanujan integer basis, and similarly from the Natural Basis subspace. For randomly generated NPSs as well, we would need 10 basis vectors with probability 1. So for each of these other NPSs, we have $K' = 10$ in (6.21). Fig. 6.6 shows the pseudo-spectra obtained from each of these NPSs using (6.20). As is evident, $K < K'$ wasn't really an issue when using Ramanujan subspaces, even for the non-Vandermonde integer basis vectors. There were no spurious peaks or missing peaks. For the natural basis and the randomly generated NPSs, we can see spurious peaks at multiples of the true period, which is 10. So the period can only be estimated up to a multiple of the true period. Hence in practice, for signals such as pure sinusoids, it is recommended to use the Ramanujan subspaces instead of more general NPSs.

This completes the formulation of the NPS based iMUSIC algorithms.

6.4 Experiments

In this section, we present several examples and comparisons to highlight various aspects of the proposed methods.

Comparison of Estimation Accuracies

Figs. 6.7 and 6.8 compare the accuracy of period estimation for several techniques as a function of SNR. For each SNR shown, 200 Monte Carlo trials were carried out with randomly generated signals, each signal being an additive mixture of periods 10 and 14. The total datalength for each signal was $L = 500$ samples. The snapshot length (N in (6.4)) was chosen as 150. K , the number of complex exponentials in (6.3), turns out to be 23 in this case. Since the signals were randomly generated, K' in (6.21) turns out to be equal to K for all NPSs. For simplicity, we assumed that the value of the total signal space dimension K is known to all the methods here, including MUSIC and HMUSIC. P_{max} , the maximum period that is searched for in the signal, was chosen to be 20. The probability of correct estimation is plotted for each method¹, which is the fraction of trials in which the detected periods were

¹Mean Squared Error (MSE), a popular metric in general, is not appropriate here due to two reasons: (a) Different methods could detect different number of component periods. It is not straightforward to compare vectors of different lengths using MSE. (b) Estimating the period upto a multiple may be acceptable in many applications. For example, in Fig. 6.2, proteins with Ankyrin

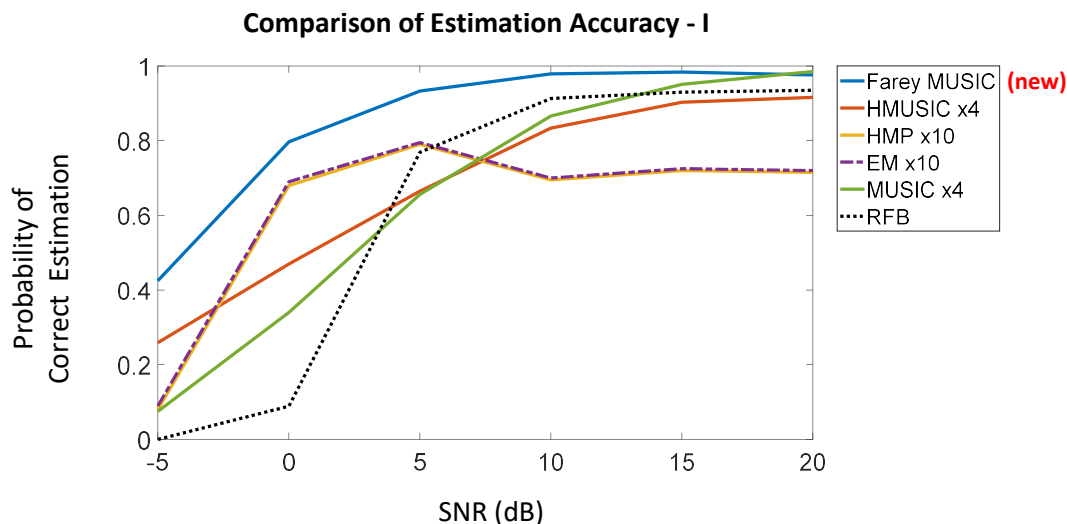


Figure 6.7: Probability of Estimating both the component periods exactly. Comparison of the proposed Farey-MUSIC with other techniques. See

exactly equal to the set $\{10, 14\}$. For visual clarity, we have split the different methods into two figures, Figs. 6.7 and 6.8. Our observations are as follows:

1. Traditional MUSIC and HMUSIC: Both MUSIC and HMUSIC [16], [17], [94], evaluated on a frequency grid that has the same size as the Farey grid, yield probability of correct estimation close to 0 even at high SNRs. They required at least four times denser grids than the Farey grid to reach the performances shown in Figs. 6.7 and 6.8. Further increase in the grid size did not improve their accuracy significantly. For a fair comparison, the period estimates of both these methods were rounded to the nearest integers. In fact, HMUSIC here was given the true value of the number of hidden periodic components in the mixture (Q in (6.17)) as an input. It was also given the number of harmonics for each component signal (K_l in (6.17)) as inputs. Still, both HMUSIC and MUSIC do not perform as well as the proposed iMUSIC methods (unless we use randomly generated NPSs for iMUSIC). With four times denser grids, HMUSIC and MUSIC are able to outperform random NPSs.

2. Prior Ramanujan-Subspace based methods: An alternate way of using NPSs for period estimation is using compressed-sensing based dictionary methods [23], [26]. While they work very well compared to other methods for very short

repeats are known to have periods in the range 30 - 40. So it might be more acceptable to estimate 66 as the period, instead of say 40, since we can readily deduce that 66 might actually indicate period 33 repeats. MSE on the other hand penalizes 66 more than 40. In any case, probability of correct estimation is in fact a stricter metric than MSE.

datalengths [23], [37], for the parameters considered here, the dictionaries turn out to be tall. Least squares based approach using such dictionaries can be efficiently implemented as a filter bank called the Ramanujan Filter Bank (RFB) [103], [104]. As seen in Fig. 6.7, the proposed iMUSIC methods outperform RFB, especially at smaller SNRs. It is useful to note however that the most appropriate applications for the RFB are signals exhibiting localized or time varying periodicity such as chirps [103], [104].

3. Harmonic Matching Pursuit and Expectation Maximization: Fig. 6.7 also shows the performance of two other multi-pitch methods: Expectation Maximization [94], [98] and the Harmonic Matching Pursuit [94], [97] algorithms. Once again, both these methods give close to 0 probability of correct estimation when their frequency grid sizes are comparable to those of the Farey dictionary. Both of them required at least 10 times denser grids than the Farey grid to achieve the accuracies shown in Fig. 6.7, which made them significantly more computationally expensive.

4. The proposed iMUSIC methods: As seen in Figs. 6.7 and 6.8, iMUSIC using the Farey dictionary clearly outperforms all other methods considered here. iMUSIC based on natural basis performs equally well at higher SNRs (Fig. 6.8). Recall that the natural basis vectors are very sparse with a few 1's and mostly 0's (Fig. 6.3). An interesting observation is that, although Farey and Ramanujan dictionaries both span the same (Ramanujan) subspaces, Farey based iMUSIC performs better. This might be because the Farey columns tend towards orthogonality for large enough N , while the Ramanujan integer basis vectors do not.

Apart from the methods considered above, there are several other techniques that are popular in the literature such as [3]–[7], [95], [105]. While being the state-of-the-art for applications such as pitch estimation, these methods cannot be directly used in the above example since they are not easily extended to the case of mixtures of periodic signals. It is important to note that, although iMUSIC outperforms the other techniques in Figs. 6.7 and 6.8, these other methods, including the aforementioned papers on single pitch estimation, can handle the more general case of non-integer periods. Whether the iMUSIC algorithms can be adapted to such applications requires a detailed analysis in itself, and will be a part of our future research. In a following subsection, we will compare the iMUSIC methods with the state-of-the-art for an application with inherently integer periods: namely, protein repeats.

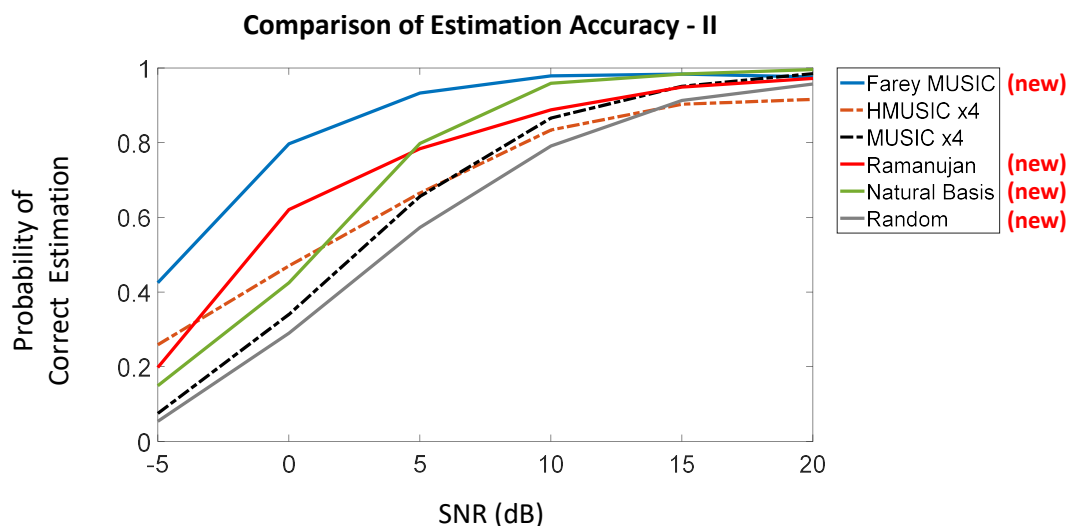


Figure 6.8: Probability of Estimating both the component periods exactly. Comparison of the various NPSs. HMUSIC and MUSIC have been included for reference.

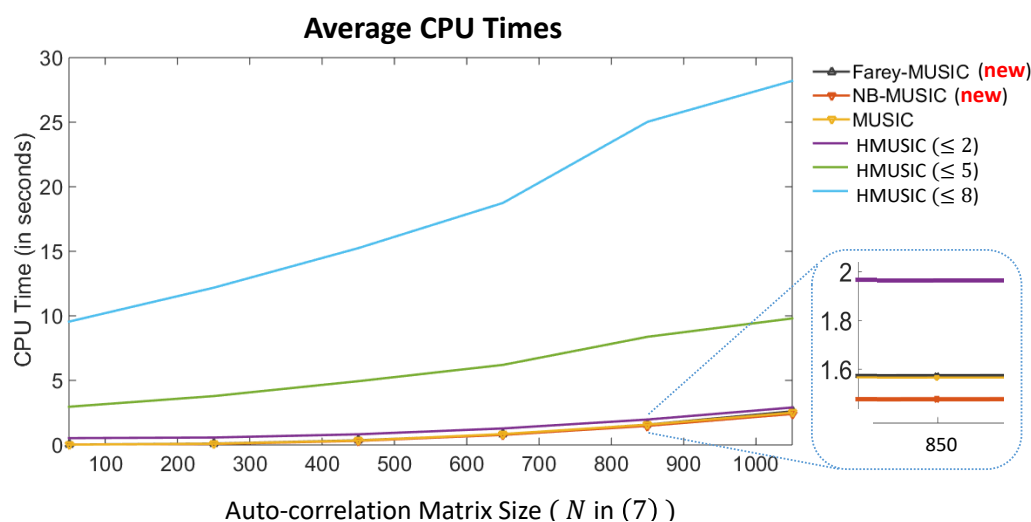


Figure 6.9: A comparison of the CPU Times. See Sec. 6.4 for details.

Comparison of CPU times for Eigenspace Methods

To show the computational savings that iMUSIC algorithms achieve over prior variants of MUSIC, Fig. 6.9 compares the average CPU times (MATLAB 2014b on a 2.4GHz CPU with 8GB RAM) as a function of the size of the autocorrelation matrix (which is also the size of the snapshots N in (6.4)). The total datalength of the signal, L was chosen as $3N$, and the dimension of the signal subspace K was fixed at 25 for simplicity. MUSIC and HMUSIC were implemented with a uniform frequency grid of the same size as the Farey grid. Recall however that both these methods typically require much more denser grids than Farey MUSIC (Sec. 6.4).

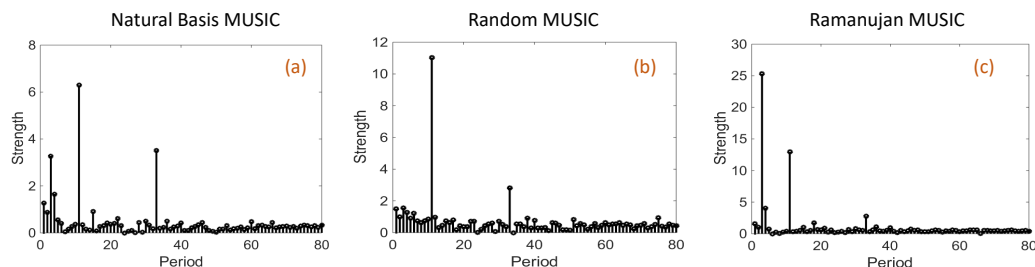


Figure 6.10: Pseudospectra of the proposed NPS based techniques for the Ankyrin protein repeats shown in Fig. 6.2.

Notice that our natural basis (NB) MUSIC is the fastest in Fig. 6.9. Farey-MUSIC and MUSIC are similar to each other in terms of CPU time due to identical grid sizes. In Fig. 6.9, $\text{HMUSIC}(\leq T)$ denotes using (6.17) with the prior knowledge that the number of hidden periodic components in the signal $Q \leq T$. As mentioned previously, the complexity of $\text{HMUSIC}(\leq T)$ increases exponentially with T . In contrast, since we check the NPS basis vectors one-by-one in (6.19), the complexity of our proposed techniques does not depend on T . From Figs. 6.8 and 6.9, it is evident that our methods offer much better accuracy for integer period estimation than prior variants of MUSIC, while keeping the computational complexity low at the same time.

Examples of Protein Repeats and DNA Microsatellites

We will now demonstrate the proposed iMUSIC algorithms on repeats in proteins. We refer the reader to Sec. 5.4 for an introduction to protein repeats. To the best of our knowledge, this is the first time any MUSIC based approach is being used for this application. In the following examples, we used the popular Kyte-Doolittle hydrophobicity scale [62] to map amino acids to numbers. To estimate the signal space dimension, we used the same 5% criteria mentioned in Sec. 6.3 (Fig. 6.4).

In our first example, we consider the protein AnkyrinR (PDB 1n11) that enables red blood cells to resist shear forces during circulation. The period 33 repeats in AnkyrinR can easily be identified in the pseudo-spectra shown in Fig. 6.10. These plots show the results of applying the proposed methods using Ramanujan (integer basis), Natural Basis and Random Integer NPSs (The Farey basis can also be used; it was shown earlier in Fig. 1). All four plots have clear peaks at 33 and its divisors. Notice that the Ramanujan (integer basis) plot in Fig. 6.10(c) has a weak peak at 33. However, the LCM of the peaks at 11 and 3 indicate the presence of the period 33 repeats.

Fig. 6.11 shows a second example, the protein Ribonuclease Inhibitor, which contains 15 luciferase-rich repeats, alternately 28 and 29 residues long [106]. So according to our definition of periodicity in (7.19), the period of these repeats is 57. This can once again be easily identified in the pseudo-spectrum plots shown in Fig. 6.11. In all our experiments, we observed that the Ramanujan Subspaces (both Farey and the integer basis) gave the cleanest plots, followed by randomly generated NPSs. The Natural Basis Subspaces, although showing the tallest peaks at the correct periods, often had smaller spurious peaks. Recall that the natural basis subspaces performed well in Sec. 6.4 at higher SNRs. But the noise there was additive, whereas here we have substitution and insertion-deletion errors. It seems that the natural basis subspaces are more sensitive than the other NPSs for such errors.

For comparison, Table I shows the period estimates of various techniques for three examples of protein repeats². Apart from the examples shown in Figs. 6.10 and 6.11, we also consider the protein HetL (PDB:3du1), which is 237 amino acids long and contains tandem pentapeptide (period 5) repeats. This protein is known to play an important role in the nitrogen fixation process in cyanobacteria [47]. In Table I, notice that the estimates of HMUSIC and MUSIC are not as accurate as those of iMUSIC. Moreover, it can be seen that HMUSIC and MUSIC were very sensitive to errors in the estimation of signal space dimension. On the other hand, iMUSIC was very robust in this regard. While Figs. 6.10 and 6.11 used a cut-off of 5% for identifying the noise eigenvalues, the plots were very similar at 10% as well. The ‘(s)’ next to natural basis iMUSIC’s estimates in Table I indicates the presence of smaller spurious peaks in its pseudospectrum (such as the one at period 38 in Fig. 6.11).

Table 1 also compares three state of the art techniques used for protein repeats. RADAR [48] and TRUST [49] algorithms are based on self-alignment techniques (trace matrices and dynamic programming), while HHrepID **HHrepID** uses Hidden Markov Models. Once again, iMUSIC performs well in comparison to these methods.

As a final example, we show an instance of tandem repeats in the human DNA in Fig. 6.12. Such repeats in the DNA are of significance in a number of contexts(see

²HMUSIC, as described in [16], [94], uses a computationally intensive discrete optimization to find the signal space dimension. However, the implementation of HMUSIC provided by its author Christensen (as a part of [94]) requires the user to specify the signal space dimension as an input. For simplicity, we used the same 5% (and 10%) rule that we used with iMUSIC, for estimating the signal space dimension for HMUSIC and traditional MUSIC as well.

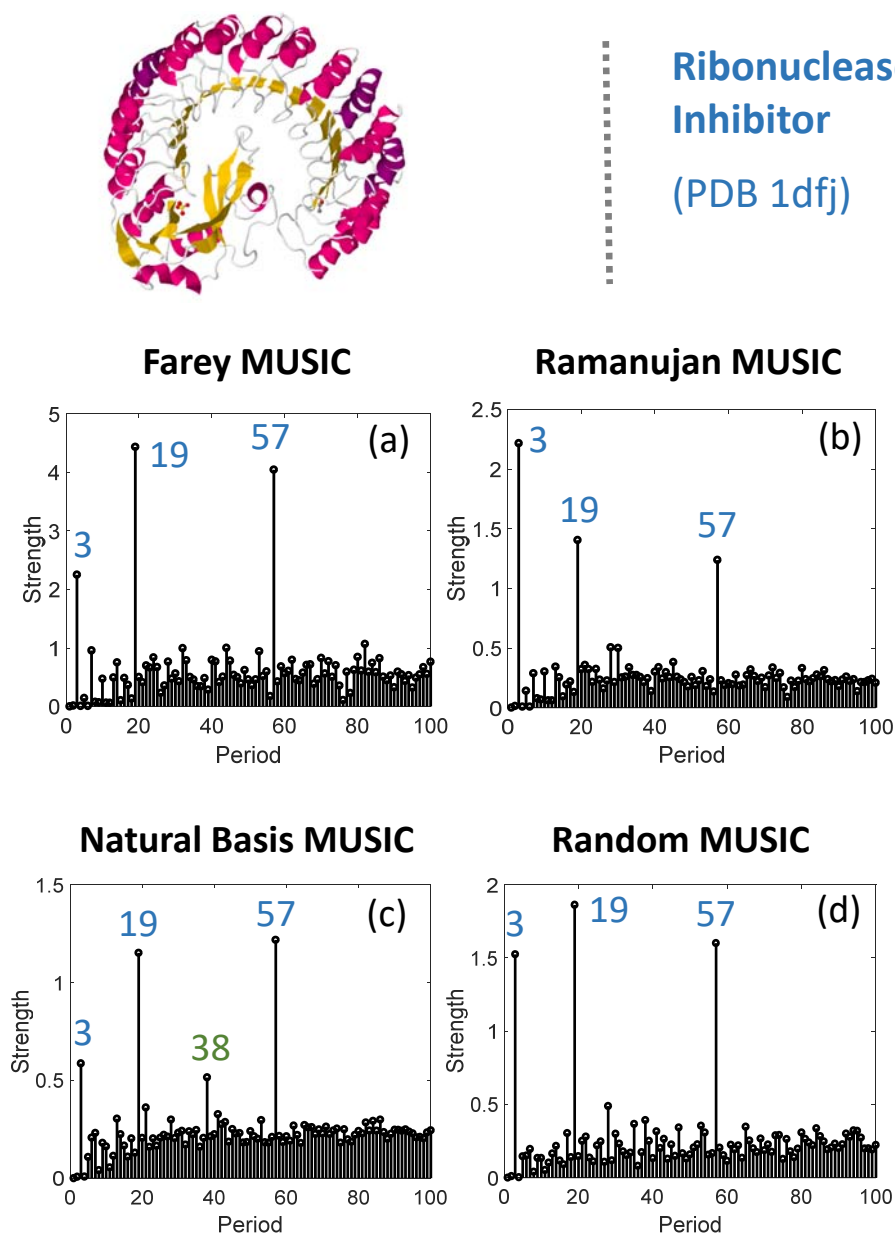


Figure 6.11: Top: The protein Ribonuclease Inhibitor (PDB: 1dfj) exhibiting luciferin-rich repeats. The pseudo-spectra obtained from (a) Ramanujan Subspaces (Farey basis) (b) Ramanujan Subspaces (Integer Basis) (c) Natural Basis Subspaces and (d) a randomly generated NPS, are shown. See

