Matrix Estimation using Shrinkage of Singular Values with Applications to Signal Denoising

A

thesis submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

by

Santosh Kumar Yadav

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039, INDIA

JAN 2018

To **Late Mr. Somnath Pal**

for his heavenly support and encouragement

Certificate

This is to certify that the thesis entitled as "**Matrix Estimation using Shrinkage of Singular Values with Applications to Signal Denoising**", submitted by **Santosh Kumar Yadav** (Roll. 11610213), a research scholar in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by him under our supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the Institute and in our opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. Rohit Sinha	Dr. Prabin Kumar Bora
Professor	Professor

Department of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati
Guwahati - 781 039, India.

Dated:
Place: IIT Guwahati

Acknowledgements

At the outset, I would like to express my whole hearted and deep sense of gratitude to my thesis supervisors Prof. Rohit Sinha and Prof. Prabin Kumar Bora for their guidance, help, and encouragement throughout my research work. I greatly admire their critical analysis, creative thinking, hard work, and dedication in my research work. These are the great source of inspiration for me in all my endeavours. I am highly grateful to them for patiently sitting with me while checking all my manuscripts and thesis. Writing this thesis would not have been possible without their bounteous effort. More than guides, they are my mentors for shaping my personal and professional life, without whom I would not have been where I am today. As friends and guides their help are immeasurable. I owe my profound gratitude to Prof. Sinha and Prof. Bora for their supports in all respects.

I am sincerely thankful to my doctoral committee members Dr. Manas Kamal Bhuyan, Dr. Amit Sethi, and Dr. Arijit Sur for their support, encouragement, and suggestions during my research work. Many faculty members of the department including Prof. Samarendra Dandapat, Prof. S. R. Mahadeva Prasanna, Prof. Chitralekha Mahanta, and Dr. Suresh Sundaram have directly or indirectly helped me to grow as an independent individual. Their kind support and valuable suggestions helped me a lot to improve myself academically and personally. I would like to thank Dr. Amit Sethi, Dr. Brijesh Kumar Rai, and Dr. Tony Jacob for their friendly discussions, suggestions, and encouragement that helped me continue this journey.

I would like to acknowledge the support shown by Mr. Sanjib Das, Mr. P. J. Goswami, Mr. Mukut Barua, and Mr. Uday Shankar Uzir. Their indirect help during my tenure at the department are immeasurable. I would like to thank my colleagues Dr. Saroj Mondal, Dr. Sikandar Kumar, Dr. Avishek Vahadane, Dr. Anurag Singh, Dr. Ankit Dalal, Dr. Brajesh Rawat, Mr. Gaurav Yadav, Mr. Karam Singh Mallik, Mr. Kukil Khanikar, Dr. Sunil Kumar, Mr. Biplab K. C., Dr. Dharmdev, and others without whom my stay at the institute would have been completely monotonous. I would like to sincerely acknowledge the friendly supports by Saroj, Sikandar, Ankit, and Gaurav during my odd days. My sincere gratitude goes to all the research scholars of Image Processing and Computer Vision Laboratory who worked with me.

My greatest debt of gratitude is owed to my parents, whose hard works and sacrifices gave me the opportunities I have today. I am extremely grateful to have cousins like Ritesh, Abhishek, Ashutosh, Jay Shree, Deepika, and Preeti who have always been by my side. I sincerely acknowledge the help and support provided by them on many occasions.

Finally, I feel extremely grateful to acknowledge the heavenly support and encouragement by Late Mr. Somnath Pal, my favourite teacher at University Institute of Technology, Burdwan University, who motivated me to take this career path and made me to dream beyond my capability before his sudden demise. My happiness would have been many fold if he were with me today. This dissertation is dedicated to him for being a great and eternal admirer for me forever.

S. K. Yadav

Abstract

This thesis documents investigations on developing the better spectral shrinkage functions for matrix estimation. It proposes two effective spectral shrinkage estimators. The first estimator exploits the correlation present in the data matrix by decoupling the shrinkage and the truncation of singular values. It employs a logistic function based shrinkage of singular values and shows better rank estimation than the existing methods. The parameters of this estimator are tuned by grid-search solution of a SURE formulation. However, the method is computationally expensive and the applicability of this estimator is limited. The second estimator addresses this problem. It expresses the thresholding function as a linear combination of the derivatives of a Gaussian so that the corresponding SURE admits a closed form solution for the optimal parameters of the shrinkage function. Hence, it avoids the huge computational cost involved in the first method in parameter search. In addition, this linearization of thresholding function allows to increase the number of parameters arbitrarily, providing a sufficient flexibility in the shape of the shrinkage function. The estimator performs better matrix estimation than the state-of-the art methods.

The thesis further investigates the effectiveness of the proposed estimators in signal denoising applications in a collaborative filtering framework. In this framework, the noisy signals under investigation are rearranged in the form of several data matrices. The denoising of electrocardiogram (ECG) signal is presented as the first application. For this purpose, the data matrices are formed by grouping similar blocks from the noisy ECG signal. These data matrices are then denoised by the proposed matrix estimation methods. The denoised signal is obtained by inverse operation of the matrix formation. The quality of the denoised signal is measured using different performance metrics and it is found to be better than that of the existing methods. Following that the denoising of two-dimensional photographic images and three-dimensional magnetic resonance images are also explored. On applying the proposed matrix estimation methods in collaborative filtering framework the competitive denoising performances are achieved in both the applications.

Keywords: Matrix denoising, collaborative filtering, ECG denoing, MRI denoising, image denoising, spectral shrinkage estimator, etc.

Contents

List of Figures xiv

List of Tables xix

List of Acronyms xxi

List of Symbols xxiv

1 Introduction 1

1.1 Matrix estimation using SVD . . . 3

1.1.1 Data model . . . 4

1.1.2 Spectral regularization . . . 5

1.1.3 Spectral estimator of a matrix . . . 6

1.2 Random matrix theory . . . 7

1.2.1 Empirical singular value distribution . . . 8

1.2.2 An asymptotically optimal shrinkage estimator . . . 11

1.3 Stein's unbiased risk estimator of the MSE . . . 12

1.4 Optimal shrinkage using the RMT and the SURE . . . 13

1.5 Progress and challenges . . . 14

1.5.1 Problem formulation and objectives . . . 15

1.6 Contributions of the thesis . . . 16

1.7 Organization of the thesis . . . 17

2 SVLT: Decoupling Shrinkage and Truncation for Matrix Rank Estimation 19

2.1 Matrix rank estimation . . . 21

2.1.1 SURE based rank estimation methods . . . 22

2.1.2 Drawbacks of the existing rank estimators . . . 24

2.2 Proposed method . . . 24

2.3 Experimental results and discussion . . . 25
    2.3.1 SURE for SVLT vs. MSE . . . 26
    2.3.2 Rank estimated by SVLT vs. true rank . . . 26
    2.3.3 Quantitative performance of the SVLT based rank estimator . . . 28
    2.3.4 Quantitative performance of the SVLT based matrix estimator . . . 28
2.4 Summary . . . 30

3 SVLET: A Fast and Adaptive Shrinkage Method for Matrix Estimation 33

3.1 Shrinkage of SVs for matrix estimation . . . 34
3.2 Proposed method . . . 36
    3.2.1 Linearized shrinkage function . . . 36
    3.2.2 Asymptotic behavior of the SVLET . . . 38
3.3 Results and discussion . . . 41
    3.3.1 Selection of parameters . . . 41
    3.3.2 Matrix denoising performance . . . 44
        3.3.2.1 Comparison with RMT based methods . . . 44
        3.3.2.2 Comparison with SURE based methods . . . 45
3.4 Summary . . . 45

4 Application to ECG Denoising 47

4.1 Non-local filtering for ECG denoising . . . 50
4.2 Proposed ECG denoising method . . . 51
    4.2.1 SDM extraction . . . 52
    4.2.2 Noise filtering . . . 54
        4.2.2.1 Denoising by thresholding of 2D DWT coefficients . . . 54
        4.2.2.2 Denoising by singular values shrinkage estimators . . . 55
    4.2.3 Aggregation . . . 56
4.3 Selection of parameters . . . 56
    4.3.1 Block size and overlap . . . 57
    4.3.2 Search window . . . 58
    4.3.3 Block matching threshold and SDM size . . . 58

4.3.3.1 Hard-threshold for NLWT 58

4.3.4 Order of linearization and transition parameter of the SVLET 59

4.4 Experimental results 59

4.4.1 Performance measures 60

4.4.2 Evaluation on MIT-BIH Physionet database 60

4.4.3 Evaluation on PTB Diagnostic database 64

4.5 Discussion 65

4.6 Summary 69

5 Application to Image Denoising **71**

5.1 Image denoising by collaborative filtering 73

5.1.1 Block matching based 3D filtering (BM3D) 75

5.1.2 Higher order singular value decomposition based CF 76

5.1.3 Weighted nuclear norm minimization (WNNM) 77

5.1.4 CF with low-rank approximation 78

5.1.5 Low-rank collaborative filtering (LRCF) 79

5.2 Proposed methods 79

5.2.1 Denoising with FRIT in CF 79

5.2.2 Denoising by matrix estimation in CF 81

5.3 Experimental set-up and parameter tuning 82

5.3.1 Parameters of the proposed CF methods 83

5.3.1.1 Parameter tuning for the FRIT-CF method 84

5.3.1.2 Parameter tuning for the SVLET-CF method 87

5.4 Results and discussion 90

5.5 Summary 92

6 Application to MR Image Denoising **93**

6.1 Denoising of MR images 94

6.2 Noise in MR image 97

6.2.1 Variance stabilization 98

6.3 Proposed MR image denoising method 99

6.3.1 Block grouping . . . 100

6.3.2 Noise filtering . . . 100

6.3.2.1 Noise filtering by SVLET . . . 101

6.3.2.2 Noise filtering by OptShrink . . . 101

6.3.3 Aggregation . . . 102

6.4 Experimental set-up and parameter tuning . . . 103

6.4.1 Performance measures . . . 103

6.4.2 Selection of parameters . . . 104

6.4.3 Effect of iterations . . . 113

6.5 Results and discussion . . . 113

6.6 Summary . . . 115

7 Conclusions and Future Directions 117

7.1 Conclusions . . . 118

7.2 Future directions . . . 119

Bibliography 121

List of Publications 131

List of Figures

1.1 Comparison of different shrinkage functions 9

1.2 Comparison of SURE and MSE for the SVST for a range of μ 9

2.1 Given a 20-rank data matrix $\mathbf{X}$ of size 100×100 and its WGN added noisy version $\widetilde{\mathbf{X}}$. (a) Demonstration of the SVST (pink) and the SVHT (black) of the SVs of the noisy data (red) with those of the original data (blue). (b) Plots of the logistic function $\psi(a, b)$ embedded in the proposed SVLT for varying parameters. (c) Demonstration of the improved singular value shrinkage (black) with the application of the proposed SVLT. 21

2.2 Comparison of SURE (tick marks) and Monte Carlo (solid lines) simulations for four different matrices under varying SNRs for tuning the parameter b. 27

2.3 Comparison of the performance of the SVLT with that of the EYM, the US, and the SVST. The diagonal green line corresponds to the oracle estimator of the rank. 27

2.4 Comparison of the performance of the SVLT with that of the EYM, the US, and the SVST methods in terms of the inaccuracy for different SNR (in dB) levels and ranks. 29

2.5 Comparison of the performance of the SVLT with that of the EYM, the US, and the SVST methods in terms of the MSE for different SNR (in dB) levels and ranks. 29

3.1 Comparison of existing singular value shrinkage functions. A typical example of (a) asymptotically optimal bulk-shrinkage function [1], (b) ATN [2] ($\mu_1 = 2.2,\ \mu_2 = 5$), (c) SVLT [3] ($a = 0.8,\ b = 2.2,\ c = 0.1$), and (d) proposed SVLET shrinkage functions ($\mathbf{a} = [1,\ -1]^{\mathrm{T}},\ K = 2,\ T = 3\tau$) plotted with SVHT ($\mu = 2.18$) and SVST ($\mu = 1.98$). The values in the braces are the parameters of corresponding shrinkage function. 42

3.2 Effect of the parameters of the proposed SVLET shrinkage estimator on the denoising performance. Effect on the NMSE of varying (a) C and (b) K. In each case, the averaged NMSE computed over all possible ranks of the underlying clean matrix is reported. 42

3.3 Comparison of average denoising performance of the proposed SVLET estimator with the OptShrink and other estimators for different ranks of 50×50 matrix corrupted by AWGN at (a) SNR $= 0.5$, (b) SNR $= 1.0$, (c) SNR $= 2.0$, and (d) SNR $= 4.0$. 43

3.4 Comparison of average denoising performance of the proposed SVLET estimator with SURE based contrast estimators for different ranks of 50×50 matrix corrupted by AWGN at (a) SNR $= 0.5$, (b) SNR $= 1.0$, (c) SNR $= 1.5$, and (d) SNR $= 2.0$. 43

4.1 Block diagram of the proposed NLWT method. The blocks of samples that are marked in same colors are similar and grouped in an SDM. 53

4.2 Illustration of an SDM extracted from (a) a noisy ECG signal, and its denoised versions (b) estimated by the NLWT, and (c) estimated by the SVLET. 53

4.3 Illustration of parameter tuning in the NLWT and in the SVLET. The plots showing the variation of normalized MSE (NMSE) at different noise levels with respect to parameters (a) δ_{2D}, (b) C, and (c) K. 62

4.4 Comparison of the denoising performances of the proposed methods along with those of a few existing methods at 20dB noise level in terms of (a) SNR improvement, (b) mean squares error, and (c) PRD measures. In each case, the performance averaged over all the 12-leads of the ECG signal chosen from MIT-BIH database, is reported. 62

4.5 The impact of noise level on the denoising performances of the proposed methods along with those of a few existing methods measured separately in terms of (a) MSE, (b) PRD, and (c) SNR improvement. 63

4.6 Illustration of a typical ECG signal showing its different features. Courtesy: www.wikidata.org 63

4.7 Qualitative performance comparison of the proposed methods with the contrast method on the noisy ECG corresponding to the noise-free ECG signal numbered as 215 from MIT-BIH Arrythmia database. 65

4.8 Comparison of the denoising performance of the proposed methods and the existing method averaged over all the channels of the chosen ECG signals at 20dB noise level. 66

4.9 Illustration of preservation of the depressed ST-segments while denoising the lead-I of the signal number s0354lrem derived from the PTB diagnostic database. 67

5.1 Block diagram of the two-step collaborative filtering in the BM3D. In the figure, the patch marked as "R0" represents a reference patch and the color of the boxes indicate the non-locality of the photometrically similar patches. 74

5.2 Denoising performances for global processing (with threshold value, $\sigma\sqrt{2\,log\,N}$). (a) Original synthetic image ($N = 127 \times 127$), (b) Noisy image, $\sigma = 80$, (c) 2D DWT (PSNR: 19.89 dB), (d) FRIT (PSNR: 21.19 dB), (e) FRIT + Wiener (21.37 dB). 81

5.3 Block diagram of the proposed SVLET based image denoising approach. 82

5.4 Performance comparison of salient existing and the proposed denoising methods in global and block-based processing cases at a noise level of $\tau = 25$. 84

5.5 Performance comparison of salient existing and the proposed denoising methods in global and block-based processing cases at a noise level of $\tau = 25$. 85

5.6 Tuning of hard-threshold parameter coefficients λ_{3D} of the FRIT-CF on two different images. 86

5.7 Tuning of the parameters of the SVLET-CF: (a) the transition parameter T, and (b) the order of linerization K. 86

5.8 Effect of iteration on denoising performance, in terms of PSNR (in dB), of the SVLET-CF in case of different noise levels. 86

5.9 Comparison of performance on a selected portion of (a) the Straw image, (e) the Grass image and (i) the House image. 89

5.10 Comparison of local denoising performance on chosen image segments at noise level $\tau = 35$. Monarch image segments denoised by (a) BM3D: 22.95 dB, (b) FRIT-CF: 23.28 dB, (c) BM3D: 29.33 dB, and (d) FRIT-CF: 28.67 dB. Barbara image segments denoised by (e) BM3D: 24.57 dB, (f) FRIT-CF: 24.97 dB, (g) BM3D: 29.61 dB, and (h) FRIT-CF: 29.05 dB. 89

5.11 Comparison of qualitative and quantitative denoising performance on different images at noise level $\tau = 30$. 90

6.1 Block diagram of collaborative filtering for MR image denoising. The iterative CF block in the above is same as the one in Figure 5.3. 100

6.2 Tuning of parameter T and K of the SVLET-CF on an MR image volume cropped from the 'normal' SBD at different noise levels. 105

6.3 Effect of the iteration on the PSNR of denoised MR image corresponding to a volume cropped for the 'normal' case from BrainWeb corpus at different noise levels. 105

6.4 Qualitative performance comparison on 'normal' brain MR image extracted from Simulated BrainWeb Database, in case of Gaussian noise at 15% noise levels. . . 107

6.5 Qualitative performance comparison on 'MS Lesion' brain MR image extracted from Simulated BrainWeb Database, in case of Gaussian noise at 15% noise levels. 109

6.6 Qualitative performance comparison on 'normal' brain MR image extracted from Simulated BrainWeb Database, in case of Rician noise at 15% noise levels. . . . 111

6.7 Qualitative performance comparison on 'MS Lesion' brain MR IMAGE data extracted from Simulated BrainWeb Database, in case of Rician noise at 15% noise levels. 114

List of Tables

1.1 List of various estimators proposed/explored for different denoising applications. Abbreviations used in the table are defined in the list of acronyms. 16

2.1 Comparison of the computational complexity/time of different methods for estimating the rank of a 200 × 400 matrix. The notation double-bar over the symbols denotes the cardinality of the parameter tuning-set. 30

3.1 Comparison of the execution time of different methods for estimating a 50 × 50 matrix. 45

4.1 Different parameters used in the proposed NLWT algorithm, their tuning ranges, and the optimal values. 57

4.2 Pathologies known for the signals used in experiment reported in Figure 4.8. . . . 67

4.3 Comparison of the execution time of different denoising methods for estimating an ECG signal of length $N = 10,000$ samples. For all methods employing collaborative filtering, the SDM extraction time of 30 seconds has been shown separately. 68

5.1 Parameter set used in the first step of BM3D and the proposed CF with FRIT . 87

5.2 Comparison of denoising performance of the different methods in terms of PSNR (dB). The numbers in the boldface are the highest in the columns at given noise levels and indicate the best performance by the method in the given rows. . . . 88

6.1 Comparison of performances on 'normal' brain MR image extracted from BrainWeb corpus, in case of Gaussian noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB. 106

6.2 Comparison of performances on 'MS Lesion' brain MR image extracted from BrainWeb corpus, in case of Gaussian noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB. 108

6.3 Comparison of performances on 'normal' brain MR image extracted from BrainWeb corpus, in case of Rician noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB. 110

6.4 Comparison of performances on 'MS Lesion' brain MR image extracted from BrainWeb corpus, in case of Rician noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB. 112

List of Acronyms

1D/2D/3D/4D	One/Two/Three/Four Dimensional
AIC	Akaike Information Criterion
AMSE	Asymptotic Mean Squared Error
ATN	Adaptive Trace Norm
AWGN	Additive White Gaussian Noise
BIC	Bayesian Information Criterion
BM3D	Block Matching and 3D Filtering
BM4D	Block Matching and 4D Filtering
CF	Collaborative Filtering
CV	Cross Validation
coif	Coiflet (wavelet)
DOG	Derivative of Gaussian
DWT	Discrete Wavelet Transform
dB	Decibel
db2	Daubechies 2 (wavelet)
ECG	Electrocardiogram
EM	Efron and Morris (shrinkage estimator)
EMD	Empirical Mode Decomposition
EYM	Eckart-Young-Mirsky
EWF	Empirical Wiener Filter
FIR-LPF	Finite Impulse Response-Low Pass Filter
FRIT	Finite Ridgelet Transform
FRIT-CF	FRIT Collaborative Filtering

GEM	Generalized Efron and Morris (shrinkage estimator)
GHz	Giga Hertz
HOSVD	Higher Order Singular Value Decomposition
HT	Hard Thresholding
IDWT	Inverse Discrete Wavelet Transform transform
IID	Independent and Identically Distributed
IMF	Intrinsic Mode Function
JS	James and Stein (shrinkage estimator)
KLT	Karhunen-Loeve Transform
LLMSE	Locally Linear Mean Squared Error
LRMA	Low Rank Matrix Approximation
LRASVD	Low Rank Approximation with Singular Value Decomposition
LRCF	Low Rank Collaborative Filtering
MAD	Median Absolute Deviation
MDL	Minimul Description Length
MI	Myocardial Infarction
MIT-BIH	Massachusetts Institute of Technology-Beth Israel Hospital
MLE	Maximum Likelihood Estimator
MS	Multiple Sclerosis
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NLM	Non-Local Means
NLWT	Non-Local Wavelet Transform
NNM	Nuclear Norm Minimization
NMSE	Normalized Mean Squared Error
OptShrink	Optimal Shrinkage
PC	Principal Component
PCs	principal Components

PRD	Percentage Root Mean Squared Difference
PTB	Physikalisch-Technische Bundesanstalt
PSNR	Peak Signal-to-Noise Ratio
Ref.	Reference
RF	Radio Frequency
RMT	Random Matrix Theory
SDM	Similarity Data Matrix
SDMs	Similarity Data Matrices
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
SURE	Steins' Unbiased Risk Estimation
SURELET	Steins' Unbiased Risk Estimation based Linearly Expanded Thresholding
SV	Singular Value
SVs	Singular Values
SVD	Singular Value Decomposition
SVHT	Singular Value Hard-Thresholding
SVST	Singular Value Soft-Thresholding
SVSE	Singular Value Shrinkage Estimator
SVLT	Singular Value Logistic Thresholding
SVLT-CF	SVLT Collaborative Filtering
SVLET	Singular Value Linearly Expanded Thresholding
SVLET-CF	SVLET Collaborative Filtering
sym	symlet (wavelet)
US	Ulfarsson and Solo
VST	Variance Stabilization Trasnform
WGN	White Gaussian Noise
WNNM	Weighted Nuclear Norm Minimization

List of Symbols

Boldface Lower-case Letters	Vectors
Boldface Upper-case Letters	Matrices
$\mathbf{X}$	A matrix with real elements
x_{ij}	An element at ith row and jth column of the matrix
$\mathbf{x}_i$	A column vector extracted from ith column of the matrix
$\mathbf{X}^{\mathrm{T}}$	Transpose of the matrix
$\mathbf{X}^{-1}$	Inverse of the matrix
$\mathrm{diag}(\mathbf{x}_i)$	A diagonal matrix with the elements from $\mathbf{x}_i$ at its principal diagonal
$\mathbf{I}_M$	Identity matrix of dimension $M \times M$
$\forall$	For all
$\in$	Set membership
$\mathrm{E}[\cdot]$	Expectation
$\lvert \cdot \rvert$	Modulus
$\lVert\cdot\rVert_{\mathrm{Frob}}$	Frobenius norm
$\lVert\cdot\rVert_{\mathrm{Nuc}}$	Nuclear norm
$\sum\limits_{i=1}^{n}$	Summation
$exp(\cdot)$	Exponential
$\mathsf{N}\left(m, \tau^2\right)$	Gaussian with mean m and variance τ^2
$\mathsf{R}\left(m, \tau\right)$	Rician with parameters m and τ
$\log_{10}(\cdot)$	Logarithm with base 10
$(\alpha)_+$	Maximum of $\{0, \alpha\}$
$\max\{\cdot, ..., \cdot\}$	Maximum

$\min\{\cdot, ..., \cdot\}$	Minimum
$\lim$	Limits
$\rightarrow$	Tends to
$\xrightarrow{a.s.}$	Converges almost surely
$\overset{a.s.}{=}$	Equals almost surely
$\mathbf{1}_{(\cdot)}(\cdot)$	Indicator function

1

Introduction

Contents

1.1 Matrix estimation using SVD 3

1.2 Random matrix theory 7

1.3 Stein's unbiased risk estimator of the MSE 12

1.4 Optimal shrinkage using the RMT and the SURE 13

1.5 Progress and challenges 14

1.6 Contributions of the thesis 16

1.7 Organization of the thesis 17

Signal processing is a branch of electrical engineering that deals with extracting information form the signals acquired from natural and man-made sources. In general, it consists of three stages: *acquisition*, *processing*, and *inferencing*. In the process of acquisition, the signals get corrupted with different types of sensor and channel noises. As a result, the inferencing becomes more challenging. One of the objectives in the processing stage is to reduce the effect of noise. It can be achieved either by removing the noise from the observation or by making the information extraction robust to the noise. In this thesis, our interest lies in the former and we propose a few methods for noise removal. Further, the signals used for different applications have different nature and are affected by different types of noise. Thus, suitable algorithms are required to be developed for each of the applications.

The problem of removing noise from the signals, also known as *signal denoising*, has been a classical research problem in signal processing domain. It finds applications in numerous fields such as multivariate statistical signal processing [4–8], machine learning [9,10], wireless communication [11,12], etc. A signal can be treated as a time series in which each sample represents a measured value of the underlying physical quantity, possibly at a given time instant. However, two/three dimensional signals such as digital images/videos can be represented as arrays of sample values at spatial/spatio-temporal locations. The noise removal from such signals is often recast as the matrix estimation problem by transforming the observed multidimensional signals into the form of data matrices.

To estimate the matrix corresponding to the original signal, a number of methods have been proposed in the literature [1,3,13–24]. In almost all of these methods, the observed data are first decorrelated using spectral factorization techniques, the spectral components corresponding to the signal and the noise are identified, and then the noise dominant components are removed to obtain an estimate of the underlying signal matrix. Quite often, the singular value decomposition (SVD) is employed for spectral factorization. In the following, we review some of the methods for matrix estimation that uses the SVD, their drawbacks, the recent developments to mitigate these drawbacks, the motivation of our contribution in this line of works, and finally the organization of the thesis.

1.1 Matrix estimation using SVD

The SVD, being one of the most popular matrix factorization methods, is used in numerous applications. It is basically used for the spectral decomposition of an $m \times n$ signal matrix $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$ where $\mathbf{x}_i \in \mathbb{R}^m$. For this purpose, the eigenvalue decomposition of either row- or column- sample covariance matrices, denoted respectively by $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, is performed. Specifically,

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}} \quad \text{and} \quad \mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}, \tag{1.1}$$

where $\Lambda = \mathrm{diag}([\lambda_1, ..., \lambda_n])$ is a diagonal matrix with the eigenvalues on its principal diagonal and the orthogonal matrices $\mathbf{V}$ and $\mathbf{U}$, with their columns, as the eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, respectively. The set $\{\lambda_1, \lambda_2, ..., \lambda_n\}$ is the spectrum of the data matrix $\mathbf{X}$. It can be easily shown that

$$\mathbf{U}^{\mathrm{T}}\mathbf{X}\mathbf{V} = \mathbf{\Sigma} \qquad \Rightarrow \qquad \mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} = \sum_{i=1}^{L} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}, \tag{1.2}$$

where $L = \min(n, m)$ and $\mathbf{\Sigma} = \mathrm{diag}([\sigma_1, ..., \sigma_n])$ with the diagonal elements σ_is being referred to as the singular values (SVs) whereas the columns of matrix $\mathbf{U}$ ($\mathbf{V}$) are called left (right) singular vectors of $\mathbf{X}$. A simple algebraic adjustment shows that $\lambda_i = \sigma_i^2$.

The rank of a matrix is defined as the number of non-zero SV. In case of a low-rank matrix, its spectrum is given as $\mathbf{\Sigma} = \mathrm{diag}([\sigma_1, ..., \sigma_r, 0, ..., 0])$ where r is the rank. The SVD efficiently exploits the correlation present in the signal vectors and packs most of the signal energy in a few spectral components enabling data compression [4, 5]. Generally, the SVs are ordered in non-increasing order such that $\sigma_1 \geq \sigma_1 \geq ... \geq \sigma_r \geq 0$. Thus, by considering only a few SVs along with corresponding singular vectors, a good estimate of the signal matrix can be obtained as

$$\widehat{\mathbf{X}}_{\hat{r}} = \sum_{i=1}^{\hat{r}} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}, \tag{1.3}$$

where $\hat{r} << L$ is the rank of the estimate [25]. In many cases, the elements of the signal matrix $\mathbf{X}$ may not be exactly known or noise may be present in the measurements. For example, the occluded pixels in an image can be thought of as unknown entries in the image matrix. Similarly, the noisy image or video can be considered as the noisy measurement of an underlying clean image matrix or tensor. The study of effectiveness of the SVD in such scenarios is a popular research topic in signal denoising.

1.1.1 Data model

Let the data model for $m \times n$ noisy observation matrix $\widetilde{\mathbf{X}}$ be

$$\widetilde{\mathbf{X}} = \mathbf{X} + \overline{\mathbf{X}}, \tag{1.4}$$

where elements of $\overline{\mathbf{X}}$ are independent and identically distributed (IID) Gaussian random variables. Thus, $\bar{x}_{j,i} \overset{\mathrm{IID}}{\sim} \mathsf{N}(0, \tau^2)$ where τ is the standard deviation of noise. Without the loss of generality, the column (row) $\bar{\mathbf{x}}_{i(j)}$ can be thought of as a multivariate Gaussian random vector, i.e., $\bar{\mathbf{x}}_{i(j)} \overset{\mathrm{IID}}{\sim} \mathsf{N}(\mathbf{0}, \tau^2 \mathbf{I}_{m(n)})$. We can consider it as the source of nuisance in the columns (rows) of the data matrix $\widetilde{\mathbf{X}}$. To address this, usually the correlation between columns (rows) of underlying signal matrix $\mathbf{X}$ is exploited by employing the SVD of the noisy data matrix $\widetilde{\mathbf{X}}$. Hence, it is interesting to see the effect of noise on the SVD of $\mathbf{X}$ with the hope that it helps in recovering $\mathbf{X}$ from given $\widetilde{\mathbf{X}}$. Let us perform the SVD of $\widetilde{\mathbf{X}}$ and $\overline{\mathbf{X}}$, respectively, as

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{U}}\ \widetilde{\boldsymbol{\Sigma}}\ \widetilde{\mathbf{V}}^{\mathrm{T}} = \sum_{i=1}^{L} \tilde{\sigma}_i \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}} \quad \text{and} \quad \overline{\mathbf{X}} = \overline{\mathbf{U}}\ \overline{\boldsymbol{\Sigma}}\ \overline{\mathbf{V}}^{\mathrm{T}} = \sum_{i=1}^{L} \bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^{\mathrm{T}} \tag{1.5}$$

where $\{\tilde{\sigma}_i\}_{i=1}^{L}$ is referred to as *data SVs* (in contrast to the *signal SVs* $\{\sigma_i\}_{i=1}^{L}$) with $\{\tilde{\mathbf{u}}_i(\tilde{\mathbf{v}}_i)\}_{i=1}^{L}$ being associated left (right) singular vectors. The corresponding SVD factors of the noise matrix are identified by putting a bar over the symbols. It is worth mentioning that, for the SVD it does not matter whether we consider a data vector as a column or a row of a matrix, hence, this fact is left understood in the rest of the thesis.

The problem of finding the signal matrix $\mathbf{X}$ from the data matrix $\widetilde{\mathbf{X}}$ can have infinitely many solutions, in general. It can be solved in many different ways by using different regularizing priors on the signal matrix. In this thesis, we seek the solution to the matrix estimation problem by applying suitable prior on the signal SVs. To this end, the low rank prior suggested by Eckart-Young-Mirsky [25, 26] is one of the oldest formulations.

1.1.2 Spectral regularization

For the data model in (1.4), the signal matrix $\mathbf{X}$ can be recovered from the given data matrix $\widetilde{\mathbf{X}}$ by solving the objective function given as

$$\widehat{\mathbf{X}} = \min_{\mathbf{X}} \|\widetilde{\mathbf{X}} - \mathbf{X}\|_{\text{Frob}}^2 + \mu\ J(\mathbf{X}), \tag{1.6}$$

where $\|\cdot\|_{\text{Frob}}$ denotes the Frobenius norm, $J : \mathbb{R}^{m\times n} \mapsto \mathbb{R}$ is a regularization function that imposes some structure on the spectrum of $\mathbf{X}$, and $\mu > 0$ is the parameter of regularization. For $J(\mathbf{X}) = \text{Rank}(\mathbf{X})$, the objective function in (1.6) reduces to the *low-rank matrix approximation* (LRMA) problem. For rank being a non-convex function, the LRMA is an NP-hard problem and does not admit a closed form solution. An approximate solution to it has been suggested in the form of the *singular value hard-thresholding* (SVHT) and is defined as

$$\widehat{\mathbf{X}}_{\mu}^{\text{SVHT}} = \sum_{i=1}^{L} \tilde{\sigma}_i \mathbf{1}_{(\tilde{\sigma}_i > \mu)} \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\text{T}}, \tag{1.7}$$

where $\mathbf{1}_{(\tilde{\sigma}_i > \mu)}$ is an indicator function. The spectral regularization problem (1.6), however, can be easily solved if $J(\mathbf{X}) = \|\mathbf{X}\|_{\text{Nuc}}$, where $\|\cdot\|_{\text{Nuc}}$ denotes the nuclear norm defined as the sum of SV. Clearly, $\|\cdot\|_{\text{Nuc}}$ is a convex function [27]. The solution of the resulting optimization problem is given by the *singular value soft-thresholding* (SVST) [28] as

$$\widehat{\mathbf{X}}_{\mu}^{\text{SVST}} = \sum_{i=1}^{L} (\tilde{\sigma}_i - \mu)_{+} \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\text{T}}, \tag{1.8}$$

where $(\alpha)_+ = \max(\alpha, 0)$. It can be noticed that in both the SVHT and the SVST estimators, the SVs of the data matrix undergo modification while the singular vectors remain unaltered. The reason behind the same would become obvious when we will discuss the random matrix theory (RMT) in the following section. Moreover, we are interested in a more general class of estimators that estimate the signal matrix by applying some function on the data SVs only, without resorting to a specific form of the regularizer $J(\mathbf{X})$. Such estimators are commonly referred to as the *spectral estimators*.

1.1.3 Spectral estimator of a matrix

A typical spectral estimator $\mathrm{F} : \mathbb{R}^{m\times n} \mapsto \mathbb{R}^{m\times n}$ is an unitary invariant matrix valued function expressed as

$$\mathrm{F}(\widetilde{\mathbf{X}}) = \widetilde{\mathbf{U}}\mathrm{F}(\widetilde{\mathbf{\Sigma}})\widetilde{\mathbf{V}}^{\mathrm{T}} = \sum_{i=1}^{L} f(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\mathrm{T}}, \tag{1.9}$$

where $f : \mathbb{R} \mapsto \mathbb{R}$ is a real valued scalar function. Intuitively, on considering WGN in (1.4), the signal SVs get positively biased. For this reason, the function $f(\tilde{\sigma}_i)$ must shrink the SVs of the data matrix towards zero, hence, it is named as the shrinkage function. Clearly, for the SVHT and the SVST, the shrinkage function has the form $f_\mu^{\mathrm{SVHT}}(\tilde{\sigma}_i) = \tilde{\sigma}_i \mathbf{1}_{(\tilde{\sigma}_i > \mu)}$ and $f_\mu^{\mathrm{SVST}}(\tilde{\sigma}_i) = (\tilde{\sigma}_i - \mu)_+$, respectively.

Note that, the first term in (1.6) ensures the fidelity of the solutions while the second term regularizes or prunes the solution space. The low-rank regularization enables to suppress the noise energy as the underlying signal matrix is low-rank and the noise matrix is full-rank. The extent of noise suppression is determined by the regularization parameter μ. For smaller values of μ, the estimates remain too close to the data, thus leading to lower noise suppression. For larger values of μ, substantial noise suppression is achieved at the cost of significant loss in the signal energy. In a nutshell, the performances of the SVHT and the SVST based estimators are sensitive to the value of μ and hence it must be chosen judiciously. Intuitively, one may try to determine the optimal value of μ by minimizing the mean square error (MSE) of the estimator. But, it is not possible since the MSE of an estimator $\widehat{\mathbf{X}}_\mu^{\mathrm{F}}$ with $\mathrm{F} = \{\mathrm{SVHT}, \mathrm{SVST}\}$ is defined as

$$\mathrm{MSE}\left(\widehat{\mathbf{X}}_\mu^{\mathrm{F}}, \mathbf{X}\right) = \mathbb{E}\left[\|\widehat{\mathbf{X}}_\mu^{\mathrm{F}} - \mathbf{X}\|_{\mathrm{Frob}}^2\right] \tag{1.10}$$

where $\mathbf{X}$ is the unknown true signal matrix. Thus, for being dependent of unknown quantity, the MSE can not be used for determining the parameter of the spectral estimators. The question now arises whether there is a way to compute the MSE or its variant without knowledge of the true value. The direct answer to the question is not affirmative. But we could invoke either the random matrix theory or the Stein's unbiased risk estimator, for circumventing this challenge. These approaches are outlined below.

1.2 Random matrix theory

Random matrix theory (RMT) [29] provides an alternative to infer the asymptotic MSE (AMSE). We refer to the asymptotic setting (equivalently, the large matrix limit) in the data model (1.4) as $\lim_{m\to\infty} \frac{m}{n_m} \to \beta$ such that $0 < \beta \leq 1$ and $\tau = 1/\sqrt{n_m}$, where n_m is an increasing sequence of integers depending on the dimensionality m of the signal vectors $\mathbf{x}_i$s. For this large matrix limits, the AMSE is defined as

$$\text{AMSE}\left(\widehat{\mathbf{X}}^{\text{F}}, \mathbf{X}\right) = \lim_{m\to\infty} \text{MSE}\left(\widehat{\mathbf{X}}^{\text{F}}, \mathbf{X}\right). \tag{1.11}$$

Following this, the recent works in [1, 19–21] tried to find the optimal shrinkage functions for matrix denoising. In [20], a data driven estimator which forms the dominant low-rank estimator is proposed. In [21], the authors showed that the optimal value of hard-threshold is a deterministic quantity equal to $(4/\sqrt{3})\sqrt{n}\tau$ that minimizes the corresponding AMSE. In [1], the same authors proposed a generalized framework to derive the optimal shrinkage function for a number of loss functions.

Note that,the RMT is an independently developing field of mathematics. Efforts have been made for a long time to study the empirical distribution of eigenvalues of the sample covariance of random matrices. One of the goals of the RMT is to establish the relationship between the eigenvalues of the sample covariance matrix with that of the corresponding population covariance matrix. Loosely speaking, the main results in the RMT state that the empirical eigenvalue distribution of the sample covariance matrix of uncorrelated data follows the semicircular distribution with a finite support. Introducing any correlation in the data samples causes some of the eigenvalues to lie beyond the support. This phenomenon is exploited for noise suppression. More specifically, for the data model in (1.4), the noise matrix is uncorrelated whereas the signal matrix is highly correlated. As a result, the sample covariance matrix of the data matrix has non-zero off-diagonal elements. In this case, the eigenvalues within and beyond the support of the semicircular distribution correspond to the noise matrix and the signal matrix, respectively. Thus, the recovery of the signal matrix by preserving the eigenvalues beyond the support while discarding the rest leads to denoising. Such methods that retain some of the eigenvalues and throw away the rest is analogous to the SVHT. In the following, we will be interested in a more general class of shrinkage functions than the parameterized ones like the SVHT and the SVST discussed earlier. The following results are equally applicable to the eigenvalues as they are directly related to the SV.

1.2.1 Empirical singular value distribution

In order to recover a signal matrix from the observation corrupted by a random matrix using SVD, we need to understand the behavior of the SVs of random matrices and its effect on the SVs of the signal matrix when added to it. To this end, following lemmas are worth noting.

Lemma 1.2.1 (Generalized Quarter-circle Distribution [29,30]). *For an* $m \times n_m$ *random matrix* $\overline{\mathbf{X}}$ *with elements* $\bar{x}_{j,i} \overset{IID.}{\sim} \mathsf{N}(0, 1/n_m)$, *where* n_m *is an increasing sequence of integers dependent on* m, *the empirical distribution of its singular values* $\{\bar{\sigma}_i\}_{i=1}^{m}$ *in large matrix limits* ($\lim_{m\to\infty} \frac{m}{n_m} \to \beta$ *such that* $0 < \beta \leq 1$) *converges to a non-random distribution with probability density function*

$$p(\bar{\sigma}) = \frac{\sqrt{(\bar{\sigma}^2 - \beta_-^2)(\beta_+^2 - \bar{\sigma}^2)}}{\pi \beta \bar{\sigma}} \mathbf{1}_{[\beta_-, \beta_+]}(\bar{\sigma}), \tag{1.12}$$

almost surely (*a.s.*), *as* $m \to \infty$ *where* $\beta_- = 1 - \sqrt{\beta}$, $\beta_+ = 1 + \sqrt{\beta}$, *and* $\mathbf{1}_{[\cdot,\cdot]}$ *is an indicator function. Moreover,* $\bar{\sigma}_1 \xrightarrow{a.s.} \beta_+$ *and* $\bar{\sigma}_n \xrightarrow{a.s.} \beta_-$ *as* $m \to \infty$.

The above lemma states that no SV of $\overline{\mathbf{X}}$ lies beyond the support $[\beta_-, \beta_+]$. The SVs in (β_-, β_+) are often referred to as the *bulk* with its two extremes β_- and β_+ being called the *bulk-edges* [19]. Clearly, any SV beyond the bulk-edges indicates the presence of correlation in the data and so they are attributed to the signal[1].

Def. 1 (Bulk Shrinker of order r, f_{β_+}). We define a *bulk-shrinker* of order r as

$$f_{\beta_+}(\tilde{\sigma}_i) = \begin{cases} 0, & \sigma_i \leq \beta^{1/4} \text{ and } 1 \leq i \leq r, \\ 0, & \text{for all } \sigma_i \geq 0 \text{ and } i \geq r, \\ f(\tilde{\sigma}_i), & \text{otherwise.} \end{cases} \tag{1.13}$$

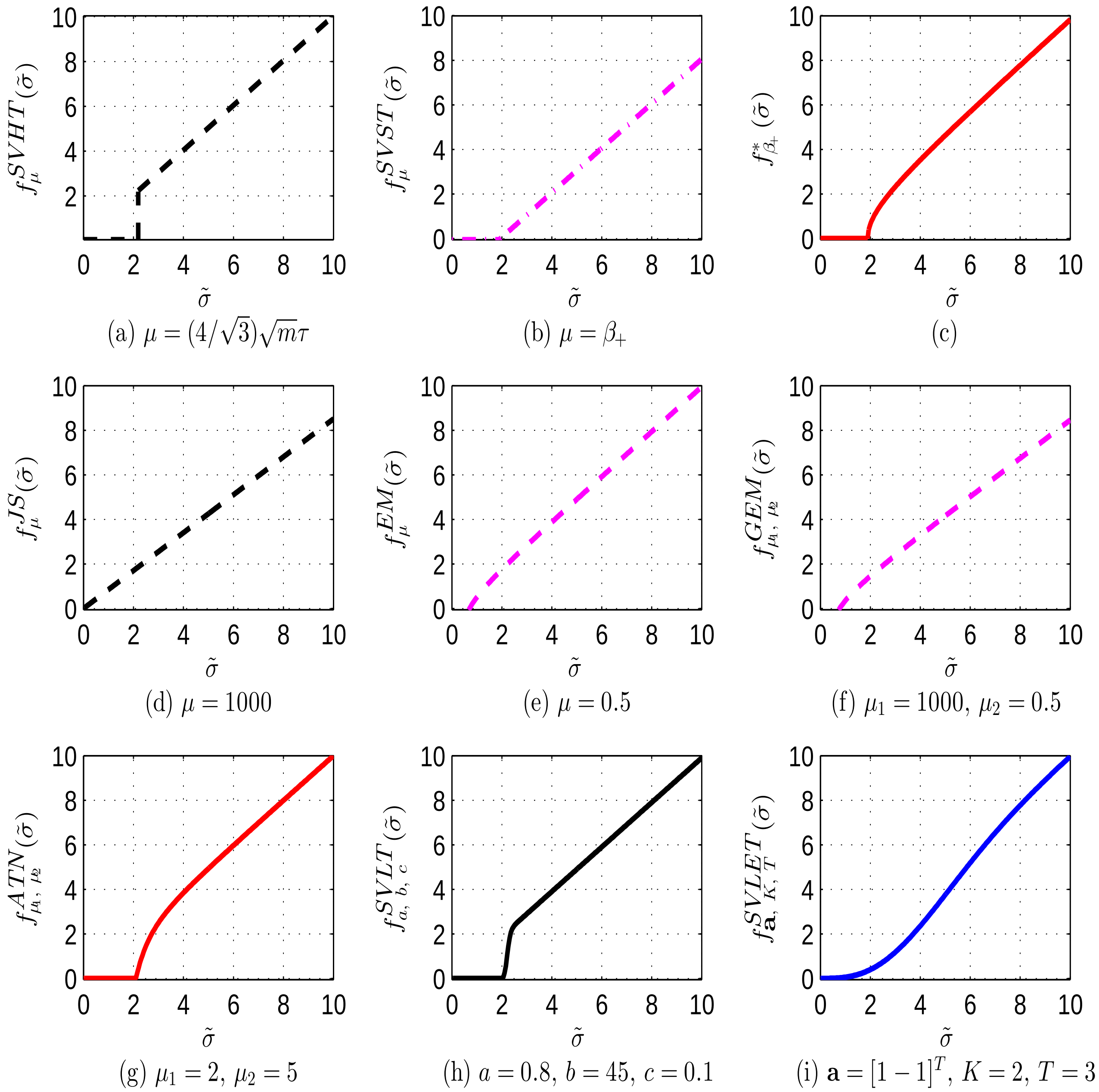
It will soon become clear that how it collapses the entire bulk to zero.

Equipped with these notations, we can find an optimal bulk-shrinker $f^*_{\beta_+}$, which minimizes the AMSE for a signal matrix $\mathbf{X}$ of rank r as

$$f^*_{\beta_+} = \min_{f_{\beta_+}} \underbrace{\lim_{m\to\infty} \frac{1}{mn_m} \left\| \sum_{i=1}^{L} f_{\beta_+}(\tilde{\sigma}_i) \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}} - \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}} \right\|_{\mathrm{Frob}}^2}_{\mathrm{AMSE}\left(\widehat{\mathbf{X}}^{\mathrm{F}}_{\beta_+}, \mathbf{X}\right)}. \tag{1.14}$$

---

[1] Here we have ignored an infinitesimal margin ϵ near the bulk edges.



(a) $\mu = (4/\sqrt{3})\sqrt{m}\tau$ (b) $\mu = \beta_+$ (c)

(d) $\mu = 1000$ (e) $\mu = 0.5$ (f) $\mu_1 = 1000,\ \mu_2 = 0.5$

(g) $\mu_1 = 2,\ \mu_2 = 5$ (h) $a = 0.8,\ b = 45,\ c = 0.1$ (i) $\mathbf{a} = [1\ -1]^T,\ K = 2,\ T = 3$

Figure 1.1: Comparison of different shrinkage functions

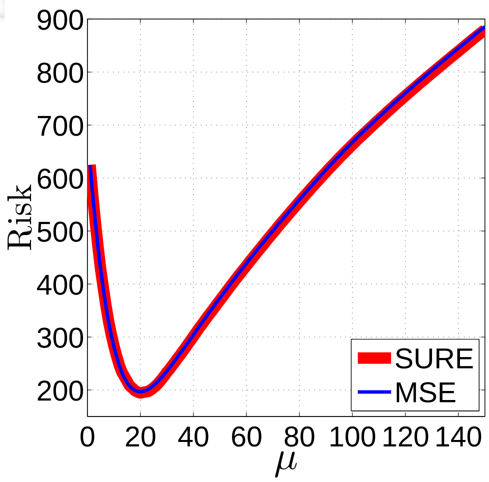


Figure 1.2: Comparison of SURE and MSE for the SVST for a range of μ

The authors in [1, 19] showed that a continuous bulk-shrinker given as

$$f^*_{\beta_+}(\tilde{\sigma}_i) = \sigma_i \langle \mathbf{u}_i, \tilde{\mathbf{u}}_i \rangle \langle \mathbf{v}_i, \tilde{\mathbf{v}}_i \rangle, \quad \forall\, i \le r \ \text{ and } \sigma_i > \beta^{1/4} \tag{1.15}$$

admits the optimal solution of (1.14) where $\langle \cdot, \cdot \rangle$ denotes the inner product. Clearly, the optimal bulk-shrinker depends not only on the true SVs of signal matrix but also on the cosine of the angular deviation (caused by the noise matrix added to it) in their singular vectors. Hence, it is important to study the effects of the noise matrix on the signal SVs as well as on the corresponding singular vectors. Lemma 1.2.2 presents the asymptotic locations of the data SVs relative to the signal SV, thereby providing a relation between $\tilde{\sigma}_i$ and σ_i. Lemma 1.2.3 presents a relative phase transition in the singular vectors of noise at bulk edges due to the presence of the signal components providing the cosine of the angles required in (1.15).

Lemma 1.2.2 (Asymptotic Location of SVs [19, 31])**.** *Consider the data model in (1.4) with the data SVs as* $\tilde{\sigma}_1 \ge \tilde{\sigma}_2 \ge \ldots \ge \tilde{\sigma}_{n_m}$ *and the signal SVs as* $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_r \ge 0$*. For* $1 \le i \le r$*, in large matrix limits, the asymptotic locations of SVs are related as*

$$\lim_{m\to\infty} \tilde{\sigma}_i \overset{a.s.}{=} \begin{cases} \rho(\sigma_i), & \sigma_i > \beta^{1/4} \\ \beta_+, & 0 < \sigma_i \le \beta^{1/4} \end{cases} \tag{1.16}$$

where

$$\rho(\sigma) = \sqrt{\left(\sigma + \frac{1}{\sigma}\right)\left(\sigma + \frac{\beta}{\sigma}\right)}. \tag{1.17}$$

Lemma 1.2.3 (Phase-Transition [19, 32])**.** *Suppose* $\mathbf{X}$ *in the data model (1.4) has distinct singular values* $\sigma_1 > \sigma_2 > \ldots > \sigma_r > 0$*. Then, for* $1 \le i, j \le r$ *and* $\sigma_i > \beta^{1/4}$

$$\lim_{m\to\infty} |\langle \mathbf{u}_i, \tilde{\mathbf{u}}_j \rangle| \overset{a.s.}{=} \begin{cases} \theta_u(\sigma_i), & i = j \\ 0, & i \ne j \end{cases} \tag{1.18}$$

and

$$\lim_{m\to\infty} |\langle \mathbf{v}_i, \tilde{\mathbf{v}}_j \rangle| \overset{a.s.}{=} \begin{cases} \theta_v(\sigma_i), & i = j \\ 0, & i \ne j \end{cases} \tag{1.19}$$

where

$$\theta_u(\sigma) = \sqrt{\frac{\sigma^4 - \beta}{\sigma^2(\sigma^2 + \beta)}} \quad \textit{and} \quad \theta_v(\sigma) = \sqrt{\frac{\sigma^4 - \beta}{\sigma^2(\sigma^2 + 1)}}. \tag{1.20}$$

However, if $\sigma_i \le \beta^{1/4}$ *or for* $i, j \ge r$ *we have* $\lim_{m\to\infty} |\langle \mathbf{u}_i, \tilde{\mathbf{u}}_j \rangle| \overset{a.s.}{=} \lim_{m\to\infty} |\langle \mathbf{v}_i, \tilde{\mathbf{v}}_j \rangle| \overset{a.s.}{=} 0.$

Corollary 1.2.4 (Equivalence)**.** *For the data model (1.4), as a direct consequence of Lemma 1.2.2, the conditions* $\sigma_i > \beta^{1/4}$ *and* $\tilde{\sigma}_i > \beta_+$ *are equivalent.*

Evidently, one can interchangeably use these conditions, however in the rest of the thesis, we prefer the latter as $\tilde{\sigma}_i$ is a known quantity. Also, it is clear now that the signal SVs smaller than $\beta^{1/4}$ are impossible to recover using the RMT. Often, the data SVs greater than β_+ are considered as active SVs and the number of active SVs are taken as an estimate of the rank, i.e., $r^* = \sum_{i=1}^{L} \mathbf{1}_{(\tilde{\sigma}_i > \beta_+)}$. Moreover, it is to note from (1.17) that the data SVs are always positively biased with respect to the signal SVs validating the 'shrinking toward zero' nature of the shrinkage estimators. In contrast to (1.15), we show that it is possible to achieve the optimal estimate by the shrinkage of data SVs while leaving the singular vectors unchanged. However, by no means, it should be understood that AWGN does not effect the signal singular vectors.

1.2.2 An asymptotically optimal shrinkage estimator

Using Lemma 1.2.2 and Lemma 1.2.3, the asymptotically optimal bulk-shrinker in (1.15) can be written as

$$f^*_{\beta_+}(\tilde{\sigma}_i) = \begin{cases} \frac{\sqrt{(\tilde{\sigma}_i^2 - \beta - 1)^2 - 4\beta}}{\tilde{\sigma}_i}, & \tilde{\sigma}_i > \beta_+ \text{ and } i \geq r^* \\ 0, & \text{otherwise.} \end{cases} \tag{1.21}$$

On plugging this optimal bulk-shrinker in (1.9), it yields an asymptotically optimal shrinkage estimator, referred to as the *OptShrink* estimator and denoted as $\widehat{\mathbf{X}}^*_{\beta_+}$, that achieves the lower bound on the AMSE. Figure 1.1 presents a comparison of different shrinkage functions. From Figure 1.1(c), we can see that the OptShrink estimator truncates the SVs smaller than β_+ and shrinks the SVs greater than β_+ towards zero. On comparing Figure 1.1(c) with Figure 1.1(a), it can be seen that $f_\mu^{\mathrm{SVHT}}(\tilde{\sigma})$ and $f^*_{\beta_+}(\tilde{\sigma})$ are quite close to each other, specially for the larger values of SV. Hence, a similar bias-variance trade-off is made by these estimators in the asymptotic setting. However, in the finite matrix limits or in relatively higher rank cases, this optimality is lost as the truncation of smaller SVs may introduce a small bias. In practice, we often deal with the finite limits where an exact low-rank modeling is too wishful to serve the purpose. Hence, the quest for an efficient shrinkage estimator in finite matrix limits is the recent interest for many researchers [2, 22, 33].

Coming back to (1.10), we were in pursuit of an estimator, more specifically its parameters, which minimize an unobservable quantity, i.e., the MSE. Fortunately, there exists Stein's unbiased risk estimator (SURE) of the MSE for the observation model in (1.4). SURE was originally developed for estimating the mean of the multivariate normal distribution [34]. It is a powerful method for unbiased estimation of the MSE and considered as one of the milestones in the development of statistical signal processing. With the help of SURE, the parameters of shrinkage estimators can be determined. For the same, the value that achieves the minimum of SURE is taken as the optimal parameter value, thus SURE serves as a surrogate to the MSE. In the following, we discuss about SURE and its usage in the determination of parameters of spectral estimators.

1.3 Stein's unbiased risk estimator of the MSE

SURE is based on the famous multivariate lemma proposed by Stein in 1981. The lemma is stated below.

Lemma 1.3.1 (Stein's Multivariate Lemma [34]). *For a matrix* $\widetilde{\mathbf{X}} \overset{IID}{\sim} \mathsf{N}(\mathbf{X}, \tau^2\mathbf{I}_{mn})$ *and a weakly differentiable function* F*,*

$$\mathbb{E}\left[(\widetilde{\mathbf{X}} - \mathbf{X})\mathrm{F}(\widetilde{\mathbf{X}})\right] = \tau^2\mathbb{E}\left[\nabla\mathrm{F}(\widetilde{\mathbf{X}})\right], \tag{1.22}$$

where $\nabla\mathrm{F}(\widetilde{\mathbf{X}})$ *denotes the divergence of* $\mathrm{F}(\widetilde{\mathbf{X}})$*.*

In what follows, we can see how remarkable this result is. For the spectral function $\mathrm{F}(\widetilde{\mathbf{X}})$, the MSE in (1.10) can be written as

$$\begin{aligned}\mathrm{MSE}\left(\mathrm{F}(\widetilde{\mathbf{X}}), \mathbf{X}\right) &= \mathbb{E}\left[\|\mathrm{F}(\widetilde{\mathbf{X}}) - \widetilde{\mathbf{X}} + \widetilde{\mathbf{X}} - \mathbf{X}\|_{\mathrm{Frob}}^2\right] \\ &= \mathbb{E}\left[\|\widetilde{\mathbf{X}} - \mathbf{X}\|_{\mathrm{Frob}}^2\right] + \mathbb{E}\left[\|\mathrm{F}(\widetilde{\mathbf{X}}) - \widetilde{\mathbf{X}}\|_{\mathrm{Frob}}^2\right] + 2\mathbb{E}\left[(\mathrm{F}(\widetilde{\mathbf{X}}) - \widetilde{\mathbf{X}})(\widetilde{\mathbf{X}} - \mathbf{X})\right] \\ &= -mn\tau^2 + \mathbb{E}\left[\|\mathrm{F}(\widetilde{\mathbf{X}}) - \widetilde{\mathbf{X}}\|_{\mathrm{Frob}}^2\right] + 2\mathbb{E}\left[(\widetilde{\mathbf{X}} - \mathbf{X})\mathrm{F}(\widetilde{\mathbf{X}})\right] \\ &= -mn\tau^2 + \mathbb{E}\left[\|\mathrm{F}(\widetilde{\mathbf{X}}) - \widetilde{\mathbf{X}}\|_{\mathrm{Frob}}^2\right] + 2\tau^2\mathbb{E}\left[\nabla\mathrm{F}(\widetilde{\mathbf{X}})\right]. \qquad \text{(Using (1.22))}\end{aligned}$$

Note that, no term in the right depends on the unknown quantity $\mathbf{X}$. Thus, Stein's lemma enables us to infer the MSE. Hence, SURE for a spectral function is given as

$$\mathrm{SURE}_{(\mathrm{F})}(\widetilde{\mathbf{X}}) = -mn\tau^2 + \|\mathrm{F}(\widetilde{\mathbf{X}}) - \widetilde{\mathbf{X}}\|_{\mathrm{Frob}}^2 + 2\tau^2\nabla(\mathrm{F}(\widetilde{\mathbf{X}})), \tag{1.23}$$

provided its divergence $\nabla\mathrm{F}(\widetilde{\mathbf{X}})$ exists. Clearly, $\mathrm{MSE} = \mathbb{E}\left[\mathrm{SURE}_{(\mathrm{F})}(\widetilde{\mathbf{X}})\right]$. We show this fact with a toy example in Figure 1.2 where we plot the MSE and SURE for the SVST against different values of threshold parameter μ. It suggests how SURE can be used to find the parameter without needing $\mathbf{X}$, as SURE replicates the MSE as a function of μ. Specifically,

$$\mu_{\text{opt}} = \min_{\mu} \text{SURE}_{(\text{SVST})}(\widetilde{\mathbf{X}}, \mu). \tag{1.24}$$

However, computing the divergence of the spectral function is non-trivial. Various works [35–38] have tried to compute this quantity explicitly for the estimation functions of their interests. For the nuclear norm regularized problem, the authors in [17] derive the divergence in closed form. In [39], the authors derive the divergence for the reduced rank regression problem. Recently, in [18], the divergence of the SVST is derived and corresponding SURE is used as a surrogate to the MSE for determining the parameters of the SVST. The following lemma states a useful result from the same paper and provides a recipe for the works presented in this thesis.

Lemma 1.3.2 (Divergence of Spectral Function [18])**.** *Suppose* $\mathfrak{F}$ *is a set of full-rank simple matrices such that the Lebesgue measure of its complementary set is zero. Consider a unitary-invariant spectral function* $\mathrm{F} : \mathfrak{F} \mapsto \mathbb{R}^{n\times m}$ *of the form (1.9). If* $\mathrm{F}(\widetilde{\mathbf{X}})$ *is weakly differentiable, its divergence can be given as*

$$\nabla \mathrm{F}(\widetilde{\mathbf{X}}) = \sum_{i=1}^{L} f'(\tilde{\sigma}_i) + |n-m| \sum_{i=1}^{L} \frac{f(\tilde{\sigma}_i)}{\tilde{\sigma}_i} + 2 \sum_{i=1}^{L} \sum_{\substack{j=1,\\ i\neq j}} \frac{\tilde{\sigma}_i \, f(\tilde{\sigma}_i)}{\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2}, \tag{1.25}$$

where $f' = \partial f / \partial \tilde{\sigma}_i$*.*

The differentiability assumption in (1.25) can be relaxed, without any loss, to the differentiability in a spectral neighborhood which in turns depends on the differentiability of all factors of the SVD of $\widetilde{\mathbf{X}}$ [40–42]. Hence, we only require $\widetilde{\mathbf{X}}$ to be a simple (i.e., having no repeated SVs) and full-rank matrix. For the data model (1.4), due to the random Gaussian noise, $\widetilde{\mathbf{X}}$ is full-rank with high probability irrespective of the rank of the signal matrix $\mathbf{X}$.

1.4 Optimal shrinkage using the RMT and the SURE

In the previous section, we saw that the RMT finds an optimal shrinkage estimator in case of large matrix limits and SURE infers the MSE enabling us to find the optimal value of shrinkage parameters. As the expectation of SURE achieves the MSE asymptotically, we can rely on it when the number of samples are sufficiently large. Hence, one would naturally tend to examine the behavior of SURE in case of large matrix limits. The following lemma helps in establishing the link between the RMT and SURE.

Lemma 1.4.1 (Asymptotic SURE [24]). *In asymptotic setting, for a data matrix* $\widetilde{\mathbf{X}}$ *sampled from (1.4) with the singular values* $\tilde{\sigma}_1 > \tilde{\sigma}_2 > ... > \tilde{\sigma}_L \geq 0$ *and underlying signal singular values* $\sigma_1 > \sigma_2 > ... > \sigma_r \geq 0$ *with* $\sigma_i > \beta^{1/4}$ *for all* $i \leq r$*, we have*

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j=1;j\neq i}^{m} \frac{\tilde{\sigma}_i}{\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2} = \frac{1}{\rho(\sigma_i)} \left(1 + \frac{1}{\sigma_i^2}\right), \tag{1.26}$$

as $m \to \infty$*, almost surely.*

This lemma will be used to establish the asymptotic performance of the proposed estimator in Chapter 3.

1.5 Progress and challenges

Shrinkage estimators, though have gained much popularity in signal processing very recently, date back to the seminal work [34] by James and Stein published in 1981. The shrinkage estimator proposed in their work can be given as

$$\widehat{\mathbf{X}}_{\mu}^{\text{JS}} = \sum_{i=1}^{L} \tilde{\sigma}_i \left(1 - \frac{\mu}{\sum_{i=1}^{L} \tilde{\sigma}_i^2}\right) \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}}. \tag{1.27}$$

It dominates the maximum likelihood estimator of $\mathbf{X}$ and referred to as the *JS* estimator in the rest of the thesis. The JS estimator applies the shrinkage on all SVs equally. However, all the signal SVs are not equally informative in practice and hence they need to be treated differently. Improving on the JS estimator, Efron and Morris [43] proposed an estimator referred to as the *EM* estimator and given as

$$\widehat{\mathbf{X}}_{\mu}^{\text{EM}} = \sum_{i=1}^{L} \left(\tilde{\sigma}_i - \frac{\mu}{\tilde{\sigma}_i}\right) \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}}. \tag{1.28}$$

Later, in [44], they proposed a generalized shrinkage estimators as

$$\widehat{\mathbf{X}}_{\mu_1,\mu_2}^{\text{GEM}} = \sum_{i=1}^{L} \left(\left(1 - \frac{\mu_1}{\sum_{i=1}^{L} \tilde{\sigma}_i^2}\right) \tilde{\sigma}_i - \frac{\mu_2}{\tilde{\sigma}_i} \right) \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}}. \tag{1.29}$$

We refer to this estimator as the *GEM* estimator. The parameters of these estimators were used to be fixed heuristically until Candès *et al.* derived SURE for the SVST. The drawback of the SVST is its rigid shape that equally shrinks the larger SVs and truncates the smaller SVs. Also, for being coupled, the shrinkage and the truncation in the SVST can not be controlled separately. Recently, a few works have been reported to make the shape of the shrinkage function more flexible [2,33]. In [2], an adaptive trace norm (ATN) estimator given as

$$\widehat{\mathbf{X}}^{\mathrm{ATN}}_{\mu_1,\mu_2} = \sum_{i=1}^{L} \tilde{\sigma}_i \left(1 - \left(\frac{\mu_1}{\tilde{\sigma}_i}\right)^{\mu_2}\right)_{+} \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}} \tag{1.30}$$

was proposed. Here, μ_1 controls the amount of shrinkage or threshold applied whereas μ_2 controls the type of the shrinkage. For $\mu_2 = 1$, it reduces to the SVST and as $\mu_2 \to \infty$ it works as the SVHT. With an additional parameter, the ATN estimator undoubtedly provides a more flexible shrinkage estimator than the SVST. As the parameters of the ATN estimator are also tuned using SURE, it is computationally more complex than the SVST. Hence, the challenge in this line of work is two-fold: firstly, the design of a better shrinkage estimator and secondly, the determination of the shrinkage parameters.

1.5.1 Problem formulation and objectives

Above discussions emphasize the fact that designing a better shrinkage estimators is an active research problem. Form the literature review, we identify the following two difficulties of the existing estimators: (1) they lack in control over the shape of the shrinkage functions, therefore, they are unable to adaptively shrink the SVs, (2) the efforts made to design adaptive estimators having multiple parameters are marred by the complexity involved in tuning them.

The reason of limited flexibility and hence the suboptimal bias-variance trade-off lies in the rigidity of the shape of the existing shrinkage functions. For an instance, the truncation and the shrinkage of SVs are performed by the single threshold parameter in the SVST and can not be performed independently. The ATN estimator tried to separately control these operations by introducing an additional parameter into the estimator. However, it could not achieve the optimal truncation of the SVHT and the optimal shrinkage of the SVST, simultaneously. The parameters of these estimators are tuned by minimizing SURE which does not admits a closed form solution. Alternatively, the grid search is employed. But the complexity of the same increases with the increase in the number of parameters as well as the resolution of the grids.

Motivated by the above mentioned challenges, we intend to design an estimator that completely decouples the truncation and the shrinkage of the SVs so that they can be separately controlled. The estimator should adopt its shape according to the statistics of the data. The decoupling is achieved by using two extra parameters. As a result, the tuning becomes cum-

Table 1.1: List of various estimators proposed/explored for different denoising applications. Abbreviations used in the table are defined in the list of acronyms.

Signals	Methods
1D ECG	NLWT
	EYM
	US
	SVST
	SVLT
	ATN
	SVLET
	OptShrink
2D Image	FRIT
	SVLET
3D MR Image	SVLET
	OptShrink

bersome. To address this problem, we linearize the shrinkage function so that SURE can be optimized in closed form. This not only yields in a drastic reduction in computational cost but also provides a way to enhance the parameter space. In Figure 1.1, the proposed shrinkage functions are compared with the various existing shrinkage functions. We further want to explore the proposed SV shrinkage based estimators for signal denoising tasks involving electrocardiograms (ECG), photographic images, and magnetic resonance (MR) images.

1.6 Contributions of the thesis

In this thesis, we proposed two shrinkage estimators and applied them for the denoising of practical signals of varying dimensionality. In addition, we have also explored some existing shrinkage estimators for the same. In particular, the usage of the EYM, the US, the SVST, the ATN, and the OptShrink estimators in collaborative filtering based ECG signal denoising is reported. Table 1.1 summarises the different existing/proposed estimators evaluated in this thesis for various denoising applications.

The salient contributions made in the thesis are as follows:

- A thorough review of different singular value shrinkage estimators (SVSEs) proposed in the literature are presented while critically analysing their pros and cons.
- A logistic function based SVSE is proposed and its efficacy for the matrix rank estimation

in the presence of the AWGN is analysed.

- A shrinkage estimator by linear approximation of the threshold function is proposed and a theoretical analysis on the asymptotic optimality of this estimator is presented. The proposed SVLET estimator is more flexible and data-adaptive, and it performs better than the state-of-the-art shrinkage estimators for the matrices of any rank. Additionally, the tuning of its parameter is comparatively faster than that of the existing estimators.
- The proposed two matrix estimators are applied for the denoising of ECG signals, photographic images, and MR images in a collaborative filtering framework.

1.7 Organization of the thesis

This chapter introduced the matrix estimation problem and the existing solution methods using spectral shrinkage. It also highlighted the drawbacks of the existing shrinkage estimators. The rest of the thesis is organized in six chapters.

In Chapter 2, we propose the SVLT estimator for rank estimation that decouples the shrinkage and the truncation. The estimator provides better estimation accuracy than the existing rank estimators. The parameters of this estimator are tuned by minimizing SURE using grid search over a given search range. The tuning of SURE for the optimal value of various parameters is computationally intensive. Hence, despite improved estimation performance, its applications are limited.

In Chapter 3, we propose the comparatively faster SVLET estimator that does not require any grid search and provides better matrix estimation performance. This is achieved by the linearization of the non-linear threshold function so that the corresponding SURE admits a closed form solution and allows to directly find the optimal parameters of the estimator. We also analyse and establish the optimality of this estimator in the asymptotic case.