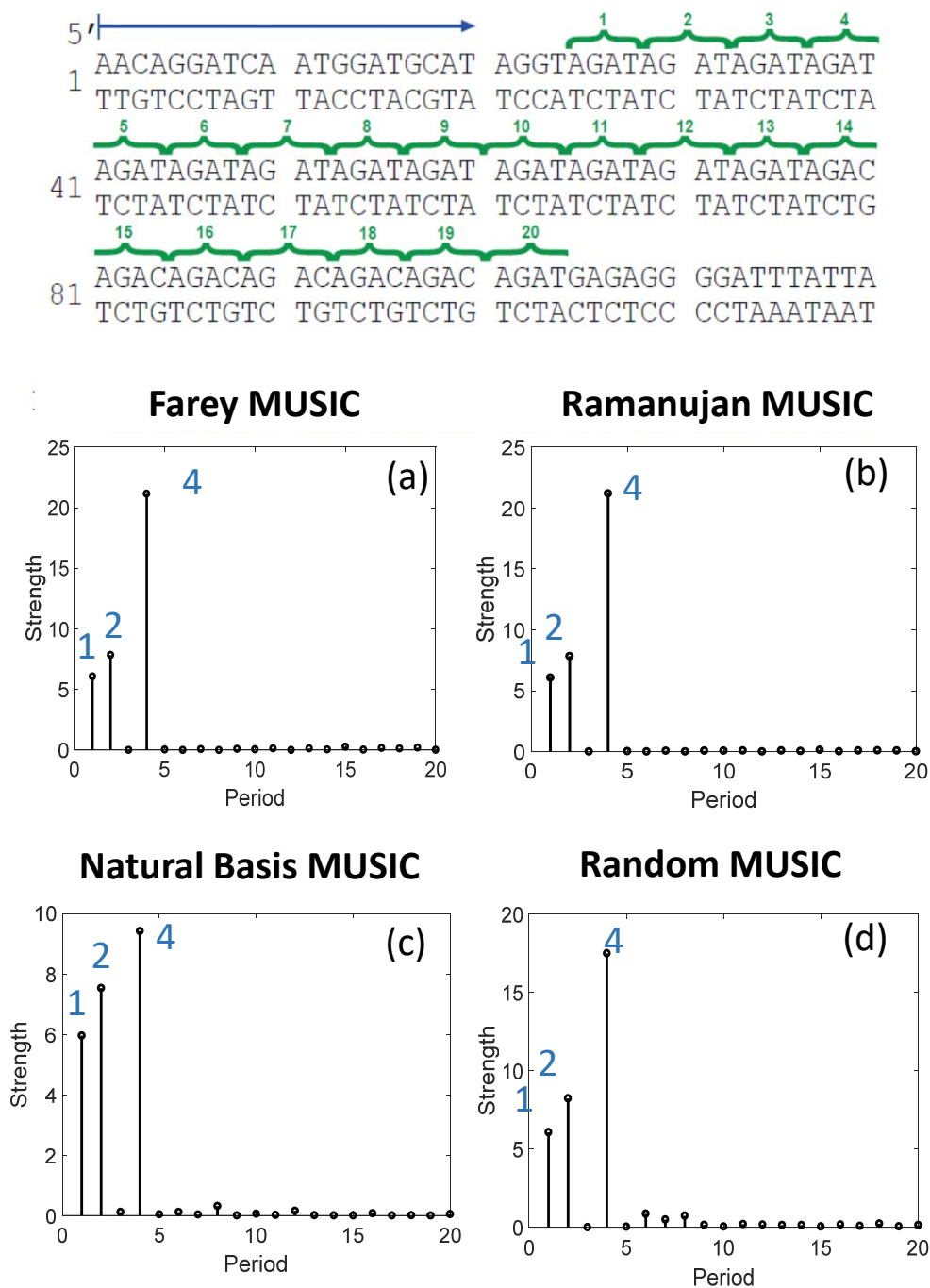


Figure 6.12: Top: An example of DNA microsatellites that are used in DNA fingerprinting. The pseudo-spectra obtained from (a) Ramanujan Subspaces (Farey basis) (b) Ramanujan Subspaces (Integer Basis) (c) Natural Basis Subspaces and (d) a randomly generated NPS, are shown. See

Table 1: Protein Repeats Comparison

Protein	1n11	1dfj	3du1
True Period	33	57	5
HMUSIC (5%)	66	82	40
HMUSIC (10%)	66	59	5
MUSIC (5%)	33	28	39
MUSIC (10%)	33	28	5
RADAR	27	26	23
TRUST	33	57	10
HHrepID	33	57	10
iMUSIC Farey	33	57	5
iMUSIC Ramanujan	33	57	5
iMUSIC Natural Basis	33	57 (s)	5 (s)
iMUSIC Random	33	57	5

Sec. 5.5 for details). For example, they are the primary bio-markers used today in DNA fingerprinting, kinship analysis etc. [69], [77], [78]. They are also associated with several genetic disorders such as the fragile X syndrome, myotonic dystrophy, Huntington’s disease and Friedreich’s ataxia [69]. Fig. 6.12 shows an example of repeats (GenBank G08921) that are used in DNA fingerprinting. We mapped nucleotides to numbers using a randomly generated mapping. The iMUSIC psuedospectra can easily identify the period 4 repeats.

While the above examples do demonstrate the proposed methods as good candidates for these applications, a more thorough experimental evaluation of their performance in comparison with prior works in these application domains is still necessary. Such an analysis merits a much broader discussion than the scope of this chapter. Our focus here has been to introduce and establish these methods on a sound theoretical footing. Tailoring them for specific applications such as DNA and protein repeats will be a part of our future work.

6.5 Conclusion

This chapter presents a new family of MUSIC-like algorithms for integer period estimation, based on Ramanujan subspaces [18] and nested periodic subspaces [23]. These new algorithms offer very simple integer valued basis vectors for spanning the signal space, and result in significantly better accuracy and computational simplicity than existing techniques for integer periods. The non-Vandermonde nature of the basis vectors introduces a number of subtle differences from the traditional MUSIC formulation. These were carefully addressed in this chapter. A number of simulation

experiments were presented demonstrating these algorithms, including examples from protein and DNA repeats.

While the model in (7.19) is especially relevant to applications with inherent integer periodicity, many state of the art methods for conventional periodicity applications such as in speech [12], [95], [96] also use integer period approximations. Adapting our techniques for such applications will be of interest to us in our future work. Even for integer period applications such as proteins and DNA repeats, we are interested in specifically tailoring our algorithms for each of these applications in a more thorough fashion by optimizing over larger databases, and comparing their performance with the existing state of the art methods in those domains. Finally, apart from MUSIC, techniques such as ESPRIT [107] and the recent atomic norm based methods [108], [109] are also popularly used for various line spectral applications. While we specifically focused on MUSIC in this chapter, it will be very interesting to see if we can similarly adapt these other techniques for harmonic spectra in the future.

6.6 Chapter Appendix

Proof of (6.12): Substituting from (6.7), the signal component of (6.11) is

$$\mathbf{V} \left(\frac{1}{M} \sum_{i=1}^M \Lambda_{\omega}(i) \mathbf{c} \mathbf{c}^{\dagger} \Lambda_{\omega}^{\dagger}(i) \right) \mathbf{V}^{\dagger}. \quad (6.30)$$

The noise component is

$$\frac{1}{M} \sum_{i=1}^M \mathbf{e}(i) \mathbf{e}^{\dagger}(i), \quad (6.31)$$

and the cross terms are

$$\mathbf{V} \left(\frac{1}{M} \sum_{i=1}^M \Lambda_{\omega}(i) \mathbf{c} \mathbf{e}^{\dagger}(i) \right), \quad (6.32)$$

and its transpose conjugate. The matrix inside brackets in Eq. (6.30) has ml -th element

$$c_m c_l^* \frac{1}{M} \sum_{i=1}^M e^{j(\omega_m - \omega_l)i} \quad (6.33)$$

For $m \neq l$ we have $\omega_m - \omega_l \neq 0 \pmod{2\pi}$, so (6.33) approaches zero for large M . So, (6.30) has the form $\mathbf{V} \Lambda_c \mathbf{V}^{\dagger}$ where Λ_c is a diagonal matrix with diagonal elements $|c_k|^2$. Secondly, for large M Eq. (6.31) approaches $\sigma_e^2 \mathbf{I}$. Thirdly the matrix inside brackets in Eq. (6.32) has ml -th element $c_m \sum_{i=1}^M e^{j\omega_m i} e_l^*(i) / M$. This is a zero-mean random variable with variance $|c_m|^2 \sigma_e^2 / M \rightarrow 0$ for large M . These three observations justify (6.12). $\nabla \nabla \nabla$

Chapter 7

MINIMUM DATALENGTH FOR INTEGER PERIOD ESTIMATION

7.1 Introduction

Let us consider again the integer periodicity model that we started with in Chapter 1: A signal $x(n)$ is said to be periodic if

$$x(n + P) = x(n) \tag{7.1}$$

for all n for some integer P . The smallest such nonzero integer P is said to be the period of $x(n)$.

While this model has been very popular in many applications such as speech, ECG, EEG, protein and DNA repeats, there is one fundamental question which has surprisingly never been addressed before. Namely, given a sequence $x(n)$ with integer period, what is the absolute lower bound on the data-length required to be able to identify its period? More generally, given a mixture of periodic signals, what is the absolute minimum data-length required to identify the periods of the hidden components? Notice that the bounds we seek are generic, i.e., independent of any particular technique we may choose to estimate the periods. We will also see that the definition of “hidden” integer periods is rather tricky, if we have to get unique and meaningful answers (Sec. 7.3). Some effort is therefore spent in this chapter to develop a formal definition, and also to study some interesting properties of hidden integer periods. It is rather surprising that none of these questions has been raised in the signal processing literature in the past.

Even though the bounds we seek are generic for the most part (i.e., independent of any particular technique we may choose to estimate the periods), we also address the special case of the dictionary based techniques reported in Chapter 3. For these specific methods the minimum number of samples is, not surprisingly, larger than the theoretical minimum. These method-dependent-bounds are also given in this chapter, in view of the practical usefulness of the dictionary based methods.

At the outset we would like to make it clear that development of faster and better methods for estimation of periods (integer or otherwise) is *not the main goal* of this chapter. Rather, the purpose is to develop algorithm-independent theoretical

bounds on the number of samples required for estimation of integer periods and integer hidden periods in discrete time data. True, the proof techniques that are involved in this chapter (to justify the expressions for minimum data length) are constructive, and therefore place in evidence some procedures for estimation of integer periods with minimum number of samples. But these should be regarded as proof-of-concept algorithms. They are not computationally efficient, nor robust to noise.

Chapter Outline

A general rule of thumb in prior works has been the following: To estimate the period, we need a data-length that is at least twice the largest expected period (for e.g., [95], [110]). While this was observed to be true for some particular methods such as [95], it is not a fundamental bound in the sense that it need not apply to all possible period estimation techniques. A natural question is, is there a fundamental lower bound on the data-length that applies regardless of which technique we choose?

In Sec. 7.2 we show that such a lower bound does in fact exist: To estimate the period from the plausible set $\mathbb{P} = \{P_1, P_2, \dots, P_N\}$ of integers, the absolute minimum necessary and sufficient data-length is

$$L_{min} = \max_{P_i, P_j \in \mathbb{P}} P_i + P_j - (P_i, P_j) \quad (7.2)$$

where (P_i, P_j) denotes the greatest common divisor. For example, to estimate the period from the set $\{7, 24, 100\}$, classical theory tells us that we need $N_{min} = 200$ samples, whereas the actual bound is just 120 samples. In the special case when the set of plausible periods is:

$$\mathbb{P} = \{1, 2, \dots, P_{max}\} \quad (7.3)$$

L_{min} turns out to be $2P_{max} - 2$, close to the previously used rule of thumb of $2P_{max}$ samples [95].

In Sec. 7.3 and 7.4, we derive similar results for the case of mixtures of periodic signals, where each component signal satisfies (7.1). Defining the component periods (or hidden integer periods) in a mixture in a unique and meaningful manner is rather subtle. The existing literature on period estimation has been rather informal in this regard, lacking a careful mathematical analysis on the uniqueness and identifiability of such component periods. One of our contributions in this work is to fill this gap in the theory. This is done in Sec. 7.3. Following this, in Sec. 7.4 we derive precise bounds on the minimum data-length required to estimate each of the component

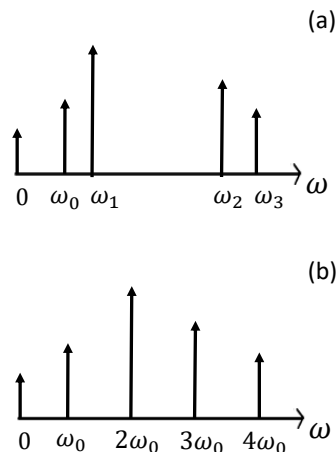


Figure 7.1: Part (a) - An arbitrary line spectrum; Part (b) - The harmonic line spectrum of a periodic signal. Can we use this additional structure in the spectrum of a periodic signal to reduce the data length required for period estimation?

periods (Theorems 7.4.2 and 7.4.3). In this case, the fundamental lower bound significantly differs from the $2P_{max}$ rule of thumb.

While the primary goal of this chapter is to derive algorithm-independent bound on data lengths for period estimation, we also extend these results to one family of algorithms in Sec. 7.5. This is for the recently proposed dictionary based integer period estimation techniques (Chapter 3, [23]). Such dictionaries were shown to offer useful advantages compared to traditional methods, especially for mixtures of periodic signals. However, the minimum data-length required for the dictionary based algorithms in [23] was not reported earlier. It should be clear that any algorithm-specific bound on the data length that might have been reported earlier in the literature is necessarily at least as large as the generic bound derived here.

In Sec. 7.6, we briefly explore the datalength requirements when the period of $x(n)$ is *not exactly an integer*. Such signals typically arise as sampled versions of continuous time signals. Even though they might not be strictly periodic according to (1), their spectrum still has a harmonic structure as shown in Fig. 7.1 (b). An analysis of the minimum required data-length for estimating the period of such signals yields some rather interesting results. For instance, it will be shown (Theorem 7.6.1) that the minimum datalength depends only on the number of harmonics expected in the signal, with more datalength being required as the number of harmonics increases. Connections to the classical result by Caratheodary and Fejer [111] on the datalength

requirements for the identifiability of complex exponentials is also discussed. In Sec. 7.7, we demonstrate some of the theorems derived in this chapter by using tandem repeats in DNA sequences as examples.

All the above mentioned sections deal with contiguous datalengths. Sec. 7.8 addresses the following question: If we are allowed to pick the samples in a non-contiguous fashion, what is the least number of samples needed to estimate the period? And how should we choose those samples? This question is quite difficult to answer in general, but the smaller case of resolving between two periods is analyzed in this section.

Special Notations

1. The divisor set of a set of integers \mathbb{P} is defined as:

$$D.S.(\mathbb{P}) = \{d : \exists P \in \mathbb{P}, d|P\} \quad (7.4)$$

That is, $D.S.(\mathbb{P})$ is the union of the divisors of each $P \in \mathbb{P}$. For example, if $\mathbb{P} = \{6, 8\}$, then $D.S.(\mathbb{P}) = \{1, 2, 3, 4, 6, 8\}$.

7.2 Minimum Datalength for the Single Period Case

In this section we study the minimum data length required to identify a single integer period. The main result of this section is the following:

Theorem 7.2.1. Minimum Required Data-Length for Period Estimation: *Let $x(n)$ be a periodic signal, whose period is known to lie in the integer set $\mathbb{P} = \{P_1, P_2, \dots, P_K\}$. To estimate the period using L consecutive samples, it is both necessary and sufficient that:*

$$L \geq L_{min} = \max_{P_i, P_j \in \mathbb{P}} P_i + P_j - (P_i, P_j) \quad (7.5)$$

◇

Remark: By necessity, we mean that for any $L' < L_{min}$, there exist vectors of length L' , that can be segments of both period P_i and period P_j signals, for some $P_i \neq P_j$. For example, if $\mathbb{P} = \{6, 15\}$, then $L_{min} = 6 + 15 - (6, 15) = 18$. Fig. 7.2 shows a segment of length 17 (indicated in black) that can be expressed as a part of both period 6 and period 15 signals. So given such a segment, one can never tell whether it belongs to a period 6 or a period 15 signal. Sufficiency on the other hand implies

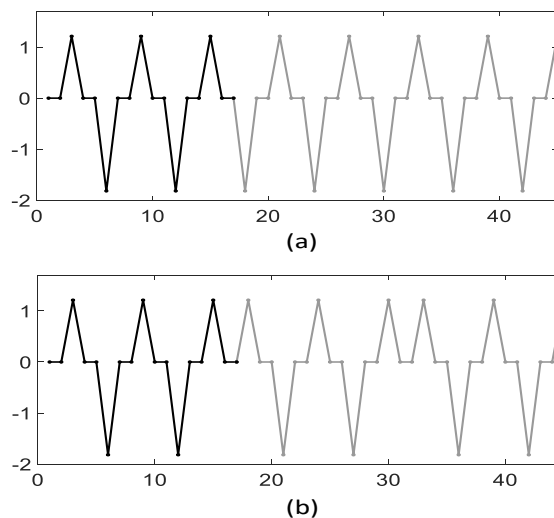


Figure 7.2: An example illustrating the necessity of L_{min} (Eq. (7.5)) samples. Part (a) - A period 6 signal; Part (b) - A period 15 signal. The initial 17 samples shown in black are common to both signals.

that, given an additional 18th sample, one can immediately estimate whether the period is 6 or 15.

In order to derive the above result, we first need a smaller result. This is discussed next.

Identifying The Period From a Set of Size Two

Let us first consider the case where the set \mathbb{P} in Theorem 7.2.1 has just two integers. That is, $\mathbb{P} = \{P_1, P_2\}$. While this case is too simple to be of practical importance, the result itself is nontrivial, and lays the foundation for more general results in the next few sections. We start with the following theorem, which follows from Theorem 12 of [9].

Theorem 7.2.2. A Reinterpretation of Fourier Series: *The decomposition of a periodic signal $x(n)$ into a sum of complex exponentials¹ is unique. Further,*

1. *This decomposition can be written in the following form:*

$$x(n) = \sum_{\substack{1 \leq d \leq P \\ d|P}} \sum_{\substack{1 \leq k \leq d \\ (k,d)=1}} \alpha_{k,d} e^{j\frac{2\pi k}{d}n} \quad (7.6)$$

where P is the period of $x(n)$.

¹In this theorem and the rest of the thesis, the term “complex exponential” denotes a signal of the form $e^{j\omega n}$, where $\omega \in \mathbb{R}$.

2. Let \mathbb{P}_C be the set of periods of all those complex exponentials that have non-zero coefficients in (7.6). Then, the following holds true:

$$P = \text{lcm}(\mathbb{P}_C) \quad (7.7)$$

◇

Remark 1: The complex exponential $e^{j\omega n}$ has period d iff $\omega = \frac{2\pi k}{d}$ for some $(k, d) = 1$. So there are precisely $\phi(d)$ complex exponentials with period d , and these are the terms in the inner summation of (7.6). Further, since $\sum_{d|P} \phi(d) = P$ [24], there are P complex exponentials in total in (7.6).

Remark 2: The second part of the theorem is called the **the LCM property**. In general, when signals with periods P_1, P_2, \dots, P_N are added together, then the resulting signal can have any divisor of $\text{lcm}\{P_1, P_2, \dots, P_N\}$ as its period. So Theorem 7.2.2 is special, since the period of $x(n)$ in (7.6) is exactly equal to the lcm and not a proper divisor of it.

We are now ready to derive the following result:

Theorem 7.2.3. Distinguishing between two periods: *Let $x(n)$ be a periodic signal, whose period is known to lie in the integer set $\mathbb{P} = \{P_1, P_2\}$. To estimate the period using R consecutive samples, it is both necessary and sufficient that:*

$$R \geq T = P_1 + P_2 - (P_1, P_2) \quad (7.8)$$

◇

Proof of sufficiency: Let \mathbb{B} be the set of all complex exponentials whose periods belong to $D.S.(\mathbb{P})$ (defined in Eq. (7.4)). Since, for every d , there are precisely $\phi(d)$ complex exponentials with period d (Remark 1 above), the total number of complex exponentials in \mathbb{B} is equal to $\sum_{d \in D.S.(\mathbb{P})} \phi(d)$. This number is in fact equal to T in (7.8) (see Lemma 7.11.3 in the Appendix). Further, $x(n)$ can be uniquely represented as a linear combination of the signals in \mathbb{B} , as given by Eq. (7.6). We can re-write this in matrix notation as follows:

$$\mathbf{x}_{\infty \times 1} = \mathbf{B}_{\infty \times T} \mathbf{z} \quad (7.9)$$

Here, $\mathbf{x}_{\infty \times 1} = [\dots, x(-2), x(-1), x(0), x(1), x(2), \dots]^T$ denotes the signal $x(n)$ written as a column vector, and $\mathbf{B}_{\infty \times T}$ is an infinitely tall matrix whose columns are the

signals in \mathbb{B} . According to Theorem 7.2.2, the *lcm* of the periods of those columns of $\mathbf{B}_{\infty \times T}$ that have non-zero entries in \mathbf{z} , is equal to the period of $x(n)$.

Notice however that we obtain the same \mathbf{z} as in (7.9) by solving the following system of equations instead:

$$\mathbf{x}_{T \times 1} = \mathbf{B}_{T \times T} \mathbf{z} \quad (7.10)$$

where $\mathbf{x}_{T \times 1}$ is any contiguous segment of $\mathbf{x}_{\infty \times 1}$ of length T , and $\mathbf{B}_{T \times T}$ is the corresponding $T \times T$ submatrix of $\mathbf{B}_{\infty \times T}$. This is because, $\mathbf{B}_{T \times T}$ is a Vandermonde matrix with distinct columns, and hence has full rank. So any T consecutive samples of $x(n)$ are sufficient to recover \mathbf{z} , and hence to estimate the period.

Proof of necessity: Let $\mathbf{B}_{(T-1) \times T}$ be the $(T-1) \times T$ sub-matrix of $\mathbf{B}_{T \times T}$, obtained by dropping the last row of $\mathbf{B}_{T \times T}$. Clearly, there exists a vector $\mathbf{s} \neq \mathbf{0}$ such that,

$$\mathbf{B}_{(T-1) \times T} \mathbf{s} = \mathbf{0} \quad (7.11)$$

This can be re-written as follows:

$$\mathbf{B}_{(T-1) \times T}^{(P_1)} \mathbf{s}^{(P_1)} = -\mathbf{B}_{(T-1) \times T}^{(P_2)} \mathbf{s}^{(P_2)} \quad (7.12)$$

where $\mathbf{B}_{(T-1) \times T}^{(P_i)}$ consists of those columns of $\mathbf{B}_{(T-1) \times T}$ that are complex exponentials with divisors of P_i as periods. Those columns of $\mathbf{B}_{(T-1) \times T}$ that have common divisors of P_1 and P_2 as their periods are included in both $\mathbf{B}_{(T-1) \times T}^{(P_1)}$ and $\mathbf{B}_{(T-1) \times T}^{(P_2)}$, by dividing the corresponding entries in \mathbf{s} in half.

Let $x_1(n)$ be the signal obtained by considering the infinitely tall version of the L.H.S. of (7.12). That is,

$$x_1(n) = \mathbf{B}_{\infty \times T}^{(P_1)} \mathbf{s}^{(P_1)} \quad (7.13)$$

Then, from Theorem 7.2.2, the period of $x_1(n)$ must be the *lcm* of the periods of those columns of $\mathbf{B}_{\infty \times T}^{(P_1)}$ that have non-zero entries in $\mathbf{s}^{(P_1)}$. Now, $\mathbf{s}^{(P_1)}$ cannot have any zero entry, since \mathbf{s} cannot have any non-zero entries (this is a subtle point; see Lemma 7.11.1 in the Appendix). Hence, the period of $x_1(n)$ has to be P_1 . Similarly, the period of $x_2(n) = \mathbf{B}_{\infty \times T}^{(P_2)} \mathbf{s}^{(P_2)}$ has to be P_2 .

So given any $T' < T$, let us construct a $T' \times 1$ vector \mathbf{y} by choosing it to be any T' consecutive samples of $\mathbf{B}_{(T-1) \times T}^{(P_1)} \mathbf{s}^{(P_1)}$. Evidently, for such a \mathbf{y} , we can never say whether it came from a period P_1 signal, or a period P_2 signal. This shows that at least T samples are necessary to estimate the period from the set $\{P_1, P_2\}$.

Identifying The Period From a Set of Arbitrary Size

We will now use the above theorem to prove Theorem 7.2.1.

Theorem 7.2.1 - Proof of necessity: For every $L' < L_{min}$, there exists a pair $P_a, P_b \in \mathbb{P}$, $P_a \neq P_b$, such that

$$L' < \sum_{d \in D.S.(\{P_a, P_b\})} \phi(d) \quad (7.14)$$

So from Theorem 7.2.3, there must exist a vector $\mathbf{y} \in \mathbb{C}^{L' \times 1}$ that can be expressed as a segment of both period P_a and period P_b signals. So the period of \mathbf{y} can never be estimated. This shows that L_{min} samples are necessary to estimate the period.

Theorem 7.2.1 - Proof of sufficiency: Let $P_i, P_j \in \mathbb{P}$. Let $T = P_i + P_j - (P_i, P_j)$. Let $\mathbf{x}_{T \times 1}$ be any $T \times 1$ contiguous segment of $x(n)$. If we try to solve the following system of linear equations as in the proof of Theorem 7.2.3,

$$\mathbf{x}_{T \times 1} = \mathbf{B}_{T \times T} \mathbf{z} \quad (7.15)$$

then, since $\mathbf{B}_{T \times T}$ is a full rank matrix, there is a unique solution \mathbf{z} . Further, there are three possible outcomes. The *lcm* of the periods of those columns of $\mathbf{B}_{T \times T}$ that have non-zero coefficients in \mathbf{z} can be

1. P_i ,
2. P_j , or
3. neither P_i nor P_j .

Now, P (which is the period of $x(n)$) is the unique element in \mathbb{P} such that, for any pair $P, P' \in \mathbb{P}$, solving (7.15) always results in the *lcm* being P . This follows from Theorem 7.2.3. This can be used to identify P . Clearly, L_{min} samples are sufficient for this approach.

Remark 1: The method we used in the above proof of sufficiency is to be viewed only as a ‘proof technique’. There are much easier ways to estimate the period. For example, we may use the DFT or the autocorrelation. Although such techniques are computationally much better, it is not clear how they can be used to prove an algorithm-independent result such as Theorem 7.2.1.

Remark 2: Even though the theoretical bounds on the minimum data length derived here are insightful, the accuracy of estimation in the presence of noise will be poor

if we just use this minimum number. This is demonstrated in Fig. 7.3. In this plot, $\mathbb{P} = \{1, 2, \dots, 6\}$. Theorem 7.2.3 shows that at least 10 samples of the signal $x(n)$ are needed for period estimation. However, $x(n)$ now is contaminated by AWGN with 0 dB SNR. For each data-length shown, 10,000 independent realizations of $x(n)$ with period 3 were generated. A least squares modification of (7.15) was used to estimate the period, where, if z_{max} is the entry in \mathbf{z} with the maximum absolute value, then all the entries in \mathbf{z} less than a threshold times $|z_{max}|$ were set to zero. The error rate (the fraction of times the period was estimated incorrectly) is plotted for three different values of the threshold. As shown, the error is large when data length is close to the theoretical minimum, and improves significantly as more data length becomes available. Once again, there are many excellent works such as [16], [95] for estimating the period under noise. Fig. 7.3 is just a simple demonstration that one may typically need more than the minimum data-length for reliable estimation under noise.

A special case of Theorem 7.2.1 is the following result:

Corollary 7.2.1. *For the special case when $\mathbb{P} = \{1, 2, 3, \dots, P_{max}\}$, the minimum necessary and sufficient data length in Theorem 7.2.1 becomes:*

$$L_{min} = 2P_{max} - 2 \quad (7.16)$$

◇

Proof. It is easy to see that P_{max} and $P_{max} - 1$, being co-prime, maximize (7.5) in Theorem 7.2.1. ▽ ▽ ▽

So when $\mathbb{P} = \{1, 2, 3, \dots, P_{max}\}$, L_{min} in (7.16) turns out to be close to the $2P_{max}$ rule of thumb used in prior works such as [95]. The difference is that, we have now proved that necessity of (7.44) in fact is an algorithm-independent fundamental bound.

7.3 Mixtures of Periodic Signals

Our goal in this and the next sections is to answer the following question: Given a mixture of periodic signals, what is the absolute minimum data-length required to estimate all its hidden periods²? Clearly, to begin answering this question, we first

²Keeping in line with our papers [9], [18], [23], [37], we will refer to the ‘component’ periods in a mixture as the ‘hidden’ periods in the mixture.

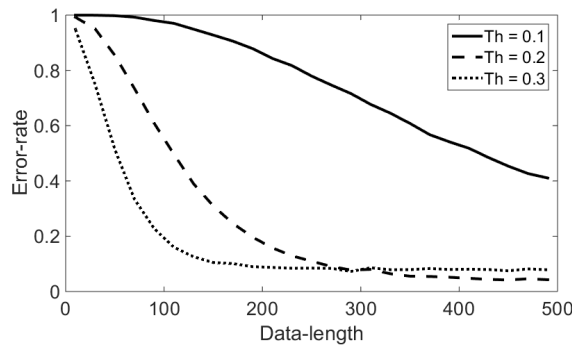


Figure 7.3: Error Rate vs Data-Length for a fixed SNR. “Th” refers to the threshold. SNR = 0dB. See text for details.

need a precise mathematical definition of what we mean by the ‘hidden periods’ in a mixture. Intuitively, we may visualize hidden periods in the following way. Given a signal, if we can decompose it as a sum of periodic signals with integer periods $\{P_1, P_2, \dots, P_N\}$, then we may want to call these integers as the hidden periods. However, there are some subtle issues with this, that have never been formally addressed in prior literature.

For example, let $x_P(n)$ be a period P signal. Let d be any divisor of P . Then $x_P(n)$ can be trivially rewritten as

$$x_P(n) = x_d(n) + \tilde{x}_P(n) \quad (7.17)$$

where $x_d(n)$ is almost any arbitrary period d signal, and $\tilde{x}_P(n) = x_P(n) - x_d(n)$ is still a period P signal. Thus, for a period P signal, it is meaningless to declare any of its divisors d as another hidden period.

By insisting that the distinct hidden periods in a mixture should constitute an M-set as defined below, we can eliminate this redundancy:

Definition 7.3.1. M-set: A set of numbers \mathbb{F} is defined to be an M-set, if the following holds:

$$f_i \nmid f_j \quad \forall f_i, f_j \in \mathbb{F}, f_i \neq f_j \quad (7.18)$$

That is, no member of an M-set is a divisor of another member³. \diamond

For example, $\{6, 7, 10\}$ is an M-set, whereas $\{3, 6, 10\}$ is not.

³The ‘M’ stands for Multiple-free.

There is a second type of ambiguity that arises while defining hidden periods. We can always associate different M-sets to any mixture of periodic signals. For example, consider adding period 3 and period 8 signals to give a period 24 signal. Since $\{3, 8\}$ is an M-set, we may want to call them as the hidden periods. However, given the periodic components $\{3, 8\}$, we can add and subtract any arbitrary period 5 signal from these two components, to get periods $\{15, 40\}$. These new signals with periods 15 and 40 will also add to give the same period 24 signal. Moreover, $\{15, 40\}$ is also an M-set. So now, which among $\{15, 40\}$ and $\{3, 8\}$ should one declare as the hidden periods of the period 24 signal?

In the above example, it turns out that periods $\{3, 8\}$ represent the *finest periodic structure* in the mixture. That is, one cannot further decompose the period 3 signal into a sum of signals with periods are strictly smaller than 3. Similarly, one cannot further decompose the period 8 signal into a sum of signals with periods strictly smaller than 8. On the other hand, in this example, both the period 15 and the period 40 signals can be further decomposed into signals with periods strictly smaller than 15 and 40 respectively (namely, into periods 5 and 3, and periods 5 and 8 respectively). This can also be seen when above period 15 signal is decomposed as in (7.6), where it will not have any terms of the form $e^{j\frac{2\pi k}{15}n}$ with $(k, 15) = 1$. And the same holds for the period 40 signal.

We will now show that if we define the hidden periods of a signal as those representing the *finest periodic structure* (in the above sense), then the hidden periods become unique for any mixture of periodic signals. That is, we define hidden periods formally in the following way:

Definition 7.3.2. Hidden Periods: A signal $x(n)$ is said to have hidden periods $\mathbb{P}_H = \{P_1, P_2, \dots, P_N\}$ if the following hold true:

1. \mathbb{P}_H is an M-set (see Notations in Sec. 7.1).
2. $x(n)$ can be written as follows:

$$x(n) = x_{P_1}(n) + x_{P_2}(n) + \dots + x_{P_N}(n) \quad (7.19)$$

where each $x_{P_i}(n)$ has period P_i , and cannot be further decomposed into a sum of periodic signals, all of whose periods are strictly smaller than P_i .

◇

The following uniqueness result can now be proved:

Theorem 7.3.1. Uniqueness of Hidden Periods: *Let $x(n)$ satisfy Definition 7.3.2, with \mathbb{P}_H as its hidden periods. Further, let $x(n)$ be decomposed as in (7.6)⁴. If \mathbb{P}_C is the set of periods of all those complex exponentials that have non-zero coefficients in (7.6), then $\mathbb{P}_H \subseteq \mathbb{P}_C$. More precisely:*

$$\mathbb{P}_H = \{P_i \in \mathbb{P}_C : MP_i \notin \mathbb{P}_C \forall M > 1\} \quad (7.20)$$

◇

Remark 1: In (7.20), the R.H.S. is the set of all those numbers in \mathbb{P}_C that do not have any multiples also present in \mathbb{P}_C . For example, if $\mathbb{P}_C = \{1, 3, 5, 10\}$, then the R.H.S. of (7.20) is $\{3, 10\}$. We will call this as the **extracted M-set** of \mathbb{P}_C .

Remark 2: The extracted M-set of \mathbb{P}_C is unique, since \mathbb{P}_C itself is unique for any signal (Theorem 7.2.2). Hence, the set of hidden periods, \mathbb{P}_H , is also unique.

Remark 3: As an example, consider the following three signals:

1. $e^{j\frac{2\pi n}{6}}$ has period 6. There are no hidden periods (other than 6 itself).
2. $e^{j\frac{2\pi n}{2}} + e^{j\frac{2\pi n}{3}}$ has period 6. The hidden periods are 2 and 3, but 6 is not a hidden period.
3. $e^{j\frac{2\pi n}{2}} + e^{j\frac{2\pi n}{3}} + e^{j\frac{2\pi n}{6}}$ has period 6. There are no hidden periods other than 6 itself, since this signal can never be written as a sum of signals, all of whose periods are strictly smaller than 6. Also, as discussed earlier, calling 2 and 3 as hidden periods in addition to 6 is redundant in this case, since any period 6 signal $x_6(n)$ can always be written as

$$x_6(n) = x_2(n) + x_3(n) + \tilde{x}_6(n) \quad (7.21)$$

for some non-zero signals $x_2(n)$, $x_3(n)$ and $\tilde{x}_6(n)$, with periods 2, 3 and 6 respectively. So Definition 7.3.2 simply gives 6 as the only hidden period.

Proof of Theorem 7.3.1: The proof involves three steps:

1. Showing that $\mathbb{P}_H \subseteq \mathbb{P}_C$.

⁴Any mixture of periodic signals must also be a periodic signal.

2. Showing that $\mathbb{P}_C \subseteq D.S.(\mathbb{P}_H)$.
3. Using Lemma 7.11.2 from the Appendix to conclude that \mathbb{P}_H is as in (7.20).

Proof of Step 1: Since $x(n)$ has hidden periods \mathbb{P}_H , it will have a decomposition similar to (7.19). Let us now decompose the L.H.S. and each term in the R.H.S. of (7.19) in a manner similar to (7.6). While the decomposition of the L.H.S. is given in (7.6), each $x_{P_i}(n)$ in the R.H.S. can be written as:

$$x_{P_i}(n) = \sum_{\substack{1 \leq d_{P_i} \leq P_i \\ d_{P_i} | P_i}} \sum_{\substack{1 \leq k \leq d_{P_i} \\ (k, d_{P_i}) = 1}} \alpha_{k, d_{P_i}} e^{j \frac{2\pi k}{d_{P_i}} n} \quad (7.22)$$

This decomposition of $x_{P_i}(n)$ must have at least one complex exponential with period P_i with a non-zero coefficient. Else, $x_{P_i}(n)$ can be written as a sum of signals, all of whose periods are strictly smaller than P_i , which contradicts Condition 2 in Definition 7.3.2.

Further, if $x_{P_i}(n)$ has a complex exponential with period P_i , then that complex exponential cannot be canceled in the R.H.S. of (7.19) by the complex exponentials in some other $x_{P_j}(n)$. This is because, such a cancellation can occur only if $x_{P_j}(n)$ has a period P_i complex exponential, meaning $P_i | P_j$. This is not possible because \mathbb{P}_H is an M-set. Hence, when the R.H.S. of (7.19) is expressed as a sum of complex exponentials, there has to be at least one complex exponential with period P_i , for every $P_i \in \mathbb{P}_H$.

Since complex exponentials are linearly independent signals, when both the L.H.S. and the R.H.S. of (7.19) are expressed in terms of complex exponentials, we must have the same complex exponentials on both sides. This implies that the decomposition of $x(n)$ in (7.6) must also have at least one complex exponential with period P_i , with a non-zero coefficient. Hence, $\mathbb{P}_H \subseteq \mathbb{P}_C$.

Proof of Step 2: Let $q \in \mathbb{P}_C$. From (7.6), the DTFT of $x(n)$ must have an impulse at the frequency $\frac{2\pi r}{q}$, for some $(r, q) = 1$. But if $q \notin D.S.(\mathbb{P}_H)$, then none of the $x_{P_i}(n)$ in (7.19) have an impulse in their DTFT at the frequency $\frac{2\pi r}{q}$ (This is because, from Theorem 7.2.2, a period P signal can have non-zero DTFT only at frequencies that are of the form $\frac{2\pi m}{d}$, where $(m, d) = 1$ and $d | P$). Hence, we obtain a contradiction. So $q \in D.S.(\mathbb{P}_H)$, and hence $\mathbb{P}_C \subseteq D.S.(\mathbb{P}_H)$.

Finally, using Lemma 7.11.2 from the Appendix, we can conclude that $\mathbb{P}_H = \{P_i \in \mathbb{P}_C : MP \notin \mathbb{P}_C \forall M > 1\}$.

The above result has an intuitively reassuring corollary:

Corollary 7.3.1. *Let $x(n)$ be a signal with period P and hidden periods \mathbb{P}_H . Then, the following holds true:*

$$P = \text{lcm}(\mathbb{P}_H) \quad (7.23)$$

i.e., the lcm of the hidden periods of a signal is exactly equal to the period of the signal. \diamond

Proof. From Theorem 7.2.2, it follows that $P = \text{lcm}(\mathbb{P}_C)$. Notice that \mathbb{P}_C and the R.H.S. of (7.20) have the same *lcm*. And so (7.23) follows. $\nabla \nabla \nabla$

Remark 1: Readers familiar with Ramanujan Subspaces and Nested Periodic Subspaces (NPSs) [18], [37] may recall the following: Suppose we decompose $x(n)$ along a set of NPSs [37]. If \mathbb{P}_{NPS} is the set of periods of all the non-zero components in the decomposition, then the extracted M-set of \mathbb{P}_{NPS} was intuitively interpreted as the hidden periods in those works. It can be proved that this in fact turns out to be exactly equal to the set of hidden periods as defined in Definition 7.3.2. Moreover, Theorem 7.3.1 shows that although \mathbb{P}_{NPS} for a given signal may depend on which NPS is used, the extracted M-set would always be the same.

Remark 2: In many applications such as speech, the signal of interest may arise naturally as a sum of periodic signals. While the hidden periods in Definition 7.3.2 represent the *finest periodic structure* in a mixture, how do they relate to the original set of component periods that generated that mixture? It can be shown that if $x(n)$ is a mixture of randomly generated periodic signals, say by periodically extending Gaussian random vectors with periods $\mathbb{P} = \{P_1, P_2, \dots, P_K\}$, then the extracted M-set from \mathbb{P} is in fact exactly equal to the hidden periods as given by Definition 7.3.2 with probability 1. We will skip the proof in this regard. Once again, determining the divisors of the integers in the extracted M-set as also component periods in the mixture is mathematically meaningless.

7.4 Minimum Datalength for Estimating The Hidden Periods

Now that we have shown that the hidden periods in a mixture can be uniquely defined, the following question can be posed unambiguously: What is the minimum required data length to estimate all the hidden periods in a signal? Similar to Sec. 7.2, we start with a simple case first.

Theorem 7.4.1. Distinguishing between two M-sets: Let $x(n)$ be a periodic signal, whose set of hidden periods is known to be one of the following two M-sets: \mathbb{P}_1 or \mathbb{P}_2 . To estimate the hidden periods of $x(n)$ using R consecutive samples, it is both necessary and sufficient that:

$$R \geq T = \sum_{d \in D.S.(\mathbb{P}_1 \cup \mathbb{P}_2)} \phi(d) \quad (7.24)$$

where $D.S.$ is the divisor set defined in Eq. (7.4). \diamond

Proof of sufficiency: Similar to (7.9), we can write the following equation:

$$\mathbf{x}_{\infty \times 1} = \mathbf{B}_{\infty \times T} \mathbf{z} \quad (7.25)$$

where, $\mathbf{B}_{\infty \times T}$ now has as its columns the set of all complex exponentials whose periods belong to $D.S.(\mathbb{P}_1 \cup \mathbb{P}_2)$.