In Chapter 4, we evaluate the proposed SV shrinkage estimators for the denoising of ECG signals in a collaborative filtering framework. For the same, the similar blocks of a noisy 1D ECG signal are grouped together in the form of the data matrices. Such matrices are denoised using the proposed SV shrinkage estimators. The denoised matrices are then rearranged to obtain the estimate of the underlying noise-free signal. The proposed approach is compared with the state-of-the-art nonlocal means based ECG denoising method. The effectiveness of the proposed approach is also demonstrated for the denoising of pathological ECG signals.

In Chapter 5, we explored the proposed SV shrinkage estimators for 2D image denoising. The similar 2D patches within a noisy image are searched and their vectorised versions are put in the columns to form the data matrices. The denoising of these data matrices are then performed with various proposed and existing methods. The intermediate denoised image is obtained by aggregating the patch estimates. For the further enhancement, the denoising steps are iterated a couple of times unlike the ECG signal denoising case. The denoising performance of the proposed approaches are found to be competitive with the state-of-the-art image denoising methods, both quantitatively and qualitatively.

In Chapter 6, the proposed estimators are evaluated for denoising the MR images. For the same, the similar 3D blocks of a noisy MR image are grouped together and their vectorised versions are put in the columns to form the data matrices. Rest of the process remains similar to that of the image denoising case. Both quantitative as well as qualitative assessment of the denoising performances are presented.

Finally, the contributions of the thesis and the possible future works are discussed in Chapter 7.

2

SVLT: Decoupling Shrinkage and Truncation for Matrix Rank Estimation

Contents

2.1 Matrix rank estimation 21
2.2 Proposed method 24
2.3 Experimental results and discussion 25
2.4 Summary 30

Rank estimation is a classical problem with many applications in statistical signal processing, e.g., dimensionality reduction [45], image denoising [46], signal subspace tracking [47], motion estimation [48], system model selection [49], and communication engineering [50]. The rank estimation and the matrix estimation problems are interrelated. With an estimated rank, one can find out the corresponding estimate of the matrix. However, such an estimate of the signal matrix is not optimal. Also, the rank estimation problem has a strong connection with factor analysis if the noise variances in all the factors are assumed to be same [51]. Traditionally, the information theoretic criteria like Akaike information criterion (AIC) [52], Bayesian information criterion (BIC) [53], and minimum description length (MDL) [54] have been used for solving this problem. However with these approaches, good performance can only be expected if the number of observation is large compared to dimensionality of the data vectors. In practice, the dimensionality of the data in modern signal processing applications (medical imaging, meteorology, oceanography, etc.) is very large and acquisition of more observations is often costly. In such cases, cross-validation (CV) methods [55] were the only choice until recently. The CV methods are however computationally expensive.

Recently, an effective method for the selection of principal components (PCs) is proposed, in [56], by employing SURE [34]. It is noted to result in better rank estimation accuracy at a very low computational cost. It forms the current state of the art for the rank estimation problem. Other methods for matrix estimation proposed in [14, 57] can also be used for rank estimation. In these methods, either noise in signal dominated SVs or signal in noise dominated SVs is generally neglected. This is because, the signal dominated SVs and the noise dominated SVs are treated equally. As a result, the selection of PCs may be erroneous. In this chapter, we propose an efficient rank estimation method using a logistic function based thresholding of the SVs. It treats signal dominated and noise dominated SVs differently. Like in [56], the parameters of the proposed estimator are tuned using SURE. The proposed method is able to approximate both the SVHT and the SVST as special cases; hence it generalizes them. In addition to rank estimation, the efficacy of the proposed approach is also demonstrated for the recovery of the original data matrix from its noisy observation.

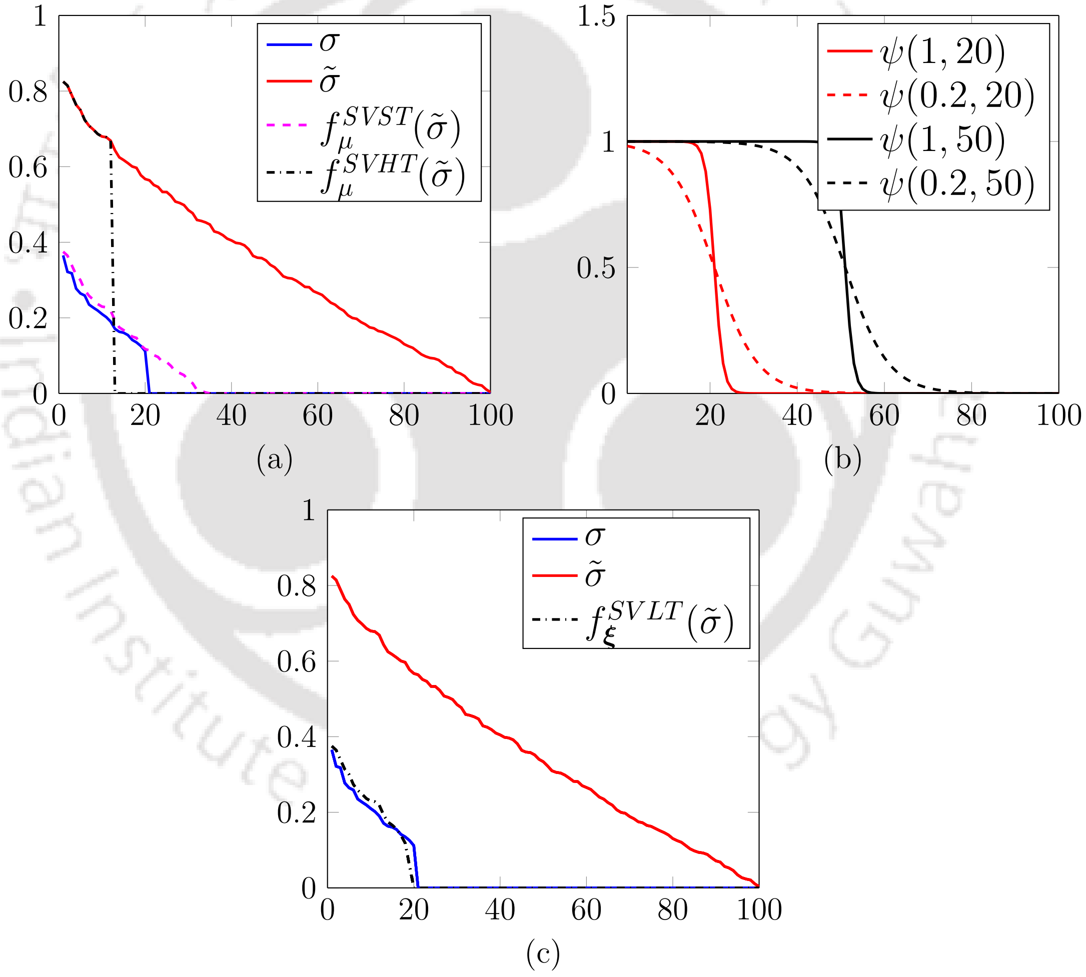
2.1 Matrix rank estimation

For an observation model in (1.4), the rank estimation problem tries to find out the rank of the underlying signal matrix $\mathbf{X}$ from the given data matrix $\widetilde{\mathbf{X}}$. It basically aims at minimizing the MSE for the rank of the matrix estimate

$$\widehat{\mathbf{X}}_{\hat{r}} = \sum_{i=1}^{\hat{r}} \tilde{\sigma}_i \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}} \tag{2.1}$$

as

$$\hat{r} = \arg\min_{r} \mathrm{MSE}\left(\widehat{\mathbf{X}}_r, \mathbf{X}\right). \tag{2.2}$$



(a) (b)

(c)

Figure 2.1: Given a 20-rank data matrix $\mathbf{X}$ of size 100×100 and its WGN added noisy version $\widetilde{\mathbf{X}}$. (a) Demonstration of the SVST (pink) and the SVHT (black) of the SVs of the noisy data (red) with those of the original data (blue). (b) Plots of the logistic function $\psi(a, b)$ embedded in the proposed SVLT for varying parameters. (c) Demonstration of the improved singular value shrinkage (black) with the application of the proposed SVLT.

The rank of a matrix is equal to the number of non-zero SV. However, adding the white

Gaussian noise (WGN) to the elements of the matrix makes it full rank. This is illustrated with an example in Figure 2.1(a). Thus, inferring the rank of the original matrix by the number of non-zero SVs of the noisy matrix is not viable. If the noise energy is not too high, the SVs that correspond to noise are small. In such cases, one can consider the number of dominant SVs as an estimate of the rank of underlying data matrix. Various ad hoc methods traditionally used for selecting the dominant SVs include the approaches such as (i) locating an *elbow* in the Scree plot of the noisy data [58]; (ii) finding r largest SVs that accumulate 80-90% of total variation of the noisy observation [59]; and (iii) using the CV methods [55].

In the Scree plot of a noisy version of a low-rank matrix, the eigenvalues on the right-side of the elbow are assumed to correspond to noise. Thus, the number of eigenvalues on the left-side of the elbow is taken as the rank of the underlying (low-rank) data matrix. However, locating the elbow in a Scree plot is sometimes quite difficult task. The second approach assumes that the percentage cumulative variation captured by r largest SVs, computed as $\left(\sum_{i=1}^{r} \tilde{\sigma}_i^2 / \sum_{i=1}^{L} \tilde{\sigma}_i^2\right) \times 100\%$, should lie in 80-90% of the total variation of the observation. The minimum number of SVs satisfying this range can be considered as an estimate of the rank of underlying data matrix. This is a widely used method to choose the number of PC in various signal processing applications. The CV approach [60] selects the rank by cross-validating the r-rank approximation using the SVD of the reduced version the data matrix until one comes up with the best estimate or the rank. For the heuristic nature of the CV approach, it is computationally intensive. The notable drawback of the first and the second approach is the subjectivity in the rank estimation process. In [56, 61], Stein's lemma [34] is used for matrix rank estimation that mitigates this drawback. In the following, we briefly discuss some of the SURE based rank estimation methods.

2.1.1 SURE based rank estimation methods

Ulfarsson and Solo, in [56], proposed a rank estimator by minimizing SURE corresponding to $\widehat{\mathbf{X}}_r$ as

$$\hat{r} = \arg\min_{r} \mathrm{SURE}_{\widehat{\mathbf{X}}_r}(\widetilde{\mathbf{X}}). \tag{2.3}$$

This approach of rank estimation is referred to as the *US* method in this thesis. The US method is analogous to the SVHT if its parameter μ is selected properly. Specifically, $\hat{r}$ is equal to the number of non-zero SVs of $\widehat{\mathbf{X}}_{\mu}^{\mathrm{SVHT}}(\widetilde{\mathbf{X}})$ provided μ is selected to minimize SURE corresponding to the SVHT. As the SVHT is a discontinuous function, computing its divergence is non-trivial, specially in case of repeated SV. Thus, the use of SURE for finding optimal μ for the SVHT is theoretically not profound.

Candès *et al.* in [18] derived the expression for SURE corresponding to the weakly differentiable SVST estimator given in (1.8), and found the optimal value of μ as

$$\mu_{\mathrm{opt}} = \arg\min_{\mu} \mathrm{SURE}_{\mathrm{SVST}}(\widetilde{\mathbf{X}}) \tag{2.4}$$

where

$$\mathrm{SURE}_{\mathrm{SVST}}(\widetilde{\mathbf{X}}) = -mn\tau^2 + \sum_{i=1}^{L} \min(\mu^2, \tilde{\sigma}_i^2) + 2\tau^2 \mathrm{div}(\widehat{\mathbf{X}}_{\mu}^{\mathrm{SVST}}) \tag{2.5}$$

with

$$\mathrm{div}(\widehat{\mathbf{X}}_{\mu}^{\mathrm{SVST}}) = \sum_{i=1}^{L} \mathbf{1}_{(\tilde{\sigma}_i > \mu)} + |m-n| \sum_{i=1}^{L} \frac{(\tilde{\sigma}_i - \mu)_+}{\tilde{\sigma}_i} + \sum_{i=1}^{L} \sum_{\substack{j=1,\\ i \neq j}}^{L} \frac{\tilde{\sigma}_i(\tilde{\sigma}_i - \mu)_+}{\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2}. \tag{2.6}$$

Once μ_{opt} is obtained, the estimated rank is given by the number of data SVs greater than μ_{opt}. In the rest of the chapter, this method of rank estimation is referred to as the *SVST.*

Recently, Guo *et al.* [57] used the *Eckart-Young-Mirsky* (EYM) theorem to estimate the rank of a matrix for the denoising application. According to the EYM theorem, for a real matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ with SVs $\{\sigma_i\}_{i=1}^{L}$ and an r-rank matrix $\mathbf{Z}$, the Frobenius norm of error is lower bounded as

$$||\mathbf{X} - \mathbf{Z}||_{\mathrm{Frob}}^2 \geq \sum_{i=r+1}^{L} \sigma_i^2 \tag{2.7}$$

where the equality is attained when $\mathbf{Z}$ is an r-rank approximation of $\mathbf{X}$. Now, for the data model in (1.4), we ideally have

$$||\widetilde{\mathbf{X}} - \mathbf{X}||_{\mathrm{Frob}}^2 = mn\tau^2. \tag{2.8}$$

Using (2.7) and (2.8), we can write

$$\sum_{i=r+1}^{L} \tilde{\sigma}_i^2 \leq mn\tau^2 < \sum_{i=r}^{L} \tilde{\sigma}_i^2. \tag{2.9}$$

The value of r satisfying the inequality in (2.9) is taken as the estimate of the rank. This approach of rank estimation, in the rest of the chapter, is referred to as the *EYM* method. The rank estimation methods discussed above have been used to provide the contrast to the method proposed in this chapter.

2.1.2 Drawbacks of the existing rank estimators

The methods discussed above result in a sub-optimal estimate of the rank. In the US method, only the cardinality of SVs is used to minimize SURE. The US and the EYM method do not give much importance to the noise in the signal dominant components while minimizing the SURE. In Figure 2.1(a), it can be noted that the SVHT hardly selects the true rank. This is because it applies truncation to minimize the MSE without any discrimination between the noise and the signal in the larger SV. When the SNR is low and the underlying matrix is of very small rank, the SVST also is unable to shrink the noise dominated SVs to zero in order to preserve the signal energy in the larger SV. Hence, the truncation of noisy SVs and the shrinkage—the lesser is the shrinkage the more is the preservation of signal energy—are needed to be decoupled while minimizing SURE. In this chapter, we suggest a logistic function based methodology to decouple the shrinkage and the truncation. The parameters of this logistic function provide the independent control over the amount of the shrinkage and the truncation applied while minimizing the MSE.

2.2 Proposed method

The shrinkage and the truncation can be easily decoupled by an indicator function and the resulting estimator can be given as

$$\widehat{\mathbf{X}}_{(\mu_1,\mu_2)} = \sum_{i=1}^{L} \mathbf{1}_{(\tilde{\sigma}_i \geq \mu_1)} f_{\mu_2}^{\mathrm{SVST}}(\tilde{\sigma}_i) \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}}. \tag{2.10}$$

As the shrinkage estimator in (2.10) is not differentiable, the determination of the parameters μ_1 and μ_2 is challenging. However, if the indicator function in (2.10) is replaced by a logistic function, the parameter estimation problem can be simplified. This fact motivates us to propose a *singular value logistic thresholding* (SVLT) estimator defined as

$$\widehat{\mathbf{X}}_{\boldsymbol{\xi}}^{\mathrm{SVLT}} = \sum_{i=1}^{L} (\tilde{\sigma}_i \psi_i(a,b) - c)_+ \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}} \tag{2.11}$$

where $\psi_i(a,b) = \frac{1}{1+\exp\{a(i-b)\}}$ is a logistic function and $\boldsymbol{\xi} = [a,b,c]^{\mathrm{T}}$ is the parameter vector. The non-negative parameters a, b, and c control the slope, the shift, and the scale factors, respectively. Note that,for a very large value of a, the SVLT reduces to the SVHT, whereas for a very small value of a it works as the SVST. The shift parameter b tries to track the elbow in the Scree plot, whereas the scale parameter c determines the amount of shrinkage applied on the signal dominant SVs. The parameter b basically determines the transition region in the Scree-plot. For $i < b$, the shrinkage is applied by the parameter c and for $i \geq b$, the truncation is done. Parameter a determines the slope of transition between these two regions. Thus, the proposed SVLT decouples the shrinkage and the truncation and results in a more flexible shrinkage estimator of matrix rank.

Using (1.23) and (1.25), SURE corresponding to the SVLT estimator (2.11) can be easily derived as

$$\mathrm{SURE}_{\mathrm{SVLT}}(\widehat{\mathbf{X}}) = -mn\tau^2 + \sum_{i=1}^{L} \left((\tilde{\sigma}_i(1-\psi_i)+c)^2 \mathbf{1}_{(\tilde{\sigma}_i\psi_i > c)} + \tilde{\sigma}_i^2 \mathbf{1}_{(\tilde{\sigma}_i\psi_i < c)} \right) + 2\tau^2 \nabla(\mathrm{SVLT}_{\boldsymbol{\xi}}(\widetilde{\mathbf{X}})) \tag{2.12}$$

where

$$\nabla(\widehat{\mathbf{X}}_{\boldsymbol{\xi}}^{\mathrm{SVLT}}(\widetilde{\mathbf{X}})) = \sum_{i=1}^{L} \psi_i \mathbf{1}_{(\tilde{\sigma}_i\psi_i > c)} + |m-n| \sum_{i=1}^{L} \left(\psi_i - \frac{c}{\tilde{\sigma}_i} \right)_+ + 2 \sum_{i=1}^{L} \sum_{\substack{j=1, \\ i \neq j}}^{L} \frac{\tilde{\sigma}_i(\tilde{\sigma}_i\psi_i - c)_+}{\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2}. \tag{2.13}$$

The parameters a, b, and c can be optimally obtained by minimizing SURE as

$$\hat{\boldsymbol{\xi}} = \{\hat{a}, \hat{b}, \hat{c}\} = \underset{a,b,c}{\arg\min}\ \mathrm{SURE}_{\mathrm{SVLT}}(\widetilde{\mathbf{X}}). \tag{2.14}$$

Then, the estimated rank is obtained as $\hat{r}_{\mathrm{SVLT}} = \sum_{i=1}^{L} \mathbf{1}_{(\tilde{\sigma}_i > \hat{c}(1+\exp\{\hat{a}(i-\hat{b})\}))}$ and corresponding estimate of underlying signal matrix is obtained using (2.11). Thus, the proposed method is expected to yield a better estimate of the rank and hence that of the matrix.

2.3 Experimental results and discussion

In the following, we present a comparative study on the experimental performance of the proposed and the existing rank estimators through a number of simulations on artificially generated data. The original low-rank data matrix $\mathbf{X}_{m\times n}$, having a rank $r \leq L$, is generated by multiplying two random Gaussian matrices $\mathbf{L}_{m\times r}$ and $\mathbf{R}_{n\times r}$ as $\mathbf{X} = \mathbf{L}\mathbf{R}^T$. For simplicity and generality, no constraint on the SV profile of $\mathbf{X}$ is imposed unless specifically stated. The noisy observation $\widetilde{\mathbf{X}}$ of the original data matrix $\mathbf{X}$ is obtained by adding with it the WGN of different variances. To demonstrate the efficacy of the SVLT, a 20-rank matrix and its noisy version are created and their SV plots for different methods are shown in Figure 2.1. From Figure 2.1(a), we note that neither the SVHT nor the SVST is able to achieve the Scree plot of the data matrix $\mathbf{X}$. Figure 2.1(b) shows the logistic function embedded in the SVLT for different values of the parameters. From Figure 2.1(c), it is clear that the SVLT closely matches the Scree plot of $\mathbf{X}$, if the parameters are tuned properly. For this reason, the SVLT is expected to outperform the existing methods.

In the following, we present the results of the various experiments to validate the proposed SVLT based rank estimation method.

2.3.1 SURE for SVLT vs. MSE

To show the closeness of SURE for the SVLT to the MSE, we conducted an experiment similar to the one reported in [18]. We took four matrices $\{\mathbf{X}^{(k)}\}_{k=1}^{4}$ each of size 200×400. $\mathbf{X}^{(1)}$ is full rank, $\mathbf{X}^{(2)}$ is 100-rank, and $\mathbf{X}^{(3)}$ is 10-rank whereas $\mathbf{X}^{(4)}$ has the SV profile given as $\sigma_i = \sqrt{200}/(1 + \exp\{\frac{i-100}{20}\})$. Each of these matrices is normalized and then WGN at varying SNRs (0.5, 1, 2, and 4 dB) is added to obtain the corresponding noisy matrices $\widetilde{\mathbf{X}}^{(k)}$. Thereafter, the $\text{SURE}_{\boldsymbol{\xi}}(\widetilde{\mathbf{X}}^{(k)})$ is computed for $a = 1$ and $c = 0$ while varying b. From Figure 2.2, it can be noted that $\text{SURE}_{\boldsymbol{\xi}}(\widetilde{\mathbf{X}}^{(k)})$ is able to achieve the Monte-Carlo simulation (with known $\mathbf{X}^{(k)}$) for all the considered matrices and SNRs. Similar performances are noted when other parameters (a and c) are varied. Hence, we can rely on SURE for the optimal parameter selection.

2.3.2 Rank estimated by SVLT vs. true rank

In this experiment, to obtain the noisy observations the matrices of size 200×400, the signal matrix having different ranks are generated as in the above experiment and zero mean WGN at the SNR levels of -2, 0, 2, and 4 dB are added. The rank is estimated using all the considered methods and the performances averaged over 50 realizations are reported. In all experiments, hereafter, a is fixed as 1 being the best choice found experimentally. The rest of the two

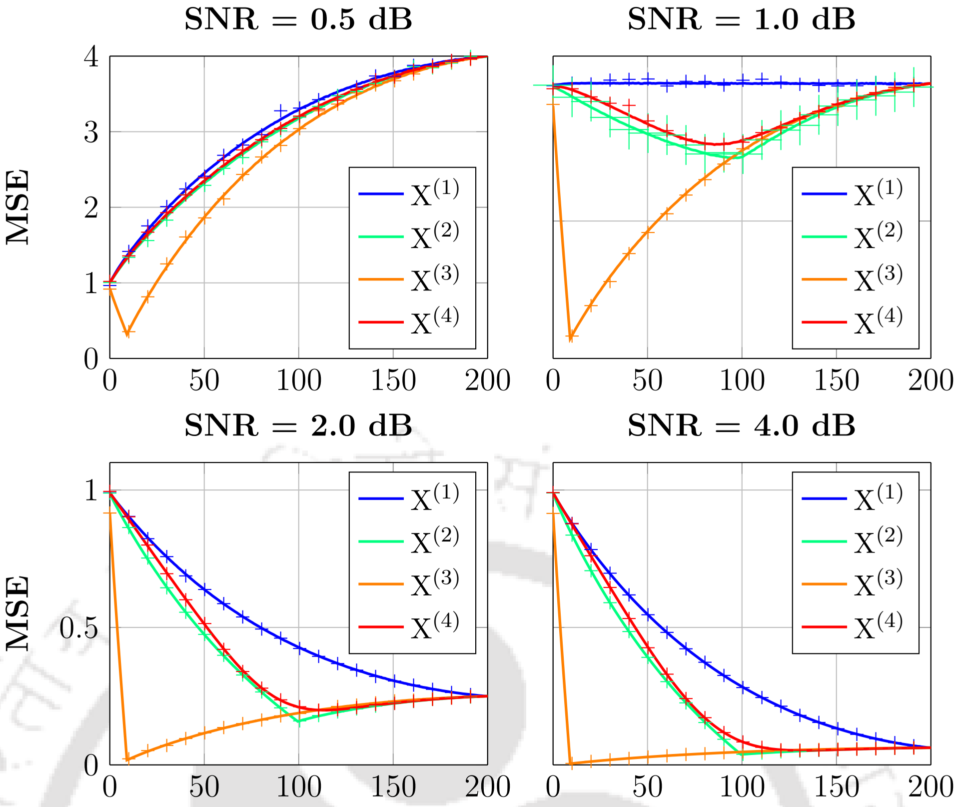


Figure 2.2: Comparison of SURE (tick marks) and Monte Carlo (solid lines) simulations for four different matrices under varying SNRs for tuning the parameter b.

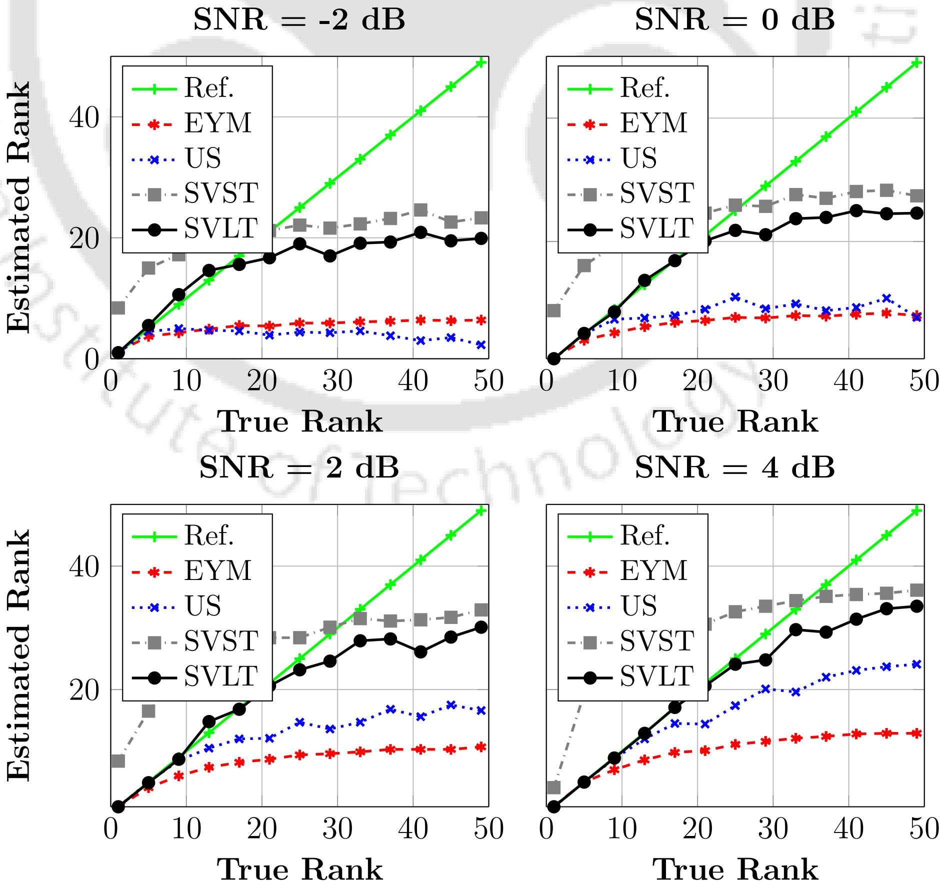


Figure 2.3: Comparison of the performance of the SVLT with that of the EYM, the US, and the SVST. The diagonal green line corresponds to the oracle estimator of the rank.

parameters are adaptively tuned using SURE. The plots of the estimated ranks are shown with the true rank in Figure 2.3. A better estimator should be aligned as much as possible with the true value represented by the green diagonal line in plots. Note that,the SVLT based rank estimator is better aligned with the true rank compared to the contrast methods. In case of 4 dB SNR, the SVLT is able to recover the rank up to 35, whereas the US method and the EYM method could recover the rank only up to 15 and 5, respectively. The SVST is noted to produce quite inferior rank estimate when the underlying true rank is low (i.e., less than 35). However, when the underlying rank is high, the estimate produced by the SVST is close to that of the SVLT as the former being the special case of the latter. Similar trends are noted for other three lower SNR cases.

2.3.3 Quantitative performance of the SVLT based rank estimator

In this experiment, we compare the quantitative performance of the SVLT based rank estimator with the existing methods. We consider, the inaccuracy in the rank estimation defined by

$$\text{Inaccuracy} = \left| 1 - \frac{\text{Estimated Rank}}{\text{True Rank}} \right| \tag{2.15}$$

as the quality measure. The inaccuracy plots for all the estimators, averaged over 50 different realizations, are shown in Figure 2.4. From these plots, following observations are made.

- For low-rank matrices, the US method and the proposed SVLT method perform better than the other two methods.
- In low noise cases, except EYM method, all the three methods are able to recover the true ranks of underlying signal matrices.
- For all the considered cases, the SVLT method consistently shows lower inaccuracy than all the three contrast methods.

2.3.4 Quantitative performance of the SVLT based matrix estimator

With the data set used in the above experiment, we intend to show the efficacy of the SVLT for matrix recovery. For the same, in Figure 2.5, we plot the MSE of the SVLT and the contrast methods for varying noise levels. From the figure, following salient inferences can be drawn.

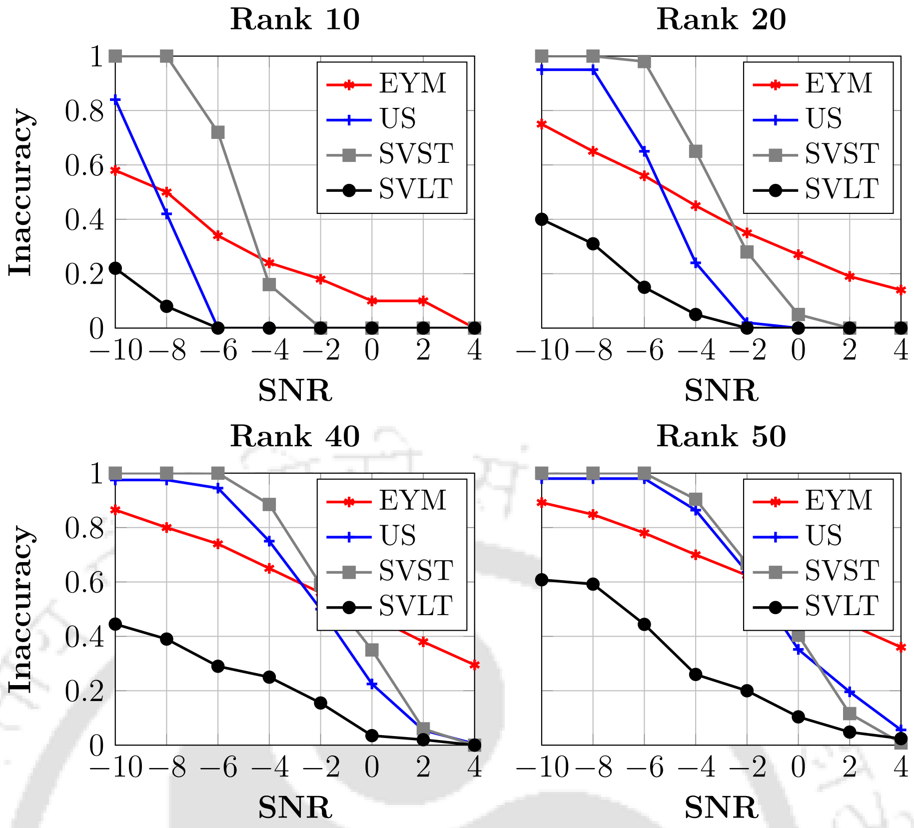


Figure 2.4: Comparison of the performance of the SVLT with that of the EYM, the US, and the SVST methods in terms of the inaccuracy for different SNR (in dB) levels and ranks.

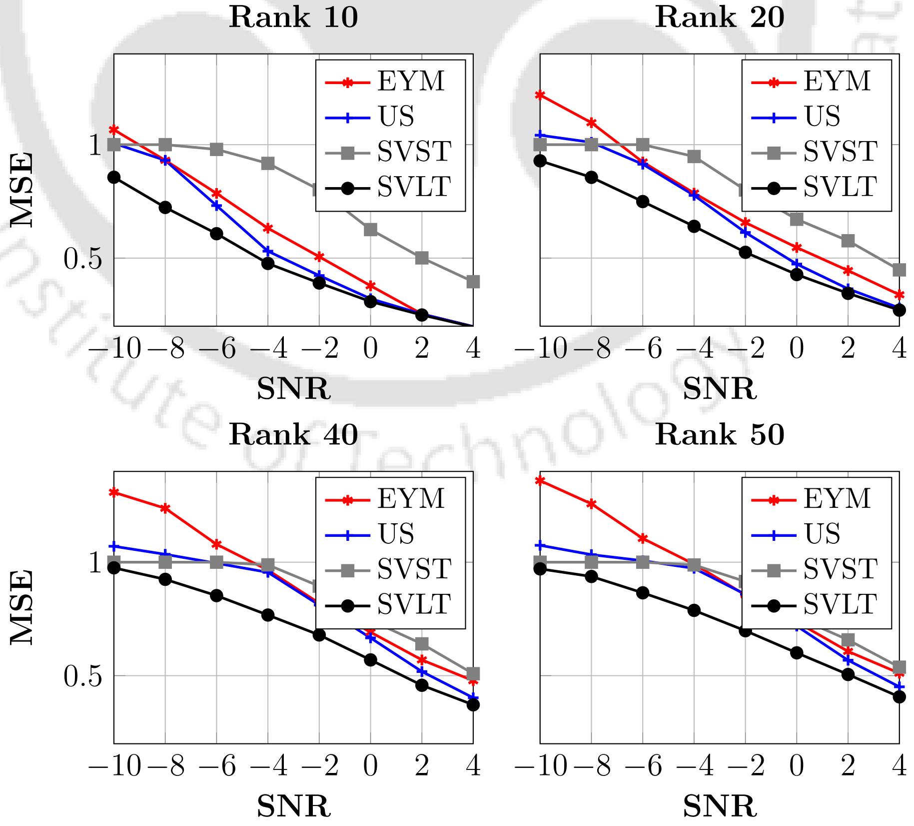


Figure 2.5: Comparison of the performance of the SVLT with that of the EYM, the US, and the SVST methods in terms of the MSE for different SNR (in dB) levels and ranks.

- In the SVST method, a sub-optimal trade-off between the preservation of the signal dominant SVs and the suppression of the noise-dominant SVs is made. As a result, a poor matrix recovery performance is noted in the SVST method, despite the efficient estimation of true rank.

- In the US method and in the EYM method, the noise components in signal dominant SVs are ignored resulting in a comparatively higher MSE than the SVLT method.

- The SVLT method consistently outperforms the contrast methods for low-rank matrix recovery in terms of the MSE.

In Table 2.1, the proposed method is compared against the contrast methods in terms of the order of computational complexity. It is to note that the computational complexity of the full SVD, i.e., $\mathcal{O}(m^2n + n^2m + n^3)$, is not included in the table as it is the common step across all the methods.

All reported experiments in this chapter are carried out in MATLAB R2012a running on Microsoft Window 7 machine having Intel® i5 (3.8 GHz) processor with 8 GB of RAM. The execution time for the different methods are also given in Table 2.1. Note that, the SVLT achieves a better estimation accuracy at the cost of some increase in the computational cost. This is because of the decoupling of the shrinkage and truncation achieved by using additional parameters in the proposed SVLT method.

Table 2.1: Comparison of the computational complexity/time of different methods for estimating the rank of a 200×400 matrix. The notation double-bar over the symbols denotes the cardinality of the parameter tuning-set.

	EYM	US	SVST	SVLT
Complexity Order	$\mathcal{O}(r)$	$\mathcal{O}(m^2r)$	$\mathcal{O}(m^2\bar{\bar{\mu}})$	$\mathcal{O}(m^2\bar{\bar{a}}\bar{\bar{b}}\bar{\bar{c}})$
Time (in sec.)	0.05	25	30	40

2.4 Summary

In this chapter, a novel matrix rank estimator by employing a logistic function based thresholding of SVs is proposed. The parameters of the shrinkage function are chosen using SURE. Different experiments were conducted to show the better rank estimation performance by the proposed method compared to the contrast methods. The experiments show the effectiveness of the proposed SVLT based rank estimation method at different noise levels and for the matrices of different ranks. The SVLT is also noted to possess better denoising capability than the traditional PC selection based methods. This is a hopeful indication of improving the denoising performance of existing signal denoising methods employing the shrinkage estimators that constitutes the course of further study presented in the next chapter.

3

SVLET: A Fast and Adaptive Shrinkage Method for Matrix Estimation

Contents

3.1 Shrinkage of SVs for matrix estimation 34
3.2 Proposed method 36
3.3 Results and discussion 41
3.4 Summary 45

In Chapter 2, we discussed various methods for estimating the rank of a matrix $\mathbf{X}$ from its noisy observation $\widetilde{\mathbf{X}}$. It was also argued that, given the estimated rank $\hat{r}$, one can easily obtain an estimate of underlying matrix $\mathbf{X}$ using (2.1). However, a close inspection of this estimator reveals that this is nothing but an $\hat{r}$-rank approximation of the noisy data matrix $\widetilde{\mathbf{X}}$ [25]. For the Gaussian noise model in (1.4), it turns out to be the r-rank maximum likelihood estimate (MLE) of $\mathbf{X}$. Certainly, one would quest for the better estimators of the signal matrix, possibly the nonlinear ones, than the low-rank approximation of the observation matrix [62,63]. Also, as it uses no additional structures on the matrix $\mathbf{X}$ other than the low-rankness, the better estimation methods exploiting the additional structures such as non-negativity [64, 65], sparsity [66], Hankel [67], and Toeplitz nature [68] also exist. This is why the SVLT not only performs as a better rank estimator but also as a better matrix estimator. It uses three parameters that are tuned using SURE. Similarly, the ATN also uses SURE to tune both the parameters. Tuning the parameters of these shrinkage estimators involves a considerable computational burden. It becomes a more severe problem when the number of parameters further increases. For example, the SVLT and the ATN are comparatively more complex than the SVST as they use more number of parameters. For this reason, they are unsuitable for many practical applications where the computational speed matters.

In this chapter, our goal is to speed up and automate the parameter selection process for denoising the observed data matrix. Specifically, we propose a linearization approach by means of which the grid search for parameter tuning is completely eliminated. The proposed approach is then tested for matrix denoising and shown to perform better than the existing estimators.

3.1 Shrinkage of SVs for matrix estimation

The rationale behind the matrix approximation by the truncation of SVs is the low rankness of the underlying signal matrix. The heuristic solution to this LRMA problem is given by the SVHT. As discussed in previous chapters, the SVHT has several disadvantages. In case of noisy observation, we know that the SVs of data matrix become positively biased. For this reason, the shrinkage of SVs can be expected to yield a better performance. The simplest shrinkage estimator obtained by solving the nuclear norm regularized problem is the SVST. However, the SVST not only thresholds the SVs smaller than μ but it also applies the shrinkage by the same amount. The selection of μ is crucial as it determines the extent of shrinkage applied and hence the trade-off between the bias and the variance. Higher values of μ result in smaller variance but introduce larger bias, whereas lower values of μ cause smaller bias at the cost of larger variance. One may naturally choose a value of μ that admits the minimum MSE. As $\mathbf{X}$ is unknown, the MSE is an unobservable quantity. In [18], the authors proposed to determine this parameter using SURE. For involving only a single parameter, the SVST approach has very low computational complexity. It suffers from two drawbacks. Firstly, for the shrinkage and thresholding being coupled, it is unable to provide optimal trade-off between the bias and the variance. Secondly, it shrinks all the dominant SVs equally despite the fact that each of them may not be equally important [69].

To deal with the first drawback, a logistic function based shrinkage that decouples the shrinkage and the thresholding by introducing additional parameters is proposed in Chapter 2. Towards addressing the second drawback, in [33], a non-linear estimator which applies less shrinkage on larger SVs compared to the smaller ones has been proposed. The authors, in [2], proposed an ATN estimator with two parameters. It deals with both the drawbacks of the SVST and provides more flexible shrinkage. The parameters of these estimators are often selected by minimizing SURE. On account of the inherent non-linearity, the methods employing SURE lack the closed form solution for the optimal value of the parameters. As a result, a grid search over fairly large search range is done for finding the optimal value of the parameters, thus limiting the use of such estimators in practical applications. Apart from these SURE-tuned matrix estimators, the RMT [29] provides an alternative by which the AMSE can be inferred without the knowledge of matrix $\mathbf{X}$. Following this, the recent works in [1, 19–21] tried to find the optimal shrinkage functions for matrix denoising but in an asymptotic case, where the size of the matrices grows infinitely keeping the rank and the aspect ratio fixed. In [20], a data driven estimator which dominates the EYM estimator is proposed. In [21], the authors showed that the optimal value of hard-threshold is a deterministic quantity equal to $(4/\sqrt{3})\sqrt{n}\tau$ where τ is the standard deviation of Gaussian entries in $\overline{\mathbf{X}}$. In [1], the same authors proposed a general framework to derive the optimal shrinkage function for a number of loss functions. However, the RMT has the limitation of being confined in asymptotic cases and also the RMT based estimators happen to suppress all the SVs below the threshold β_+. Departing from this large matrix low-rank setting, our focus is on estimating the signal matrix of finite size which is not necessarily the low-rank. Moreover, using the results from the RMT [24], we show that the proposed estimator, when forced to be a bulk-shrinker, maintains its asymptotic optimality in large matrix limits too.

In this chapter, we propose a linearly expandable singular value thresholding/shrinkage function which admits a closed form solution to corresponding SURE thereby circumventing the limitation of existing shrinkage functions outlined above and reducing the computational cost drastically. As comparing with the RMT based estimators, it achieves the similar performance in case of large matrix limits and outperforms in case of finite limits. It outperforms all the SURE based estimators in terms of estimation accuracy as well as in terms of computational cost.

3.2 Proposed method

3.2.1 Linearized shrinkage function

What hinders in obtaining the closed form optima of SURE is the nonlinearity present in the shrinkage functions employed in existing estimators. This non-linearity is necessary for optimal trade-off between bias and variance of estimators. However, it can be linearised so as to get the closed form minima of SURE. For this reason, we propose a linearly parameterized nonlinear shrinkage function as

$$f(\sigma) = \sum_{k=1}^{K} a_k \phi_k(\sigma) = \mathbf{\Phi}^{\mathrm{T}} \mathbf{a} \tag{3.1}$$

where $\mathbf{a} = [a_1, a_2, ..., a_K]^{\mathrm{T}}$ is the shrinkage parameter vector, $\mathbf{\Phi} = [\phi_1(\sigma), \phi_2(\sigma), ..., \phi_K(\sigma)]^{\mathrm{T}}$ is hitherto unspecified bases, and K is the order of linearization. Plugging this in (1.9) and using (1.23), the optimal value of parameter $\mathbf{a}$ of the proposed shrinkage function can be obtained by minimizing SURE in closed form as given below. At optima, we have

$$\begin{aligned} 0 &= \frac{1}{2}\frac{\partial}{\partial a_k} \mathrm{SURE}_{\mathrm{F}}(\widetilde{\mathbf{X}}) \\ &= \underbrace{\frac{1}{2}\frac{\partial}{\partial a_k} \|\widetilde{\mathbf{X}} - \mathrm{F}(\widetilde{\mathbf{X}})\|^2_{\mathrm{Frob}}}_{\mathrm{R}_{k,1}} + \sigma^2 \underbrace{\frac{\partial}{\partial a_k} \nabla(\mathrm{F}(\widetilde{\mathbf{X}}))}_{\mathrm{R}_{k,2}}. \end{aligned} \tag{3.2}$$

Solving for $\mathrm{R}_{k,1}$ and $\mathrm{R}_{k,2}$ as

$$\begin{aligned}\mathrm{R}_{k,1} &= \sum_{i=1}^{L} \frac{1}{2}\frac{\partial}{\partial a_k}\left(\tilde{\sigma}_i - \sum_{k=1}^{K} a_k \phi_k(\tilde{\sigma})\right)^2 && \text{(Using (3.1))}\\ &= \sum_{l=1}^{K}\sum_{i=1}^{L} \phi_k(\tilde{\sigma}_i)\phi_l(\tilde{\sigma}_i) a_l - \sum_{i=1}^{L} \tilde{\sigma}_i \phi_k(\tilde{\sigma}_i),\\ \mathrm{R}_{k,2} &= \sum_{i=1}^{L}\left(\phi_k'(\tilde{\sigma}_i) + |n-m|\frac{\phi_k(\tilde{\sigma}_i)}{\tilde{\sigma}_i} + 2\sum_{\substack{j=1,\\ j\neq i}}^{L} \frac{\tilde{\sigma}_i \phi_k(\tilde{\sigma}_i)}{\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2}\right)\end{aligned}$$

and collecting the terms in (3.2), we have

$$0 = \sum_{l=1}^{K} \underbrace{\sum_{i=1}^{L} \phi_k(\tilde{\sigma}_i)\phi_l(\tilde{\sigma}_i)}_{M_{k,l}}\; a_l - c_k \tag{3.3}$$

where

$$c_k = \sum_{i=1}^{L}\left(\tilde{\sigma}_i - \frac{|n-m|\tilde{\sigma}^2}{\tilde{\sigma}_i} - \sum_{j=1, j\neq i}^{L} \frac{2\,\tilde{\sigma}^2\tilde{\sigma}_i}{\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2}\right)\phi_k(\tilde{\sigma}_i) - \sum_{i=1}^{L} \sigma^2 \phi_k'(\tilde{\sigma}_i) \tag{3.4}$$

is free from any unknown parameter. In 3.4, $\phi_k'(\tilde{\sigma}_i)$ denotes the derivative of $\phi_k(\tilde{\sigma}_i)$ with respect to $\tilde{\sigma}_i$. Repeating the above differentiation for all $k \in \{1, 2, ..., K\}$ and collecting the corresponding equations, as in (3.3), in matrix form we have

$$\mathbf{0} = \mathbf{Ma} - \mathbf{c} \Longrightarrow \mathbf{a}_{\text{opt}} = \mathbf{M}^{-1}\mathbf{c}, \tag{3.5}$$

where

$$\mathbf{M} = \begin{bmatrix} \sum_{i=1}^{L} \phi_1(\tilde{\sigma}_i)\phi_1(\tilde{\sigma}_i) & \cdots & \sum_{i=1}^{L} \phi_1(\tilde{\sigma}_i)\phi_K(\tilde{\sigma}_i) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{L} \phi_K(\tilde{\sigma}_i)\phi_1(\tilde{\sigma}_i) & \cdots & \sum_{i=1}^{L} \phi_K(\tilde{\sigma}_i)\phi_K(\tilde{\sigma}_i) \end{bmatrix}$$

and $\mathbf{c} = [c_1, c_2, ..., c_K]^{\mathrm{T}}$. Thus, for a given K and $\phi_k(\cdot)$, finding the value of shrinkage parameter vector **a** has reduced to solving a system of linear equations. Note that,the order of linearization K can be arbitrarily increased, as long as the matrix **M** is invertible. This provides sufficient flexibility in the shape of shrinkage function. What remains still unspecified is the choice of basis function $\phi_k(\cdot)$. Authors in [70] applied similar shrinkage function on multiscale DWT coefficients and named it as the SURELET. On experimenting over several bases, they found the derivative of Gaussian (DOG) to be the best choice. Following that, in this work, we adopt the DOG as the basis for linearization of the unknown shrinkage function. The functional form of the DOG is

$$\phi_k(\sigma) = \sigma \exp\left\{-(k-1)\frac{\sigma^2}{2T^2}\right\} \tag{3.6}$$

where T is a constant that depends on the strength of noise, i.e., $T = C\sigma$ with $C > 0$ being a constant. The value of constant C determines a smooth transition between noise dominated and signal dominated SVs. It depends on the dynamic range of SVs of underlying noise free signal matrix and hence can be experimentally fixed for a data set. In Figure3.1(d), a typical example of the proposed shrinkage function with the DOG basis is compared with the few existing shrinkage functions. Note that, unlike the existing shrinkage functions, the proposed shrinkage function does not abruptly truncate the SVs, thus providing a better trade-off between the bias and the variance. For larger SVs, it is able to approximate the SVHT well. Hence, it introduces a lower bias while keeping all the advantages of the SVST intact in the case of lower SVs. As it does not necessarily truncate the smaller SVs to achieve the bias-variance trade-off, it can denoise the matrices of any rank.

We term the proposed estimator with such linearization of shrinkage function as the SVLET. Just as the SURELET shrinks the DWT coefficients, the proposed SVLET shrinks the SVs using the same principle. Further, we denote the optimal SVLET shrinkage function and the corresponding shrinkage estimator as $f^{\text{SVLET}}(\tilde{\sigma}_i)$ and $\widehat{\mathbf{X}}^{\text{SVLET}}_{\mathbf{a},T,K}$, respectively. Clearly, $\widehat{\mathbf{X}}^{\text{SVLET}}_{\mathbf{a},T,K} = \sum_{i=1}^{L} f^{\text{SVLET}}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}}$. Further, $\widehat{\mathbf{X}}^{\text{SVLET}}_{\beta_+} = \sum_{i=1}^{r^*} f^{\text{SVLET}}_{\beta_+}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}}$ denotes an estimator when we enforce the SVLET to be a bulk-shrinker of order r^*. Such an estimator is termed as the *bulk-shrinker generated by the SVLET* in the rest of the thesis.

3.2.2 Asymptotic behavior of the SVLET

In this subsection, we analyze the behavior of the SVLET in an asymptotic setting. In the following, we show that the bulk-shrinker generated by proposed SVLET boils down to the asymptotically optimal estimator (1.21). For that, we apply the asymptotic conditions ($\tau = 1/\sqrt{n}$, and $m/n \to \beta$, with $0 < \beta \leq 1$ as $m \to \infty$) on SURE in (1.23). Thus, the asymptotic SURE corresponding to the bulk-shrinker f_{β_+} can be given as

$$
\begin{aligned}
\mathrm{ASURE}_{f_{\beta_+}}(\widetilde{\mathbf{X}}) &= \lim_{m\to\infty} \mathrm{SURE}_{f_{\beta_+}}(\widetilde{\mathbf{X}}) \\
&= \lim_{m\to\infty} \left\{ -m + \|\widetilde{\mathbf{X}} - \widehat{\mathbf{X}}_{\beta_+}\|^2_{\mathrm{Frob}} + \frac{2}{n}\,\nabla(\widehat{\mathbf{X}}_{\beta_+}) \right\} \\
&= \lim_{m\to\infty} -m + \sum_{i=1}^{r^*} \left((\tilde{\sigma}_i - f_{\beta_+}(\tilde{\sigma}_i))^2 + \frac{2(1-\beta) f_{\beta_+}(\tilde{\sigma}_i)}{\tilde{\sigma}_i} + \lim_{m\to\infty} \frac{4\beta}{m} \sum_{j=1, i\neq j}^{r^*} \frac{\tilde{\sigma}_i f_{\beta_+}(\tilde{\sigma}_i)}{\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2} \right) \\
&= \lim_{m\to\infty} -m + \sum_{i=1}^{r^*} \left((\tilde{\sigma}_i - f_{\beta_+}(\tilde{\sigma}_i))^2 + \frac{2(1-\beta) f_{\beta_+}(\tilde{\sigma}_i)}{\tilde{\sigma}_i} + \frac{4\beta(\sigma_i^2+1) f_{\beta_+}(\tilde{\sigma}_i)}{\rho(\sigma_i)\sigma_i^2} \right).
\end{aligned}
\tag{3.7}
$$

(using (1.4.1))

Note that, in the above we used the fact that the term $\lim_{m\to\infty} \frac{f'_{\beta_+}(\tilde{\sigma}_i)}{n}$ vanishes as $n \to \beta^{-1}m$. The following theorem guarantees that, in an asymptotic setting, the bulk-shrinker generated by the SVLET is optimal as it achieves the true asymptotic mean square error and the resulting estimator exactly turns out to be the same as the optimal bulk-shrinker given in (1.21).

Theorem 3.2.1. *In the asymptotic setting, the bulk-shrinker generated by the SVLET achieves the true mean-squared error estimator and it boils down to an optimal bulk-shrinkage estimator. Specifically,*

$$
\min_{f_{\beta_+}^{\mathrm{SVLET}}} \mathrm{ASURE}_{f_{\beta_+}^{\mathrm{SVLET}}} \left(\widetilde{\mathbf{X}}\right) = \min_{f_{\beta_+}^{\mathrm{SVLET}}} \mathrm{AMSE}_{f_{\beta_+}^{\mathrm{SVLET}}} \left(\widetilde{\mathbf{X}}\right). \tag{3.8}
$$

Equivalently,

$$
\lim_{m\to\infty} f_{\beta_+}^{\mathrm{SVLET}} \left(\widetilde{\mathbf{X}}\right) = f_{\beta_+}^{*} \left(\widetilde{\mathbf{X}}\right). \tag{3.9}
$$

Proof. We first solve the left-hand-side of (3.8) for the minima of ASURE. At optima, we have

$$
\begin{aligned}
0 &= \frac{1}{2}\frac{\partial}{\partial a^{\infty}_{k,opt}} \mathrm{ASURE}_{f^{\mathrm{ASURE}}_{\beta_+}}(\widetilde{\mathbf{X}}) \\
0 &= \sum_{i=1}^{r^*} \left\{ \frac{1}{2}\frac{\partial}{\partial a^{\infty}_{k,opt}} \left(\tilde{\sigma}_i - f^{\mathrm{ASURE}}_{\beta_+}(\tilde{\sigma}_i)\right)^2 + \frac{(1-\beta)}{\tilde{\sigma}_i}\frac{\partial f^{\mathrm{ASURE}}_{\beta_+}(\tilde{\sigma}_i)}{\partial a^{\infty}_{k,opt}} + \frac{2\beta(\sigma_i^2+1)}{\rho(\sigma_i)\sigma_i^2}\frac{\partial f^{\mathrm{ASURE}}_{\beta_+}(\tilde{\sigma}_i)}{\partial a^{\infty}_{k,opt}} \right\} \\
0 &= \sum_{i=1}^{r^*} f^{\mathrm{ASURE}}_{\beta_+}(\tilde{\sigma}_i)\phi_k(\tilde{\sigma}_i) - \sum_{i=1}^{r^*} \left\{ \tilde{\sigma}_i - \frac{(1-\beta)}{\tilde{\sigma}_i} + \frac{2\beta(\sigma_i^2+1)}{\rho(\sigma_i)\sigma_i^2} \right\} \phi_k(\tilde{\sigma}_i).
\end{aligned}
$$

Using definition of the proposed SVLET in (3.1), with a restriction of being a bulk-shrinker of order r^* and using the fact from (1.16) that $\rho(\sigma_i) = \tilde{\sigma}_i$ for $\sigma_i > \beta^{1/4}$, we have

$$\underbrace{\sum_{l=1}^{K} \sum_{i=1}^{r^*} \phi_k(\tilde{\sigma}_i)\phi_l(\tilde{\sigma}_i)}_{M^{\infty}_{k,l}} \, a^{\infty}_{l,\text{opt}} = \sum_{i=1}^{r^*} \frac{\sigma_i^2(\tilde{\sigma}_i^2 - \beta - 1) - 2\beta}{\sigma_i^2 \tilde{\sigma}_i} \phi_k(\tilde{\sigma}_i).$$

Let $b_i = \tilde{\sigma}_i^2 - \beta - 1$, then from (1.17) $\sigma_i^2 = \frac{b_i \pm \sqrt{b_i^2 - 4\beta}}{2}$. Thus,

$$\sum_{l=1}^{K} M^{\infty}_{k,l} \, a^{\infty}_{l,\text{opt}} = \sum_{i=1}^{r^*} \frac{1}{\tilde{\sigma}_i} \sqrt{b_i^2 - 4\beta} \; \phi_k(\tilde{\sigma}_i).$$

From (1.21),

$$\sum_{l=1}^{K} M^{\infty}_{k,l} a^{\infty}_{l,\text{opt}} = \underbrace{\sum_{i=1}^{r^*} f^*_{\beta_+}(\tilde{\sigma}_i)\phi_k(\tilde{\sigma}_i)}_{c^{\infty}_k} . \tag{3.10}$$

Now, differentiating the ASURE with respect to $a^{\infty}_{k,\text{opt}}$ for all k and collecting the terms in matrix form yields

$$\mathbf{M}^{\infty}\mathbf{a}^{\infty}_{\text{opt}} = \mathbf{c}^{\infty} \quad \Longrightarrow \mathbf{a}^{\infty}_{\text{opt}} = \left(\mathbf{M}^{\infty}\right)^{-1}\mathbf{c}^{\infty} \tag{3.11}$$

where $\mathbf{c}^{\infty} = \left(\boldsymbol{\Phi}^{\infty}\right)^{\mathrm{T}} \mathbf{f}^*_{\beta_+}(\tilde{\sigma})$. On solving (3.8) from the right-hand-side, we have

$$\begin{aligned} 0 &= \frac{1}{2}\frac{\partial}{\partial a^{\infty}_{k,\text{opt}}} \text{AMSE}_{f^{\text{SVLET}}_{\beta_+}}\left(\widetilde{\mathbf{X}}\right) \\ &= \frac{1}{2}\frac{\partial}{\partial a^{\infty}_{k,\text{opt}}} \sum_{j=1}^{r^*} \left\| f^{\text{SVLET}}_{\beta_+}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\mathrm{T}} - \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}} \right\|^2_{\text{Frob}} \\ &= \sum_{i=1}^{r^*} f^{\text{SVLET}}_{\beta_+}(\tilde{\sigma}_i)\phi_k(\tilde{\sigma}_i) - \sum_{i=1}^{r^*} \sigma_i \langle u_i, \tilde{u}_j\rangle\langle v_i, \tilde{v}_j\rangle \phi_k(\tilde{\sigma}_i) \end{aligned} \tag{3.12}$$

$$= \sum_{l=1}^{K}\sum_{i=1}^{r^*} \phi_k(\tilde{\sigma}_i)\phi_l(\tilde{\sigma}_i) a^{\infty}_{l,\text{opt}} - \sum_{i=1}^{r^*} \sigma_i \theta_u(\sigma_i)\theta_v(\sigma_i)\phi_k(\tilde{\sigma}_i) \qquad \text{(Using Lemma (1.2.3))}$$

Using (1.21), we can write

$$\sum_{l=1}^{K}\sum_{i=1}^{r^*} \phi_k(\tilde{\sigma}_i)\phi_l(\tilde{\sigma}_i) a^{\infty}_{l,\text{opt}} = \sum_{i=1}^{r^*} f^*_{\beta_+}(\tilde{\sigma}_i)\phi_k(\tilde{\sigma}_i) \tag{3.13}$$

Now, differentiating the AMSE with respect to $a^{\infty}_{k,\text{opt}}$ for all k and collecting the terms in matrix form again gives (3.11), thus (3.8) is proved. With this result, to prove (3.9), let us find the optimal bulk-shrinker for the SVLET estimator as

$$\begin{aligned}\lim_{m\to\infty} f_{\beta_+}(\widetilde{\mathbf{X}}) &= \sum_{i=1}^{r^*} \phi_k^{\infty}(\tilde{\sigma}_i) a_{k,\text{opt}}^{\infty} \\ &= \mathbf{\Phi}^{\infty}\mathbf{a}_{\text{opt}}^{\infty} \\ &= \mathbf{\Phi}^{\infty}(\mathbf{M}^{\infty})^{-1}\mathbf{c}^{\infty} \\ &= \mathbf{\Phi}^{\infty}(\mathbf{M}^{\infty})^{-1}\mathbf{\Phi}^{\infty^{\mathrm{T}}}\mathbf{f}_{\beta_+}^{*} \\ &= \mathbf{f}_{\beta_+}^{*} \qquad \text{(Since, } \mathbf{\Phi}^{\infty}(\mathbf{M}^{\infty})^{-1}\mathbf{\Phi}^{\infty^{\mathrm{T}}} = \mathbf{I}_n\text{)}\end{aligned}$$

This completes the proof. ☐

In the following, we will empirically show the asymptotic optimality by observing the closeness of the SVLET to the OptShrink estimator.

3.3 Results and discussion

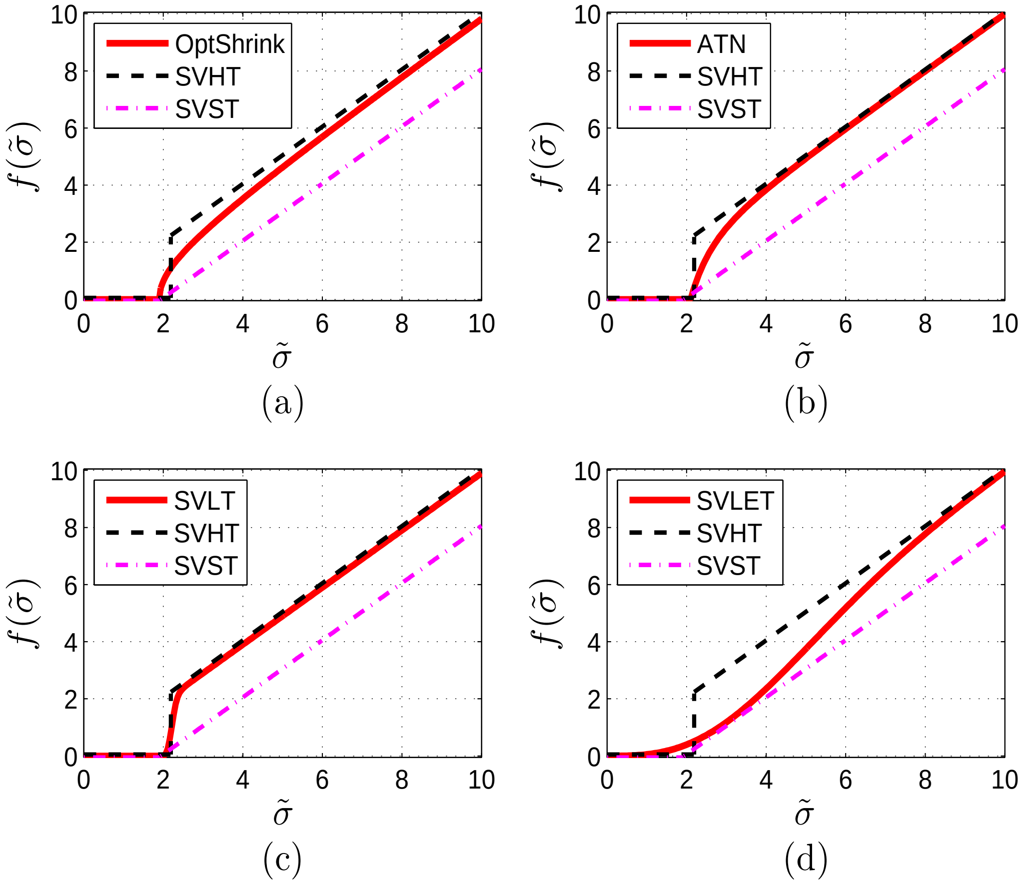
In this section, we empirically evaluate the performance of the proposed SVLET for matrix estimation through simulations. We compare its performance with some of the recently proposed shrinkage estimators. The empirical evaluation is primarily done on artificially generated matrices of different ranks. For that, an r-rank signal matrix is generated as $\mathbf{X} = \mathbf{L}\mathbf{R}^{\mathrm{T}}$ where $\mathbf{L}$ and $\mathbf{R}$ are, respectively, $m \times r$ and $n \times r$ random matrices whose entries are independently drawn from zero mean unit variance Gaussian distribution. To obtain the corresponding observation matrix $\widetilde{\mathbf{X}}$, we add zero mean white Gaussian noise (WGN) of known variances to each element of $\mathbf{X}$. Denoising is then performed on this observation matrix and the performance is measured in terms of the normalized MSE given as

$$\text{NMSE} = \frac{1}{P}\sum_{p=1}^{P} \frac{\|\widehat{\mathbf{X}}_{\text{F}}^{(p)} - \mathbf{X}^{(p)}\|_{\text{Frob}}^2}{\|\mathbf{X}^{(p)}\|_{\text{Frob}}^2} \tag{3.14}$$

where P is the number of realizations. All the performances reported are averaged over 10 noise realizations, i.e., $P = 10$.

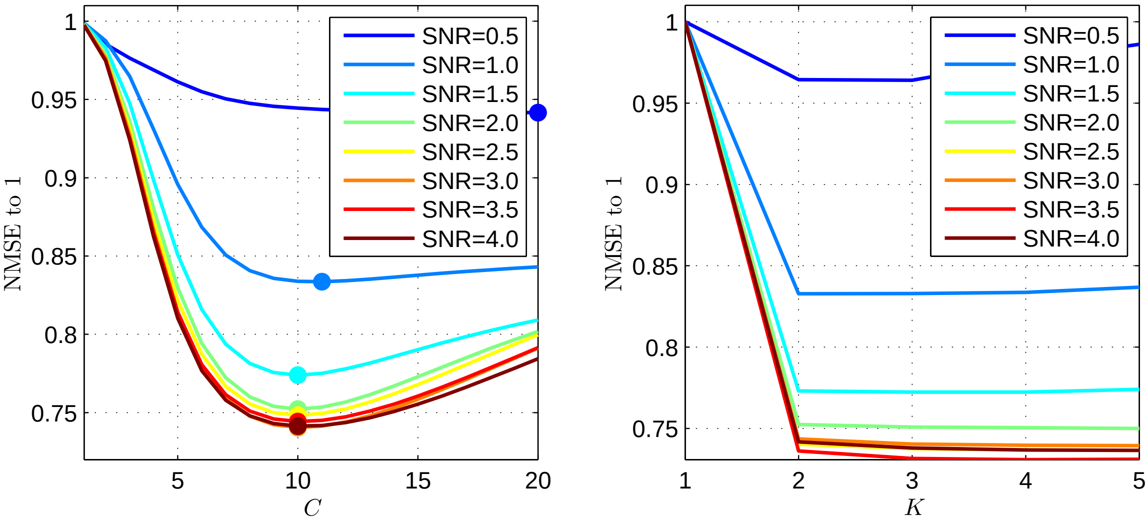
3.3.1 Selection of parameters

Before comparing the SVLET with existing methods, we examine the effects of its two fixed parameters C and K on denoising performance. For that, we generate the signal matrix of size 50×50 with $1 \leq r \leq 50$ as discussed above and add noise to it at various signal-to-noise ratio



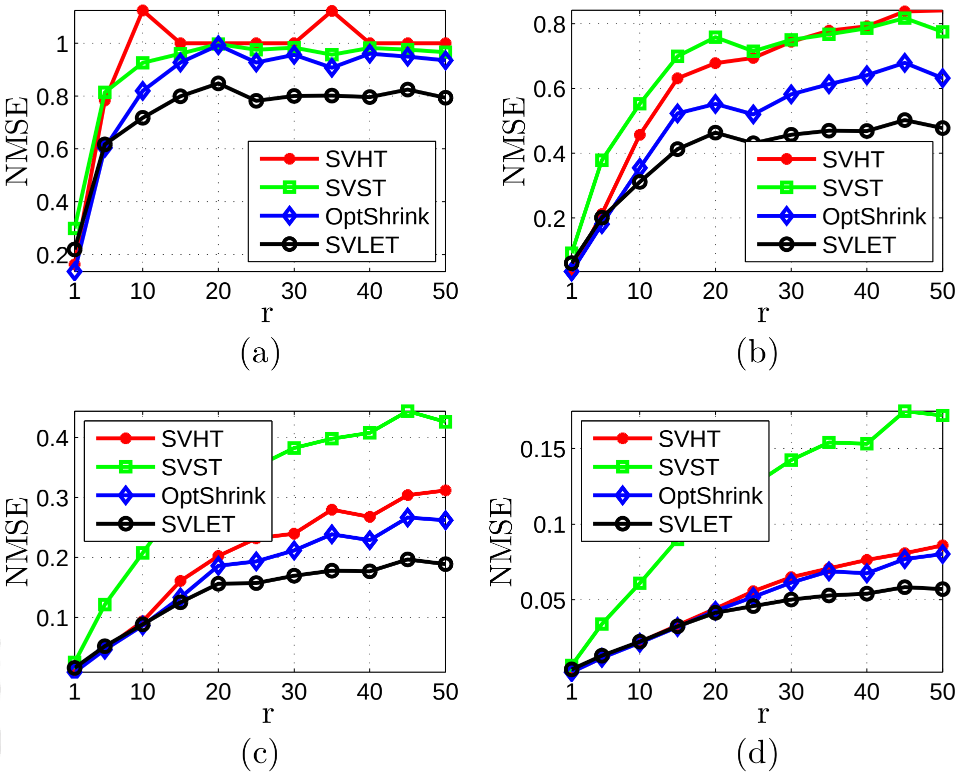
(a) (b) (c) (d)

Figure 3.1: Comparison of existing singular value shrinkage functions. A typical example of (a) asymptotically optimal bulk-shrinkage function [1], (b) ATN [2] ($\mu_1 = 2.2,\ \mu_2 = 5$), (c) SVLT [3] ($a = 0.8,\ b = 2.2,\ c = 0.1$), and (d) proposed SVLET shrinkage functions ($\mathbf{a} = [1,\ -1]^{\mathrm{T}},\ K = 2,\ T = 3\tau$) plotted with SVHT ($\mu = 2.18$) and SVST ($\mu = 1.98$). The values in the braces are the parameters of corresponding shrinkage function.