Let \mathbb{P}_C be the set of periods of those columns of $\mathbf{B}_{\infty \times T}$, that are multiplied by non-zero entries in \mathbf{z} . Then, from Theorem 7.3.1, the set of hidden periods of $x(n)$ is given by:

$$\mathbb{P}_H = \{P \in \mathbb{P}_C : MP \notin \mathbb{P}_C \ \forall M > 1\} \quad (7.26)$$

However, using similar arguments as in the proof of Theorem 7.2.3, we can obtain the same solution for \mathbf{z} as in (7.25) by solving the following system instead:

$$\mathbf{x}_{T \times 1} = \mathbf{B}_{T \times T} \mathbf{z} \quad (7.27)$$

where $\mathbf{x}_{T \times 1}$ is any contiguous segment of $\mathbf{x}_{\infty \times 1}$ of length T , and $\mathbf{B}_{T \times T}$ is the corresponding $T \times T$ sub-matrix of $\mathbf{B}_{\infty \times T}$. So we have proved that T samples are sufficient to find the hidden periods of $x(n)$.

Proof of necessity: Similar to the proof of Theorem 7.2.3, we can write the following equation:

$$\mathbf{B}_{(T-1) \times T}^{(\mathbb{P}_1)} \mathbf{s}^{(\mathbb{P}_1)} = -\mathbf{B}_{(T-1) \times T}^{(\mathbb{P}_2)} \mathbf{s}^{(\mathbb{P}_2)} \quad (7.28)$$

where, $\mathbf{s}^{(\mathbb{P}_1)}, \mathbf{s}^{(\mathbb{P}_2)}$ do not contain any zero entries, and $\mathbf{B}_{(T-1) \times T}^{(\mathbb{P}_1)}$ and $\mathbf{B}_{(T-1) \times T}^{(\mathbb{P}_2)}$ consist of complex exponentials with periods in $D.S.(\mathbb{P}_1)$ and $D.S.(\mathbb{P}_2)$ respectively. Now, let $x_1(n) = \mathbf{B}_{\infty \times T}^{(\mathbb{P}_1)} \mathbf{s}^{(\mathbb{P}_1)}$. Then from Theorem 7.3.1, the set of hidden periods of $x_1(n)$ is given by the following set:

$$\{P \in D.S.(\mathbb{P}_1) : MP \notin D.S.(\mathbb{P}_1) \ \forall M > 1\} \quad (7.29)$$

Using Lemma 7.11.2 from the Appendix (with $\mathbb{G} = D.S.(\mathbb{P}_1)$ and $\mathbb{S} = \mathbb{P}_1$), we can conclude that the above set is in fact equal to \mathbb{P}_1 . This shows that $x_1(n)$ has hidden periods \mathbb{P}_1 .

Similarly, $x_2(n) = \mathbf{B}_{\infty \times T}^{(\mathbb{P}_2)} \mathbf{s}^{(\mathbb{P}_2)}$ has hidden periods \mathbb{P}_2 . So given any $T' < T$, we can construct a vector \mathbf{y} by choosing it to be any T' consecutive samples of $\mathbf{B}_{(T-1) \times T}^{(\mathbb{P}_1)} \mathbf{s}^{(\mathbb{P}_1)}$. For such a \mathbf{y} , we can never estimate its hidden periods. This shows that T samples are necessary to find the hidden periods.

Using the above result, we next address the more general case:

Theorem 7.4.2. Min. Data-Length for Hidden Periods Estimation: *Suppose $x(n)$ is known to have N hidden periods belonging to the integer set $\mathbb{P} = \{P_1, P_2, \dots, P_K\}$. To estimate the hidden periods of $x(n)$ using M consecutive samples, it is both necessary and sufficient that:*

$$M \geq M_{min} = \max_{\substack{\mathbb{P}_i, \mathbb{P}_j \subset \mathbb{P} \\ \mathbb{P}_i, \mathbb{P}_j \text{ are} \\ M\text{-sets of size } = N}} \sum_{d \in D.S.(\{\mathbb{P}_i \cup \mathbb{P}_j\})} \phi(d) \quad (7.30)$$

◇

Proof of necessity: By definition of M_{min} (7.30), it follows that for every $M' < M_{min}$, there must exist a pair of M -sets, \mathbb{P}_1 and \mathbb{P}_2 in \mathbb{P} , such that

$$M' < \sum_{d \in D.S.(\mathbb{P}_1 \cup \mathbb{P}_2)} \phi(d) \quad (7.31)$$

So from Theorem 7.4.1, there must exist a vector $\mathbf{y} \in \mathbb{C}^{M' \times 1}$ that can be expressed as a segment of a signal with hidden periods \mathbb{P}_1 , as well as a signal with hidden periods \mathbb{P}_2 . Hence, at least M_{min} samples are necessary to find the hidden periods.

Proof of sufficiency: Let \mathbb{P}_i and \mathbb{P}_j be two M -sets of size N in \mathbb{P} . Let $T = \sum_{d \in D.S.(\mathbb{P}_i \cup \mathbb{P}_j)} \phi(d)$. Let $\mathbf{x}_{T \times 1}$ be any $T \times 1$ segment of $x(n)$. If we try to solve the following system of linear equations as in the proof of Theorem 7.4.1,

$$\mathbf{x}_{T \times 1} = \mathbf{B}_{T \times T} \mathbf{z} \quad (7.32)$$

then, since $\mathbf{B}_{T \times T}$ is a full rank matrix, there is a unique solution \mathbf{z} . Further, there are three possible outcomes. If \mathbb{P}_C is the set of periods of all those columns of $\mathbf{B}_{T \times T}$ that have non-zero coefficients in \mathbf{z} , then the set $\{P_i \in \mathbb{P}_C : MP \notin \mathbb{P}_C \forall M > 1\}$ can be any of the following:

1. \mathbb{P}_i .
2. \mathbb{P}_j .
3. Neither \mathbb{P}_i nor \mathbb{P}_j .

Now, \mathbb{P}_H (which is the actual set of hidden periods in $x(n)$) is the unique M-set of size N in \mathbb{P} such that, when \mathbb{P}_H is paired with any other M-set of size N from \mathbb{P} , solving (7.32) always results in the following outcome: $\{P_i \in \mathbb{P}_C : MP \notin \mathbb{P}_C \forall M > 1\} = \mathbb{P}_H$. This follows directly from the proof of Theorem 7.4.1. Clearly, M_{min} samples are sufficient for this algorithm.

Interestingly, (7.30) is very similar to (7.5). The latter can be re-written in terms of the Euler-totient function using the following result (see Lemma 7.11.3 in the Appendix):

$$P_i + P_j - (P_i, P_j) = \sum_{d \in D.S.(\{P_i \cup P_j\})} \phi(d) \quad (7.33)$$

So the quantity L_{min} in (7.5) can be written as:

$$L_{min} = \max_{P_i, P_j \in \mathbb{P}} \sum_{d \in D.S.(\{P_i \cup P_j\})} \phi(d) \quad (7.34)$$

This indeed resembles (7.30) in form. We maximize over all pairs of allowed M-sets of size N in (7.30), just as we maximize over all pairs of periods in (7.34).

Theorem 7.4.2 assumed that the number of hidden periods is known (equal to N). This can be easily generalized to the case when the number of hidden periods is unknown. We state the following result without proof, since the idea is similar to that in the proof of Theorem 7.4.2.

Theorem 7.4.3. *Let $x(n)$ be a signal whose number of hidden periods is known to belong to the set \mathbb{N}_H . Further, suppose the hidden periods are known to belong to the integer set $\mathbb{P} = \{P_1, P_2, \dots, P_K\}$. To estimate these hidden periods using R consecutive samples of $x(n)$, it is both necessary and sufficient that:*

$$R \geq \max_{\substack{P_i, P_j \subset \mathbb{P} \\ P_i, P_j \text{ are} \\ M\text{-sets whose size} \in \mathbb{N}_H}} \sum_{d \in D.S.(\{P_i \cup P_j\})} \phi(d) \quad (7.35)$$

◇

7.5 Connection to Dictionaries Spanning Periodic Signals

As discussed in the previous sections, the period or the hidden periods in a signal can be identified from (7.6), using (7.7) or (7.20) respectively. In a recent work [26], a dictionary known as the Farey dictionary was proposed to directly obtain the decomposition in (7.6) for arbitrary periodic signals. For every period P , the Farey dictionary consists of the following complex exponentials as columns: $\{e^{j\frac{2\pi k}{P}n} : (k, P) = 1\}$. Given an input signal, one can solve the following system:

$$\mathbf{x}_{L \times 1} = \mathbf{B}_{L \times W} \mathbf{s} \quad (7.36)$$

where, $\mathbf{x}_{L \times 1}$ is the signal vector $[x(0), x(1), x(2), \dots, x(L-1)]^T$, and $\mathbf{B}_{L \times W}$ is the Farey dictionary with L rows and W columns, where W depends on P_{max} in the following way [24]:

$$W = \sum_{P=1}^{P_{max}} \phi(P) = \mathcal{O}\left(\frac{3P_{max}^2}{\pi^2}\right) \quad (7.37)$$

If $L > W$, $\mathbf{B}_{L \times W}$ has full rank. The set of periods of those columns in $\mathbf{B}_{L \times W}$ that are multiplied by non-zero entries in \mathbf{s} , is equal to the set \mathbb{P}_C in (7.7) and (7.20). Hence, the period and the hidden periods of $x(n)$ can be estimated from \mathbf{s} .

What happens when $L < W$? One of the ways to recover a signal's support in that case is to solve the following problem:

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad s.t. \quad \mathbf{x}_{L \times 1} = \mathbf{B}_{L \times W} \mathbf{s} \quad (7.38)$$

In [23] and [26], (7.38) was relaxed to l_1 norm based convex programs. Further, [23] proposes a computationally efficient weighted l_2 norm based convex program as well. These approaches were shown to offer important advantages over traditional period estimation techniques in [23]. From a theoretical perspective however, one question remained unanswered in those works. Namely, it was not known how much datalength L these techniques required to recover the support of \mathbf{s} .

In this regard, we are able to derive the following theorems for the Farey dictionary, using tools similar to those used in the previous sections:

Theorem 7.5.1. Farey Dictionary: Single Period Case *Let $\mathbf{x}_{L \times 1}$ be a segment of a signal whose period is known to lie in the set $\{1, 2, \dots, P_{max}\}$. Then, for the support recovered from solving (7.38) using a Farey dictionary, to involve the same complex exponentials as those with non-zero coefficients in (7.6), it is both necessary and sufficient that:*

$$L \geq L_{min}^{(D)} = 2P_{max} \quad (7.39)$$

Proof of sufficiency: The Kruskal rank R of $\mathbf{B}_{L \times W}$ is the largest integer such that, any set of R columns is linearly independent. Let $\mathbf{x}_{L \times 1} = \mathbf{B}_{L \times W} \mathbf{s}_0$ represent the decomposition in (7.6). Then, it is well known [11] that solving (7.38) yields \mathbf{s}_0 as the unique optimum solution if the support size of \mathbf{s}_0 is $\leq R/2$. Now, if $x(n)$ had period P , then the support size of \mathbf{s}_0 is $\leq P$ from Theorem 7.2.2. So if $R \geq 2P_{max}$, then for any $x(n)$ with period in $1 \leq P \leq P_{max}$, we can recover its support as given by (7.6) using (7.38).

Now, the Kruskal rank of $\mathbf{B}_{L \times W}$ is equal to L , as long as $L \leq W$. This is because, for $L \leq W$, any $L \times L$ submatrix of $\mathbf{B}_{L \times W}$ is Vandermonde with distinct columns, and hence, will have full rank. Hence, $L \geq 2P_{max}$ is sufficient to recover the support of $x(n)$.⁵

Proof of necessity: Let $L' = 2P_{max} - 1$. Consider an $L' \times 2P_{max}$ matrix, call it \mathbf{A} , whose column set can be partitioned as $\mathbf{A} = [\mathbf{A}^{(P_{max})}, \hat{\mathbf{A}}]$, where $\mathbf{A}^{(P_{max})}$ consists of the columns of $\mathbf{B}_{L' \times W}$ that have periods equal to divisors of P_{max} , and $\hat{\mathbf{A}}$ consists of any other P_{max} columns of $\mathbf{B}_{L' \times W}$. Since \mathbf{A} is a fat matrix, there exists an $\mathbf{r} \neq \mathbf{0}$ (with non-zero entries, using Lemma 7.11.1 from the Appendix), such that $\mathbf{A}\mathbf{r} = \mathbf{0}$. We can re-write this as:

$$\mathbf{A}^{(P_{max})} \mathbf{r}^{(P_{max})} = -\hat{\mathbf{A}} \hat{\mathbf{r}} \quad (7.40)$$

Clearly, if $x(n)$ was the period P_{max} signal obtained by extending the complex exponentials in the L.H.S. of (7.40), then, its support on $\mathbf{B}_{L' \times W}$ cannot be recovered using (7.38), since the R.H.S. of (7.40) also has the same support size. Notice that this holds true for any $L < L'$ as well, by considering the first L samples of the L.H.S. of (7.40). This shows that $2P_{max} - 1$ samples are necessary.

Remark: The above proof of necessity might remind some readers of the well known Kruskal rank result from [11], which says that, there will always exist a set of columns in $\mathbf{B}_{L \times W}$ of size $> R/2$, such that, if $\mathbf{x}_{L \times 1}$ is a linear combination of these columns, then (7.38) cannot identify the true support of this signal. However, in our case, notice that $\mathbf{x}_{L \times 1}$ in (7.38) cannot be a linear combination of any arbitrary columns in $\mathbf{B}_{L \times W}$. It must necessarily be periodic, with period in the range $1 \leq P \leq P_{max}$. So, is it possible that the counter-examples in the general Kruskal rank result do not

⁵This proof is a generalization of Lemma 1 from [26], which states that the Kruskal rank of $\mathbf{B}_{P_{max} \times W}$ is equal to P_{max} .

Table 7.1: Comparison between (7.30) and (7.41) for $P_{max} = 20$.

Number of hidden periods	M_{min} (7.30)	from $M_{min}^{(D)}$ (7.41)
2	70	76
3	94	108
4	116	140 ★
5	128	164 ★

correspond to any periodic signal, with period in the range $1 \leq P \leq P_{max}$? The above necessity proof shows that this is in fact not the case.

Theorem 7.5.1 shows that (7.39) requires only 2 additional samples than (7.16) in Corollary 7.2.1, and this indeed matches with the datalength requirements of some of the popular period estimation techniques in the literature such as [95]. This was for the estimation of one period. The equivalent of Theorem 7.5.1 for the case of hidden periods is as follows:

Theorem 7.5.2. Farey Dictionary: Hidden Periods Case: *Let $\mathbf{x}_{L \times 1}$ be a segment of a signal with N hidden periods from the set $\mathbb{P} = \{1, 2, 3, \dots, P_{max}\}$. Then, for the support recovered from solving (7.38) using a Farey dictionary, to involve the same complex exponentials as those with non-zero coefficients in (7.6), it is both necessary and sufficient that:*

$$L \geq M_{min}^{(D)} = \max_{\substack{\mathbb{P}_i \subseteq \mathbb{P} \\ \mathbb{P}_i \text{ is an} \\ M\text{-set of size } N}} 2 \times \sum_{d \in D.S.(\{\mathbb{P}_i\})} \phi(d) \quad (7.41)$$

◇

The proof is similar to Theorem 7.5.1.

For a quantitative perspective on these results, Table 7.1 compares (7.41) with the algorithm-independent bound from (7.30) when $\mathbb{P} = \{1, 2, \dots, 20\}$. For this case, the number of columns in the Farey dictionary is $W = 128$ (from (7.37)). Notice that the minimum required datalength $M_{min}^{(D)}$ from (7.41) turns out to be greater than W when the number of hidden periods is more than 3 in Table 7.1 (indicated by ★). But if we do have more than W samples of the signal, (7.36) becomes over-determined, and we do not need a sparsity based approach to solve for the signal's support \mathbf{s} .

$M_{min}^{(D)} > W$ simply means that (7.38) cannot be used to find the hidden periods when the available datalength L is smaller than W .

There are two important extensions still required for Theorem 7.5.1 and Theorem 7.5.2. First, for practical implementations, (7.38) is often relaxed to l_1 norm and l_2 norm based convex programs [23]. As expected, we found that such l_1 and l_2 norm based approaches typically require more data-length than (7.38). Precise expressions for such relaxations are yet to be derived, and are of interest to us in our future research. A second extension needed is to the case of the more general Nested Periodic Dictionaries (NPDs), proposed in [23], [37]. The Farey dictionary was shown to be only a special case of NPDs, with the more general NPDs offering several advantages, particularly in terms of computations.

7.6 Non-Integer Periodicity and Connections to Caratheodory's Results

A classical result by Caratheodary and Fejer [111] shows that, if $s(n)$ is a sum of exponentials as follows:

$$s(n) = \sum_{k=0}^{K-1} c_k e^{j\omega_k n} \quad (7.42)$$

then, $2K$ contiguous samples of $s(n)$ are necessary and sufficient to estimate $\{c_k, \omega_k\}_{k=0}^{K-1}$. Now, if a signal $x(n)$ satisfies (1), then it's Fourier series expansion is given by:

$$x(n) = \sum_{k=0}^{P-1} c_k e^{j\frac{2\pi k}{P}n} \quad (7.43)$$

So, using Caratheodary's result, we can deduce the following: if the period of $x(n)$ is to be estimated from the integer set $\mathbb{P} = \{P_1, P_2, \dots, P_N\}$, then the following number of samples are sufficient:

$$N_{min} = \max_{P_i \in \mathbb{P}} 2P_i \quad (7.44)$$

While this is a sufficiency bound, it is not also a necessity bound unlike the results of this chapter. For periodic signals with integer periods, the N_{min} above is strictly larger than the actual necessary and sufficiency bound derived in Theorem 7.2.1. For example, if $\mathbb{P} = \{7, 24, 100\}$, then $N_{min} = 200$, while Theorem 7.2.1 shows that just 120 samples are enough. This is because, in using Caratheodary's result directly, we have completely ignored the special harmonic structure between the frequencies ω_k 's in the case of integer-periodic signals.

Caratheodary's result, although not giving the tightest bounds for integer periods, offers useful insights for non-integer periodicity. When a continuous time periodic

signal is sampled, and the sampling rate is sufficiently high, the spectrum of the sampled signal will be as shown in Fig. 7.1(b). The lines in this spectrum are spaced at integer multiples of a fundamental frequency ω_0 . Such a sampled signal need not satisfy (1) for any integer P . As discussed in Sec. 7.1, the *approximate* period of this sampled signal is usually considered to be:

$$P = \frac{2\pi}{\omega_0} \quad (7.45)$$

which is typically not an integer, and there are several excellent algorithms for its estimation [3]–[7], [16], [95], [105]. However, one can again ask a similar question as in the previous sections of this chapter: Namely, is there a fundamental, generic, lower bound on the minimum required data-length for the identifiability of such non-integer periods? While a generally agreed upon rule of thumb in prior works has been that we need a data-length of at least twice the largest period (sometimes empirically justified for specific algorithms such as [95]), an algorithm-independent analysis on the identifiability issue itself has been missing in prior literature to the best of our knowledge.

For simplicity in analysis, let us consider the signal in Fig. 7.1(b) to have no $k = 0$ component. In time domain, it can be expressed as:

$$x(n) = \sum_{k=1}^K c_k e^{jk\omega_0 n} \quad (7.46)$$

Once again, applying Caratheodary's result directly without considering the harmonic relationship between the frequencies, tells us that $2K$ samples are sufficient to estimate the fundamental frequency. But can we leverage the harmonic structure to use fewer samples? The following result gives us a surprising answer:

Theorem 7.6.1. Non-Integer Periods: *Let ω_0 and ω_1 be distinct frequencies in $[0, 2\pi)$ such that the following two sets do not intersect:*

$$\{\omega_0, 2\omega_0, \dots, K\omega_0\} \quad \text{and} \quad \{\omega_1, 2\omega_1, \dots, K\omega_1\}$$

Then, there will always exist signals $x_0(n)$ and $x_1(n)$, with fundamental frequencies ω_0 and ω_1 respectively, such that:

$$x_0(n) = x_1(n) \quad \forall n \in \{0, 1, 2, \dots, 2K - 2\}$$

That is, the true fundamental frequency cannot be uniquely identified from these $2K - 1$ samples. ◇

Proof: The proof is similar to that of Theorem 7.2.3, where we need to consider null-space vectors of the $(2K - 1) \times 2K$ Vandermonde matrix constructed using the following complex exponentials as its columns: $\{e^{jk\omega_0 n}\}_{k=1}^K$ and $\{e^{jk\omega_1 n}\}_{k=1}^K$. Since the arguments are similar, we will skip further details here.

The above result reveals some interesting facts: Firstly, notice that the number of variables has reduced significantly as we go from (7.42) to (7.46). While we have K unknown frequencies in (7.42), we only have one frequency parameter ω_0 in (7.46). But there seems to be no benefit in terms of the minimum required data-length to identify the frequencies. The number of necessary and sufficient samples is still $2K$ in both cases. Secondly, one might intuitively expect that having more harmonics in the signal should help in estimating the fundamental frequency. However Theorem 7.6.1 suggests the exact opposite, showing that we in fact require more data-length when additional harmonics are present in the signal.

One of the most important outcomes of Theorem 7.6.1 is the following: For a signal satisfying (7.46), the minimum required data-length to estimate the fundamental is *not really twice the largest expected period*. Rather, the precise datalength depends only on the number of expected harmonics in the signal. For instance, consider the following two cases:

- If the original continuous time signal is sampled close to its Nyquist rate, the expected number lines in the signal's spectrum, K , is approximately $\lfloor 2\pi/\omega_0 \rfloor$, which is the closest integer approximation to the period. So in this special case, the minimum necessary and sufficient datalength using Caratheodary's result and Theorem 7.6.1 does come close to $2P_{max}$, where P_{max} is the largest expected period of the discrete time signal.
- However, if the sampling rate is sufficiently above the Nyquist rate, then the number of lines in the spectrum will be less than $\lfloor 2\pi/\omega_0 \rfloor$. So the minimum required samples can be significantly smaller than $2P_{max}$. For instance, if the sampling rate is twice the Nyquist rate, then K cannot be more than $\lfloor \pi/\omega_0 \rfloor$. This means that the minimum required datalength is only about P_{max} samples.

How do these results generalize to mixtures of periodic signals with non-integer periods? Once again, it can be shown that the minimum necessary and sufficient datalength will depend only on the total number of lines in the spectrum. For example, if $x(n)$ is a mixture of N periodic signals with at most K harmonics

(a) Period 4 sequence:

AGAT AGAT AGAT AGAT AGAT ...

(b) Period 10 sequence:

AGATAGATAG AGATAGATAG AGATAGATAG ...

Figure 7.4: Demonstration of Theorem 7.2.3 using DNA repeats. (a) Period 4 repeats used in DNA fingerprinting applications (GenBank G08921). (b) A period 10 sequence constructed such that its first 11 samples (shown in black) are the same as in the sequence (a).

each, then $2NK$ samples are both necessary and sufficient to find all the component periods. The necessity proof is very similar to that on Theorem 7.6.1, while sufficiency follows from Caratheodary's result [111]. Here too, in the special case when the sampling rate of the continuous time signal is close to the Nyquist rate, this limit comes close to $2NP_{max}$, where P_{max} is the largest expected period in the mixture.

7.7 Examples of DNA Repeats

To put the theorems derived in this chapter into perspective, let us consider examples of tandem repeats in DNA. Fig. 7.4 and 7.6 show two examples from the human DNA. Such repeats are today the primary bio-markers used for fingerprinting, tracking ancestry, studying population evolution and so on [77], [78]. They are also associated with a number of genetic disorders such as Huntington's disease and Friedreich's ataxia [69]. Given such a DNA sequence with repeats, what is the minimum data-length needed to be able to identify the correct period? Our results in this chapter can be used to address this question.

Let us start with a simple demonstration of Theorem 7.2.3. Consider the repeats shown in Fig. 7.4(a). These particular repeats occurs at several places in the human DNA (e.g. GenBank: X14720, G08921, M26434, M84567, X77751 etc.) and are widely used in fingerprinting applications [77]. Suppose we were to determine the true period of these repeats from the set of possible answers $\{4, 10\}$. Theorem 7.2.3 claims that, in general, we need at least $4 + 10 - (4, 10) = 12$ samples. That is, if there were < 12 samples, there could be sequences of periods 4 or 10 for which the true period cannot be determined by any technique. Indeed, as shown in Fig. 7.4 for this DNA sequence, a segment of length < 12 samples can be extended to both period 4 and period 10 repeats. The first 11 samples for the sequences shown in Fig. 7.4(a) and (b) are common. So given those 11 samples, one cannot determine

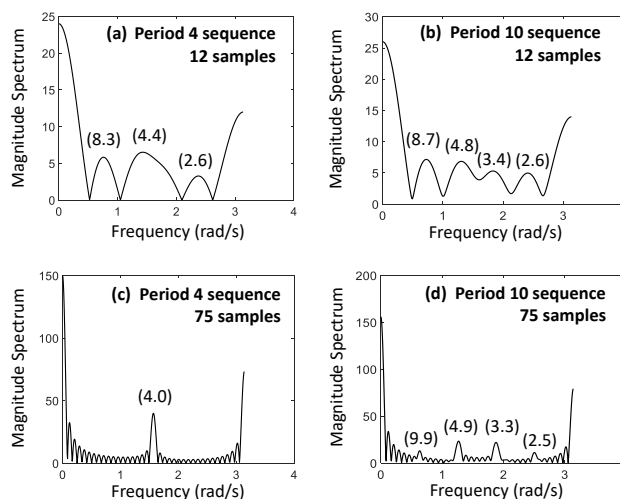


Figure 7.5: DFT spectra for the sequences in Fig. 7.4. The periods ($2\pi/\omega$ values) corresponding to the peak frequencies are shown in parenthesis. See text for details.

whether the true period was 4 or 10 no matter which method is used.

To give a quantitative idea of how insightful this knowledge of minimum data is, Fig. 7.5 shows the DFT spectra for the sequences in Fig. 7.4. A simple mapping of $A = 1$, $T = 2$, $C = 3$ and $G = 4$ was used to compute the DFT. As seen, a data-length of 12 is too short for DFT to reveal the true periods for either sequence. As the data-length increases to 75, the peak at period 4 in Fig. 7.5(c) is very distinct, and the harmonics of the period 10 in Fig. 7.5(d) are clearer. On the other hand, the technique used in the proof of Theorem 7.2.3 can provably identify the period from just 12 samples. It may be quite possible to devise other simpler techniques that can identify the period from data-lengths shorter than DFT's requirements. However, this chapter's focus is not to develop such techniques, but rather to show that there exists a fundamental mathematical bound (12 samples here), below which no possible technique can identify the true period. The existence of, and the precise expressions for such bounds were not known in prior literature.

Fig. 7.6 shows another example of human DNA repeats, demonstrating Theorem 7.2.1. For the repeats shown in Fig. 7.6(a), suppose we were to estimate the period from the set of possible values $\{1, 2, 3, 4, 5, 6\}$. Theorem 7.2.1 says that at least 10 samples are required in general. And indeed, Figs. 7.6(a) and (b) show two sequences, with period 5 and 6, for which the first 9 samples are common. So given those 9 samples, one can never estimate what the true underlying period is.

(a) Period 5 sequence:

TTTTC TTTTC TTTTC TTTTC ...

(b) Period 6 sequence:

TTTTCT TTTTCT TTTTCT TTTTCT ...

Figure 7.6: Demonstration of Theorem 7.2.1 using DNA repeats. (a) Period 5 repeats used in DNA fingerprinting applications (GenBank M86525). (b) A period 6 sequence constructed such that its first 9 samples (shown in black) are the same as in the sequence (a). See text for details.

7.8 Minimal Non-Contiguous Sampling For Period Estimation

An interesting question at this point is, if we are allowed to use non-contiguous samples, can we estimate the period using fewer samples than (7.5) or (7.30)? Notice that if this were possible, then those fewer samples must necessarily be spread out over a larger data-length than (7.5) or (7.30) (because of Theorem 7.2.1 and Theorem 7.4.2). Such a generalization of Theorems 4.3.1 and 7.4.2 is yet unknown. However, in this section, a smaller result is derived in this regard. It is shown that:

Theorem 7.8.1. *Given a periodic signal $x(n)$ whose period P lies in the set $\mathbb{P} = \{P_1, P_2\}$, where $P_1 < P_2$, the following number of samples is necessary and sufficient to identify P :*

$$M_{min} = \begin{cases} P_2 & \text{if } P_1 \text{ divides } P_2 \\ P_1 & \text{otherwise} \end{cases} \quad (7.47)$$

◇

When $\mathbb{P} = \{P_1, P_2\}$, the quantity L_{min} in Theorem 7.2.1 becomes $P_1 + P_2 - \gcd(P_1, P_2)$, which is clearly $\geq M_{min}$ from (7.47). In the following, the above theorem is proved using ideas from classical number theory, and the Nested Periodic Matrices of Chapter 3.

Proving Sufficiency

In order to prove the sufficiency part of Theorem 7.8.1, we will need the following two lemmas.

Lemma 7.8.1. *Let $x(n)$ be a periodic signal with period P . Let $y(n)$ be the M -fold decimated version of $x(n)$, that is,*

$$y(n) = x(Mn) \quad \forall n \in \mathbb{Z} \quad (7.48)$$

Then, $y(n)$ is also a periodic signal, with its period being a divisor of $\frac{P}{(M,P)}$. \diamond

Proof.

$$\begin{aligned} y\left(n + \frac{P}{(M,P)}\right) &= x\left(Mn + M\frac{P}{(M,P)}\right) \\ &= x(Mn) \\ &= y(n) \end{aligned}$$

So $\frac{P}{(M,P)}$ is a repetition index of $y(n)$. A repetition index of a signal must always be a multiple of its period (see Lemma 3 in [9]). $\nabla \nabla \nabla$

We also need the following lemma.

Lemma 7.8.2. *Let $x(n)$ be a periodic signal with period $P \neq 1$. Let M be an integer coprime to P , and let $y(n) = x(Mn)$. Then, $y(n)$ is a periodic signal with period $\neq 1$.* \diamond

Proof. From Lemma 7.8.1, it follows that $y(n)$ must be a periodic signal whose period is a divisor of P . So $y(n)$ can have period 1 if and only if all the entries in the following set are equal:

$$\mathbb{Y} = \{y(0), y(1), \dots, y(P-1)\} \quad (7.49)$$

Using (7.48), we can re-write \mathbb{Y} as:

$$\mathbb{Y} = \{x(0), x(M), \dots, x(MP-M)\} \quad (7.50)$$

Since $x(n)$ is periodic with period P , we can re-write this further as:

$$\mathbb{Y} = \{x(0 \bmod P), x(M \bmod P), \dots, x(MP-M \bmod P)\}$$

It is a well known result in Number Theory that when $(M, P) = 1$, the following two sets are the same (permuted versions of each other):

$$\begin{aligned} \{(0 \bmod P), (M \bmod P), \dots, (MP-M \bmod P)\} \\ = \{0, 1, \dots, P-1\} \end{aligned}$$

Hence the set \mathbb{Y} is in fact (a permuted version of) the following set:

$$\{x(0), x(1), \dots, x(P-1)\} \quad (7.51)$$

Since $x(n)$ has period $P \neq 1$, all the elements in the above set cannot be equal to each other. Hence, there are at least two unequal entries in (7.49), due to which, the period of $y(n)$ cannot be 1. $\nabla \nabla \nabla$

We will now prove the sufficiency part of Theorem 7.8.1. We shall do this by considering two distinct cases as follows:

Sufficiency when $P_1|P_2$

In this case, Theorem 7.8.1 claims that P_2 samples are sufficient. To prove this, we can use the Nested Periodic Matrices proposed in Chapter 3. Let $x(n)$ be a signal whose period is either P_1 or P_2 , and let \mathbf{x} be a $P_2 \times 1$ vector consisting of P_2 consecutive samples of $x(n)$. Further, let \mathbf{A} be a $P_2 \times P_2$ Nested Periodic Matrix. Then, since \mathbf{A} is a full rank matrix, the following system of equations has a unique solution for \mathbf{s} :

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (7.52)$$

Moreover, it follows from Lemma 3.2.3 that the Least Common Multiple of the periods of those columns of \mathbf{A} that are multiplied by non-zero entries in \mathbf{s} , is equal to the period of $x(n)$ (either P_1 or P_2 in this case). Hence, P_2 samples of $x(n)$ are sufficient to find its period.

Sufficiency when $P_1 \nmid P_2$

When $P_1 \nmid P_2$, P_1 samples are sufficient to find the period. We will first prove this for the case when P_1 and P_2 are coprime, and then extend the proof to the more general case.

When P_1 and P_2 are coprime, consider downsampling $x(n)$ by $M = P_2$ to yield $y(n)$. That is, $y(n) = x(Mn)$. In this case, if $x(n)$ had period P_2 , then, the period of $y(n)$ is 1 from Lemma 7.8.1. However, if $x(n)$ had period P_1 , then the output $y(n)$ cannot have period 1 according to Lemma 7.8.2. To check whether $y(n)$ has period 1, we need at most P_1 samples of $y(n)$, since its period can be any divisor of P_1 . This proves that P_1 samples are sufficient to find the period in this case.

Let us now consider the more general case when $(P_1, P_2) = G$ ($G \neq P_1$, since $P_1 \nmid P_2$). We now propose to use the setup shown in Fig. 7.7. Let us first consider

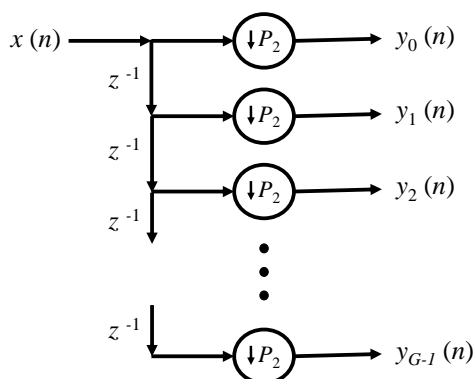


Figure 7.7: Finding the period of $x(n)$ when $(P_1, P_2) = G \neq P_1$. See text for details.

the case when $x(n)$ has period P_2 . Then, it is easy to see that all the outputs $y_0(n), y_1(n), \dots, y_{G-1}(n)$ will have period 1.

However, if $x(n)$ has period P_1 , then atleast one of $y_0(n), y_1(n), \dots, y_{G-1}(n)$ will have period > 1 . To prove this, we re-draw Fig. 7.7 as Fig. 7.8. If $x(n)$ had period P_1 , and if $P_1 \neq G$, then at least one of $u_0(n), u_1(n), \dots, u_{G-1}(n)$ must have period > 1 . This is because, if all of them are period 1 signals, then $x(n)$ must satisfy:

$$x(n + G) = x(n) \quad \forall n \in \mathbb{Z} \quad (7.53)$$

which then necessitates that $P_1 | G$ (since the period must always divide any repetition index). This is possible only if $G = P_1$, which contradicts our assumption that $P_1 \nmid P_2$.

Let us assume that $u_i(n)$ has period > 1 . Because of Lemma 7.8.1, the period of $u_i(n)$ must be a divisor of $\frac{P_1}{G}$. Further, since $\frac{P_1}{G}$ and $\frac{P_2}{G}$ are always co-prime, any divisor of $\frac{P_1}{G}$ is also coprime to $\frac{P_2}{G}$. So using Lemma 7.8.2, we can conclude that the period of $y_i(n)$ must be > 1 .

Notice that, using Lemma 7.8.1, $u_0(n), u_1(n), \dots, u_{G-1}(n)$ can have their periods as any divisors of $\frac{P_1}{G}$, and so the outputs $y_0(n), y_1(n), \dots, y_{G-1}(n)$ can have their periods as any divisors of $\frac{P_1}{G}$. So $\frac{P_1}{G}$ samples of each output are sufficient to check if they have period 1. Since there are G such outputs, $\frac{P_1}{G} \times G = P_1$ samples of $x(n)$ are sufficient to determine whether the period is P_1 or P_2 . This completes the proof of the sufficiency side of Theorem 7.8.1.

We will now prove the necessity part of Theorem 7.8.1.

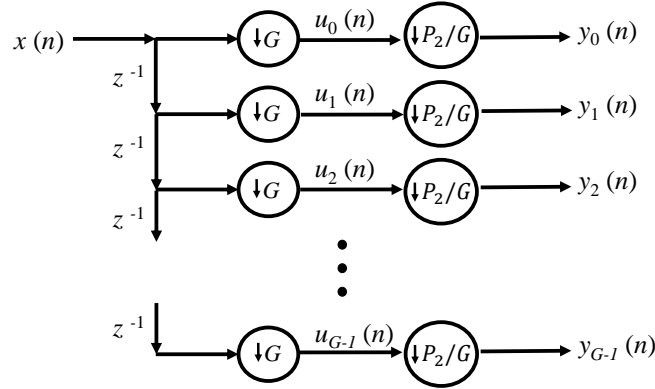


Figure 7.8: Re-drawing Fig. 7.7 for analysis. See text for details.

Proving Necessity

We will first show that at least P_1 samples are necessary to find the period from the set $\{P_1, P_2\}$, irrespective of whether $P_1|P_2$ or $P_1 \nmid P_2$. Later, we will show that when $P_1|P_2$, P_2 samples are necessary.

Theorem 7.8.2. *Given any set of $L < P_1$ time indices $\mathbb{N}_T = \{n_1, n_2, \dots, n_L\} \subset \mathbb{Z}$, there exist periodic signals $x_{P_1}(n)$ and $x_{P_2}(n)$ with periods P_1 and P_2 respectively such that*

$$x_{P_1}(n) = x_{P_2}(n) \quad \forall n \in \mathbb{N}_T \quad (7.54)$$

◇

Proof. Since $L < P_1$, there exists at least one integer in the set $\{0, 1, \dots, P_1 - 1\}$ that does not belong to the set $\{(n_1 \bmod P_1), (n_2 \bmod P_1), \dots, (n_L \bmod P_1)\}$. Let m be such an integer. We define $x_{P_1}(n)$ as follows:

$$x_{P_1}(n) = \begin{cases} 0 & \text{if } n \bmod P_1 = m \\ 1 & \text{otherwise} \end{cases} \quad (7.55)$$

It is easy to see that $x_{P_1}(n)$ has period P_1 (a zero occurs only once every P_1 samples). Notice that $x_{P_1}(n) = 1 \quad \forall n \in \mathbb{N}_T$. In the same way, we can construct a period P_2 signal $x_{P_2}(n)$ that satisfies $x_{P_2}(n) = 1 \quad \forall n \in \mathbb{N}_T$. Clearly, for these $x_{P_1}(n)$ and $x_{P_2}(n)$, (7.54) is satisfied. ▽ ▽ ▽

We will now prove that when $P_1|P_2$, one needs at least P_2 samples to estimate the period.

Theorem 7.8.3. Let $P_1|P_2$. Given any set of $L < P_2$ time indices $\mathbb{N}_T = \{n_1, n_2, \dots, n_L\} \subset \mathbb{Z}$, and any period P_1 signal $x_{P_1}(n)$, there exists a period P_2 signal $x_{P_2}(n)$ such that

$$x_{P_1}(n) = x_{P_2}(n) \quad \forall n \in \mathbb{N}_T \quad (7.56)$$

◇

Proof. Let $x_{P_2}(n)$ be defined to be equal to $x_{P_1}(n)$ for all $n \in \mathbb{N}_T$. Doing so will not violate the following condition:

$$x_{P_2}(n + P_2) = x_{P_2}(n) \quad \forall n \in \mathbb{N}_T \quad (7.57)$$

since, $n + P_2 = n + kP_1$ for some integer k .

Further, since $L < P_2$, there exists at least one integer in the set $\{0, 1, \dots, P_2 - 1\}$ that does not belong to the set $\{(n_1 \bmod P_2), (n_2 \bmod P_2), \dots, (n_L \bmod P_2)\}$. Let m be such an integer. Moreover, let u and v be integers such that $u > \max_n x_{P_1}(n)$ and $v < \max_n x_{P_1}(n)$. We define $x_{P_2}(n)$ as follows:

$$x_{P_2}(n) = \begin{cases} x_{P_1}(n) & \text{if } n \bmod P_2 \in \mathbb{N}_T \\ u & n \bmod P_2 = m \\ v & \text{otherwise} \end{cases} \quad (7.58)$$

It is easy to see that $x_{P_2}(n)$ has period P_2 (u occurs only once every P_2 samples). Hence, we have constructed a period P_2 signal $x_{P_2}(n)$ satisfying the conditions of the theorem. ▽ ▽ ▽

This completes the proof of Theorem 7.8.1.

7.9 Simulations Under Noise

We can easily adapt the period estimation algorithms proposed in Sec. 7.8 to deal with noisy inputs. For the case when $P_1|P_2$, one may use least squares to solve for the vector \mathbf{s} in (7.52). We refer the reader to Chapter 3 for more details. Here, we analyze the case when $P_1 \nmid P_2$. Recall that the period estimation techniques presented earlier in this section for this case involve, apart from downsampling, checking whether certain signals have period 1 or not. When there is noise, we can compute the variance of $u_i(n)$ (in Fig. 7.8) to check whether its period is 1. If the variance is less than a suitably chosen threshold, we may hypothesize that period is 1. To illustrate this, we consider the simple case when $(P_1, P_2) = 1$ here.

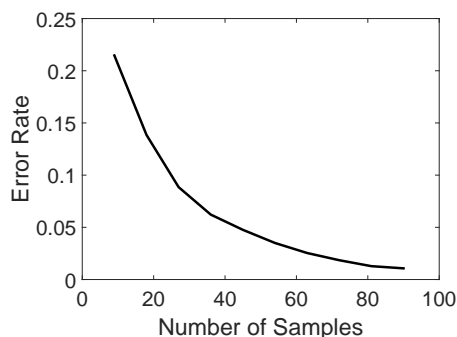


Figure 7.9: Error rate vs. Number of Samples. $P_1 = 9$, $P_2 = 13$, SNR = 0dB. See Sec. 7.9 for details.