(a) (b)

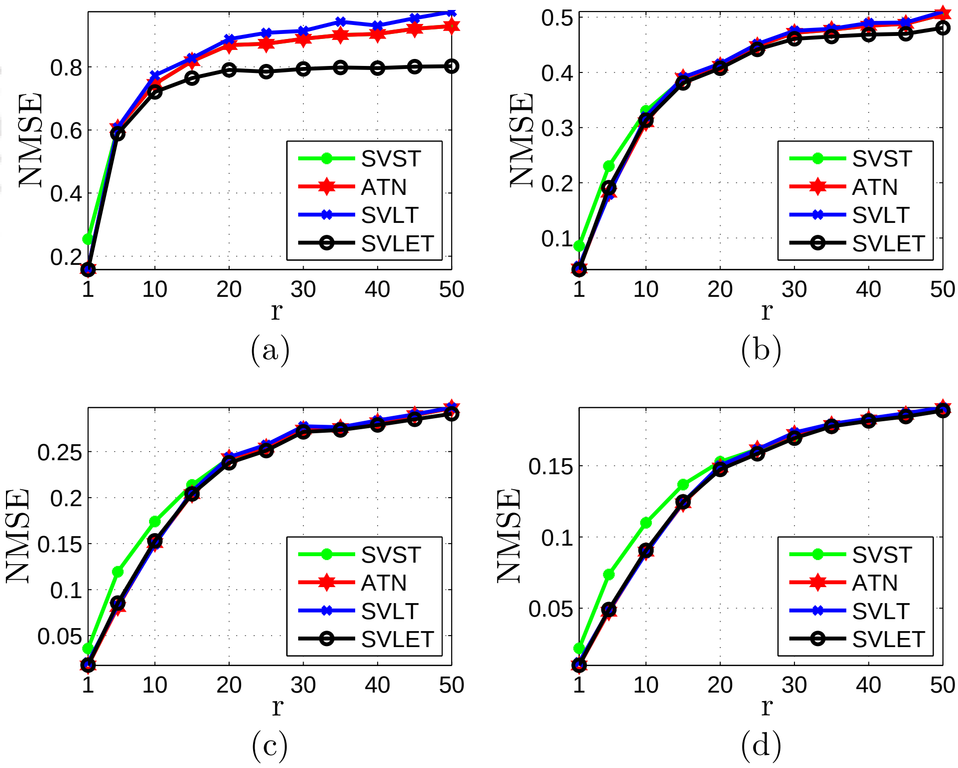
Figure 3.2: Effect of the parameters of the proposed SVLET shrinkage estimator on the denoising performance. Effect on the NMSE of varying (a) C and (b) K. In each case, the averaged NMSE computed over all possible ranks of the underlying clean matrix is reported.



(a) (b)

(c) (d)

Figure 3.3: Comparison of average denoising performance of the proposed SVLET estimator with the OptShrink and other estimators for different ranks of 50×50 matrix corrupted by AWGN at (a) SNR = 0.5, (b) SNR = 1.0, (c) SNR = 2.0, and (d) SNR = 4.0.



(a) (b)

(c) (d)

Figure 3.4: Comparison of average denoising performance of the proposed SVLET estimator with SURE based contrast estimators for different ranks of 50×50 matrix corrupted by AWGN at (a) SNR = 0.5, (b) SNR = 1.0, (c) SNR = 1.5, and (d) SNR = 2.0.

(SNR) ranging from 0.5-4.0 to obtain the observation matrix. The SVLET with different values of C and K is employed for denoising. In Figure 3.2(a), the NMSE, averaged over all ranks and $1 \leq K \leq 5$, is plotted against different values of C. In Figure 3.2(b), the normalized MSE, averaged over all ranks and $1 \leq C \leq 20$, is plotted for $1 \leq K \leq 5$. We see that increasing the value of K beyond 2 does not help much. The rationale behind this behavior is the sufficient adaptation of shrinkage function in capturing the singular value profile of the original signal matrix by the parameter $\mathbf{a}$ only. Hence, the choice $C = 10$ and $K = 2$ are fair enough to be fixed in the rest of the experiments in this chapter.

3.3.2 Matrix denoising performance

In the following, the denoising performance of the SVLET is compared with state-of-the-art methods using the artificially generated data set as discussed above.

3.3.2.1 Comparison with RMT based methods

The denoising performance of the SVLET is compared with that of the asymptotic estimators proposed in [1, 21] for SNR = 0.5, 1.0, 2.0, and 4.0. As the RMT based methods are calibrated for standard deviation $\tau = 1/\sqrt{n}$ of noise in asymptotic setting, we adjust the scale of given observed matrix before applying these estimator. Thus, the final estimate is obtained as $\sqrt{n}\tau\widehat{\mathbf{X}}^*_{\beta_+}(\widetilde{\mathbf{X}}/(\sqrt{n}\tau))$. For the SVHT, the optimal value of hard-threshold is $\mu^* = 4/\sqrt{3}$ and for the SVST, we take $\mu^* = (1 + \sqrt{\beta})$ as suggested in [1].

Figure 3.3 compares the NMSE of the proposed SVLET with that of the existing estimators. Form Figure 3.3(a)-(d), it can be seen that for comparatively lower ranks, where the asymptotic conditions hold, the performance of the SVLET and the OptShrink coincide, as expected from Theorem 3.2.1. As the rank increases the SVLET dominates over the RMT-based shrinkage estimators, at all SNR levels. At low SNR, the SVLET performs better than the SVHT and the SVST, even at very low ranks. This is because, the SVLET does not make any assumption about the rank of underlying signal matrix. The overall improved performance by the SVLET can be attributed to its better adaptability to the SV structure of the underlying signal matrix and corrupting noise.

3.3.2.2 Comparison with SURE based methods

Figure 3.4 compares the SVLET with SURE based state-of-the-art shrinkage estimators for SNR = 0.5, 1.0, 1.5, and 2.0. For the SVST, the optimal soft-threshold value μ is searched at 100 equally spaced grid points. For the ATN and the SVLT, the optimal values of parameters are obtained by minimizing SURE, in the range given in Table 3.1. From Figure 3.4(a), it is clear that the SVLET outperforms all the methods, especially when the rank of underlying signal matrix is large and the noise level is high. Figures 3.4(b), 3.4(c) and 3.4(d) show that the NMSE of the SVLET is coinciding with the ATN and the SVLT. This shows that it as good as both of these in low noise cases.

All the methods are implemented in MATLAB 8.3 (R2014a), running on Intel® i3 processor with Window 8.1 (8 GB RAM) machine. Although CPU time is not a standard measure of computational cost, we use it to roughly compare the execution-time of the SVLET with other methods. In this setup, the average execution-time for denoising a 50 × 50 matrix is compared in Table 3.1 for different methods. Note that, as the number of parameters increases the execution-time also increases. The SVLET is the fastest one as it solves the system of linear equation for determining the optimal value of data dependent parameters.

Table 3.1: Comparison of the execution time of different methods for estimating a 50 × 50 matrix.

Method	Parameter (Value/Range)	Complexity Order	Time (in sec.)
SVST	$\mu \in (0, 0.5\tilde{\sigma}_1)$	$\mathcal{O}(m^2\bar{\bar{\mu}})$	0.10
ATN	$\mu \in (0, 0.5\tilde{\sigma}_1)$, $\gamma \in [1, 20]$	$\mathcal{O}(m^2\bar{\bar{\mu}}\bar{\bar{\gamma}})$	0.35
SVLT	$p_1 = 100$, $p_2 \in [1, n]$, $p_3 \in (0, 0.5\tilde{\sigma}_1)$	$\mathcal{O}(m^2 n\bar{\bar{p}}_3)$	0.50
SVLET	$\mathbf{a}_{opt}$, $C = 10$, $K = 2$	$\mathcal{O}(m^2K^3)$	0.04

3.4 Summary

In this chapter, we proposed a matrix estimation method by the adaptive shrinkage of its SVs and showed its efficacy in recovering matrices of arbitrary ranks from their noisy observations. Unlike the existing methods, which select the optimal shrinkage parameters by exhaustive search, the proposed method finds one shot solution for the optimal parameters, thereby reducing the computational complexity drastically. The main contribution of the current work is in the proposition of a fast and adaptive shrinkage of the SVs based on the linear expansion of conventional SV thresholding. We have theoretically shown that the proposed method is optimal in an asymptotic framework, if restricted to be a bulk-shrinker, as it achieves the true mean-square error and results in asymptotically optimal estimators obtained using the tools from random matrix theory. Its denoising performance is as good as the asymptotically optimal estimators when the rank of underlying signal matrix is very low, whereas it performs significantly better than the state-of-the-art methods if the rank of underlying signal matrix and the level of noise are both high. Applications of the proposed shrinkage estimator in real-time data such as in the denoising of natural image and other signals are the course of investigation in the following chapters.

4

Application to ECG Denoising

Contents

4.1 Non-local filtering for ECG denoising 50

4.2 Proposed ECG denoising method 51

4.3 Selection of parameters 56

4.4 Experimental results 59

4.5 Discussion 65

4.6 Summary 69

ECG signal is one of the most popular diagnostic means which provides an electrical picture of the heart and information about different pathological conditions. These signals originate from the heart and pass through the tissues with different characteristics to reach up to the several recording leads placed on the skin of the subject. Owing to path deformities and external electrical disturbances, the recorded ECG signals become noisy. The disturbances that affect the recorded ECG signals include baseline wander, power-line interference, and muscle noise. In recent years, the biotelemetry has become a dominant means of monitoring the cardiac condition of ambulatory patients [71, 72]. Also, to detect arrhythmias and cardiac abnormalities, wireless ambulatory ECG recording is now routinely used [73]. In such cases, the ECG data are sent to the remote locations where it can be better analyzed by clinicians/specialists. In the process of transmission, the ECG signals get corrupted by the underlying channel (wired or wireless) noise. Thus, for correct diagnosis and/or assessment of the cardiac disease, the removal of channel noise from such ambulatory ECG recording is considered essential.

A number of methods have been proposed for ECG signal denoising in the literature [74–79]. Among those methods, the ones based on the hard-thresholding of discrete wavelet transform (DWT) coefficients [75, 79, 80] and the methods based on empirical mode decomposition (EMD) [74, 78] have emerged as two popular groups. In the former group, the denoising is achieved by the truncation of the lower magnitude DWT coefficients of the noisy ECG signal followed by the inverse DWT. In the latter group, the first few intrinsic mode functions (IMFs) of the noisy ECG signal are discarded to achieve the denoising as they account for the high frequency variations (i.e., noise) present in the signal. However, this process is reported to distorts the QRS complexes. In [81], the authors attempted to preserve the portions of first few IMFs those correspond to the QRS complexes by means of a *Tukey* window. In [82], a hybrid EMD-wavelet method that combines the windowed EMD with wavelet soft-thresholding has been proposed to further improve the denoising performance.

The nonlocal means (NLM) method [83] is a very successful image denoising method. The NLM method assumes that the underlying noise-free image has several pixels with similar neighborhood. It has been applied for ECG signal denoising in [84] and is shown to outperform the hybrid EMD-wavelet method for a number of ECG signals. In normal cases, the ECG signals are almost structurally repetitive and thus their samples possess the redundancy. In one-dimensional (1D) NLM denoising proposed in [84], the estimates of the underlying clean signal samples are obtained by the weighted averaging of the samples having similar neighborhoods. The applied weighting is proportional to the similarity in the neighborhood and independent of the temporal location of the samples. As a result, the samples with quite similar neighborhoods are given higher weights whereas lower weights are assigned to the samples with dissimilar neighborhoods. Thus, the method directly exploits the nonlocal similarity present in the signal.

The NLM method uses a sample-based approach in which each sample is estimated independently. In other words, the estimate of a sample at one location does not contribute to the estimation of other samples even if those are in close proximity. The nonlinear filtering methods like the shrinkage of the DWT coefficients instead rely on the inherent sparsity of the clean signal in the transform domain. However, the DWT shrinkage based methods could not exploit the nonlocal redundancy present in the signal. On combining the transform based approach and the block-based NLM approach, their relative advantages can be exploited. The similar idea has already been explored for image denoising [85–87] but is yet to be explored for 1D biomedical signals like ECG, EEG, etc. In this chapter, we propose eight ECG denoising methods which exploit the local as well as the nonlocal similarity in the signal. In the proposed methods, the similar blocks of ECG samples are estimated in a collaborative manner. Specifically, for a given reference block of noisy ECG signal, similar blocks are collected from the same signal. These are then grouped together as the columns to form a matrix called *similarity data matrix* (SDM) [88]. Having formed such matrices corresponding to each of the reference blocks, these are then subjected to the matrix/image denoising methods for removing the noise. These denoising methods can be divided into two categories. The method in the first category denoises these SDMs using hard-thresholding of the DWT coefficients, thus exploiting the sparsity of the ECG in nonlocal wavelet transform (NLWT) domain. The methods in the second category make use of the spectral estimators discussed in previous chapters to denoise the matrices. Hence, the noise suppression is achieved by exploiting the sparsity of the ECG signals in nonlocal data-dependent transform domain. With highly overlapped blocks grouped together and for allowing a block to appear in multiple groups, several estimates for a sample are obtained. The final estimate of the ECG signal is found by averaging these estimates; thus further reduction in the additive noise is achieved. For experimental evaluation, the ECG data from the MIT-BIH Arrythmia corpus and the PTB Diagnostic corpus are used. The proposed approaches are noted to yield better ECG denoising performances when contrasted with the state-of-the-art method at various noise levels.

4.1 Non-local filtering for ECG denoising

ECG signal denoising addresses the recovery of the ECG signal from its noisy observation given the partial knowledge about the corrupting noise. For simplicity, the additive white Gaussian noise (AWGN) is commonly used to model the noise in ECG. The observed ECG signal follows the data model given as

$$\tilde{y}[t] = y[t] + \bar{y}[t], \quad t = 1, 2, \ldots, N \tag{4.1}$$

where $y[t]$ is the sample of underlying noise-free ECG and $\bar{y}(t)$ is the zero mean WGN sample with variance τ^2. Natural signals, including the ECG, exhibit correlation among their samples and this permits their sparse representation in a suitable domain. On the contrary, WGN samples are uncorrelated and do not yield a sparse representation in that domain. Hence, on applying the sparsifying transforms on the noisy signals the underlying noise-free signal can be recovered. Also, the higher is the correlation, the better is the sparsity of a signal, and hence the better is the removal of noise. Traditionally, the correlation present in the signal is exploited for denoising by averaging the neighboring samples. The NLM approach extended these concepts beyond the neighbourhood support of the sample of interest, and therefore received a lot of attention in signal restoration tasks.

In NLM method, the estimate of the tth sample is found by weighted averaging of the samples with similar local neighborhood within a search window $S(t)$. As defined in [84], the estimated sample is given as

$$\hat{y}[t] = \frac{1}{Z(t)} \sum_{q \in S(t)} w(t,q)\tilde{y}[q] \tag{4.2}$$

where $w(t,q)$ is the weight of the qth neighboring sample of the tth sample and $Z(t) =$

$\sum_q w(t,q)$. Motivated by [89], in [84] the weight $w(t,q)$ is computed as

$$\begin{aligned} w(t,q) &= \exp\left\{-\frac{\sum_{l\in\Delta}(\tilde{y}[t+l]-\tilde{y}[t+l])^2}{2L_\Delta\nu^2}\right\} \\ &= \exp\left\{-\frac{d^2(\tilde{y}[t],\tilde{y}[q])}{2L_\Delta\nu^2}\right\} \end{aligned} \tag{4.3}$$

where ν is the bandwidth parameter, Δ represents the neighborhood containing L_Δ samples surrounding the sample of interest. The term $d^2(\cdot,\cdot)$ denotes the squared Euclidean distance between the set of chosen samples as the measure of similarity of their neighborhoods. The same procedure is repeated for each sample and finally we get the denoised signal.

In the NLM method, the samples of the clean ECG signal with similar neighborhoods are assumed to be quite close in value. On averaging the samples from similar neighbourhood, the signal can be denoised if the noise is zero mean IID. The assumption that the samples being averaged are of closely similar values in the clean signal may not be reasonable for the QRS complex feature present in the ECG signals. Further, the similar samples of the underlying noise-free signal, those should have been assigned higher weights, are assigned lower weights due to their dissimilar neighborhoods in the noisy signal. Both of these artifacts may lead to erroneous estimate of underlying noise-free ECG signal. Also, in the NLM method, the estimate of a sample does not contribute in the estimation of even the nearby samples. In the following, we discuss how the proposed methods address these problems.

4.2 Proposed ECG denoising method

The proposed ECG denoising approach differs from the existing NLM method in two aspects: the block-based processing and the transform domain collaborative filtering. In the block-based processing, a block of samples is estimated instead of the individual samples. As the successive blocks are kept overlapping, multiple estimates are produced for a sample. These estimates are then averaged to find the final estimate. This averaging adds an extra layer of smoothing which helps in addressing the problem of inappropriate weights assigned to the samples as mentioned above in the context of the NLM method. Further, the intra-block correlations are not explicitly used in the NLM method. In the proposed methods, the shrinkage is applied to the transform coefficients of the SDMs which consist of the similar ECG blocks in their columns. In fact, the columns of the SDMs are the blocks centered at the samples those are weighted averaged for estimating a sample at the center of reference block in the NLM method. However, in the shrinkage of the transform coefficients based noise removal methods, many averaging modes in addition to the the weighted averaging are inherently considered. For example, in case of the Haar wavelet, the approximation coefficient accounts for the weighted average whereas the detail coefficients contain the high frequency information. Thus, it has been one of the most simple yet effective denoising methods in image processing that exploits the inter-row and the inter-column correlations. Hence, we intend to use the shrinkage based estimation for ECG signal denoising. Further, the choice of transforms and the way their coefficients need to be shrunk for better denoising of ECG signals are the course of our extensive study in this chapter. The proposed 2D-filtering approach to ECG denoising consists of three steps: the *SDM extraction*, the *SDM denoising*, and the *aggregation*. In the first step, the SDM corresponding to each of the reference blocks are formed by searching the similar blocks in a given search neighborhood. In the second step, each of these SDMs is denoised by the shrinkage of its transform coefficients (or SVs in the case of learned basis transform). In the last step, the blocks from these denoised SDMs are then returned to their original location yielding the final denoised signal. The overall block diagram of the proposed methods is shown in Figure 4.1. Further details about the steps of the proposed methods are discussed below.

4.2.1 SDM extraction

The SDM extraction step starts with selecting all the possible reference blocks of size m samples. As the denoised image is to be obtained by tiling the denoised version of these blocks, an overlap between the successive reference blocks is maintained to avoid any blocking artifact. Let us parameterize this overlap by κ where $1 \leq \kappa < m$. For an ECG signal of N samples, the total number of SDMs turns out to be $1 + \lfloor \frac{N-m}{\kappa} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function. Further, the similar blocks within a search window of size M centered at the center of the reference blocks are found and then these blocks are arranged as columns of the corresponding SDM. Thus, for each reference block we get an SDM with similar columns.

Consider a block of noisy ECG signal $\tilde{\mathbf{y}}_t = [\tilde{y}[t-(m-1)/2] \ldots \tilde{y}[t] \ldots \tilde{y}[t+(m-1)/2]]^{\mathrm{T}}$ centered at the tth sample. Let us denote a set containing the center indices of blocks similar

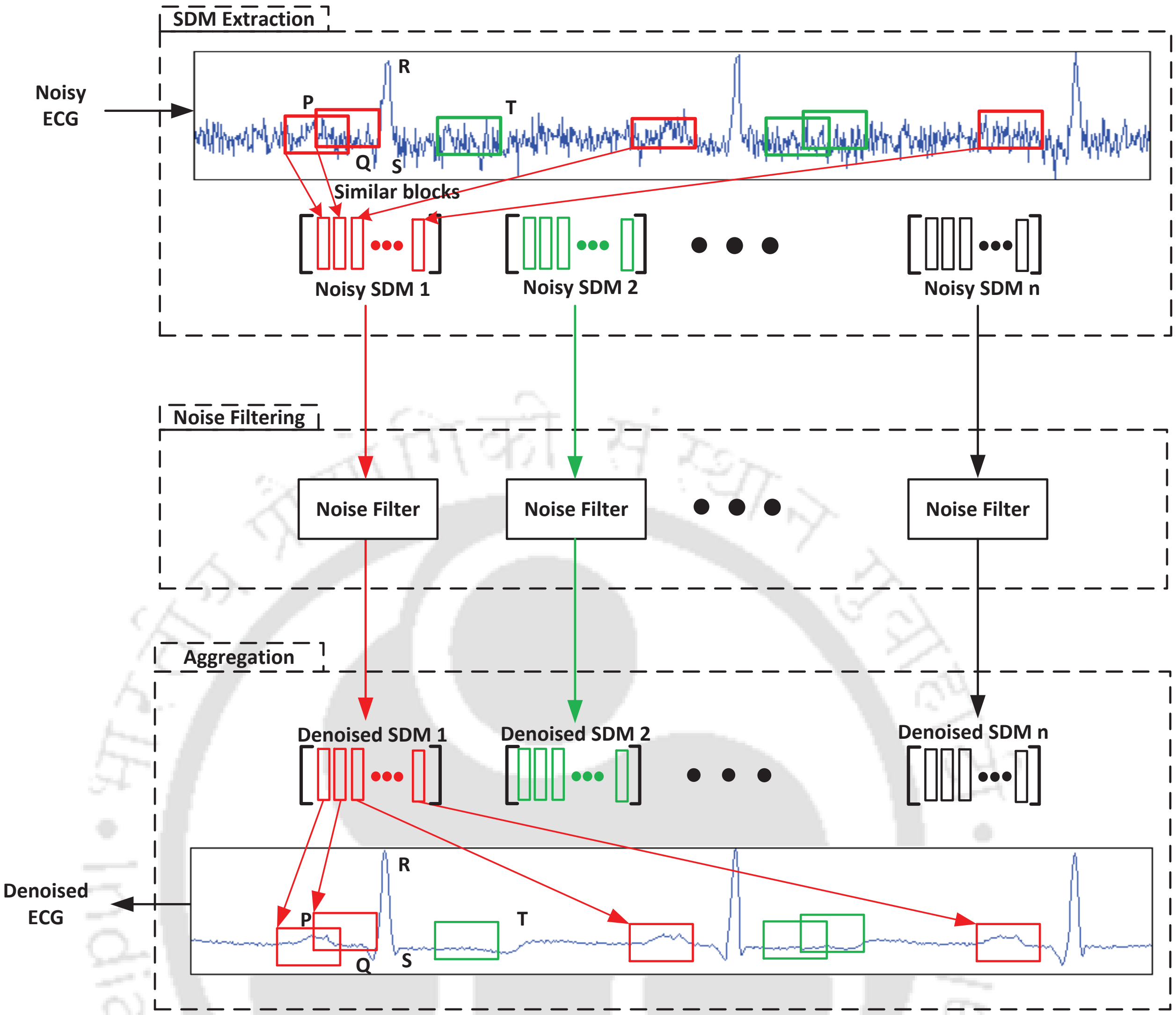
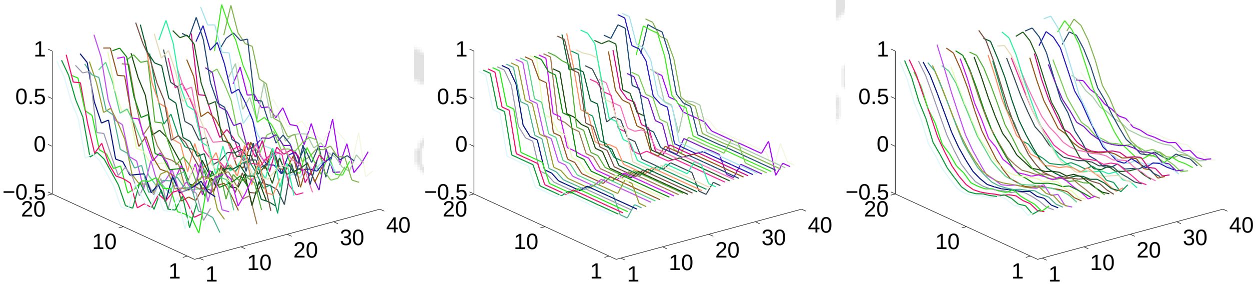


Figure 4.1: Block diagram of the proposed NLWT method. The blocks of samples that are marked in same colors are similar and grouped in an SDM.



(a) Noisy SDM (b) Denoised by NLWT (c) Denoised by SVLET

Figure 4.2: Illustration of an SDM extracted from (a) a noisy ECG signal, and its denoised versions (b) estimated by the NLWT, and (c) estimated by the SVLET.

to the $\tilde{\mathbf{y}}_t$ by $\Omega_t = \{q \in \mathbb{Z}^+ \mid d(\tilde{\mathbf{y}}_t, \tilde{\mathbf{y}}_q) \leq \tau\epsilon\}$, where τ is the standard deviation of noise and ϵ is a predetermined block-matching threshold. Further, assume a binary operator $\mathcal{R}_t$ that forms an SDM denoted by $\widetilde{\mathbf{X}}_t$ containing the blocks in its columns centered at the indices pointed by the Ω_t. Thus, $\widetilde{\mathbf{X}}_t = \mathcal{R}_t\tilde{\mathbf{y}} \in \mathbb{R}^{m\times|\Omega_t|}$ where $|\Omega_t|$ denotes the cardinality of Ω_t. The inverse of this operation that returns the extracted blocks to the location where they were extracted from is denoted by $\mathcal{R}_t^*$ such that $\mathcal{R}_t^*\mathcal{R}_t\tilde{\mathbf{y}}_t = \tilde{\mathbf{y}}_t$.

The purpose of this step is to find the blocks such that their underlying noise-free counterparts are similar. As we do not have the direct access to the clean signal, a lighter smoothing can be done before finding the similarity. However, such a pre-filtering involves extra computational cost. In this work, we rely on the Euclidean distance between noisy vectors as a measure of similarity.

4.2.2 Noise filtering

Each of these SDMs can be expressed as

$$\widetilde{\mathbf{X}}_t = \mathbf{X}_t + \overline{\mathbf{X}}_t \tag{4.4}$$

where $\mathbf{X}_t$ and $\overline{\mathbf{X}}_t$ are corresponding noise-free and noise matrices. Thus, the problem of noise suppression from 1D ECG signal is transformed into several smaller matrix denoising problems. Note that,each of the SDMs, being the collection of similar blocks, forms a highly correlated matrix. For illustrating this fact, an example SDM extracted from a noisy ECG signal is shown in Figure 4.2(a). To exploit this correlation for denoising purpose, in the following, we employ different methods including those discussed in the earlier chapters.

4.2.2.1 Denoising by thresholding of 2D DWT coefficients

The redundancy present in the SDMs can be exploited by a suitable sparsifying transform. The two dimensional DWT (2D DWT) is one of the potential candidates for this purpose. For this, each of the noisy SDMs is transformed into the DWT domain and then the transform coefficients are hard-thresholded with a predetermined threshold parameter. The corresponding denoised SDMs are then obtained by taking the inverse transform of their thresholded coefficients. The SDM denoising procedure can be mathematically expressed as

$$\widehat{\mathbf{X}}_t^{\text{NLWT}} = \mathcal{W}^{-1}\Big(\mathcal{H}(\mathcal{W}(\widetilde{\mathbf{X}}_t), h_t)\Big) \qquad \forall\ t \tag{4.5}$$

where $\mathcal{W}(\cdot)$ is the 2D DWT operator, $\mathcal{W}^{-1}(\cdot)$ is the inverse 2D DWT operator, and $\mathcal{H}(\cdot, h_t)$ is hard-thresholding operator defined as

$$\mathcal{H}(h, h_t) = \begin{cases} h, & \text{if} \quad |h| \geq h_t \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

where h_t is the threshold parameter corresponding to $\widetilde{\mathbf{X}}_t$. The hard-thresholding of the 2D DWT transform coefficients of the noisy SDMs exploits both, inter-row (local) similarity and inter-column (nonlocal) redundancy. On comparing Figures 4.2(a) and 4.2(b), the structural similarity in the blocks of SDM can be seen well preserved with the proposed 2D DWT coefficient shrinkage. The denoising method, employing such filtering of SDMs is referred to as the NLWT in the rest of the thesis.

4.2.2.2 Denoising by singular values shrinkage estimators

In the NLWT, the noise suppression is done by the hard-thresholding of 2D DWT coefficients. The DWT being a fixed bases transform is not an optimal sparsifying transform. To circumvent this issue, one can use the learned bases, possibly the redundant/overcomplete, transforms for each SDMs. However, learning the bases for each SDM is cumbersome task. The Karhunen-Loeve transform (KLT) is well known option in such cases. The KLT bases are, however, nothing but the principle components and hence hard-thresholding of KLT coefficients is the SVHT discussed in Chapter 2. Here, we take the generalized path of the spectral shrinkage estimators discussed earlier. For that, each of the SDMs is factorized using the SVD and denoising is performed by any of the SVSEs such as the EYM, the US, the SVST, the SVLT, the ATN, the SVLET, and the OptShrink estimators. The resulting denoising estimators are named as per the name of these spectral estimators for the ease of reference. Specifically, the denoised SDM using these estimators can be expressed as

$$\widehat{\mathbf{X}}_t^{\text{EYM}} = \sum_{i=1}^{\hat{r}^{\text{EYM}}} \tilde{\sigma}_i \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\text{T}} \tag{4.7}$$

$$\widehat{\mathbf{X}}_t^{\text{US}} = \sum_{i=1}^{\hat{r}^{\text{US}}} \tilde{\sigma}_i \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\text{T}} \tag{4.8}$$

$$\widehat{\mathbf{X}}_t^{\text{SVST}} = \sum_{i=1}^{L} f_{\mu}^{\text{SVST}}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}} \tag{4.9}$$

$$\widehat{\mathbf{X}}_t^{\text{SVLT}} = \sum_{i=1}^{L} f_{\boldsymbol{\xi}}^{\text{SVLT}}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}} \tag{4.10}$$

$$\widehat{\mathbf{X}}_t^{\text{ATN}} = \sum_{i=1}^{L} f_{\mu_1,\mu_2}^{\text{ATN}}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}} \tag{4.11}$$

$$\widehat{\mathbf{X}}_t^{\text{SVLET}} = \sum_{i=1}^{L} f_{T,K}^{\text{SVLET}}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}} \tag{4.12}$$

$$\widehat{\mathbf{X}}_t^{\text{OptShrink}} = \sum_{i=1}^{L} f_{\beta_+}^{*}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}} \tag{4.13}$$

where $\widetilde{\mathbf{X}}_t = \sum_{i=1}^{L} \tilde{\sigma}_i\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\text{T}}$ and $L = \min(m, |\Omega_t|)$.

At the end of this step, we get a denoised version of each SDM which need to be aggregated to obtain the final estimate of the ECG signal. In Figure 4.2(b) and (c), we show the denoised versions of a noisy SDM by the NLWT and the SVLET methods. It can be noted that the denoised SDM by the SVLET method is smoother than that of the NLWT method.

4.2.3 Aggregation

This is the final step in which the estimated blocks in the columns of $\widehat{\mathbf{X}}_t$ are sent to the temporal locations where they were extracted from as illustrated in Figure 4.1. Note that, a sample might be present in more than one block from one or more SDMs resulting in its multiple estimates. The weighted average of these multiple estimates is taken as the final estimate of that sample. Thus, the estimated signal vector can be given as

$$\hat{\mathbf{y}}^{\text{F}} = \sum_{\forall t} \frac{\mathcal{R}_t^*\widehat{\mathbf{X}}_t^{\text{SVSE}}}{\mathcal{R}_t^*\mathcal{R}_t} \tag{4.14}$$

where $\text{F} \in \{\text{NLWT}, \text{EYM}, \text{US}, \text{SVST}, \text{SVLT}, \text{ATN}, \text{SVLET}, \text{OptShrink}\}$.

4.3 Selection of parameters

The denoising performance of the proposed methods is sensitive to the choice of parameters. Those are divided into two categories based on whether they are common to all methods or method specific. The common parameters are the block size, the search-window size, and the overlap between successive reference blocks. The block-matching threshold, the hard-threshold

Table 4.1: Different parameters used in the proposed NLWT algorithm, their tuning ranges, and the optimal values.

Parameters	Notation	Range for tuning	Optimal value
Block-size	m	$0.01f_s - 0.1f_s$	20
Search-window size	M	3-5 Heart beats	2000
Block-matching threshold	ϵ	$1-5$ % of $2m$	1.2
Shrinkage threshold	δ_{2D}	± 25 % of $2\sqrt{\log(m\|\Omega_t\|)}$	3.0
Mother wavelet	-	db1-45, sym2-20, coif1-5	db2

for 2D DWT, and the weight of aggregation are the parameters specific to the NLWT method. Similarly, the SDM-size (number of similar blocks in an SDM) in all the SVSE methods, the slope parameter in the SVLT, and the order of linearization as well as the transition parameter in the SVLET are the method specific parameters. All of these parameters need to be properly selected for an optimal performance. The set of values of these parameters that produce the best results are generally chosen. However, while searching for the optimal parameter value, their physical significance has to be considered. In the following, we discuss how the various parameters of the proposed methods are fixed.

4.3.1 Block size and overlap

The block size m denotes the number of samples in a block. It should be large enough to capture the feature of interest in the signal. However, a very large value of m would create bigger blocks with reduced chances of match being found within a search window. A very small m would result in a large number of blocks which in turn lead to erroneous grouping of the blocks as well as increased overall complexity. We believe that the QRS complexes as a whole have lower chance of finding a good match in the signal than their smaller portions. Usually, the duration of the QRS complexes, in a normal sinus rhythm, lies between $0.06 - 0.1$ second. Based on that, for a signal sampled at a rate of f_s Hz, the appropriate block size is searched within the range of $0.01f_s - 0.1f_s$. Also, the overlap parameter κ depends on the block size. For a trade-off between performance and complexity, the overlap between successive reference blocks is kept as approximately 50 % in our experiments, i.e., $\kappa = (m-1)/2$.

4.3.2 Search window

The most trivial way is to search for the similar blocks over the entire length of signal but to reduce the complexity it is restricted within a window of length M samples. For a smaller value of M only the local-search is done while a larger value of M leads to the increased complexity. Hence, a moderate value of M is chosen so as to include multiple heartbeats allowing multiple QRS complexes, P-waves, and T-waves with potentially similar characteristics.

4.3.3 Block matching threshold and SDM size

The matching threshold is one of the most sensitive parameters of the proposed method as it controls the number of similar columns in an SDM (i.e., its size). It should be properly selected so that several blocks with similar features can be incorporated in an SDM. For a smaller values of ϵ no similarity for a reference block is found in the signal, leading to a single column SDM which makes the whole effort futile. A larger value of ϵ allows the dissimilar blocks being grouped in the SDMs and would lead to smoothing of the diagnostic features. For the ECG signal normalized to ±1 mV, the range distance between two blocks of size m can not exceed $2m$. In the proposed method, the value of this parameter is heuristically searched in range 0 - $2m$. In SVSE, this parameter is loosen so as to keep the size of all the SDMs greater than or equal to the block size, i.e., $|\Omega_t| \geq m$. This is done to meet the requirement of the spectral estimators used to denoise the SDMs in the proposed SVSE methods. However, such a setting has adverse effect on denoising performance in case of the proposed NLWT as it uses the fixed DWT bases and non-adaptive thresholding for the denoising of SDMs.

4.3.3.1 Hard-threshold for NLWT

The shrinkage threshold h_t is a parameter which directly affects the denoising performance of the NLWT. It should be proportional to the noise standard deviation τ so that the amount of shrinkage can be decided according to the strength of noise present in the signal. We set $h_t = \delta_t \tau$, where constant $\delta_t > 0$ controls the smoothing. It can be obtained by the *VisuShrink* soft-thresholding formula [90, 91] given as $\delta_t = \sqrt{2 \log m|\Omega_t|}$. In practice, we keep $\delta_t = \delta_{2D}$ and experimentally chose a single best performing value of δ_{2D} for all SDM instead of tuning it for each of them, separately.

4.3.4 Order of linearization and transition parameter of the SVLET

The integer parameter K determines the order of linearization of the SVLET that controls the flexibility of the shrinkage function. Increasing the value of K increases the flexibility but at the cost of increased complexity in the shape of shrinkage function. As a result, the corresponding SURE needs a larger number of similar columns in the SDMs to learn the shrinkage parameter vector **a** (refer to (3.1)) to adapt the SVLET well to the data. Increasing the range of K is prohibitive, thus, it is searched in the range $1 \leq K \leq 5$. The value of parameter $T = C\tau$ with $C > 0$ controls the transition region between the SVs to be preserved and the SVs to be truncated and hence is important for denoising. Like the synthetic example in Chapter 3, the optimal value of the parameter C can also be experimentally determined for ECG signal denoising.

4.4 Experimental results

The performance of the proposed method is evaluated on the MIT-BIH arrhythmia database [92] and the PTB diagnostic database [93] taken from the Physionet [94] data bank. The former database is used to compare the performance with the existing techniques in general whereas the latter is used to verify the performance of the proposed method on ECG signals with known pathology. Some of the channels of these raw ECG signals in these databases contain different kinds of noise such as baseline wander, 50 Hz power line, and other high frequency noises. For quantitative evaluation, the denoising methods need the noise-free signal. This is why the following preprocessing on the raw signals is done to obtain the test signals for our experiments. At first, the baseline wander noise is removed by subtracting the lower frequency components from the raw signal. Then, the notch filter is applied to remove the 50 Hz power-line noise. At last, a finite impulse response low-pass filter (FIR-LPF) is used to remove any high frequency noise.

4.4.1 Performance measures

To be consistent with the works reported in [84], we also chose following performance measures for the quantitative assessment.

Signal-to-noise-ratio improvement (SNR_{imp}):

$$\text{SNR}_{\text{imp}} = 10 \log \frac{\sum_{t=1}^{N} (\tilde{y}[t] - y[t])^2}{\sum_{t=1}^{N} (\hat{y}^{\text{F}}[t] - y[t])^2} \tag{4.15}$$

Mean square error (MSE):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}^{\text{F}}[i] - y[i])^2 \tag{4.16}$$

Percent root mean square difference (PRD):

$$\text{PRD} = 100 \sqrt{\frac{\text{MSE}}{\frac{1}{N} \sum_{i=1}^{N} y^2[i]}} \tag{4.17}$$

In the above equations, $\hat{y}[t]$ is the observed noisy ECG signal and $y[t]$ is corresponding noise-free ECG signal. Note that,a better denoising method has a higher SNR_{imp}, a lower MSE, and a lower PRD.

4.4.2 Evaluation on MIT-BIH Physionet database

The web-source[1] contains the ECG signals described by three categories of files: a binary annotation file (.atr), a binary data file (.dat), and a header file (.hea). The header file contains the information about recording environments, such as sampling frequency, patient history, and the format of the signal in the corresponding data file. This data base contains a number of single-channel ECG signal recording digitized at 360 samples per second per channels. The bit resolutions of the stored digitized signals are 11-bits over 10 millivolts (mV). However, we have normalized the signals to maximum amplitude of 1 mV for convenience. To evaluate the denoising methods, the noisy ECG signals are artificially generated by adding the WGN of chosen signal-to-noise ratio (SNR) levels to the preprocessed test signals taken from this database. These noisy ECG signals are then denoised by different methods assuming the standard-deviation τ of the noise to be knownn. The different tunable parameters involved in the proposed methods are tuned on the signal number 100. It is a normal sinus rhythm and

---

[1] https://www.physionet.org/physiobank/database/mitdb/

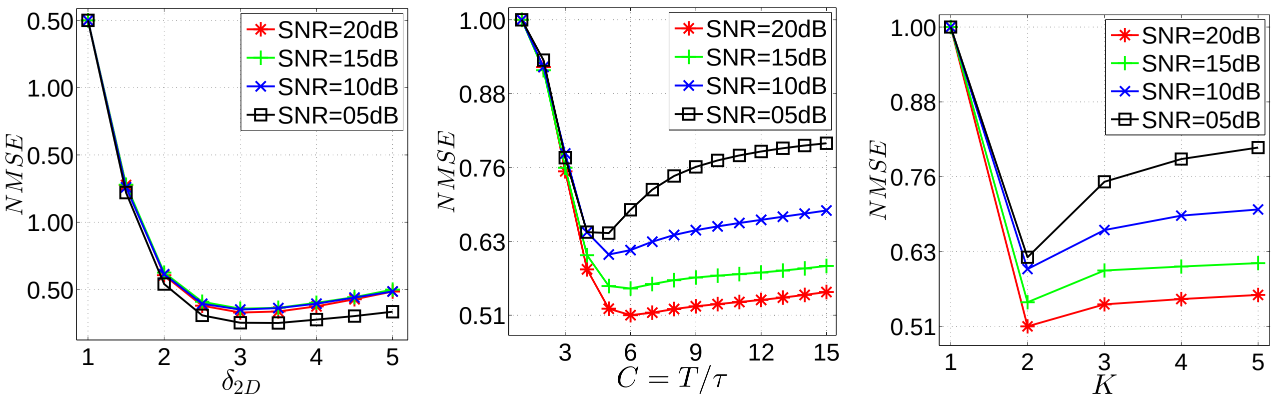
was also chosen to tune the parameters of NLM method in [84]. Tuning is done by varying the parameters within the suggested range given in the third column of Table 4.1 to minimize the MSE and their optimal values are put in the fourth column. For instance, the optimal value of parameter m is found to be 20. For a given m, M is varied from 500 to 5000 samples and $M = 2000$ is found to be the best choice. Similarly, the block matching threshold ϵ is found to be 1.5. In case of the NLWT, the different mother wavelet families such as Daubechies (db), Symlets (sym), and Coiflets (coif) have been tried. On an average, the Haar wavelet (db2) is found to the best performing. With these parameter values, the denoising parameter δ_{2D} is experimentally selected by varying it about $2\sqrt{\log m^2} = 3.23$. Figure 4.3(a) shows the sensitivity of denoising performance, in terms of the MSE, to this parameter. We see that $\delta_{2D} = 3$ can be taken as the optimal value.

Similarly, the fixed parameters of the SVSE methods are also tuned on the same test signal. For this, the size of the SDM is kept equal to the block-size while setting the block-matching threshold ϵ to a sufficiently large value. The parameters C and K of the SVLET is determined experimentally. Figures 4.3(b) and 4.3(c) show the plot of the normalized MSE for the different values C and K, respectively. Clearly, $C = 5$ and $K = 2$ can be selected as the best performing values. Rest of the parameters are kept same as in the NLWT. Though, the parameters are tuned by minimizing the MSE, we observed that on an average the parameters selected as above give the minimum PRD and the maximum $\mathrm{SNR_{imp}}$ for the chosen data set.

To evaluate the denoising performance with the parameters selected above, we have chosen different realizations of the noise at different SNR levels. The denoising is then performed on the noisy ECG by the proposed and a few chosen contrast methods. The performances, in terms of above metrics, averaged over different realizations are compared. Here, only the performance on the signals numbered as 100, 103, 104, 105, 106, 115, and 215 is reported. The performances of the different methods on these signals, at 20 dB noise level, are given in Figure 4.4.

The performances averaged over the different signals of the chosen data set, at different noise levels are shown in Figure 4.5.

Note that,the existing NLM method, which is considered to be the state-of-the-art ECG denoising method, is outperformed by the proposed methods in all three performance metrics and at all SNR levels. The NLWT method performs consistently better than the NLM. It also



(a) (b) (c)

Figure 4.3: Illustration of parameter tuning in the NLWT and in the SVLET. The plots showing the variation of normalized MSE (NMSE) at different noise levels with respect to parameters (a) δ_{2D}, (b) C, and (c) K.

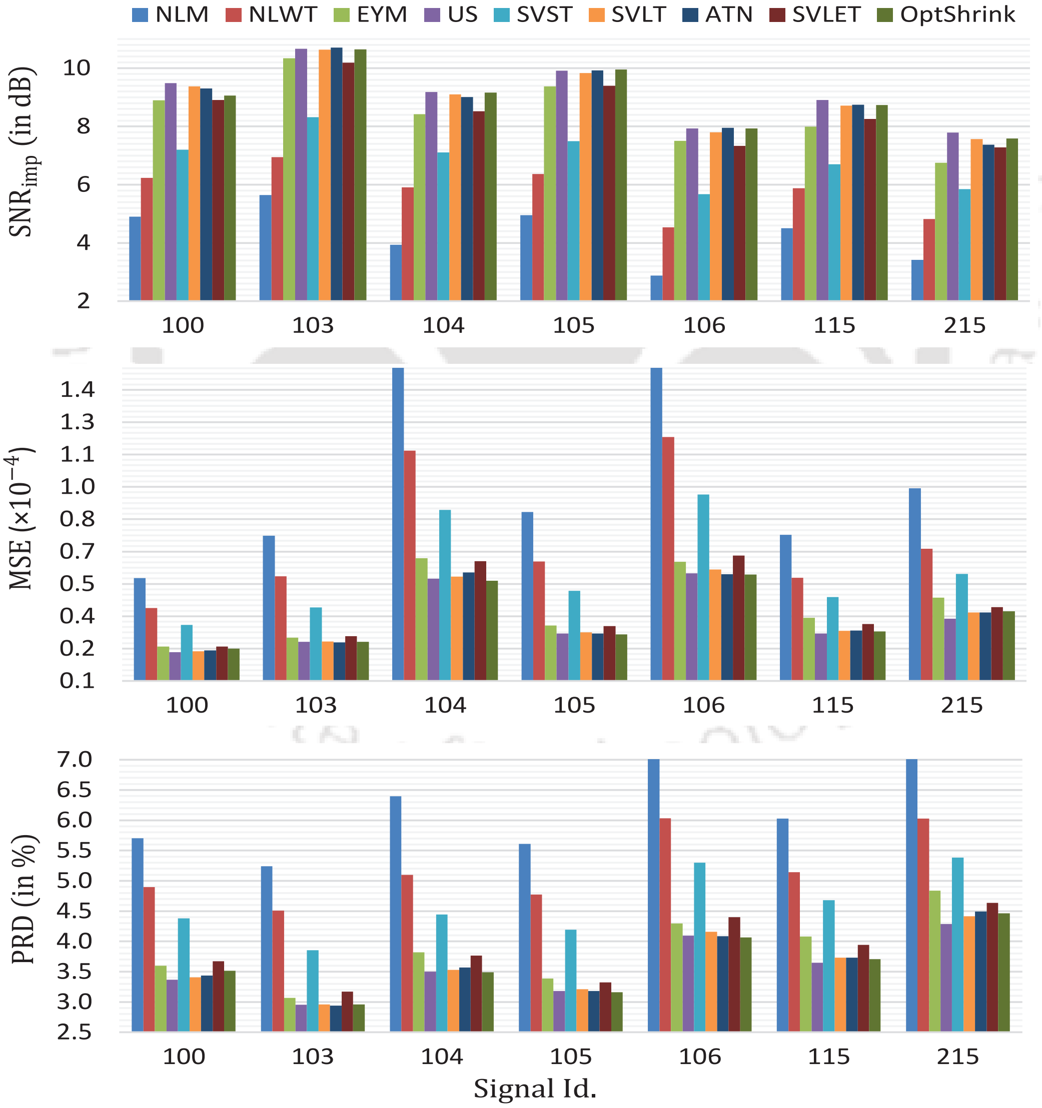
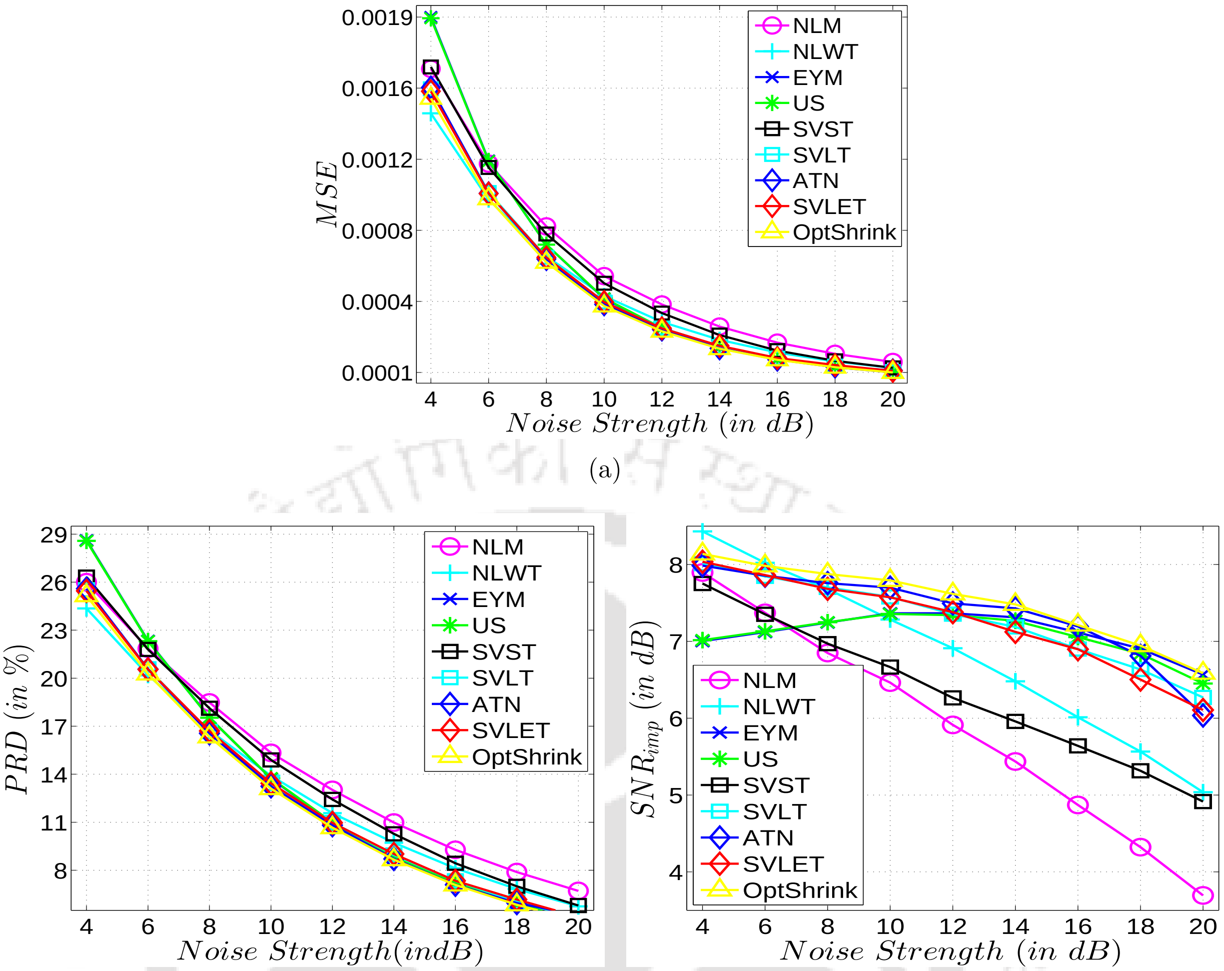


Figure 4.4: Comparison of the denoising performances of the proposed methods along with those of a few existing methods at 20dB noise level in terms of (a) SNR improvement, (b) mean squares error, and (c) PRD measures. In each case, the performance averaged over all the 12-leads of the ECG signal chosen from MIT-BIH database, is reported.



(a)

(b)

(c)

Figure 4.5: The impact of noise level on the denoising performances of the proposed methods along with those of a few existing methods measured separately in terms of (a) MSE, (b) PRD, and (c) SNR improvement.

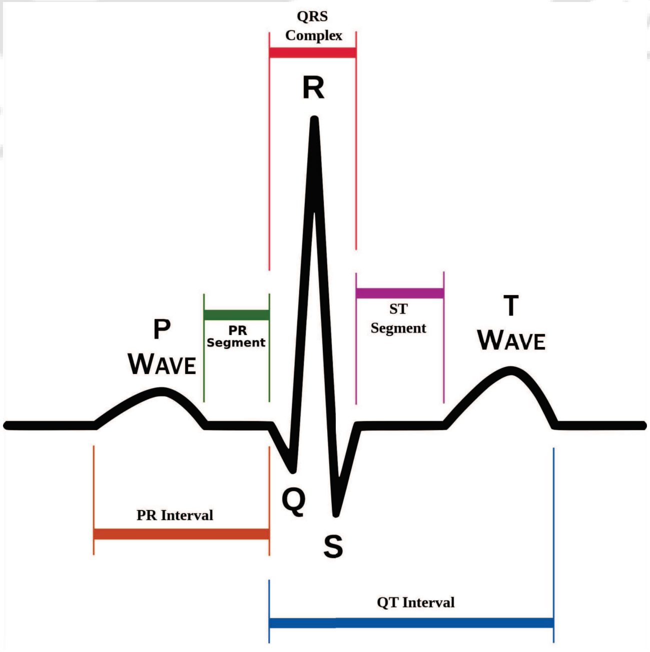


Figure 4.6: Illustration of a typical ECG signal showing its different features. Courtesy: www.wikidata.org

performs better than the SVSE methods in high noise cases, however it is inferior in case of low-noise. Among the SVSE methods, the SVLET and the OptShrink methods are quite close and they are the best performing methods.

The biomedical signals contain a number of diagnostic features which have to be visually inspected by the medical practitioners. A typical ECG signal, as shown in Figure 4.6, have various attributes that consist diagnostically relevant features. Hence, it is of prime importance to preserve those features. The important diagnostic features of an ECG signals are P-waves, QRS complexes, T-waves, U-waves, and their intervals. In denoising, these features usually undergo smoothing, thus affecting the diagnostic accuracy. Figure 4.7 shows the denoising performance of the NLM and the proposed methods for a signal number 100 at 10 dB noise level. On comparing with the noise-free signal, it can be seen that the signal denoised by the all the proposed methods better resembles the clean signal, than the noisy signal itself.

4.4.3 Evaluation on PTB Diagnostic database

This database contains the multi-channel ECG recorded on different patients with the known pathology. Each record in this database contains 15 simultaneously measured signals (12 conventional and 3 frank leads). From this data set, signals recorded on patients suffering from six different diseases are chosen. For each disease two signals are considered for the performance evaluation. To show the effectiveness of the proposed methods on these ECG signals, denoising is performed on all the leads of the chosen signals. As the sampling frequency of signals in this database is 1000 Hz, that differs from the MIT-BIH database, the parameters are accordingly adjusted. The block-size $m = 40$, the search-window size $M = 4000$, and matching threshold $\epsilon = 1.8$ are found to be optimal for this data set. The rest of the parameters are kept fixed as in the previous case. The parameters for the NLM method is also optimized in a similar way for fair comparison. For the different ECG signals with known pathology corrupted with 20 dB noise, the quantitative denoising performances are shown in Figure 4.8. The pathologies of the chosen signals are given in Table 4.2. The performance reported are averaged over all the 12 leads of a record. Note that, the NLWT method has consistently outperformed the NLM method for all the pathological ECG signals considered. However, the OptShrink performs better than the contrast methods and the proposed SVLET being quite close to it. For

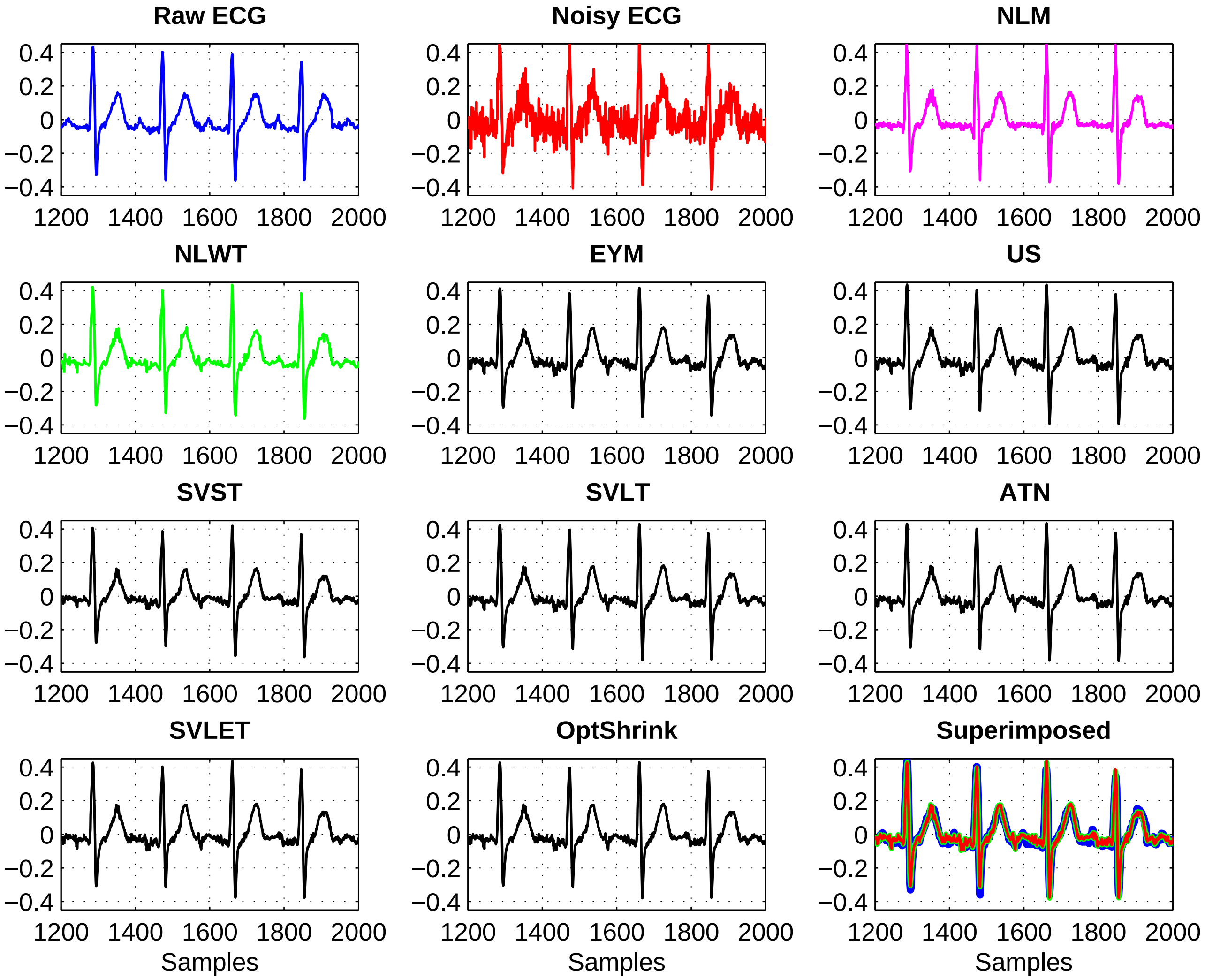


Figure 4.7: Qualitative performance comparison of the proposed methods with the contrast method on the noisy ECG corresponding to the noise-free ECG signal numbered as 215 from MIT-BIH Arrythmia database.

qualitative evaluation, the denoising on a signals from the chosen data set with the pathology myocardial infarction (MI) is shown in Figure 4.9. The diagnostically important feature for such disease is the ST-segment depression. Note that, the proposed methods preserved the depressed ST-segments as good as the NLM method.

4.5 Discussion

The proposed methods are based on the reasonable assumption that the underlying clean signal has several similar blocks of samples. The blocks with smaller amplitudes obtain more similar blocks than the higher ones, due to the lower dynamic range and the nature of ECG signals.

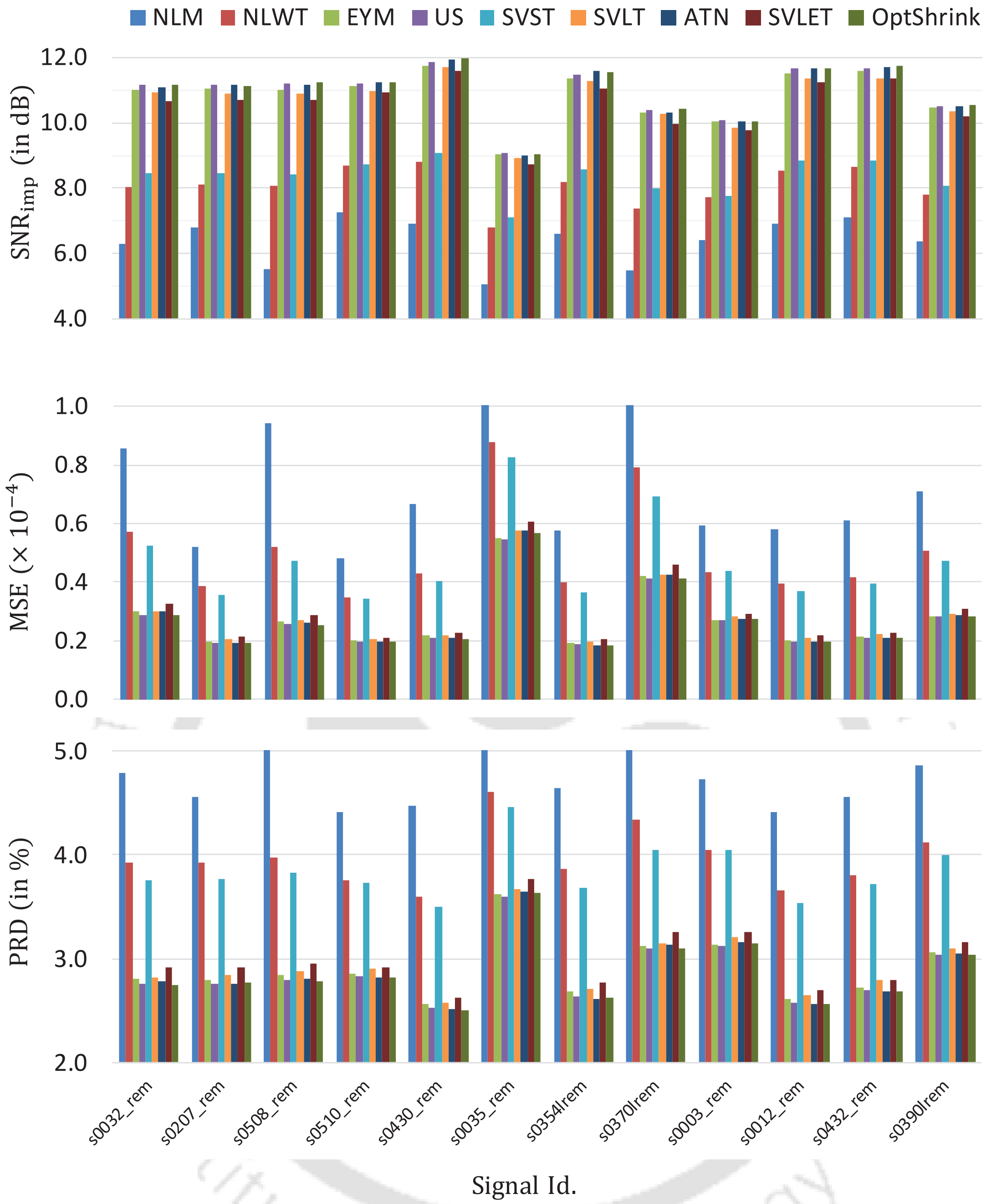


Figure 4.8: Comparison of the denoising performance of the proposed methods and the existing method averaged over all the channels of the chosen ECG signals at 20dB noise level.

The averaging is not able to remove the noise from the higher amplitude regions and hence the NLM method is found to suffer from the *rare patch* effect. The same trend is also noticed in the NLWT method, but this effect is less prominent because denoising is done in the 2D DWT domain. Further, this effect becomes insignificant in case of the SVSE methods. The rationale behind this is the sufficient adaptation of the shrinkage function according to the number of available similar patches and local statistics of the data. This makes the SVSE

Table 4.2: Pathologies known for the signals used in experiment reported in Figure 4.8.

Abnormality	Record Id in the database
Cardiac dysrhythmia	s0032_rem, s0207_rem
Myocarditis	s0508_rem, s0510_rem
Bundle branch block	s0430_rem, s0035_rem
Myocardial infraction	s0354lrem, s0370lrem
Valvular heart disease	s0003_rem, s0012_rem
Ventricular hypertrophy	s0432_rem, s0390lrem

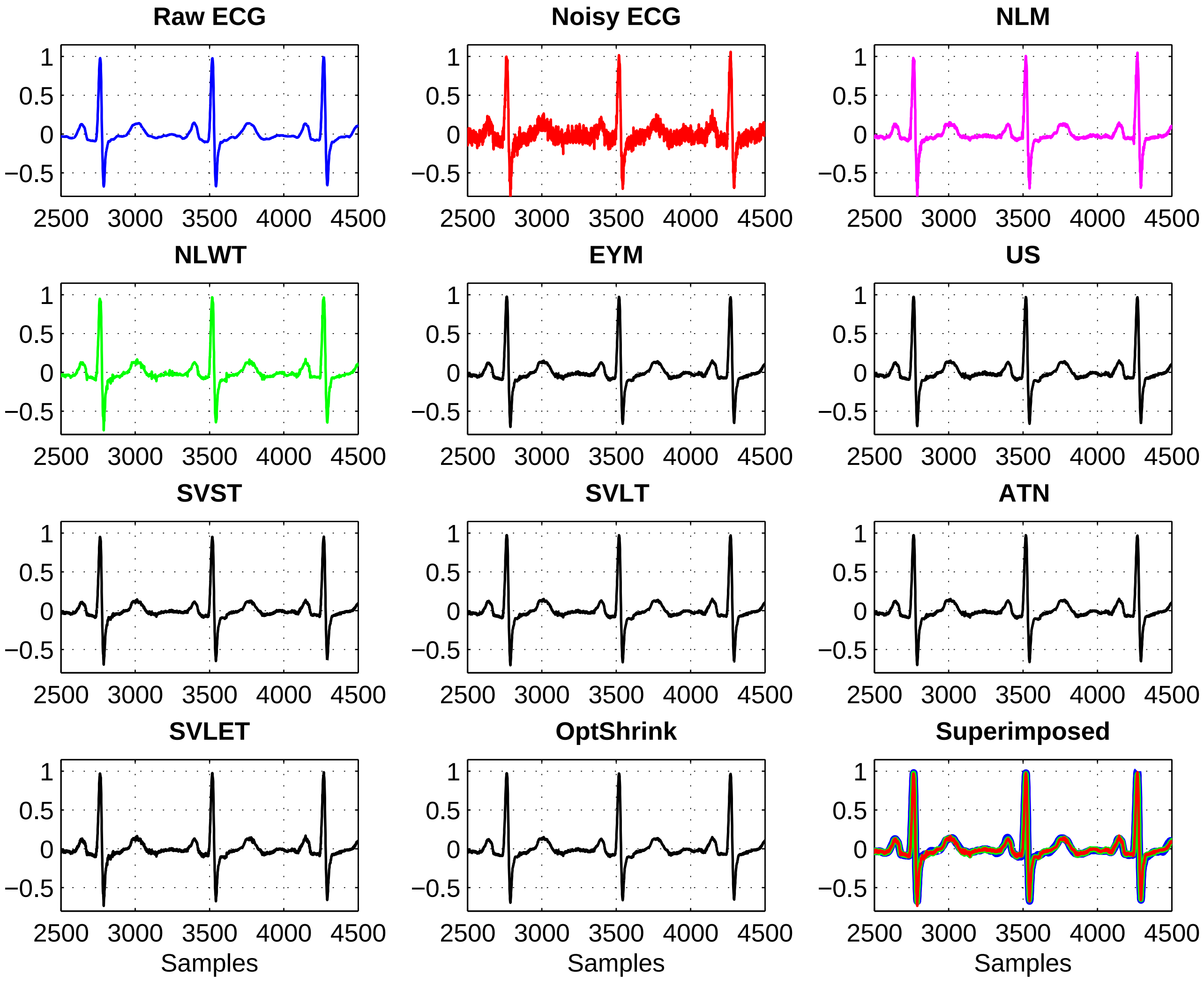


Figure 4.9: Illustration of preservation of the depressed ST-segments while denoising the lead-I of the signal number s0354lrem derived from the PTB diagnostic database.

denoising methods better performing as compared to the contrast methods. Among these, the EYM and the US methods perform quite closely whereas the SVST method outperform these two. This proves that the nuclear norm minimization performs better than the low-rank minimization for ECG denoising. Moreover, the SVLT and the ATN perform quite closely but better than the above methods, thanks to the better shrinkage function employed in them. The SVLET and the RMT based OptShrink estimators perform closely in terms of the MSE and the PRD while in terms of the SNR improvement, the latter performs better than the former. This could be because of the asymptotic assumptions made in the RMT based shrinkage method are sufficiently met in SDM due its inherent low-rankness. The SVLET as shown in Chapter 3 turns out to be near optimal in case of asymptotic case and hence is able to perform quite closely with the RMT based optimal shrinkage method.

Table 4.3: Comparison of the execution time of different denoising methods for estimating an ECG signal of length $N = 10,000$ samples. For all methods employing collaborative filtering, the SDM extraction time of 30 seconds has been shown separately.