When P_1 and P_2 are coprime, we considered the downsampler of Fig. 7.7, with $M = P_2$ (we will have only one channel in this case, since the GCD is 1). We argued that if $x(n)$ had period P_2 , then the period of $y(n)$ is 1 from Lemma 7.8.1. However, if $x(n)$ had period P_1 , then the output $y(n)$ cannot have period 1, and is in fact a permutation of the input itself (see the proof of Lemma 7.8.2). Let us now assume that the signal $x(n)$ was contaminated by an independent AWGN process $s(n)$ with sample variance σ_n^2 . Further, suppose that $x(n)$ was itself a randomly generated periodic signal with sample variance σ_x^2 .

In this case, it is easy to see that the output $y(n)$ would have sample variance σ_y^2 given by:

$$\sigma_y^2 = \begin{cases} \sigma_x^2 + \sigma_n^2 & \text{if } x(n) \text{ had period } P_1 \\ \sigma_n^2 & \text{if } x(n) \text{ had period } P_2 \end{cases} \quad (7.59)$$

So we may choose a threshold parameter $T = \sigma_n^2 + \frac{\sigma_x^2}{2}$, and predict the input to have period P_1 if the observed variance of $y(n)$ is larger than T , and predict the period as P_2 otherwise.

Using this technique, we performed the following two experiments. In the first, we study the accuracy of the period estimate as a function of the number of samples of $y(n)$ used for computing σ_y^2 . We chose $P_1 = 9$, $P_2 = 13$, and SNR = 0dB. The minimum number of samples needed in this case, as given by Theorem 7.8.1, is 9. For each value of ‘number of samples’, we generated 10000 signals with periods randomly chosen from the set $\{P_1, P_2\}$. Fig. 7.9 plots the fraction of times the period was incorrectly estimated as the error rate. It is intuitive that the error rate decreases as we have more samples.

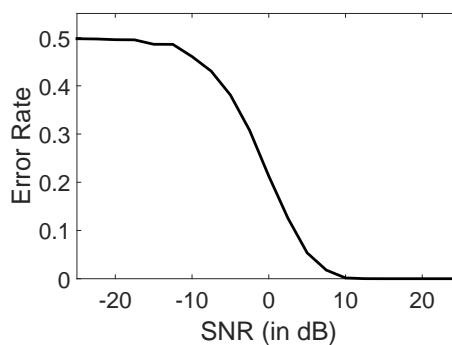


Figure 7.10: Error rate vs. SNR for the Minimum Samples Case. $P_1 = 9$, $P_2 = 13$, Number of samples = $P_1 = 9$. See Sec. 7.9 for details.

In our second experiment, we fixed the number of samples to be P_1 , and plotted the error rate for various values of SNR. This is shown in Fig. 7.10. Once again, as is consistent with intuition, we observed the error rate decrease as the SNR increases.

7.10 Concluding Remarks

This chapter derived fundamental bounds on the minimum necessary and sufficient data-length needed to estimate the integer period and the hidden periods of a sequence $x(n)$. The notion of hidden periods in a mixture was carefully formalized, including a discussion on their uniqueness and identifiability. While most of the results in this chapter are algorithm-independent, the datalength requirements for a particular case of the recently proposed dictionary based period estimators of [23] was also presented. We also briefly analyzed the case where the discrete signal's period might not exactly be an integer. Finally, the question of whether non-contiguous sampling helps to reduce the minimum number of samples required for period estimation was investigated using a simple case.

A possible extension of this chapter would be to multi-dimensional periodic signals. The extension of Caratheodary's results to the multi-dimensional setting in [112] provides useful initial insights. We would also like to theoretically study the accuracy of period estimation and its dependence on the number of samples in the presence of noise, under a rigorous statistical framework. Finally, from a practical perspective, all the algorithms used in the sufficiency proofs in this chapter are computationally intensive. This was not an issue in this chapter, since these algorithms were only used as proof techniques. However, an interesting question is as follows: Can there be algorithms that are both computationally efficient, as well as mathematically guaranteed to work with the theoretical minimum number of samples? This is an

open question we wish to address in the future.

7.11 Chapter Appendix

Lemma 7.11.1. *Let \mathbf{A} be a $N \times N$ matrix with distinct, Vandermonde columns. Let \mathbf{B} be the $(N - 1) \times N$ matrix obtained by removing the last row of \mathbf{A} . Then, any non-zero vector \mathbf{z} in the Null space of \mathbf{B} satisfies $z_i \neq 0 \forall i$.* \diamond

Proof. Assume the contrary, and thus let \mathbf{z} be a vector in $NULL(\mathbf{B})$, with say its i^{th} entry equal to 0. Let \mathbf{B}' be the $(N - 1) \times (N - 1)$ matrix obtained by deleting the i^{th} column of \mathbf{B} , and let \mathbf{z}' be the $(N - 1) \times 1$ vector obtained by deleting the i^{th} entry of \mathbf{z} . Since $\mathbf{Bz} = \mathbf{0}$, it follows that $\mathbf{B}'\mathbf{z}' = \mathbf{0}$. However, notice that \mathbf{B}' is a square Vandermonde matrix with distinct columns, and hence is full rank. So \mathbf{z}' must be $\mathbf{0}$, which then implies that $\mathbf{z} = \mathbf{0}$. $\nabla \nabla \nabla$

Lemma 7.11.2. *Let \mathbb{F} be an M -set. Let $\mathbb{G} \subseteq D.S.(\mathbb{F})$, such that $\mathbb{F} \subseteq \mathbb{G}$. Let \mathbb{H} be the following set:*

$$\mathbb{H} = \{g \in \mathbb{G} : Mg \notin \mathbb{G} \forall M > 1\} \quad (7.60)$$

Then, $\mathbb{H} = \mathbb{F}$. \diamond

Proof. Let $f \in \mathbb{F}$. This implies that $f \in \mathbb{G}$, since $\mathbb{F} \subseteq \mathbb{G}$. Suppose $f \notin \mathbb{H}$, then, $Mf \in \mathbb{G}$ for some $M > 1$. But since $\mathbb{G} \subseteq D.S.(\mathbb{F})$, $Mf|l$ for some $l \in \mathbb{F}$. This then implies that $f|l$. However, we have obtained a contradiction, since \mathbb{F} is an M -set. Hence, $f \in \mathbb{H}$. This proves that

$$\mathbb{F} \subseteq \mathbb{H} \quad (7.61)$$

Now, suppose $q \in \mathbb{G}$ such that $q \notin \mathbb{F}$. Since $\mathbb{G} \subseteq D.S.(\mathbb{F})$, $Kq = p$ for some $p \in \mathbb{F}$ and $K > 1$. But since $\mathbb{F} \subseteq \mathbb{G}$, this implies that $Kq = p$ for some $p \in \mathbb{G}$ and $K > 1$. So $q \notin \mathbb{H}$. This proves that $\mathbb{H} \subseteq \mathbb{F}$. Combining it with (7.61), we can conclude that $\mathbb{H} = \mathbb{F}$. $\nabla \nabla \nabla$

Lemma 7.11.3. *For any two positive integers P_i and P_j , the following holds true:*

$$P_i + P_j - (P_i, P_j) = \sum_{d \in D.S.(\{P_i, P_j\})} \phi(d) \quad (7.62)$$

\diamond

Proof. For any positive integer P , the following is true (from [24]):

$$\sum_{d|P} \phi(d) = P \quad (7.63)$$

Expanding each term in the L.H.S. of (7.62) using (7.63) gives us:

$$P_i + P_j - (P_i, P_j) = \sum_{d_i|P_i} \phi(d_i) + \sum_{d_j|P_j} \phi(d_j) - \sum_{d_k|G} \phi(d_k) \quad (7.64)$$

where $G = (P_i, P_j)$. Now, if $d_k|G$, then $d_k|P_i$ and $d_k|P_j$. So $D.S.(\{P_1\}) \cap D.S.(\{P_2\}) = D.S.(\{G\})$. Using the principle of inclusion and exclusion from set theory on the sets $D.S.(\{P_1\})$, $D.S.(\{P_2\})$, $D.S.(\{P_1\}) \cap D.S.(\{P_2\})$ and $D.S.(\{P_i, P_j\})$, it is now easy to see that the R.H.S. of (7.64) is equal to the R.H.S. of (7.62). $\nabla \nabla \nabla$

Chapter 8

ARBITRARILY SHAPED PERIODS IN MULTI-DIMENSIONAL PERIODICITY

Traditionally, the above notion of periodicity has been generalized for multidimensional signals in the following way [113], [114]: A signal $x(\mathbf{n})$, $\mathbf{n} \in \mathbb{Z}^D$ is said to be periodic if there exists a non-singular integer matrix \mathbf{P} such that:

$$x(\mathbf{n} + \mathbf{P}\mathbf{r}) = x(\mathbf{n}) \quad \forall \mathbf{n}, \mathbf{r} \in \mathbb{Z}^D \quad (8.1)$$

Analogous to the one dimensional case, such a \mathbf{P} can be called a repetition matrix of $x(\mathbf{n})$. For such an $x(\mathbf{n})$, the parallelepiped whose edges are represented by the column vectors of \mathbf{P} is known as a repetition region, since this parallelepiped when tiled periodically along the directions represented by the columns of \mathbf{P} , generates $x(\mathbf{n})$. Moreover, the absolute value of the determinant of \mathbf{P} gives the number of integer points inside the repetition region. If \mathbf{P} is a repetition matrix with a determinant that has the smallest absolute value among all possible repetition matrices for $x(\mathbf{n})$, then the parallelepiped whose edges are given by the columns of such a \mathbf{P} is known as a period of $x(\mathbf{n})$. In this case, \mathbf{P} is said to be a periodicity matrix of $x(\mathbf{n})$. In other words, a period is defined as the smallest parallelepiped that is a repetition region for the signal. Fig. 8.1(a) shows an example of a two dimensional periodic signal with period represented by the following matrix:

$$\mathbf{P} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \quad (8.2)$$

For simplicity, we will indicate periodic signals such as the one in Fig. 8.1(a) by plots similar to Fig. 8.1(b). (The horizontal direction in Fig. 8.1(a) and (b) represents the first coordinate.)

The question we want to address in this chapter is whether the above notion of periodicity captures all possible multidimensional periodic patterns. For example, according to the above definitions, the period and repetition regions are parallelepipeds. But there are many examples of signals with a much more diverse and rich set of geometries for the period. For example, the Dutch artist M. C. Escher had made several paintings consisting of figures that tile the two dimensional continuous

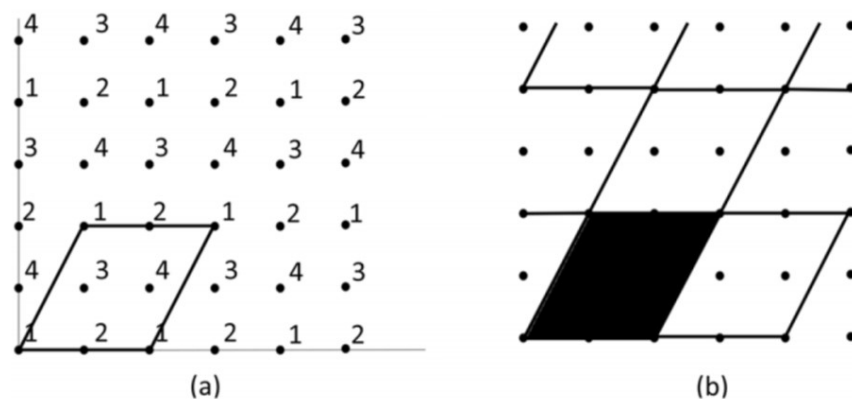


Figure 8.1: Part (a) - A two dimensional periodic signal according to the definition in (8.1), whose period is represented by the matrix in (8.2). The grid of dots is the set \mathbb{Z}^2 . The numbers shown indicate the value of the signal at those integer points. Part (b) - A convenient way to indicate repetition regions for such periodic signals.

space plane. His tiles include many intricate shapes such as reptiles, fishes, men etc. [115]. The website in the ¹, for example, investigates some of the techniques behind constructing Escher's figures. By nature, all these paintings are continuous space signals. But taking inspiration from them, we constructed the discrete space examples of Fig. 8.2 and Fig. 8.3. Such non-parallelepiped shaped tiles occur frequently in the real world, for example in honeycombs, scales of animals such as fish, in art, textile design, architecture and computer graphics. Such tilings are also useful in crystallography [116], [117].

Can we have a formal notion of periodicity with repetition regions and periods having such general shapes? Do each of these signals also have parallelepiped shaped periods? If so, how are these parallelepiped periods related to the ones shown here? Are there signals similar to Fig. 8.2 and Fig. 8.3 whose smallest repeating unit is not a parallelepiped?

To address these questions, we will formulate a new definition of periodicity in the following subsection, that is directly motivated from patterns such as in Fig. 8.2 and Fig. 8.3. We will then derive and discuss some properties of signals that satisfy

¹J. Britton, Escher in the classroom [Online]. Available: <http://britton.disted.camosun.bc.ca/jbescher.htm>

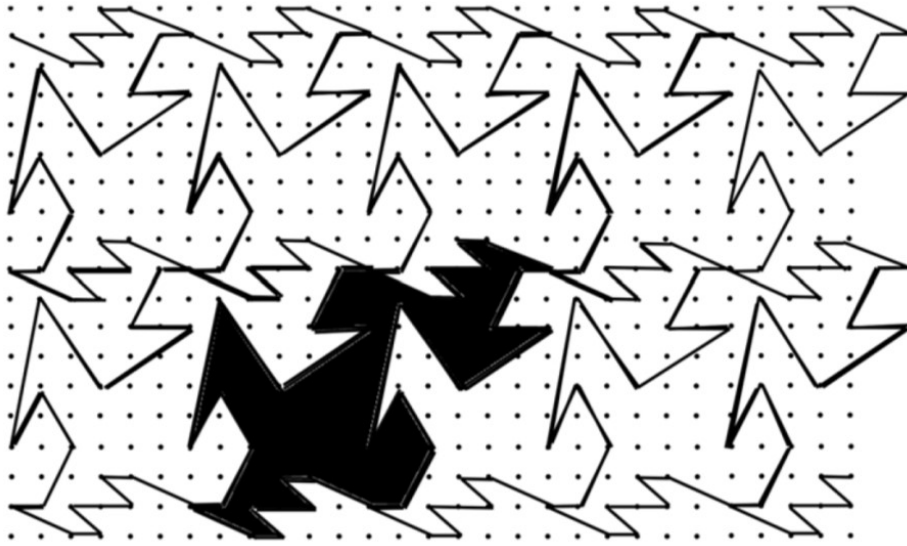


Figure 8.2: A two dimensional periodic signal with a non-parallelgram repetition region (shaded in black) tiling the plane.

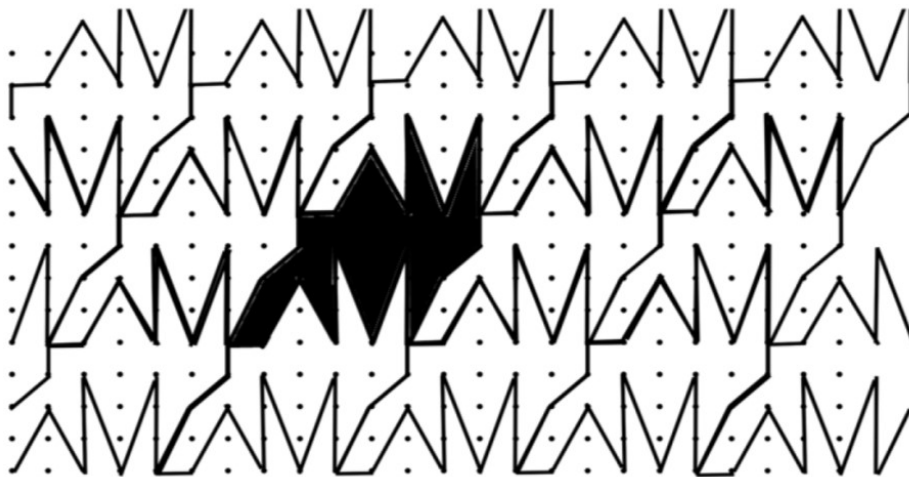


Figure 8.3: Another example of a two dimensional periodic signal with a non-parallelgram repetition region (shaded in black) tiling the plane.

this new definition. These properties show how such signals relate to the above mentioned parallelepiped formulation (Eq. 8.1).

Multi-Dimensional Periodicity - An Alternate Definition

Consider the periodic signals in Fig. 8.2 and Fig. 8.3. The period (black area) is some shape that when repeated along two directions tiles the plane. When we look at these figures, it is quite obvious to our intuition that there is a periodically repeating pattern on the plane. This perception of periodicity can be transcribed into words as follows: “copies of a shape, obtained by translating it along fixed directions, tile the plane and the signal values on that shape are replicated on its copies”. We want to model the same into a mathematical formulation. Hence the following definition:

Definition 8.0.1. A signal $x(\mathbf{n})$, is said to be periodic if there exists a non-empty set $\mathbb{S} \subset \mathbb{Z}^D$ and a matrix $\mathbf{P} \in \mathbb{Z}^{D \times N}$, such that if we define the sets:

$$\mathbb{S}_i = \mathbb{S} + \mathbf{P}\mathbf{i} \quad (8.3)$$

then,

1. $\cup_i \mathbb{S}_i = \mathbb{Z}^D$
2. $\mathbb{S}_i \cap \mathbb{S}_j = \phi$ for $\mathbb{S}_i \neq \mathbb{S}_j$
3. $x(\mathbf{n} + \mathbf{p}\mathbf{r}) = x(\mathbf{n}) \forall n \in \mathbb{S}$ and $r \in \mathbb{Z}^N$.

Here, $\mathbb{S} + \mathbf{P}\mathbf{i}$ indicates the set obtained by adding the vector $\mathbf{P}\mathbf{i}$ to every element of \mathbb{S} . As will be clear from the following discussion, for a signal $x(\mathbf{n})$ satisfying the above definition, the set $\{\mathbb{S}, \mathbf{P}\}$ is not unique. Some remarks on Definition 8.0.1:

1. *The set \mathbb{S} :* In the above definition, \mathbb{S} represents the set of points that form the shape that tiles the plane. For example, in Fig. 8.2 and Fig. 8.3, \mathbb{S} can be the shape shaded in black. We will call it a repetition region of the signal. Notice that a group of those shapes taken together also represents a valid repetition region that can tile the plane. So a signal satisfying Definition 8.0.1 will not have a unique repetition region. Also, note that unlike in the parallelepiped based periodicity framework in the beginning of this section, we are not explicitly specifying any shape for the repetition region.

2. *The sets \mathbb{S}_i and the matrix \mathbf{P} :* \mathbb{S}_i is a translated version of \mathbb{S} , translated along the directions represented by the columns of \mathbf{P} . For this reason, we will call \mathbf{P} as a translation matrix of \mathbb{S} . For a given repetition region \mathbb{S} , the translation matrix is not unique. This is because, the set of all integer vectors spanned by the columns of \mathbf{P} is the same as that spanned by the columns of \mathbf{PU} where \mathbf{U} is any integer unimodular matrix [113]. So if \mathbf{P} is a translation matrix for \mathbb{S} , then \mathbf{PU} will also be a valid translation matrix for \mathbb{S} .
3. *The size of \mathbf{P} :* The motivating idea behind Definition 8.0.1 involves a shape being translated and copied along fixed directions. But along how many directions do we need to translate? By taking a $D \times N$ translation matrix, we are essentially choosing N such directions. In Section 8.1, we will point out that if \mathbf{P} has rank $R < D$, then the repetition region must be of infinite size along $D - R$ dimensions. For simplicity, we will not consider such signals here. On the other hand, we will show that when \mathbf{P} has rank D , one may assume $N = D$ without loss of generality. So in the following discussions, we will assume \mathbf{P} to be a full rank $D \times D$ matrix.
4. *The three conditions:* The first condition in the definition just says that the set \mathbb{S} and its shifted copies must cover the entire space. The second condition requires that different shifted versions of \mathbb{S} must not overlap with each other. Finally, the third condition requires that the value of the signal $x(\mathbf{n})$ on the points in \mathbb{S} must replicate themselves on all the shifted copies of \mathbb{S} . These conditions were directly motivated from signals such as those in Fig. 8.2 and Fig. 8.3.

Notice that a repetition region \mathbb{S} might itself not be the period. For instance, in Fig. 8.2 and Fig. 8.3, a group of the indicated shapes taken together as one big shape also represent a valid repetition region. We will now formally define what we mean by a period of a periodic signal. To do so, we note that in the parallelepiped based definitions for repetition regions and periods discussed earlier, the period was defined as any parallelepiped that is a repetition region, and has the smallest number of points among all such repeating parallelepipeds. We will generalize it to more general shapes in the following definition:

Definition 8.0.2. *For a signal that is periodic according to Definition 1, a repetition region is said to be a period if for any other repetition region of the signal,*

Properties Arising From The New Definition

We will now prove some properties of signals that are periodic according to Definition 8.0.1 of the previous subsection. In the process, we will relate them to the conventional parallelepiped based formulation mentioned after (8.1). First we prove:

Theorem 8.0.1. *If $x(\mathbf{n})$ is periodic according to Definition 8.0.1, then $x(\mathbf{n} + \mathbf{Pr}) = x(\mathbf{n}) \forall \mathbf{n}, \mathbf{r} \in \mathbb{Z}^D$.*

Note the difference between Theorem 8.0.1 and condition 3 of Definition 8.0.1. Theorem 8.0.1 talks about all integer vectors $\mathbf{n} \in \mathbb{Z}^D$ whereas condition 3 required only that $\mathbf{n} \in \mathbb{S}$.

Proof. Consider any $\mathbf{n} \in \mathbb{Z}^D$. From conditions 1 and 2 in Definition 8.0.1, there must exist a unique $\mathbf{i} \in \mathbb{Z}^D$ such that $\mathbf{n} \in \mathbb{S}_i$. In that case, there must exist a point \mathbf{m} in \mathbb{S} such that $\mathbf{n} = \mathbf{m} + \mathbf{Pi}$ because of (8.3). So $x(\mathbf{n} + \mathbf{Pr}) = x(\mathbf{m} + \mathbf{Pi} + \mathbf{Pr}) = x(\mathbf{m} + \mathbf{P}(\mathbf{i} + \mathbf{r})) = x(\mathbf{m})$, where the last equality follows from condition 3 in Definition 8.0.1. Since $\mathbf{m} \in \mathbb{S}$, condition 3 in Definition 1 also implies that $x(\mathbf{m}) = x(\mathbf{m} + \mathbf{Pi}) = x(\mathbf{n})$. Hence, $x(\mathbf{n} + \mathbf{Pr}) = x(\mathbf{n})$. $\nabla \nabla \nabla$

The above theorem is important for the following reason. Let us say we have a set \mathbb{S} which was used in verifying Definition 8.0.1 to determine that a particular signal $x(\mathbf{n})$ is periodic. Irrespective of the shape of \mathbb{S} , Theorem 8.0.1 tells us that we can always have a parallelepiped repetition region for $x(\mathbf{n})$. This is because, it is well known in the literature ([113], [114] etc.) that for any signal $x(\mathbf{n})$ that satisfies:

$$x(\mathbf{n} + \mathbf{Pr}) = x(\mathbf{n}) \forall \mathbf{n}, \mathbf{r} \in \mathbb{Z}^D \quad (8.4)$$

a parallelepiped $\mathbb{P} \subset \mathbb{Z}^D$ that has the columns of \mathbf{P} as its edge vectors, when tiled periodically along the directions represented by the columns of \mathbf{P} , generates $x(\mathbf{n})$. It can easily be shown that such a parallelepiped satisfies all the conditions of Definition 8.0.1 too, and is hence a valid repetition region according to our new definitions for periodicity. It's translation matrix is in fact \mathbf{P} . Consider the signal in Fig. 8.2 for instance. Let \mathbb{S} be the repetition region shaded in black, and let \mathbf{P} be its translation matrix. If we denote the parallelogram that has its edges along the columns of \mathbf{P} by \mathbb{P} , then Fig. 8.4 shows the same signal as having \mathbf{P} (shaded part) as a repetition region.

How is such a parallelepiped \mathbb{P} related to the original repetition region \mathbb{S} ? Consider the following Theorem.

Theorem 8.0.2. *Let $x(\mathbf{n})$ be periodic according to Definition 8.0.1, with repetition region \mathbb{S} and a translation matrix \mathbf{P} . Let \mathbb{P} be another repetition region for $x(\mathbf{n})$ with the same translation matrix. Then, $|\mathbb{S}| = |\mathbb{P}| = |\det(\mathbf{P})|$.*

Proof. Let \mathbb{S} and \mathbb{P} represent two possible repetition regions for $x(\mathbf{n})$ that have the translation matrix \mathbf{P} . To every point in \mathbb{P} , we can associate a point in \mathbb{S} in the following way. Let \mathbf{m} be a point in \mathbb{P} . Then, there exists a unique point \mathbf{n} in \mathbb{S} such that $\mathbf{m} = \mathbf{n} + \mathbf{P}\mathbf{v}$ for some $\mathbf{v} \in \mathbb{Z}^D$. This is because:

Existence of \mathbf{n} : Follows from Condition 1 in Definition 8.0.1.

Uniqueness of \mathbf{n} : If there were two points \mathbf{n} and \mathbf{n}' in \mathbb{S} such that $\mathbf{m} = \mathbf{n} + \mathbf{P}\mathbf{v} = \mathbf{n}' + \mathbf{P}\mathbf{v}'$, then $\mathbf{n}' = \mathbf{n} + \mathbf{P}(\mathbf{v} - \mathbf{v}')$. Let $\mathbf{v} - \mathbf{v}' = \mathbf{i}$. Then, from Definition 8.0.1, since $\mathbf{n} \in \mathbb{S}$, \mathbf{n}' must belong to $\mathbb{S}_{\mathbf{i}}$. But $\mathbb{S} \cap \mathbb{S}_{\mathbf{i}}$ is the null set unless $\mathbf{i} = \mathbf{0}$ by Condition 2 of Definition 8.0.1. This then implies that $\mathbf{n}' = \mathbf{n} + \mathbf{P}\mathbf{i} = \mathbf{n}$. Hence, uniqueness has been proved.

From the above arguments, we can define a mapping that relates every point in \mathbb{P} to a point in \mathbb{S} . We now claim that this mapping is one-one. That is, two points in \mathbb{P} cannot be associated to the same point in \mathbb{S} . We can prove this by contradiction. Suppose there were two points \mathbf{p}_1 and \mathbf{p}_2 in \mathbb{P} such that $\mathbf{p}_1 + \mathbf{P}\mathbf{v}_1 = \mathbf{p}_2 + \mathbf{P}\mathbf{v}_2 = \mathbf{s}$, for some $\mathbf{s} \in \mathbb{S}$. Then, $\mathbf{p}_1 = \mathbf{p}_2 + \mathbf{P}(\mathbf{v}_2 - \mathbf{v}_1)$. Since \mathbf{p}_1 and \mathbf{p}_2 are distinct points, $\mathbf{v}_1 \neq \mathbf{v}_2$. But this leads to a contradiction, since \mathbb{P} itself is a repetition region that satisfies Definition 8.0.1, and using an argument similar to the uniqueness of \mathbf{n} argument above, $\mathbf{p}_1 = \mathbf{p}_2 + \mathbf{P}(\mathbf{v}_2 - \mathbf{v}_1)$ cannot be possible for $\mathbf{v}_1 \neq \mathbf{v}_2$.

So we have proved that there exists a one-one mapping from \mathbb{P} to \mathbb{S} . This means that the number of points in \mathbb{P} must be less than or equal to the number of points in \mathbb{S} .

Interchanging \mathbb{P} and \mathbb{S} in the proof above, we can similarly construct a one-one mapping from \mathbb{S} to \mathbb{P} . And hence, the number of points in \mathbb{S} must actually be equal to the number of points in \mathbb{P} .

Finally, as we remarked earlier, the parallelepiped whose edges are represented by the columns of \mathbf{P} also satisfies Definition 8.0.1 as a repetition region with the same translation matrix \mathbf{P} . The number of points in such a parallelepiped is given by $|\det(\mathbf{P})|$ [113], [114]. So from the above result, the number of points in any repetition region with translation matrix \mathbf{P} is $|\det(\mathbf{P})|$. ▽ ▽ ▽

Theorem 8.0.1 tells us that any signal $x(\mathbf{n})$ that is periodic according to Definition 8.0.1, having possibly arbitrary shaped periods such as in Fig. 8.2 or Fig. 8.3, will *also* always have a parallelepiped shaped repetition region. Theorem 8.0.2 shows that the number of points in that parallelepiped is in fact equal to the number of points in the “irregular” repetition region \mathbb{S} . This means, for example, that the parallelograms shown in Fig. 8.4 have the same number of points as the region shaded in black in Fig. 8.2. Moreover, combining Theorem 8.0.1 and Theorem 8.0.2 we can arrive at an important conclusion that the smallest repetition region for $x(\mathbf{n})$ can always be assumed to be a parallelepiped:

Theorem 8.0.3. *Let $x(\mathbf{n})$ be periodic according to Definition 8.0.1. Let \mathbb{S} be an arbitrarily shaped period of $x(\mathbf{n})$ according to Definition 8.0.2. Then, there will always exist a parallelepiped that is also a period of $x(\mathbf{n})$, having the same number of points as \mathbb{S} .*

Proof. Let \mathbb{P} be the parallelepiped whose edges are represented by the columns of the translation matrix of \mathbb{S} . As discussed earlier, from Theorem 8.0.1, \mathbb{P} is a repetition region for $x(\mathbf{n})$. And from Theorem 8.0.2, \mathbb{P} and \mathbb{S} have the same number of points. However, no repetition region can have smaller number of points than \mathbb{S} by Definition 8.0.2. And hence no other repetition region can have smaller number of points than \mathbb{P} . So \mathbb{P} is a period of according to Definition 8.0.2. $\nabla \nabla \nabla$

8.1 A Note on Translational Matrices

We had allowed the number of columns in the translation matrix \mathbf{P} in Definition 8.0.1 to be any natural number N . But we restricted our subsequent analysis to \mathbf{P} being a full rank square matrix. We will now discuss what happens when has rank $< D$, or has rank $= D$ but more than D columns.

First, consider the case when the rank of \mathbf{P} is less than D . If a shape \mathbb{S} were to tile the entire space \mathbb{Z}^D using translations along the columns of such a \mathbf{P} , then such a shape \mathbb{S} must have infinite extent along the $D - \text{rank}(P)$ dimensional subspace that is orthogonal to the space spanned by the columns of \mathbf{P} . For simplicity, we did not discuss such signals in this work.

On the other hand, if \mathbf{P} has full rank but has $N > D$, then it can be shown that the set of all copies of \mathbb{S} obtained by translations along the columns of \mathbf{P} would be the same set as that obtained by translating along the D directions represented by the

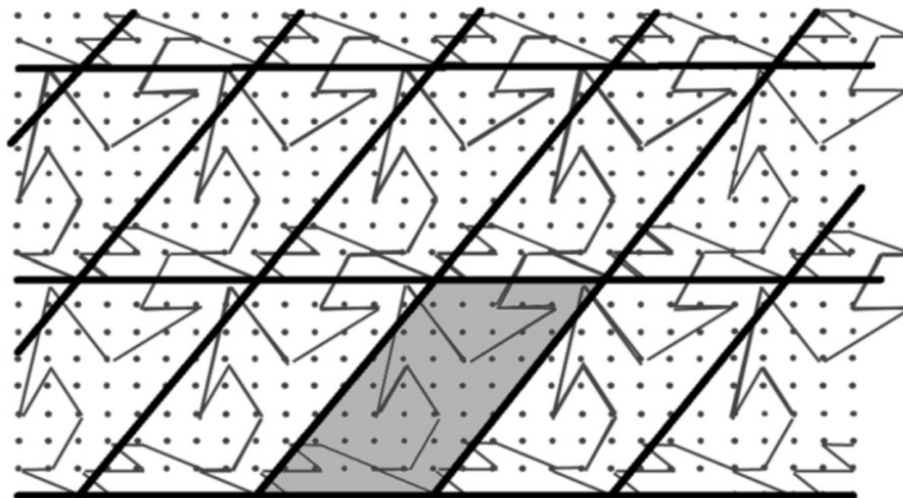


Figure 8.4: A parallelepiped repetition region always exists for any signal following Definition 8.0.1. Indicating this for the signal of Fig. 8.2.

columns of the greatest left divisor of \mathbf{P} . We will explain this in more detail in the following.

A left divisor of \mathbf{P} is defined as a matrix $\mathbf{L} \in \mathbb{Z}^{D \times D}$ that satisfies $\mathbf{P} = \mathbf{L}\mathbf{R}$ for some $\mathbf{R} \in \mathbb{Z}^{D \times N}$ [113]. A left divisor of \mathbf{P} , call it \mathbf{L}_0 , that satisfies the property that any left divisor of \mathbf{P} is a left divisor of \mathbf{L}_0 , is known as a greatest left divisor (gld) of \mathbf{P} . It can be shown that a gld of \mathbf{P} exists when \mathbf{P} has rank D . Notice that gld's are not unique, since if \mathbf{L}_0 is a gld, so is $\mathbf{L}_0\mathbf{U}$ for any integer unitary matrix \mathbf{U} . (For more details on gld's, see [118]–[120]).

Let \mathbf{L}_0 represent a gld of \mathbf{P} . We claimed above that the set $\{\mathbb{S}_i : i \in \mathbb{Z}^D\}$ obtained by using \mathbf{P} as a translation matrix is the same as that obtained by using \mathbf{L}_0 as the translation matrix. This follows from the following theorem that essentially states that the lattice generated by \mathbf{P} is the same as the lattice generated by \mathbf{L}_0 .

Theorem 8.1.1. *Let $\mathbf{P} \in \mathbb{Z}^{D \times N}$ have rank D , and let \mathbf{L}_0 be its gld. Then, there will exist $\mathbf{z} \in \mathbb{Z}^D$ and $\mathbf{i} \in \mathbb{Z}^N$ such that $\mathbf{z} = \mathbf{P}\mathbf{i}$ iff there exists a $\mathbf{j} \in \mathbb{Z}^D$ such that $\mathbf{z} = \mathbf{L}_0\mathbf{j}$.*

Proof. Let $\mathbf{z} = \mathbf{P}\mathbf{i}$. Since \mathbf{L}_0 is a left divisor of \mathbf{P} , there must exist an integer matrix \mathbf{R} such that $\mathbf{P} = \mathbf{L}_0\mathbf{R}$. Substituting this into $\mathbf{z} = \mathbf{P}\mathbf{i}$ gives $\mathbf{z} = \mathbf{L}_0\mathbf{R}\mathbf{i} = \mathbf{L}_0\mathbf{j}$ where $\mathbf{j} = \mathbf{R}\mathbf{i}$.

Conversely, let $\mathbf{z} = \mathbf{L}_0\mathbf{j}$. Let $\mathbf{P} = \mathbf{U}[\Lambda\mathbf{0}]\mathbf{V}$ be the Smith-form decomposition of \mathbf{P} [121], where \mathbf{U} and \mathbf{V} are integer unimodular matrices, $\mathbf{0}$ is a $(N - D) \times D$ zero matrix and Λ is an integer diagonal matrix. It can be shown that any gld of \mathbf{P} can be expressed as $\mathbf{L}_0 = \mathbf{U}\Lambda\mathbf{A}$ for some integer unimodular matrix \mathbf{A} . So $\mathbf{z} = \mathbf{L}_0\mathbf{j}$ can now be re-written as:

$$\mathbf{z} = \mathbf{L}_0\mathbf{j} = \mathbf{U}\Lambda\mathbf{A}\mathbf{j} = \mathbf{U}[\Lambda\mathbf{0}] \begin{bmatrix} \mathbf{A}\mathbf{j} \\ \mathbf{n} \end{bmatrix} \quad (8.5)$$

where \mathbf{n} can be any $(N - D) \times 1$ integer vector. We can further rewrite this as follows:

$$\mathbf{z} = \mathbf{U}[\Lambda\mathbf{0}] \begin{bmatrix} \mathbf{A}\mathbf{j} \\ \mathbf{n} \end{bmatrix} = \mathbf{U}[\Lambda\mathbf{0}]\mathbf{V}\mathbf{V}^{-1} \begin{bmatrix} \mathbf{A}\mathbf{j} \\ \mathbf{n} \end{bmatrix} = \mathbf{P}\mathbf{V}^{-1} \begin{bmatrix} \mathbf{A}\mathbf{j} \\ \mathbf{n} \end{bmatrix} \quad (8.6)$$

Since \mathbf{V} is a unimodular matrix, its inverse must also be an integer matrix. So if we define:

$$\mathbf{i} = \mathbf{V}^{-1} \begin{bmatrix} \mathbf{A}\mathbf{j} \\ \mathbf{n} \end{bmatrix} \quad (8.7)$$

then, we have $\mathbf{z} = \mathbf{P}\mathbf{i}$.

▽ ▽ ▽

8.2 Conclusion

The above discussion tells us that the parallelepiped based analysis for periodic signals is also suitable for signals that follow our new definition of periodicity, and hence for signals such as those in Fig. 2 and Fig. 3, where the intuitively apparent period might not be a parallelepiped to start with. This realization can have many applications. For instance, in the previous chapters, expansions of one dimensional periodic signals in terms of specially designed bases such as the Nested Periodic Subspaces) was developed into a family of algorithms that estimate unknown periods of signals. The Ramanujan subspaces have been generalized in [122] for multidimensional signals that follow the traditional parallelepiped based periodicity model. A natural question is whether those representations can be used for signals having more arbitrary shaped tilings such as those in Fig. 8.2 and Fig. 8.3. Our work here shows that this is indeed the case. Moreover, for such a signal, given any period and its translation matrix, we are able to identify the parallelepiped period that can be associated to the same signal. This is a very useful fact in determining the basis to represent the signal. More generally, these results enable us to incorporate such signals into the traditional multidimensional multirate analysis, for example [113], [114], [120]. Before we conclude, we would like to conjecture that similar results exist for continuous-space multidimensional periodic signals.

Chapter 9

CONCLUDING REMARKS

This thesis explores the classical problem of period estimation from a new perspective. We started by developing the theory of Nested Periodic Subspaces (NPSs), which were inspired from Ramanujan sums [8]. We showed that the NPSs are the least redundant sets of subspaces to span periodic sequences. It was shown that this minimality leads to several unique properties of the NPSs, which make them ideally suited for developing new period estimation algorithms.

We presented three new algorithms here: the Nested Periodic Dictionaries in Chapter 3, the Ramanujan Filter Bank in Chapter 5 and the iMUSIC family of algorithms in Chapter 6. These new methods are especially advantageous when the available datalength is short, or when there are multiple hidden periodicities in the signal. Using applications such as DNA micro-satellites, protein repeats and epileptic seizures, we compared their performances with some of the state-of-the-art methods in these application domains.

Apart from practical contributions, this thesis answered a number of fundamental questions regarding periodicity which were previously unknown. For instance, in Chapter 7, we derived precise expressions for the minimum required datalength for the identifiability of the period of a signal. It is somewhat surprising that in spite of period estimation having a rich history in DSP, this question was never studied previously in a fundamental manner. Most prior works assume (using heuristic arguments) that one needs at least twice the the number of samples as the largest expected period. However, as we showed in Chapter 7, this is only true under some very specific situations. We also derived the exact bounds for the case of mixtures of multiple periodic signals. The following extension of the minimum datalength question was also addressed: If we are allowed to pick the samples in a non-contiguous manner, can we estimate the period of a signal using a fewer number of samples than in the contiguous case? For the case of distinguishing between two potential periods, it was shown that this is indeed the case, and that the optimal sampling patten takes an interesting form of uniformly spaced bunches. Extending this non-contiguous samples analysis to more than two periods seems challenging, but will be interesting to study in the future.

There are two important directions to be pursued further. Firstly, most of the analysis we presented in this thesis did not rigorously address the issue of noisy data. For instance, how to optimally design the thresholds for the dictionary techniques of Chapter 3 under noisy data? How resilient is the LCM property of the Ramanujan and Nested Periodic Subspaces to SNR? How do the minimum datalength bounds change with SNR for the results in Chapter 7? While we empirically analyzed some of these aspects in this thesis, a thorough theoretical framework that incorporates noise models is yet to come.

A second important direction to explore is the connection with deep learning, which is currently the most popular way to approach many problems in engineering. In our ongoing research, we have noticed that Convolutional Neural Networks (CNNs), when trained for period extraction, have interesting connections with filter banks designed using Ramanujan Sums. The various trained layers in such networks can be related to DSP operations such as sub-band splitting. Even though the frequency responses of such CNNs are difficult to interpret due to non-linearities, they turn out to be strikingly similar to the characteristic coprime band structures of the Ramanujan Filter Bank (such as in Fig. 5.1). The significance of such CNNs seems limited in 1D periodicity applications, since the existing algorithms, including the ones proposed in this thesis, seem to be performing much better in our preliminary experiments. However, such CNNs can potentially make a significant difference for periodic non-separable lattices in 2D, where classical methods are not competitive computationally. This will be an important direction to follow going into the future.