Methods	Time (in sec.)
NLM	10.8
NLWT	$30 + 7.6$
EYM	$30 + 8.4$
US	$30 + 7.5$
SVST	$30 + 12.8$
SVLT	$30 + 57.2$
ATN	$30 + 18.2$
SVLET	$30 + 2.3$
OptShrink	$30 + 2.2$

Algorithms are implemented in MATLAB 7.14 on a 64-bit 3.20 GHz Intel® Core i5-3470 computer with 8 GB RAM. The execution time for denoising an ECG signal of length 10,000 samples is compared in Table 4.3. The NLM method does not need any block-matching hence it turns out to be the fastest method. Both the EYM and the US methods search the best rank of underlying noise-free SDMs and hence take quite comparable time. The number of tunable parameters in the SVST, the ATN and the SVLT methods happens to be 1, 2 and 3, respectively. Each of them is tuned by minimizing SURE as discussed in Chapter 2 and hence takes the time accordingly. As the OptShrink needs not to tune any parameter, it turns out to be the fastest one among the proposed methods. The SVLET being a linearized approximation of the above shrinkage estimators, takes very less time as compared to the non-asymptotic SURE based estimators while being quite comparable to the OptShrink in terms of the execution time.

The proposed methods target only the white Gaussian noise removal from ECG signals. However, at the time of recording, the ECG signals are mostly corrupted by non-Gaussian or structured noise caused by motion or muscle activity and other external interferences. To remove such interference noises, several methods have been proposed. To name a few, recently in [95], an adaptive filtering framework is proposed to reduce the muscle and baseline wander noise from the brain signals. The same author, in [96], proposed an artificial neural network based denoising method for ECG signal denoising. To address different kinds of interference noises, the proposed methods can be combined with these methods. But in case of ambulatory telemonitoring, ECG signals are most likely to be corrupted by channel noise which is commonly modeled as the additive white Gaussian noise. Hence, the proposed methods find a relevant scope of application.

4.6 Summary

This chapter evaluated the proposed SVLT and SVLET shrinkage methods for ECG signal denoising by exploiting the local and nonlocal redundancy present in the signal. On comparing with the state-of-the-art NLM method [84], the proposed methods are found to give better performance in terms of a number of performance measures. All ECG denoising methods using the shrinkage of SVs are novel contributions of this chapter. It is also highlighted that the proposed methods are able to preserve the diagnostic features in the signal much better than the existing algorithm. Among the proposed methods, the asymptotic optimal estimator is the fastest and the best performing, hence we finally suggest it for ECG signal denoising. At the same time, it is also emphasized that the proposed SVLET shrinkage maintains the performance quite comparable to the OptShrink estimator, hance it can also be used in such applications without any loss of diagnostic information.

5

Application to Image Denoising

Contents

5.1 Image denoising by collaborative filtering 73
5.2 Proposed methods 79
5.3 Experimental set-up and parameter tuning 82
5.4 Results and discussion 90
5.5 Summary 92

In Chapter 4, we evaluated the proposed shrinkage estimators for ECG signal denoising. In order to employ the matrix denoising methods to one dimensional (1D) ECG signals, we artificially created similarity data matrix (SDM) containing the similar blocks of the noisy ECG and denoising was performed on these matrices. Thus, 1D noisy signal was transformed into several 2D SDMs. On the contrary, the digital images are 2D signals which can be directly considered as the data matrices. Hence, it is natural to evaluate the proposed matrix denoising methods on images. However, applying the matrix denoising methods on the whole images was not found to be effective. Since a significant amount of signal details also get suppressed along with the noise, it leads to degraded denoising performance. The possible reason for this lies in the entire rows/columns of an image not being highly correlated. As a result, a slight shrinkage or truncation causes the loss of important information of the signal. Moreover, the dimensionality of the image makes the computational load of the matrix shrinkage based denoising process highly restrictive. As a possible solution to these problems, one can explore applying the matrix denoising methods on the small image patches (2D blocks), separately. This approach not only exploits the inter-patch correlation but also reduces the computational burden substantially.

In recent times, a number of image denoising methods have been proposed by exploiting the inter-patch correlations in a CF framework. In this chapter, we also choose the said framework for applying the matrix denoising methods for image denoising. Specifically, we form the SDMs from a given noisy image by grouping the similar vectorized image patches and then denoise each of these SDMs, separately. This process is almost similar to the ECG signal denoising, except that the collaborative filtering is iterated a couple of times before we get the final denoised image. The performance of the proposed SV shrinkage based image denoising methods are compared with the contrast methods and found to be quite competing with the current state-of-the-art image denoising methods. In addition, we also compare the denoising performance of the proposed shrinkage methods with that of the contrast methods considering the given image and its small overlapping patches as the data matrices, separately. The experiments show the better denoising performance by the proposed shrinkage methods in both the cases.

5.1 Image denoising by collaborative filtering

The image denoising problem has attracted a lot of attention in past and a number of methods have been reported in the literature. These methods can be broadly classified into two categories: (1) the spatial domain methods [83, 97–99] and (2) the transform domain methods [90, 100–104]. In spatial domain methods, the noise suppression is achieved by averaging the neighboring pixels with proper weights. In transform domain methods, the same is achieved by the thresholding of transform coefficients. Among the spatial domain methods, the non-local means (NLM) methods [83] that takes the weighted average of the pixels with similar neighborhood as a pixel estimate can be considered as the most effective one. Among the transform domain methods, the approach employing the adaptive thresholding of the 2D DWT coefficients [90, 103, 104] turns out to be the best performing. Combining these two broad categories, a very effective image denoising framework was proposed in [85]. This method is motivated by the fact that the natural images often have repeating structures. If we consider a small sized patch then it is more likely to find several similar patches in an image. This photometric redundancy had already been used in the NLM method that achieved quite appreciating denoising performance. In [85], it was realized that this redundancy can be better exploited in transform domain and the resulting approach was referred to as Block Matching based 3D filtering (BM3D). It not only outperformed the then existing image denoising methods, but also served as the state-of-the-art for almost a decade. At the heart of the BM3D method lies the formation of highly redundant 3D groups by stacking similar 2D patches extracted from the noisy image. Denoising is then performed on these groups by the shrinkage of their 3D DWT coefficients, thus all patches in a group collaborate in denoising. The denoised image is obtained by aggregating the overlapping pixels of the denoised groups. This process of CF has been adopted in several state-of-the-art image denoising methods [57, 86–88, 105].

The CF exploits both the intra- and the inter-patch correlations present in the natural images. The block diagram shown in the upper half of Figure 5.1 illustrates the overall processing involved in the CF. The three different stages involved in it are: (1) the block-matching, (2) the 3D-filtering, and (3) the aggregation. In the first stage, the similar image patches are searched within a search-window around the given reference patch by block matching. Usually,

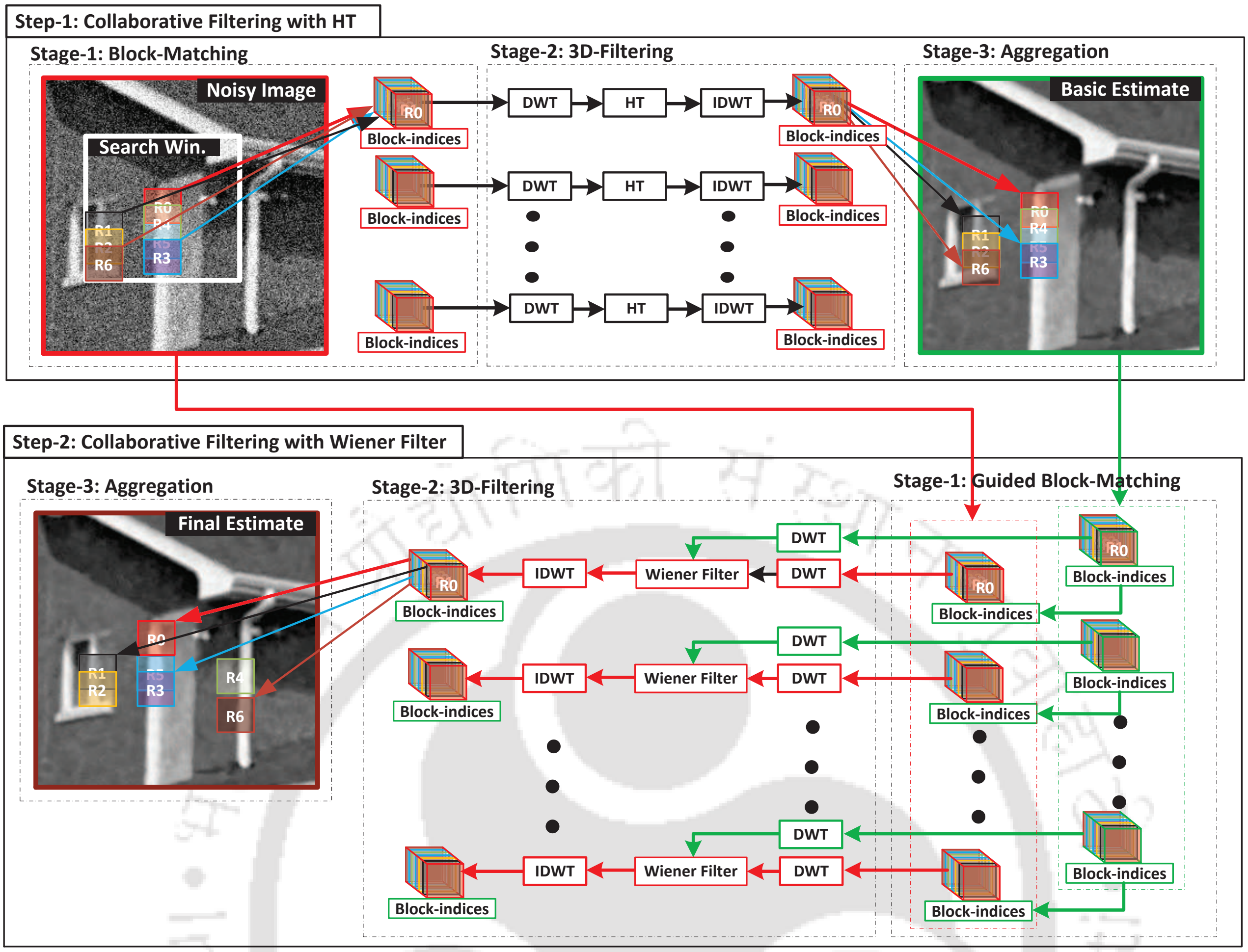


Figure 5.1: Block diagram of the two-step collaborative filtering in the BM3D. In the figure, the patch marked as “R0” represents a reference patch and the color of the boxes indicate the non-locality of the photometrically similar patches.

the Frobenius norm of the difference matrix between the patches is considered as the measure of similarity between them. However, the distance in feature-space has also been considered for the finding patch similarity in some works [88, 106]. These similar 2D patches are then stacked over one another to form a 3D array. This processes is repeated for all possible reference patches in the noisy image. Thus, at the end of this stage we obtain a 3D array for each of the reference patches. As the patches extracted for block matching are highly overlapping, this process can be thought of as non-disjoint clustering. In the second stage, in Figure 5.1, each of these arrays are denoised either by hard thresholding (HT) or the Wiener filtering of the 3D DWT coefficients. In the last stage, the denoised image is obtained by aggregating the patches from the denoised arrays. The aggregation used in this step is nothing but the weighted average of the multiple estimates of the pixels in overlapping patches. The aggregation stage adds up another layer of noise suppression as well as reduces the blocking artifacts, thereby enhancing the visual consistency in the denoised image. In the following, we discuss a few image denoising methods where the CF plays a central role.

5.1.1 Block matching based 3D filtering (BM3D)

As evident from the block-diagram of the BM3D shown in Figure 5.1, it employs two CF steps. In the first step, the CF is done by hard-thresholding of the 3D DWT coefficients yielding a basic estimate of the underlying noise-free image. The 3D DWT used in the BM3D is a separable combination of 2D and 1D DWT. Specifically, to transform a 3D array a 2D DWT is first performed on each patch of the array and then a 1D DWT is performed in the stacking dimension. In hard-thresholding of the 3D-filtering stage, all the 3D DWT coefficients smaller than a particular threshold is replaced by zero leaving the larger coefficients untouched. As the 3D array is quite homogeneous, only a few dominant coefficients capture the signal energy while rest of them represent the noise. Hence, the thresholding of the smaller coefficients suppresses the noise effectively. The weight of aggregation in the first step of the BM3D is taken inversely proportion to the number of preserved coefficients. The rationale behind this lies in the fact that a 3D array denoised by preserving a fewer coefficients are smoother and hence needed to be given a larger weight in comparison to the one preserving more number of coefficients. Thus, the basic estimate obtained at the end of this step is almost noise-free. However, in this process the minute details in the denoised image are lost. This is because of the hard-thresholding which replaces all the smaller coefficients that represents the high frequency (textured) content of the underlying noise-free image with zeros. For this reason, the basic estimate is further refined in the second CF step. Note that, in the first CF step of the BM3D the block matching is performed on the noisy image itself. As the similarity measure used in the BM3D is quite susceptible to the presence of noise, this may lead to erroneous block matching. This is why in the second CF step of the BM3D, block matching is done on basic estimate obtained in the first CF step and additional 3D arrays are formed using the corresponding blocks from the noisy image. Further, in order to prevent any loss of smaller coefficients the CF in this step is done by the empirical Wiener filtering (EWF) in 3D DWT domain. The implementation of the EWF requires the knowledge about the signal-to-noise ratio (SNR). In the BM3D, the energy of the basic estimate is considered as an approximation of the signal energy, hence the 3D arrays formed with the similar patches from the basic estimate are used for computing the SNR for applying the EWF on each of the corresponding noisy array. This SNR is also used for computing the weight of aggregation in the subsequent aggregation stage. Thus, the basic estimate of the first step works as the pilot signal for estimating the final denoised image. Further details of the BM3D can be found in [85].

The main drawback of the BM3D lies in the use of fixed transform bases. In [107], a highly competitive image denoising method by learning an over-complete dictionary from the noisy image was proposed. In the CF framework, learning a dictionary for each 3D array is computationally intensive. Nonetheless, a number of methods have been proposed making the use of the data-dependent transforms. Few of them are briefly discussed below.

5.1.2 Higher order singular value decomposition based CF

The HOSVD is the generalization of the SVD for higher order tensors [108, 109]. Recently, it has been used in various multidimensional signal restoration problems as an alternative to the SVD [110–112]. In [87], a CF based image denoising method using higher order SVD (HOSVD) is proposed. Except the use of the HOSVD in place of 3D DWT, it is exactly same as the BM3D. Specifically, in 3D filtering stage, it decomposes each 3D array using the HOSVD. The noise suppression is achieved by hard-thresholding of the higher order SVs. The basic estimate of the noise-free image is obtained by reconstructing the hard-thresholded higher order SVs followed by the aggregation. To refine this estimate a second CF step is used similar to the BM3D. In further discussion, we refer to this method as the HOSVD-CF. The quantitative denoising performance of the HOSVD-CF is found to be quite competitive with the BM3D. As expected, it could preserve the signal details better than the BM3D.

Unlike the BM3D and the HOSVD that transform the 3D arrays, one can unfold these arrays to make 2D arrays. This is particularly done by stacking the similar patch-vectors in the columns of a matrix instead of the patches themselves in the block-matching stage of the CF. Thus, the 3D-filtering of the CF is converted into a matrix denoising problem. Note that, this procedure is similar to the SDM denoising in case of ECG. Using the vectorized notations, assume that a noisy image of size $I_1 \times I_2$ is given as

$$\tilde{\mathbf{y}} = \mathbf{y} + \bar{\mathbf{y}} \tag{5.1}$$

where $\mathbf{y} \in \mathbb{R}^{I_1 I_2}$ is the underlying noise-free image and $\bar{\mathbf{y}}$ is a zero mean Gaussian noise vector with the covariance matrix $\tau^2\mathbf{I}$. Assume that $\mathcal{R}_t(\cdot)$ is an operator that extracts an SDM denoted as $\widetilde{\mathbf{X}}_t$ from the input image, corresponding to the tth reference patch of size $\sqrt{m} \times \sqrt{m}$. If an SDM contains n most similar vectorized patches in its columns then $\widetilde{\mathbf{X}}_t = \mathcal{R}_t(\tilde{\mathbf{y}}) \in \mathbb{R}^{m\times n}$ is a matrix. Clearly,

$$\widetilde{\mathbf{X}}_t = \mathbf{X}_t + \overline{\mathbf{X}}_t \tag{5.2}$$

where $\mathbf{X}_t = \mathcal{R}_t(\mathbf{y})$ and $\overline{\mathbf{X}}_t = \mathcal{R}_t(\bar{\mathbf{y}})$. Thus, the image denoising problem is converted to a set of matrix denoising sub-problems. Once all the SDMs are denoised, the denoised image is obtained by aggregation. In the following, we discuss the image denoising methods that use the matrix estimation methods in the CF framework.

A well known matrix denoising method by nuclear norm minimization (NNM) could be one of the potential candidate for this task. The NNM problem is given as

$$\widehat{\mathbf{X}} = \min_{\mathbf{X}} \|\widetilde{\mathbf{X}} - \mathbf{X}\|^2_{\text{Frob}} + \mu\|\mathbf{X}\|_* \tag{5.3}$$

where $\|\mathbf{X}\|_* = \sum_{i=1}^{L} \sigma_i$ and $L = \min(m, n)$. In (5.3) and henceforth, we drop the subscript t for the ease of notation. The solution of the NNM problem turns out to be the SVST with a threshold μ. As discussed in Chapter 2, the SVST has the drawback that it applies the shrinkage on all the SVs by an equal amount μ. Hence, it fails to achieve the optimal *bias-variance* trade-off. To mitigate this drawback of nuclear norm regularization, the weighted nuclear norm is used.

5.1.3 Weighted nuclear norm minimization (WNNM)

In [69], an image denoising method by using the WNNM in CF is proposed. The WNNM problem is given as

$$\widehat{\mathbf{X}} = \min_{\mathbf{X}} \|\widetilde{\mathbf{X}} - \mathbf{X}\|^2_{\text{Frob}} + \|\mathbf{X}\|_{\mathbf{w}} \tag{5.4}$$

where $\|\mathbf{X}\|_{\mathbf{w}} = \sum_{i=1}^{L} w_i \sigma_i$ and $\mathbf{w} = [w_1, w_2, ..., w_L]^{\mathrm{T}}$. The solution to the WNNM problem is given as

$$\widehat{\mathbf{X}}^{\mathrm{WNNM}} = \sum_{i=1}^{L} (\tilde{\sigma}_i - w_i)_{+} \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}} \tag{5.5}$$

where $\widetilde{\mathbf{X}} = \sum_{i=1}^{L} \tilde{\sigma}_i \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^{\mathrm{T}}$. Clearly, it applies different amount of shrinkage on different SVs. The choice of weights in WNNM is crucial. In [69], the authors follow an iterative weight update strategy starting with uniform weight initialization. In the subsequent iterations, they choose the weights inversely proportional to the magnitude of the corresponding SVs. The rationale behind this choice is that the larger SVs are the most important and hence needed to be shrunk less. As the smaller SVs are mostly noise dominated, this adaptive weighting of SVs helps to improve the denoising performance. In addition, the denoised image is further refined iteratively, with a back-projection procedure. The back-projection in the lth iteration is given as

$$\tilde{\mathbf{y}}^{(l)} = \hat{\mathbf{y}}^{(l-1)} + \delta(\tilde{\mathbf{y}} - \hat{\mathbf{y}}^{(l-1)}) \tag{5.6}$$

where $\hat{\mathbf{y}}^{(l-1)}$ is an estimate from the previous iteration and δ is a back-projection parameter. This process is repeated for $l > 2$ till the convergence. The WNNM in the CF framework consistently outperforms the BM3D for all noise levels and forms the state-of-the-art in image denoising.

5.1.4 CF with low-rank approximation

Recently, in [57], an image denoising method that uses low-rank approximation for matrix denoising in CF is proposed. As discussed in Chapter 2, the rank r of the underlying noise-free matrix can be obtained using the inequality

$$\sum_{i=r+1}^{L} \tilde{\sigma}_i^2 \leq mn\tau^2 < \sum_{i=r}^{L} \tilde{\sigma}_i^2. \tag{5.7}$$

In [57], once the r is obtained with this inequality, the r-rank approximation of noisy SDM $\widetilde{\mathbf{X}}$, given as

$$\widehat{\mathbf{X}}^{\mathrm{LRASVD}} = \sum_{i=1}^{r} \tilde{\sigma}_i \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i' \tag{5.8}$$

where $\widehat{\mathbf{X}}^{\mathrm{LRASVD}}$ is taken as the estimate of $\mathbf{X}$. We refer to this method by LRASVD in the rest of the chapter. Unlike the iterative refinement of the denoised image in WNNM, the LRASVD uses only a single step of refinement using the back-projection. The denoising performance of the LRASVD is quite competing with that of the WNNM.

5.1.5 Low-rank collaborative filtering (LRCF)

In the LRASVD, the SDMs are denoised by the LRMA that is analogous to the SVHT. As discussed in Chapter 3, the SVHT discards all the smaller SVs whereas it leaves the higher SVs unchanged. Hence it sometimes loses important image details and is unable to suppress the noise in the signal dominant SVs. As opposed to that, the OptShrink estimator derived using the random matrix theory was shown to perform significantly better than the SVHT [113]. Using this OptShrink estimator in the CF an iterative image denoising method is recently proposed. Like the WNNM and the LRASVD, it also employs the back-projection step in every iteration for refinement. In addition, in every refinement step it updates the standard deviation of noise as

$$\hat{\tau}^{(l)} = \gamma\sqrt{\tau^2 - \frac{1}{N}\|\tilde{y} - \tilde{y}^{(l)}\|_2^2}. \tag{5.9}$$

The denoising performance of the LRCF is quite comparable in general and better than the WNNM in most of the cases.

5.2 Proposed methods

In the following, we present image denoising methods that can be categorised as the one that uses the thresholding of 3D transform coefficients of the 3D arrays for denoising and the other that uses the shrinkage of SVs for SDM estimation. The method in the first category is same as the BM3D except the use of the *finite ridgelet transform* (FRIT) in place of the DWT in its first step. The method in the second category is more close to the WNNM and the LRCF in terms of its algorithmic implementation. It uses the SVLET in the noise filtering stage of the CF. In the following, we discuss these two methods in detail.

5.2.1 Denoising with FRIT in CF

The multi-resolution property of the DWT allows for the parsimonious representation of the smooth patches as well as the point singularities in the image. But, for the edges, which can be thought as the line singularities, the 2D DWT loses the parsimony of representation. As a result, the shrinkage of 2D DWT coefficients of the noisy image causes the blurring of the edges on reconstruction. Also, the 2D DWT exhibits large coefficients along the edges, even at finer scales. In fact, these edges are repeated in each scales [114] destroying the sparsity in DWT domain. In order to reconstruct the noise-free image, one has to estimate all these coefficients. While estimating a large number of coefficients, the balance between parsimony of representation and the accuracy in estimation gets disturbed.

Candès [115] proposed the ridgelet transform to overcome this limitation of 2D wavelet transform in the continuous domain. The continuous ridgelet transform of a continuous image function $x(i_1, i_2)$ is defined as

$$R[x](\theta, a, b) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(i_1, i_2)\varphi_{\theta,a,b}(i_1, i_2) di_1 di_2 \tag{5.10}$$

where the ridgelets $\varphi_{\theta,a,b}(i_1, i_2)$ are defined as

$$\varphi_{\theta,a,b}(i_1, i_2) = \frac{1}{\sqrt{a}}\varphi\left(\frac{i_1 \cos\theta + i_2 \sin\theta - b}{a}\right) \tag{5.11}$$

where scalars θ, a, and b are the rotation, the scaling, and the shifting parameter, respectively. The inverse ridgelet transform is given as

$$x(i_1, i_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R[x](\theta, a, b)\varphi_{\theta,a,b}(i_1, i_2)\frac{d\theta}{4\pi}\frac{da}{a^3}db. \tag{5.12}$$

The ridgelet $\varphi_{\theta,a,b}(i_1, i_2)$ is a constant functions along the line $i_1 \cos\theta + i_2 \sin\theta = b$. It enables the ridgelet transform in representing the singularity along this line in the image $x(i_1, i_2)$ very sparsely. In [116], a discrete and orthogonal implementation of the continuous ridgelet transform, known as the FRIT, is developed. In the FRIT domain, a line singularity requires a few large coefficients. However, randomly located noise samples (point singularities) are unlikely to produce large coefficients and a sparse representation in FRIT domain. Hence, better denoising performance could be achieved using hard-thresholding of the FRIT coefficients as compared to that of the 2D DWT coefficients.

This fact is illustrated with the help of a test image which is smooth away from the line singularity. The original test image and the noisy image, generated by adding white Gaussian noise to it, are shown in Figure 5.2(a) and Figure 5.2(b), respectively. It is noticed that the denoising using hard-thresholding of 2D DWT coefficients results in blurring of the edge

(Figure 5.2 (c)). On the other hand, the hard-thresholding of the FRIT coefficients is able to restore the edge excellently (Figure 5.2 (d)). On comparison, the FRIT based approach is found to give significant improvement in visual quality as well as in terms of the peak signal to noise ratio (PSNR) over the 2D DWT based approach. The drawback is that the strong edge results in the "wrap around" effect which appears as an additional weak edge seen in the smooth region. To mitigate this effect, an additional 2D Wiener filtering could be employed. The effect of such filtering can be seen in Figure 5.2(e).

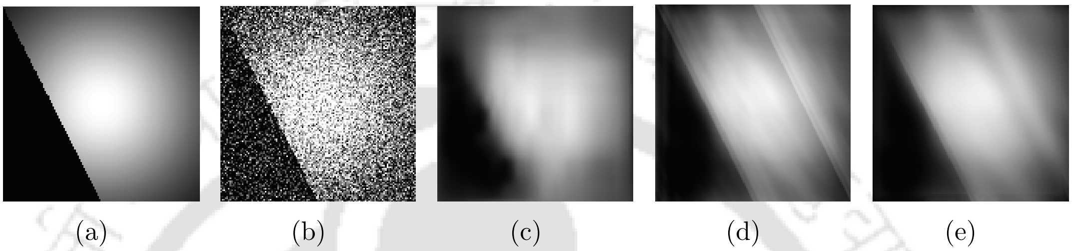


Figure 5.2: Denoising performances for global processing (with threshold value, $\sigma\sqrt{2\,log\,N}$). (a) Original synthetic image ($N = 127 \times 127$), (b) Noisy image, $\sigma = 80$, (c) 2D DWT (PSNR: 19.89 dB), (d) FRIT (PSNR: 21.19 dB), (e) FRIT + Wiener (21.37 dB).

Natural images, particularly with textured scenes, have quite high redundancy (repeating structures). Also, the curved edges are more frequent than the straight edges. The curved edges can be thought as the piece-wise straight edges if we consider very small image patches [114]. To represent such small patches sparsely, the ridgelets are more suited than the wavelets. Hence, in the first CF step of the BM3D, we replace the 2D DWT with the FRIT, while retaining the 1D Haar transform in the stacking dimension. The second CF step of the framework is exactly similar to the BM3D. Thus, in the proposed method, the patch-wise FRIT and the Haar transform in the stacking dimension forms the 3D transform. The denoising is achieved by hard-thresholding in the FRIT+Haar domain, thus exploiting both intra- and inter-patch correlations. We refer to this method as the FRIT-CF further onwards.

5.2.2 Denoising by matrix estimation in CF

In this subsection, we propose the second image denoising method using the SVSE for matrix estimation in the CF framework. Specifically, we use the proposed SVLET for the shrinkage of

SVs of noisy SDMs to estimate the corresponding noise-free SDMs as

$$\widehat{\mathbf{X}} = \sum_{i=1}^{L} f_{T,K}^{\mathrm{SVLET}}(\tilde{\sigma}_i)\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^{\mathrm{T}}. \tag{5.13}$$

Once, the estimates for all the SDMs are obtained the basic estimate of the noise-free image is obtained by aggregation. Like the BM3D, this estimate is used for the guided block matching and for computing the SNRs of the SDMs needed to apply the Wiener filter. Thus, with the help of this basic estimate a refined image is obtained using the second CF step of the BM3D. Like the LRCF, further refinement in the obtained image is done by iterating this process with a new image, at each iteration, obtained using noise back-projection given in (6.13) and with an updated variance estimation using (6.14). The detailed schematic of this method is shown in Figure 5.1. It is referred to as the SVLET-CF in the rest of the discussion.

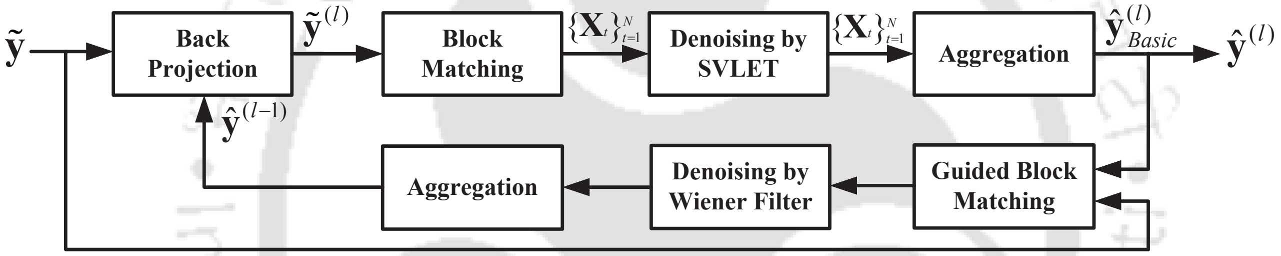


Figure 5.3: Block diagram of the proposed SVLET based image denoising approach.

5.3 Experimental set-up and parameter tuning

Our goal in this section is to evaluate the proposed denoising methods by comparing their experimental denoising performance with the existing state-of-the-art image denoising methods. We artificially add WGN to a number of publicly available gray-scale images mostly used in image processing literature to obtain the noisy image for each. Denoising is performed on these noisy images by the proposed and the contrast methods, assuming the noise level to be known. In cases where the noise level is unknown, the median absolute deviation (MAD) method can used to estimate that. The denoising performance is measured in terms of the peak signal-to-noise ratio (PSNR) defined as

$$\mathrm{PSNR}(\hat{\mathbf{y}}) = 10\,\log_{10}\left\{\frac{I_1 I_2 D^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|_2^2}\right\} \tag{5.14}$$

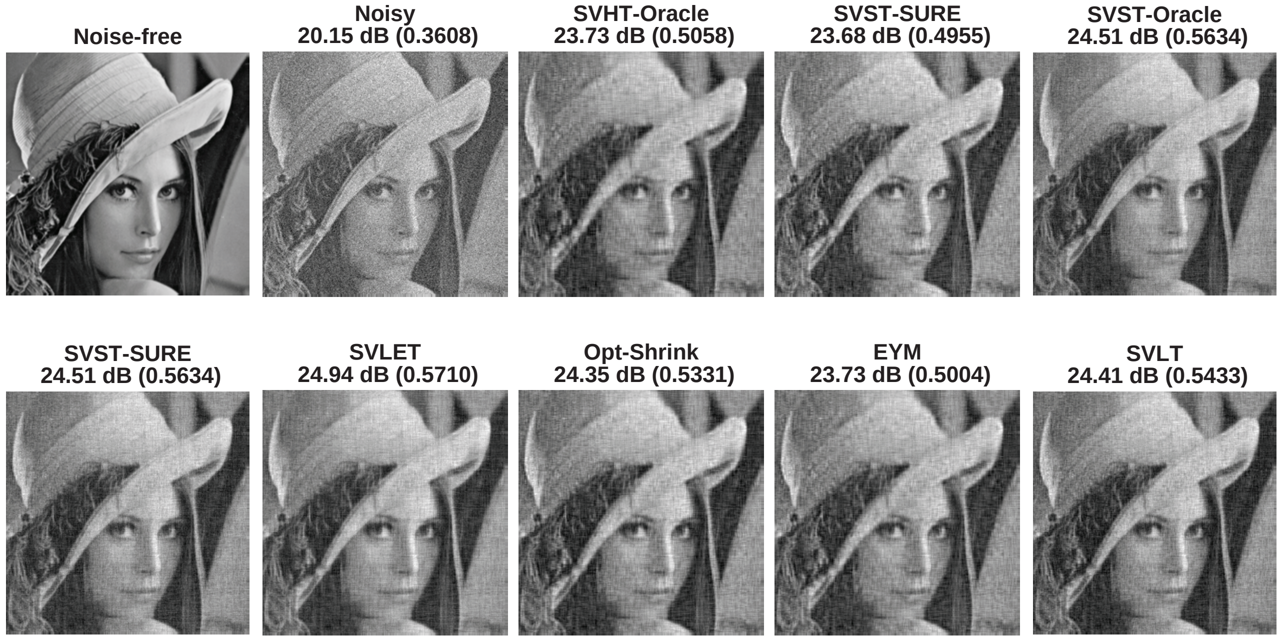
where $\hat{\mathbf{y}}$ is the denoised estimate of the noise-free image $\hat{\mathbf{y}}$ of size $I_1 \times I_2$ with the dynamic range $[0, D]$.

Before we evaluate the SVSE methods in the CF framework, we show their image denoising performance while treating a noisy image as a data matrix. This highlights the motivation behind developing the better shrinkage estimator. In Figure 5.4(a), the denoising performances of the proposed SVLT and the SVLET methods on a segment of the test image are compared with those of the different existing SVSE methods. The threshold μ of the SVHT and the SVST as well as the parameter T of the SVLET is tuned by assuming the underlying noise-free image to be known, hence the estimates correspond to the oracle cases. Specifically, the optimal value of these parameters are chosen by maximizing the PSNR. For different values of these parameters, the plot of the PSNR can be seen in Figure 5.4(b). With the best parameter values, the plots of SVs are shown in Figure 5.4(c). It can be seen from the figure that the SVLET has resulted in better SV estimates than those of the SVHT and the SVST. As a result of that it has also resulted in the best denoising performance. For highlighting the efficacy of SURE based parameter tuning, we have also shown the oracle case performances of the SVHT and the SVST methods in Figure 5.4(a). It can be noted that those have turned out to be almost same as SURE based cases. Furthermore, both the SVLT and SVLET estimators have outperformed the OptShrink estimator. This confirms the efficacy of the proposed shrinkage estimators for image denoising.