BIBLIOGRAPHY

- [1] D. H. Johnson et. al., “A thread counting algorithm for art forensics,” in *IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, 2009.
- [2] Abbott, B. P. et. al, “Gw170817: Observation of gravitational waves from a binary neutron star inspiral,” *Phys. Rev. Lett.*, vol. 119, p. 161 101, 16 2017.
- [3] M. R. Schroeder, “Period histogram and product spectrum: New methods for fundamental frequency measurement,” *J. of the Acoustical Soc. of America*, vol. 43, no. 4, pp. 829–834, 1968.
- [4] D. Hermes, “Measurement of pitch by subharmonic summation,” *J. of the Acoustical Soc. of America*, vol. 83, no. 1, pp. 257–264, 1988.
- [5] X. Sun, “Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio,” in *IEEE Intl. Conf. on Acoustics, Speech, and Sig. Proc., Orlando, USA*, 2002.
- [6] A. Camacho and J. G. Harris, “A sawtooth waveform inspired pitch estimator for speech and music,” *J. of the Acoust. Soc. of America*, vol. 124, no. 3, pp. 1638–1652, 2008.
- [7] T. Drugman and A. Alwan, “Joint robust voicing detection and pitch estimation based on residual harmonics,” *Proc. Interspeech*, pp. 1973–1976, 2011.
- [8] S. Ramanujan, “On certain trigonometrical sums and their applications in the theory of numbers,” *Trans. of the Cambridge Philosophical Society*, vol. 22, no. 13, pp. 259–276, 1918.
- [9] P. P. Vaidyanathan, “Ramanujan sums in the context of signal processing: Part i: Fundamentals,” *IEEE Trans. on Sig. Proc.*, vol. 62, no. 16, pp. 4158–4172, Aug. 2014.
- [10] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [11] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization,” *Proc. of the National Acad. of Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [12] J. D. Wise, J. Caprio, and T. W. Parks, “Maximum likelihood pitch estimation,” in *Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Proc.*, 1976.
- [13] W. A. Sethares and T. W. Staley, “Periodicity transforms,” *IEEE Trans. on Sig. Proc.*, vol. 47, pp. 2953–2964, Nov. 1999.

- [14] D. D. Muresan and T. W. Parks, "Orthogonal, exactly periodic subspace decomposition," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, Sep. 2003.
- [15] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Ant. Propag.*, vol. 34, pp. 276–280, 1986.
- [16] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. on Aud., Speech, and Lang. Proc.*, vol. 15, pp. 1635–1644, 2007.
- [17] M. G. Christensen, A. Jakobsson and S. H. Jensen, "Multi-pitch estimation using harmonic music," *Asilomar Conf. on Sig., Sys. and Comp., Pacific Grove, CA*, 2006.
- [18] P. P. Vaidyanathan, "Ramanujan sums in the context of signal processing: Part ii: Fir representations and applications," *IEEE Trans. on Sig. Proc.*, vol. 62, no. 16, pp. 4158–4172, 2014.
- [19] M. Planat, M. Minarovjeh, and M. Saniga, "Ramanujan sums analysis of long-period sequences and $1/f$ noise," *Europhys. Letters*, vol. 85, no. 4, 2009.
- [20] L. T. Mainardi, M. Bertinelli, and R. Sassi, "Analysis of α -wave alternans using the ramanujan transform," *Comput. Cardiol.*, vol. 35, pp. 605–608, 2008.
- [21] Y. Changchuan, Y. X. E., and W. Jiasong, "A novel method for comparative analysis of dna sequences by ramanujan-fourier transform," *Journal of Computational Biology*, vol. 21, no. 12, pp. 843–851, 2009.
- [22] M. Lagha, "Doppler spectrum estimation by ramanujan-fourier transform (rft)," *Digital Signal Processing*, vol. 19, no. 5, pp. 843–851, 2009.
- [23] S. V. Tenneti and P. P. Vaidyanathan, "Nested periodic matrices and dictionaries: New signal representations for period estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3736–3750, Jul. 2015.
- [24] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*. Oxford University Press Inc., 2008.
- [25] M. Nakashizuka, "A sparse decomposition for periodic signal mixtures," in *15th Intl. Conf. on Digital Sig. Proc.*, 2007.
- [26] P. P. Vaidyanathan and P. Pal, "The farey dictionary for sparse representation of periodic signals," in *Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Proc., Italy*, 2014.
- [27] W. R. Atchley, J. Zhao, A. D. Fernandes, and T. Druke, "Solving the protein sequence metric problem," *Nat. Acad. Sci.*, vol. 102, pp. 6395–6400, 2005.

- [28] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, and R. G. Mark, “Physiobank, physiotookit, and physionet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23,
- [29] B. Santhanam and P. Maragos, “Harmonic analysis and restoration of separation methods for periodic signal mixtures: Algebraic separation vs comb filtering,” *Signal Process.*, vol. 69, no. 1, pp. 81–91, 1998.
- [30] S. C. Pei and K. S. Lu, “Intrinsic integer-periodic functions for discrete periodicity detection,” *IEEE Sig. Proc. Letters*, vol. 22, no. 8, pp. 1108–1112, Aug. 2015.
- [31] D. Malioutov, M. Cetin, and A. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [32] P. Pal and P. P. Vaidyanathan, “Correlation-aware techniques for sparse support recovery,” *14th IEEE SSP workshop, Ann Arbor, MI*, 2012.
- [33] ———, “Pushing the limits of sparse support recovery using correlation information,” *IEEE Trans. on Signal Processing*, vol. 63, no. 3, pp. 711–726, 2015.
- [34] B. Recht, M. Fazel, and P. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
- [35] P. Pal and P. P. Vaidyanathan, “Gridless methods for underdetermined source estimation,” *Proc. Asil. Conf. Sig., Sys., and Comp., Monterey, CA*, 2014.
- [36] ———, “A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising,” *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 737–741, 2014.
- [37] S. V. Tenneti and P. P. Vaidyanathan, “A unified theory of union of subspaces representations for period estimation,” *IEEE Trans. on Sig. Proc.*, vol. 64, no. 20, pp. 5217–5231, 2016.
- [38] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Info. Th.*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [39] R. G. Baraniuk, “Compressive sensing,” *IEEE Sig. Proc. Mag.*, vol. 24, no. 4, pp. 118–124, 2015.
- [40] T. Blumensath and M. Davies, “Sampling theorems for signals from the union of finite-dimensional linear subspaces,” *IEEE Trans. on Inf. Th.*, vol. 55, no. 4, pp. 1872–1882, 2009.
- [41] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Info. Th.*, vol. 50, pp. 2231–2242, 2004.

- [42] E. Candes and C. F-Grandas, “Towards a mathematical theory of super resolution,” *arXiv:1203.5871*,
- [43] M. F. Duarte and R. G. Baraniuk, “Spectral compressive sensing,” *Appl. Comp. Harmon. Anal.*, vol. 35, no. 1, pp. 111–129, 2013.
- [44] M. A. Andrade, C. Perez-Iratxeta, and C. P. Ponting, “Protein repeats: Structures, functions, and evolution,” *J Struct. Biol.*, vol. 134, pp. 117–131, 2001.
- [45] M. J. Neville, E. C. Johnstone, and R. T. Walton, “Identification and characterization of ankk1: A novel kinase gene closely linked to drd2 on chromosome band 11q23.1,” *Human Mutat.*, vol. 23, pp. 540–545, 2004.
- [46] W. H. Liggett and D. Sidransky, “Role of the p16 tumor suppressor gene in cancer,” *J. Clin. Oncol.*, vol. 16, pp. 1197–1206, 1998.
- [47] S. Ni, G. M. Sheldrick, M. M. Benning, and M. A. Kennedy, “The 2 a resolution crystal structure of hetl, a pentapeptide repeat protein involved in regulation of heterocyst differentiation in the cyanobacterium nostoc sp. strain pcc 7120,” *J. of Struct. Biol.*, vol. 165, pp. 47–52, 2009.
- [48] A. Heger and L. Holm, “Rapid automatic detection and alignment of repeats in protein sequences,” *Proteins*, vol. 41, no. 2, pp. 224–237, 2000.
- [49] R. Szklarczyk and J. Heringa, “Tracking repeats using significance and transitivity,” *Bioinformatics.*, vol. 20, pp. 311–317, 2004.
- [50] M. Gruber, J. Soding, and A. Lupas, “Repper - repeats and their periodicities in fibrous proteins,” *Nucleic Acids Res.*, vol. 33, no. 2, pp. 239–243, Jul. 2005.
- [51] E. Sonnhammer, S. Eddy, E. Birney, A. Bateman, and R. Durbin, “Pfam: Multiple sequence alignments and hmm-profiles of protein domains,” *Nucleic Acids Res.*, vol. 26, pp. 320–322, 1998.
- [52] J. Schultz, F. Milpetz, P. Bork, and C. Ponting, “Smart, a simple modular architecture research tool: Identification of signaling domains,” *Proc. Natl. Acad. Sci.*, pp. 5857–64, 1998.
- [53] K. B. Murray, D. Gorse, and J. M. Thornton, “Wavelet transforms for the characterization and detection of repeating motifs,” *J. Mol. Biol.*, vol. 316, pp. 341–363, 2002.
- [54] H. M. Berman et. al, “The protein data bank,” *Nucl. Acids Res.*, vol. 28, pp. 235–242, 2000.
- [55] A. Rose, A. Bradley, Y. Valasatava, J. Duarte, A. Prlić, and P. Rose, “Web-based molecular graphics for large complexes,” *ACM Proc. of the 21st Int. Conf. on Web3D Tech.*, 2016.
- [56] A. Rose and P. Hildebrand, “Ngl viewer: A web application for molecular visualization,” *Nucl. Acids Res.*, vol. 43, 576–579, year =,

- [57] T. Smith and M. Waterman, "Identification of common molecular subsequences," *J. Mo. Biol.*, vol. 147, pp. 195–97, 1981.
- [58] O. Gotoh, "An improved algorithm for matching biological sequences," *J. of Mol. Bio.*, vol. 162, no. 3, pp. 705–708, 1982.
- [59] R. Grantham, "Amino acid difference formula to help explain protein evolution," *Science*, vol. 185, no. 4154, pp. 862–864, 1974.
- [60] B. Lee and F. M. Richards, "The interpretation of protein structures: Estimation of static accessibility," *J. Mol. Biol.*, vol. 55, pp. 379–400, 1971.
- [61] I. Cosic,
- [62] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *J. Mo. Biol.*, vol. 157, pp. 105–132, 1982.
- [63] H. Faber, C. Groom, H. Baker, W. Morgan, A. Smith, and E. Baker, "1.8 Å crystal structure of the c-terminal domain of rabbit serum haemopexin," *Structure*, vol. 3, pp. 551–559, 1995.
- [64] D. Banner, A. Bloomer, G. Petsko, D. Phillips, and I. Wilson, "Atomic coordinates for triose phosphate isomerase from chicken muscle," *Biochem. Biophys. Res. Commun.*, vol. 72, pp. 146–155, 1976.
- [65] J. Peters, M. Stowell, and D. Rees, "A leucine-rich repeat variant with a novel repetitive protein structural motif," *Nat. Struct. Biol.*, vol. 3, pp. 991–4, 1996.
- [66] M. Groves, N. Hanlon, P. Turowski, B. Hemmings, and D. Barford, "The structure of the protein phosphatase 2a p65/a subunit reveals the conformation of its 15 tandemly repeated heat motifs," *Cell*, vol. 96, pp. 99–110, 1999.
- [67] P. Michaely, D. Tomchick, M. Machius, and R. Anderson, "Crystal structure of a 12 ank repeat stack from human ankyrin," *Embo J.*, vol. 21, pp. 6387–6396, 2002.
- [68] H. Miyatake, A. Sanjoh, S. Unzai, G. Matsuda, Y. Tatsumi, Y. Miyamoto, N. Dohmae, and Y. Aida, "Crystal structure of human importin-1 (rch1), revealing a potential autoinhibition mode involving homodimerization," *PLoS One*, vol. 10, 2015.
- [69] G. Benson, "Tandem repeats finder: A program to analyze dna sequences," *Nucleic Acids Research*, vol. 27, no. 2, pp. 573–580, 1990.
- [70] B. U. B, K. Becker, D. Brock, M. Lentze, F. Bidlingmaier, and M. Ludwig, "A novel stable stable polyalanine [poly(a)] expansion in the hoxa13 gene associated with hand-foot-genital syndrome: Proper function of poly(a)-harbouring transcription factors depends on a critical repeat length?" *Hum. Gen.*, vol. 110, no. 5, pp. 488–494, 2002.

- [71] R. Kolpakov, G. Bana, and G. Kucherov, "Mreps: Efficient and flexible detection of tandem repeats in dna," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3672–3678, 2003.
- [72] D. Sharma, B. Issac, G. P. Raghava, and R. Ramaswamy, "Spectral repeat finder (srf): Identification of repetitive sequences using fourier transformation," *Bioinformatics*, vol. 20, no. 9, 2004.
- [73] Emanuele V.A. and Tran T.T. and Zhou G.T., "A fourier product method for detecting approximate tandem repeats in dna," *IEEE/SP 13th Workshop on Stat. Sig. Proc.*, 2005.
- [74] M. Buchner and S. Janjarasjitt, "Detection and visualization of tandem repeats in dna sequences," *IEEE Trans. on Sig. Proc.*, vol. 51, no. 9, pp. 2280–2287, 2003.
- [75] Gupta R. and Sarthi D. and Mittal A. and and Singh K., "Exactly periodic subspace decomposition based approach for identifying tandem repeats in dna sequences," *European Signal Processing Conference*, 2006.
- [76] Grover A. and Veenu A. and Sharma P.C., "Searching microsatellites in dna sequences: Approaches used and tools developed.," *Physiol. mol. biol. plants*, 2012.
- [77] C. M. Ruitberg, D. J. Reeder, and J. M. Butler, "Strbase: A short tandem repeat dna database for the human identity testing community," *Nucleic Acids Research*, vol. 29, no. 1, pp. 1320–1322, 2001.
- [78] A. Edwards, A. Civitello, H. A. Hammond, and C. T. Caskey, "Dna typing and genetic mapping with trimeric and tetrameric tandem repeats," *American J. of Hum. Gen.*, vol. 49, pp. 746–756, 1991.
- [79] E. Foundation, "What is a seizure," <http://www.epilepsy.com/learn/epilepsy-101/what-seizure>,
- [80] S. Scott, "Absence seizures," <http://reference.medscape.com/article/1183858-overview#aw2aab6b9>,
- [81] Gibbs FA and Davis H and Lennox WG, "The eeg in epilepsy and in conditions of impaired consciousness," *Arch Neurol Psychiat*, vol. 34, pp. 1134–1148, 1935.
- [82] Y. V. Rozas, "Eeg video monitoring," *Medscape [website]* <http://emedicine.medscape.com/article/1137908-overview>,
- [83] Hese P. V. and Martens J. P. and Boon P. and Dedeurwaerdere S. and Lemahieu I. and R. V. de Walle, "Detection of spike and wave discharges in the cortical eeg of genetic absence epilepsy rats from strasbourg," *Physics in Medicine and Biology*, vol. 48, pp. 1685–1700, 2003.

- [84] Adeli H and Zhou Z and Dadmehr N., “Analysis of eeg records in an epileptic patient using wavelet transform,” *Journal of Neuroscience Methods, Elsevier*, vol. 123, pp. 69–87, 2003.
- [85] Xanthopoulos P. and Rebennack S. and Chang-Chia Liu and Jicong Zhang and Holmes G.L. and Uthman B.M. and Pardalos P.M., “A novel wavelet based algorithm for spike and wave detection in absence epilepsy,” *Int. Conf. on BioInformatics and BioEngineering (BIBE)*, 2010.
- [86] Fanselow E. E. and Reid A. P. and Nicolelis M. A. L., “Reduction of pentylenetetrazole-induced seizure activity in awake rats by seizure-triggered trigeminal nerve stimulation,” *The Journal of Neuroscience*, vol. 20, pp. 8160–8168, 2000.
- [87] Sakkalis V. and Giannakakis G. and Farmaki C. and Mousas A. and Pediaditis M. and Vorgia P. and Tsiknakis M., “Absence seizure epilepsy detection using linear and nonlinear eeg analysis methods,” *Proc. 35th Annual International Conference of the IEEE EMBS*, pp. 6333–6336, 2013.
- [88] X. Zhang, L. Xu, L. Xu, and D. Xu, “Direction of departure (dod) and direction of arrival (doa) estimation in mimo radar with reduced-dimension music,” *IEEE Comm. Letters*, vol. 14, no. 12, pp. 1161–1163, 2010.
- [89] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.
- [90] P. Stoica and A. Nehorai, “Music, maximum likelihood, and cramer-rao bound,” *IEEE Trans. on Acoust., Speech, and Sig. Proc.*, vol. 37, no. 5, pp. 720–741, 1989.
- [91] M. Oziewicz, “On application of music algorithm to time delay estimation in ofdm channels,” *IEEE Trans. on Broadcasting*, vol. 51, pp. 249–255, 2005.
- [92] M. Hamalainen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, “Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain,” *Rev. Mod. Phy.*, vol. 65, no. 2, pp. 413–497, 1993.
- [93] J. C. Mosher, P. S. Lewis, and R. M. Leahy, “Multiple dipole modeling and localization from spatio-temporal meg data,” *IEEE Trans. on Biomed. Eng.*, vol. 39, no. 6, pp. 541–557, 1992.
- [94] M. G. Christensen and A. Jakobsson, “Multi-pitch estimation,” *Synthesis Lectures on Speech and Audio Processing*, vol. 5, pp. 1–160, 2009.
- [95] A. de Cheveigna and H. Kawahara, “Yin, a fundamental frequency estimator for speech and music,” *J. Acoust. Soc. of America*, vol. 111, pp. 1917–1930, 2002.
- [96] A. M. Noll, “Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum and a maximum likelihood estimate,” *MRI Symp. Proc., NY: Polytechnic Press*, vol. 19,

- [97] R. Gribonval and E. Bacry, “Harmonic decomposition of audio signals with matching pursuit,” *IEEE Trans. on Sig. Proc.*, vol. 51, pp. 101–111, 2003.
- [98] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, “Multi-pitch estimation,” *Elsevier Signal Processing*, vol. 88, pp. 972–983, 2008.
- [99] C. W. Therrien, *Discrete random signals and statistical signal processing*. Prentice Hall, Upper Saddle River, NJ, 1992.
- [100] S. M. Kay, *Modern spectral estimation: theory and application*. Prentice Hall, Englewood Cliffs, NJ, 1988.
- [101] G. H. Hardy and S. Ramanujan, “Asymptotic formulae in combinatory analysis,” *Proc. London Math. Soc.*, vol. 17, pp. 75–115, 1918.
- [102] M. A. Andrade, C. Perez-Iratxeta, and C. P. Ponting, “Protein repeats: Structures, functions, and evolution,” *J. Structur. Biol.*, vol. 134, pp. 117–131, 2011.
- [103] S. V. Tenneti and P. P. Vaidyanathan, “Ramanujan filter banks for estimation and tracking of periodicities,” in *Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Proc.*, 2015.
- [104] P. P. Vaidyanathan and S. V. Tenneti, “Properties of ramanujan filter banks,” *European Sig. Proc. Conf.*, 2015.
- [105] H. Duifhuis, L. F. Willems, and R. J. Sluyter, “Measurement of pitch in speech: An implementation of goldstein’s theory of pitch perception,” *J. Acoust. Soc. Amer.*, vol. 71, no. 6, pp. 1568–1580, 1982.
- [106] B. Kobe and J. Deisenhofer, “A structural basis of the interactions between leucine-rich repeats and protein ligands,” *Nature*, vol. 374, pp. 183–186, 1995.
- [107] R. Roy and T. Kailath, “Esprit-estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. on Acoust., Speech, and Sig. Proc.*, vol. 37, no. 7, pp. 984–995, 1989.
- [108] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23,
- [109] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [110] R. Goldberg and L. Riek, *A Practical Handbook of Speech Coders*. CRC Press, 2000.
- [111] C. Caratheodory and L. Fejer, “Uber den zusammenhang der extemen von harmonischen funktionen mit ihren koeffizienten und uber den picard-landauschen satz,” *Rendiconti del Circolo Matematico di Palermo*, vol. 32, pp. 218–239, 1911.

- [112] N. D. Sideropoulos, “Generalizing caratheodory’s uniqueness of harmonic parameterization to n dimensions,” *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1687–1690, 2001.
- [113] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, USA, 1984.
- [114] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Pearson, Englewood Cliffs, NJ, USA, 1993.
- [115] M. C. Escher, *Escher on Escher: Exploring the Infinite*. Harry N. Abrams, New York, NY, USA, 1989.
- [116] C. Hammond, *The Basics of Crystallography and Diffraction*. Oxford Univ. Press, New York, NY, USA, 2009.
- [117] B. Rupp, *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. Garland Science, New York, NY, USA, 2010.
- [118] C. C. MacDuffee, *The Theory of Matrices*. Chelsea, New York, NY, USA, 1946.
- [119] F. R. Gantmacher, *The Theory of Matrices*. Chelsea, New York, NY, USA, 1977.
- [120] T. Chen and P. P. Vaidyanathan, “The role of integer matrices in multidimensional multirate systems,” *IEEE Trans. Signal Process.*, vol. 41, pp. 1035–1047, 1993.
- [121] H. J. S. Smith, “On systems of linear indeterminate equations and congruences,” *Philos. Trans. Royal Soc. London*, vol. 151, pp. 293–326, 1861.
- [122] P. P. Vaidyanathan, “Multidimensional ramanujan-sum expansions on non-separable lattices,” *Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Proc., Brisbane, Australia*, 2015.