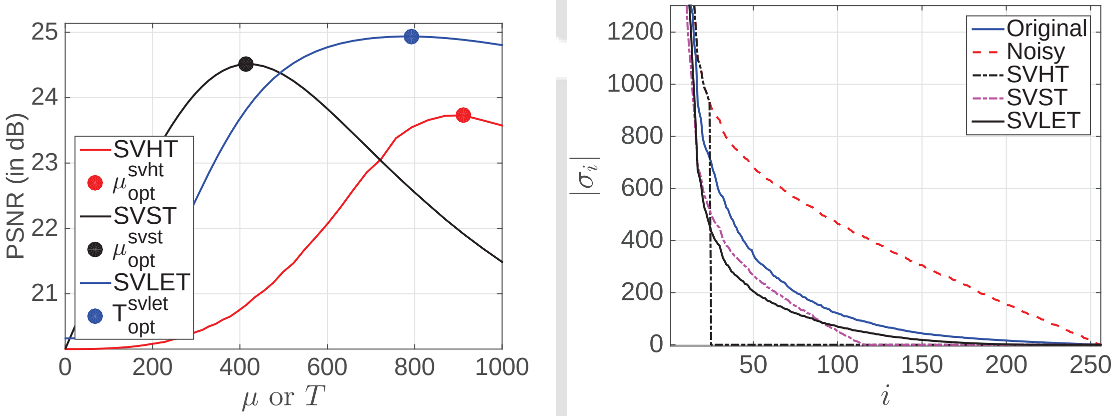
The evaluation of denoising performances of different methods with block based processing is shown in Figure 5.5(a) and the tuning of the relevant parameters for this case is shown in Figure 5.5(b). On comparing with the global case, we note that the denoising performances of all methods have considerably improved with block processing while the trends noted for the global case mostly remain the same.

5.3.1 Parameters of the proposed CF methods

Like the existing methods, the proposed methods also have several parameters that need to be selected to achieve the optimal denoising performance. In the following, we describe the parameter space and the tuning of the key parameters for both the proposed FRIT-CF and SVLET-CF methods.



(a) Performance comparison in case of global processing. The values above the images are the PSNR (in dB) along with the SSIM in round brackets.



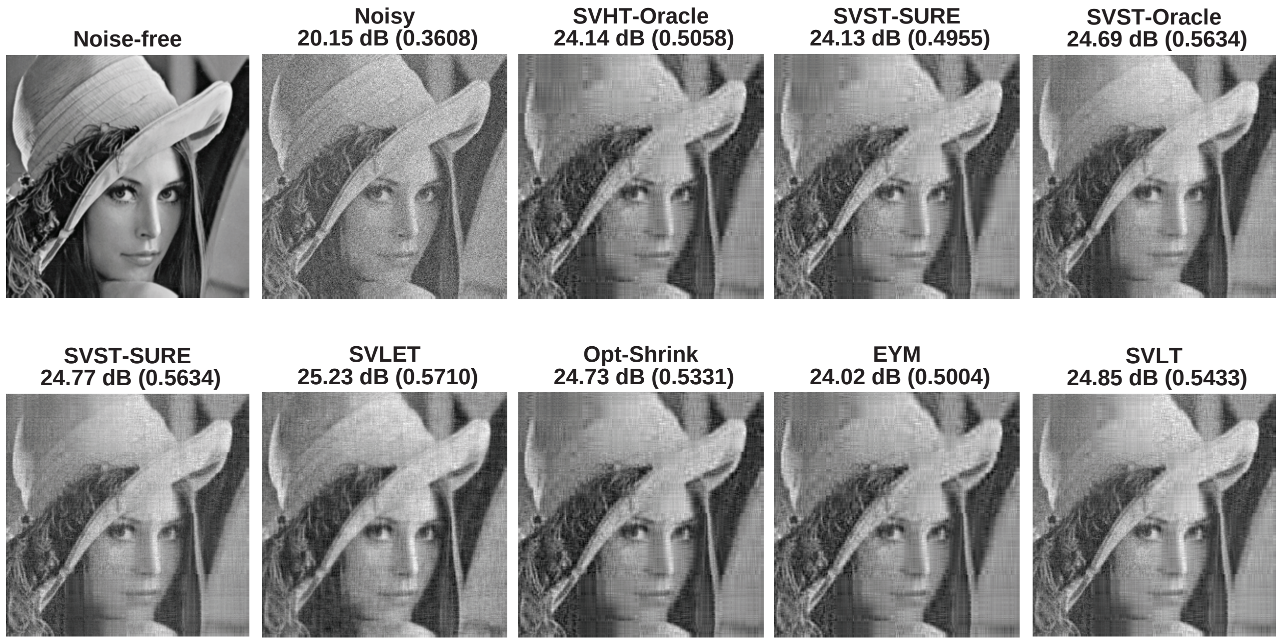
(b) Global denoising

(c) Oracle estimate of SVs.

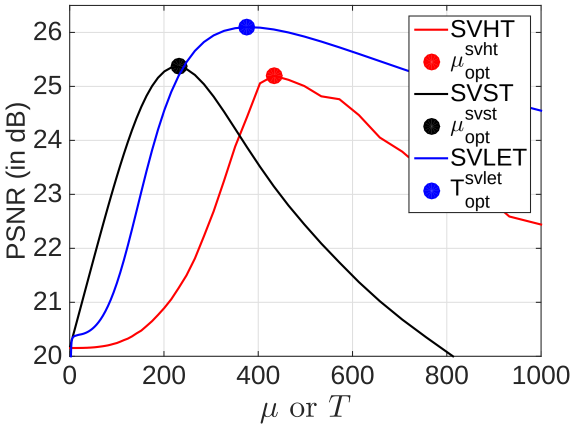
Figure 5.4: Performance comparison of salient existing and the proposed denoising methods in global and block-based processing cases at a noise level of $\tau = 25$.

5.3.1.1 Parameter tuning for the FRIT-CF method

The ridgelet transform is originally defined for the continuous 2D signals [115]. Its discrete implementation through the projection-slice theorem is limited for finite fields with prime number support size [116]. In the proposed FRIT-CF based denoising, the patches with only prime number size could be considered. For a fair comparison with the BM3D, the prime number closest to the patch-size considered in the BM3D is taken as the patch size in the FRIT-CF method. The shrinkage threshold coefficient (μ_{3D}) is chosen as 3.8 based on the experimental



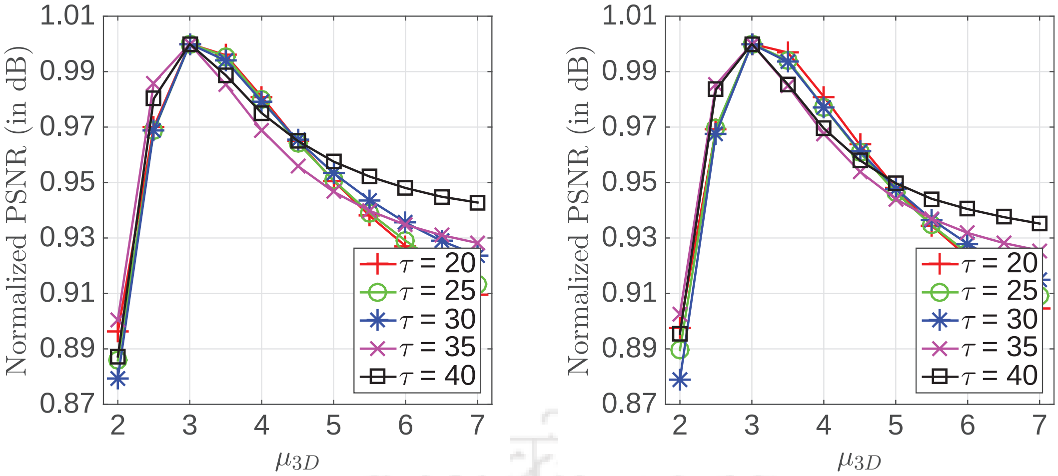
(a) Performance comparison in case of local processing. The values above the images are the PSNR (in dB) along with the SSIM in round brackets.



(b) Block-based denoising

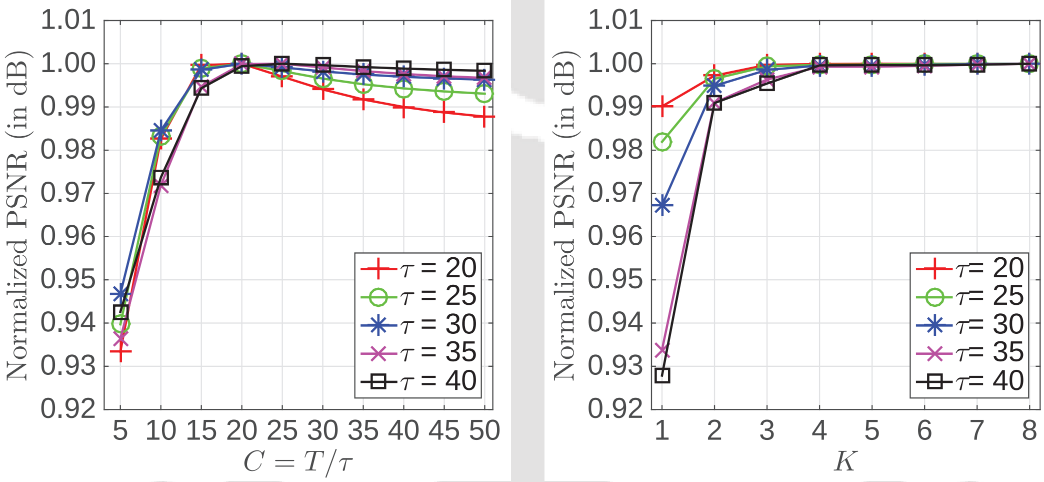
Figure 5.5: Performance comparison of salient existing and the proposed denoising methods in global and block-based processing cases at a noise level of $\tau = 25$.

tuning done on different images. The PSNR vs μ_{3D} plots on two images are shown in Figure 5.6. The FRIT uses the 1D DWT in the Radon transform domain. The mother wavelet for the FRIT is experimentally chosen to be the reverse bi-orthogonal wavelet (rbior 5.5). The value of rest of the parameters used in the two-step processing of the FRIT-CF is kept same as those in the BM3D method. For the ease of reference, they are listed in Table 5.1, for both low and high noise conditions.



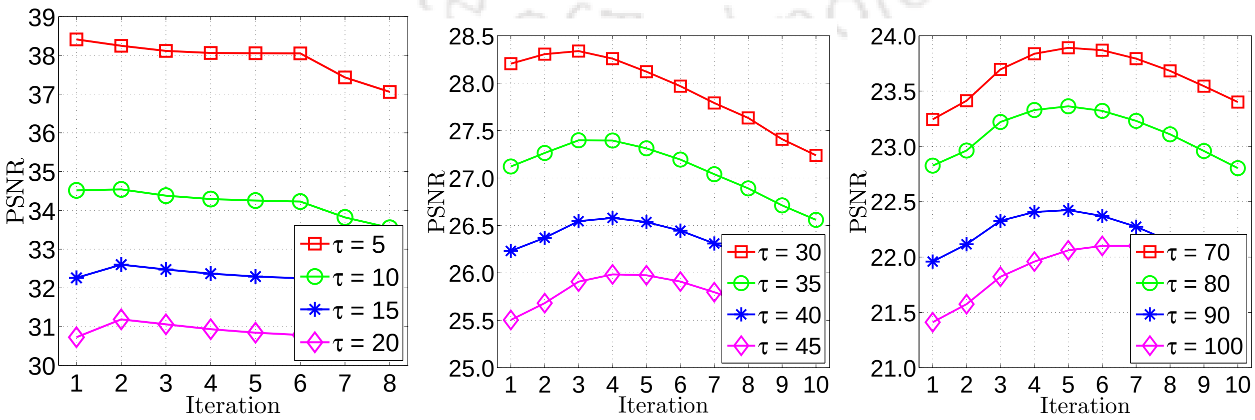
(a) Lenna Image (b) Barbara Image

Figure 5.6: Tuning of hard-threshold parameter coefficients λ_{3D} of the FRIT-CF on two different images.



(a) (b)

Figure 5.7: Tuning of the parameters of the SVLET-CF: (a) the transition parameter T, and (b) the order of linerization K.



(a) Low-Noise (b) Mid-Noise (c) High-Noise

Figure 5.8: Effect of iteration on denoising performance, in terms of PSNR (in dB), of the SVLET-CF in case of different noise levels.

Table 5.1: Parameter set used in the first step of BM3D and the proposed CF with FRIT

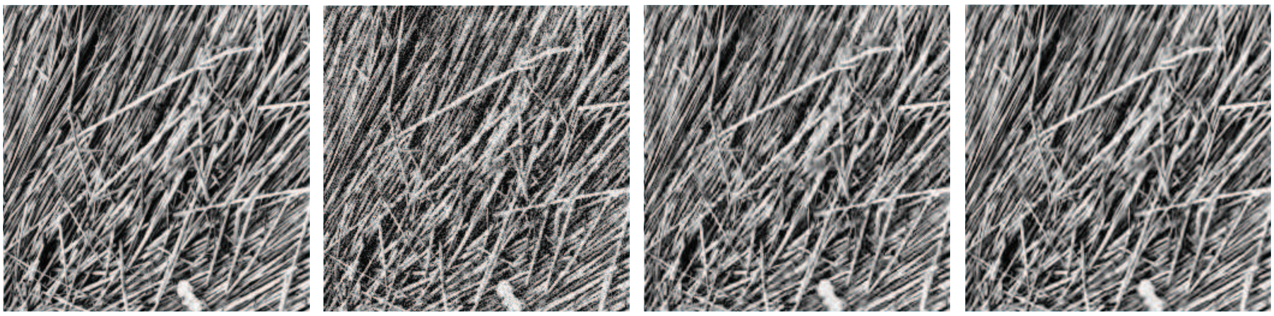
Parameters	BM3D		FRIT-CF	
	$\sigma \leq 40$	$\sigma > 40$	$\sigma \leq 40$	$\sigma > 40$
2D Transform	DWT (bior 1.5)	DCT (type-II)	FRIT (rbior 5.5)	FRIT (rbior 5.5)
1D Transform	Haar	Haar	Haar	Haar
Patch size	8×8	12×12	7×7	11×11
Search window	39×39	39×39	32×32	32×32
Group size	16	16	16	16
Group threshold	2500	5000	2500	5000
Threshold, μ_{3D}	2.7	2.8	3.0	3.0

5.3.1.2 Parameter tuning for the SVLET-CF method

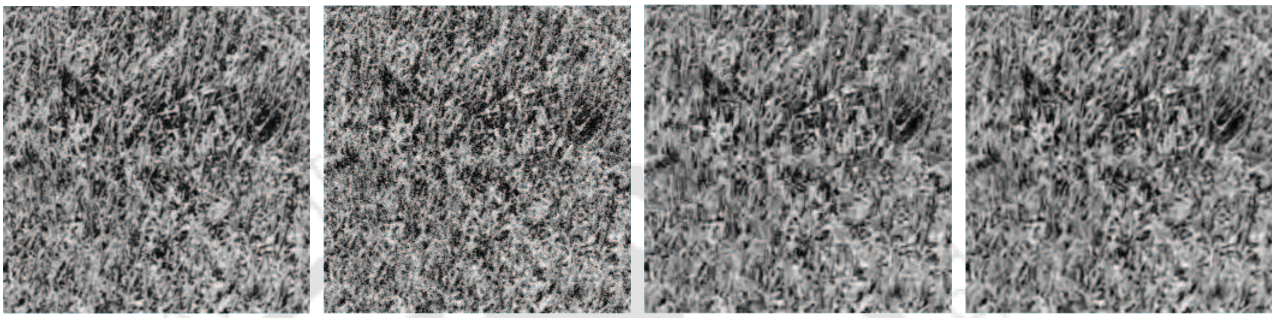
For convenience, the parameters in block-matching and aggregation stages of the proposed SVLET-CF approach are kept same as those of the LRCF method. The order of linearization K and the transition parameter T are experimentally determined. The parameter T depends on the noise level τ and hence we set it as $T = C\tau$ where coefficient $C > 0$ is experimentally tuned. The tuning of these parameters over varying noise levels on a test image are shown in Figure 5.7. Based on these tunings, we have chosen $K = 4$ and $T = 20$ as these values gives better performance over varying noise levels considered. The noise back-projection parameters δ and the variance update parameter γ are tuned around the values suggested in [113] and are fixed at 0.15 and 0.30, respectively. In Figure 5.8, we show the effect of iteration on performance of the proposed SVLET-CF method. Clearly, for low-noise levels, iterating the CF is not helping as the signal details are mostly restored in the first iteration. In case of mid-noise level, 3-4 iterations are needed. However, in high-noise cases, a few extra iterations are required. It can also be noted that, after the suggested number of iterations, the PSNR decreases as the signal details start to get smoothed out.

Table 5.2: Comparison of denoising performance of the different methods in terms of PSNR (dB). The numbers in the boldface are the highest in the columns at given noise levels and indicate the best performance by the method in the given rows.

		Low-Noise			Mid-Noise			High-Noise		
Images	Methods \ τ	5	10	15	30	35	40	70	80	90
House 	BM3D	40.81	37.45	35.68	32.62	31.79	30.79	28.33	27.23	26.94
	FRIT-CF	40.91	37.31	35.31	31.90	30.99	30.40	27.10	26.30	25.66
	WNNM	**41.18**	**37.60**	**35.86**	**33.01**	**32.36**	**31.79**	**28.93**	**28.23**	**27.79**
	LRCF	40.96	37.45	35.64	32.73	32.14	31.55	28.80	28.03	27.77
	LRA-SVD	40.99	37.30	35.42	32.43	31.64	31.04	28.02	27.16	26.62
	SVLET-CF	40.94	37.49	35.77	32.82	32.07	31.52	28.55	27.76	27.52
Lenna 	BM3D	39.30	35.46	33.28	29.64	28.95	27.76	25.76	24.91	24.38
	FRIT-CF	39.30	35.42	33.27	29.61	28.68	27.82	25.23	24.63	24.04
	WNNM	**39.61**	**35.78**	**33.57**	**29.94**	29.24	28.42	**26.04**	**25.33**	24.79
	LRCF	39.60	35.76	33.56	29.90	**29.26**	**28.44**	25.95	25.11	**24.83**
	LRA-SVD	39.44	35.65	33.47	29.83	29.13	28.36	25.42	24.71	24.44
	SVLET-CF	39.48	35.60	33.49	29.80	29.24	28.38	25.93	25.17	24.65
Barbara 	BM3D	38.66	34.76	32.57	29.11	28.49	27.35	25.29	24.59	24.20
	FRIT-CF	38.66	34.72	32.51	29.02	28.16	27.48	24.94	24.26	23.88
	WNNM	**39.30**	**35.34**	**33.01**	**29.32**	**28.77**	27.81	**25.50**	24.79	24.30
	LRCF	39.18	35.21	32.89	29.21	28.72	27.82	25.47	**24.83**	24.31
	LRA-SVD	39.03	35.12	33.00	29.21	28.68	27.70	25.04	24.47	24.02
	SVLET-CF	38.58	34.65	32.58	29.28	**28.69**	27.86	25.45	24.78	24.39
Boats 	BM3D	37.72	33.51	31.19	27.75	27.03	26.35	24.02	23.34	22.92
	FRIT-CF	37.67	33.50	31.22	27.71	26.86	26.17	23.61	23.08	22.57
	WNNM	**38.02**	**33.83**	**31.42**	**27.91**	**27.10**	**26.69**	**24.21**	**23.46**	**23.03**
	LRCF	37.91	33.69	31.33	27.83	27.03	26.70	24.05	23.34	23.02
	LRA-SVD	37.65	33.45	31.15	27.56	26.83	26.43	23.47	22.93	22.63
	SVLET-CF	37.80	33.50	31.21	27.81	26.95	26.42	23.97	23.38	22.98
Straw 	BM3D	34.45	28.92	25.99	21.05	19.88	18.99	16.73	16.11	15.57
	FRIT-CF	34.36	28.91	26.01	21.12	19.96	19.26	15.81	15.27	14.86
	WNNM	**34.64**	**29.36**	**26.57**	**22.07**	**21.16**	**20.45**	**17.44**	**16.80**	**16.33**
	LRCF	34.68	29.34	26.50	21.97	21.09	20.34	17.29	16.65	16.04
	LRA-SVD	34.59	29.26	26.43	22.00	21.05	20.39	17.05	16.30	15.41
	SVLET-CF	34.46	28.93	26.03	21.82	20.96	20.21	17.17	16.53	15.95
Bark 	BM3D	34.42	28.90	25.98	21.39	20.42	19.70	17.13	16.55	16.17
	FRIT-CF	34.30	28.79	25.86	21.22	20.27	19.66	19.48	15.91	15.69
	WNNM	**34.50**	**29.09**	**26.27**	**21.87**	**20.98**	**20.30**	**17.56**	**16.93**	**16.51**
	LRCF	**34.50**	29.04	26.18	21.74	20.83	20.13	17.42	16.82	16.38
	LRA-SVD	34.34	28.89	26.02	21.63	20.71	20.02	16.98	16.29	15.71
	SVLET-CF	34.42	28.89	25.98	21.79	20.88	20.19	17.29	16.69	16.24



(a) Noise-free Straw (b) Noised at $\sigma = 35$ (c) BM3D: 20.50 dB (d) FRIT-CF: 21.18 dB



(e) Noise-free Grass (f) Noised at $\sigma = 35$ (g) BM3D: 21.31 dB (h) FRIT-CF: 21.65 dB

(i) Noise-free House (j) Noisy $\sigma = 35$ (k) BM3D: 32.51 dB (l) FRIT-CF: 32.05 dB

Figure 5.9: Comparison of performance on a selected portion of (a) the Straw image, (e) the Grass image and (i) the House image.

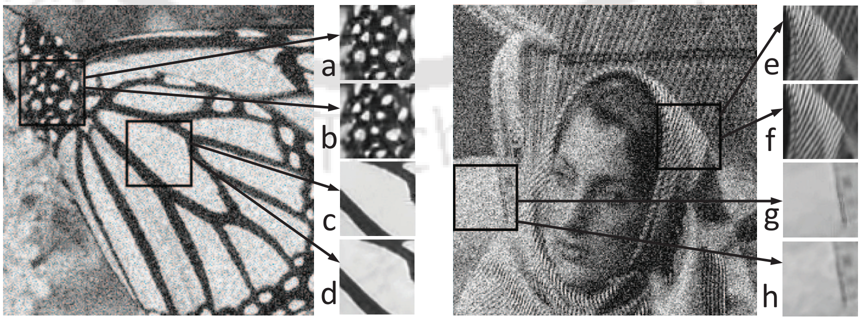


Figure 5.10: Comparison of local denoising performance on chosen image segments at noise level $\tau = 35$. Monarch image segments denoised by (a) BM3D: 22.95 dB, (b) FRIT-CF: 23.28 dB, (c) BM3D: 29.33 dB, and (d) FRIT-CF: 28.67 dB. Barbara image segments denoised by (e) BM3D: 24.57 dB, (f) FRIT-CF: 24.97 dB, (g) BM3D: 29.61 dB, and (h) FRIT-CF: 29.05 dB.


Noisy: 18.60 dB BM3D: 27.76 dB WNNM: 28.34 dB LRCF: 28.15 dB LRASVD: 28.05 dB SVLET−CF: 27.90 dB

Noisy: 17.24 dB BM3D: 29.27 dB WNNM: 29.42 dB LRCF: 29.44 dB LRASVD: 29.36 dB SVLET−CF: 29.43 dB

Noisy: 18.61 dB BM3D: 29.05 dB WNNM: 29.42 dB LRCF: 29.34 dB LRASVD: 29.03 dB SVLET−CF: 29.14 dB

Noisy: 18.57 dB BM3D: 26.23 dB WNNM: 26.40 dB LRCF: 26.36 dB LRASVD: 26.20 dB SVLET−CF: 26.25 dB

Figure 5.11: Comparison of qualitative and quantitative denoising performance on different images at noise level $\tau = 30$.

5.4 Results and discussion

In order to validate the proposed methods, the experiments are performed on various gray-scale images taken from the USC-SIPI[1] image database. The noisy images are obtained by adding zero mean white Gaussian noise of different known variances to these images. The denoising is then performed using different methods and their quantitative performances are compared in terms of the PSNR. We considered only the contrast methods for which the executable codes were made available by their authors and experiments were carried out with the parameter values suggested by them. All the performances reported in this work were averaged over 10 realizations of noise for all the methods.

---

[1] http://sipi.usc.edu/database/

In Table 5.2, we report the quantitative denoising performance of the proposed both methods (FRIT-CF and SVLET-CF) and the contrast methods on some of the selected images from the database. We compare the denoising performances under low-noise ($\tau \leq 20$), mid-noise ($20 < \tau < 50$), and high-noise ($\tau \geq 50$) levels. Form the table, we can draw a number of conclusions. The overall performance of the FRIT-CF method is quite comparable to that of the BM3D method. However, it was expected to consistently outperform the BM3D but it does that only for the texture dominated 'Straw' images for low and mid-noise levels. This fact is also supported by the qualitative assessment of the denoised images shown in Figure 5.9. The PSNR of the denoised image by the FRIT-CF is significantly higher than that of the BM3D method in case of the Straw and the Grass images. However, in case of the House image the reverse trend is noticed. This could be attributed to the presence of smooth structures in it. To confirm this, we go one level down by closely examining the denoising performance of these methods in a local region. For instance, in Figure 5.10, we can see that the FRIT-CF method outperforms the BM3D in texture dominated regions whereas in smooth regions the opposite is true. The rational behind the inferior performance in the smooth regions could be the wrap-around artifacts introduced by the FRIT algorithm as discussed earlier. In case of high-noise levels, the preservation of such a complex texture is very difficult for any method. The DWT employed in the BM3D method, however, maintains the overall intensity level. It is reflected by comparatively higher PSNR in almost all the considered images.

Table 5.2 presents a through comparison of denoising performances of the proposed FRIT-CF and the SVSE-based denoising methods with that of the BM3D method over a set of images and different noise levels. The visual performance comparison of these method can be seen in Figure 5.11. For the set of examples shown in the figure, it can be noted that the proposed SVLET-CF method is able to perform better than the BM3D and the proposed FRIT-CF methods. This improvement in the performance can be attributed to the better sparsity of the SDMs in the data-dependent basis learned by the SVD and the adaptive shrinkage of SVs used in the SVLET-CF. On overall performance comparison, the WNNM and the LRCF methods can be seen to be very closely performing methods and they dominate over the other methods.

5.5 Summary

In this chapter, we have first presented a denoising approach using the FRIT in the CF framework. On comparison, the proposed FRIT-CF approach is found to achieve significantly improved denoising performance over the BM3D method for highly textured images. The ridgelet transform is able to compete with the wavelet transform in the current era of image restoration. For non-textured images, the proposed method is found to be slightly inferior to the BM3D algorithm in homogeneous regions only. This degradation is possibly due to the "wrap-around" artifacts introduced by the implementation of the FRIT.

Another contribution lies in the exploration of various SV shrinkage based matrix estimation methods in the CF framework for image denoising. The proposed SVLET shrinkage estimator derived in Chapter 3 is found to consistently outperform the BM3D and FRIT-CF methods.

6

Application to MR Image Denoising

Contents

6.1 Denoising of MR images 94
6.2 Noise in MR image 97
6.3 Proposed MR image denoising method 99
6.4 Experimental set-up and parameter tuning 103
6.5 Results and discussion 113
6.6 Summary 115

Magnetic resonance imaging has been one of the most popular diagnostic modalities to visualize the physiological activities of tissues or organs inside the human body. During quantitative analysis, the magnetic resonance (MR) image data is subject to a number of processing such as registration, segmentation, and classification. Despite significant improvements in the acquisition systems, the obtained MR images suffer from different kinds of noise which can hamper the subsequent processing and lead to misdiagnosis. Hence, denoising forms an integral part of a typical image processing work involved in quantitative analysis of the MR images.

In this chapter, we explore the effectiveness of the proposed SVLET estimator in the CF framework for denoising of noisy MR images. In addition, we also employ the existing OptShrink estimator for this task. Thus, this chapter presents two competitive methods for MR image denoising. The experimental studies carried on the simulated BrainWeb MR database show the effectiveness of both the methods.

6.1 Denoising of MR images

An MR image is a three dimensional (3D) complex valued data in which each slice provides information about the organ(s) under test from a specific viewing angle. It is common practice to model the noise in the real and imaginary parts of the MR image to be AWGN. During the acquisition of an MR image, the noise can be suppressed by time-averaging without losing the spatial resolution. In time-averaging, each slice is replaced by the average of the multiple acquisitions from the same viewing angle. In order to get high resolution 3D representation of the tissues, one needs to acquire a large number of slices. Therefore, such averaging is not only costly but also prone to motion artifacts. The non-viability of such time-averaging necessitates the development of suitable algorithms for the denoising of the MR images. Spatial-averaging where each volume element, popularly known as voxel, is replaced by the average value of the voxels in its 3D spatial neighborhood has also been explored for the denoising of the MR images. A common example is denoising by the anisotropic diffusion filter [117]. It is widely used denoising method for its simplicity and computational speed. However, the spatial methods assume the underlying signal to be stationary as well as ergodic in the mean and the variance. The MR images exhibit strong non-stationarity characteristics and hence applying such filtering degrades the spatial-resolution.

To visualize an MR image, the voxel-wise absolute value is computed to derive the magnitude image. As the computation of absolute value is a non-linear operation, the distribution of noise present in the magnitude image no longer remains Gaussian [118, 119]. This adds another layer of difficulty in the MR image denoising. Sijbers *et al.* in [120] modeled the noise in the MR image as the Rician distributed random variable. In low signal intensity regions of the MR image, the Rician noise tends to follow the Reyleigh distribution, whereas in high intensity regions it tends to follow the Gaussian distribution. As a result, a signal dependent bias is introduced in the denoised image and hence its contrast is altered. To deal with this difficulty, Nowak [121] proposed a bias correction strategy while treating the noise in the MR image as Gaussian. In [122], an anisotropic diffusion filter, originally developed for the uniform noise level, is locally adapted using the prior information about the spatial distribution of varying noise levels. In [123], a robust noise estimation method was proposed to automatically set the parameters of the combined locally linear mean squared error (LLMSE) filter and the anisotropic diffusion filter for the MR image denoising. These spatial domain global filters suppress the noise well but smooth the signal details. In the MR images, these minute anatomical details of each slice containing important diagnostic features should necessarily be preserved. Inspired by the NLM [83] image denoising method, Majón *et al.* [124] proposed an MR image denoising method by exploiting the intra-slice correlation. However, this natural extension of the NLM method for slice-by-slice denoising of the MR images is very slow. In [125], Coupé *et al.* presented an optimized 3D implementation of the NLM filter for the denoising of the MR images. Following this, several NLM based MR image denoising methods have been proposed [126–129]. Another group of works for the MR image denoising are inspired by the DWT-domain image denoising methods. These methods exploit the sparsity of the original MR image in the DWT-domain. The noise-free DWT coefficients are estimated by thresholding the DWT coefficients of noisy MR image [90]. In [130], a preliminary classification of the DWT coefficient is proposed to identify the coefficients corresponding to the noise-free signal. In [131], the Canny edge detector is used in the DWT domain for edge preserving while applying thresholding on the DWT coefficients to suppressing the noise. In [132], the MR sinogram image obtained by the Radon transform of noisy image is denoised to deal with the Rician nature of the noise in the original MR image. A translation invariant DWT is used for multi-scale decomposition of the MR sinogram and then thresholding based noise suppression is applied. Another group of works exist which treats squared magnitude MR image as non-central scaled chi-square distribution with two degrees of freedom whose non-centrality parameter is proportional to the underlying noise-free squared magnitude. For example in [133], authors proposed the unbiased risk estimator for chi-square distribution and used the linear expansion of threshold (LET) strategy to denoise the MR image in wavelet domain.

Recent advancement in image denoising exploits the correlation of small image patches and denoising is performed jointly in spatial and transform domains. For example, in [134], an extended version of the BM3D, called the BM4D, is proposed for 3D MR image denoising. As BM3D/BM4D method uses fixed bases DWT, the learning of the bases for better sparsity of the 3D groups and hence better denoising has also been obtained [107]. In [135–137], the MR image denoising methods exploiting the sparse representation over learned dictionary are proposed. However, these methods rely on the sparsity prior of image patches or groups formed of image patches. Learning the dictionary for each group entails a huge computational burden. As a compromise between sparsity and computational cost, all of these methods learn a global dictionary for the sparse representation. As an alternative, low-rank priors for image denoising have also been explored. In [87], the HOSVD is used to decompose these 3D groups and denoising is performed by SVHT with a predetermined threshold. In recent times, low-rank modeling has been widely used for the recovery of the MR image from the reduced k-space samples as well as for the denoising of k-space or $(k,\ t)$-space raw data [138, 139]. In these applications, a low-rank prior is posed on a specially constructed Hankel or Casorati matrices. For the further details, readers may refer to [140] and the references therein.

A very few works in the literature use low-rank modeling for the denoising of MR magnitude image. In [138], similar patches extracted from the multichannel MR image is grouped together to form a low-rank matrix and denoising is performed by the iterative hard-thresholding (HT) of the SVs of the group matrices. In [141], similar to 2D patches, 3D cubes extracted from a noisy MR image volume are grouped together to form the hypercubes. The HOSVD is used for the spectral decomposition of these hypercubes and denoising is performed by hard-thresholding the higher order SVs. Recently, in [142], these hypercubes are treated as fourth order tensors and denoising is performed by the adaptive HOSVD of their tensor unfoldings. In most of the

SVD based methods, noise suppression is achieved by hard-thresholding of SVs of the group matrices (or tensors) with a predefined threshold. The MR image denoising methods based on the SVHT, however, have the following drawbacks:

- The truncated SVs represent the rare details which may correspond to important diagnostic features. Hence, throwing away these smaller SVs may result in the loss of diagnostic details.
- The noise being white Gaussian is spread over the entire spectrum. The SVHT does not remove the noise present in larger SVs.
- Each group in CF may have different SV profile that ideally should be thresholded with different thresholds. However, determining the optimal threshold for each group by the grid search methods is quite cumbersome task, whereas using a single pre-specified global threshold makes the process non-adaptive.

In order to deal with these limitations, we propose two methods for the denoising of the MR images. Both the proposed methods use adaptive shrinkage of SVs for the denoising of each group matrix in the CF framework. Specifically, in the first method, the noise filtering in the CF is performed by the SVLET method as opposed to the hard-thresholding of the DWT coefficients employed in the BM4D method. This method is referred to as the SVLET-CF method. In the second method, the noise filtering is done by the OptShrink estimator and is referred to as the OptShrink-CF method. In fact, the SVLET and the OptShrink are originally developed for matrix denoising under AWGN. To tackle with data dependent Rician noise in the MR images, a variance stabilization transform [143] is used before and after applying either the SVLET-CF or the OptShrink-CF method. The denoising performances of the proposed methods are compared with that of the state-of-the-art methods in terms of the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM).

6.2 Noise in MR image

The main source of noise in the MR image is the subjects being imaged (Brownian motion of ions in bodily fluids) and the noise is captured by the radio frequency (RF) coils where thermal noise is generated due to the stochastic motion of free electrons. The thermal noise, commonly modeled as Gaussian distribution, corrupts the k-space data. On taking the inverse Fourier transform of the k-space data, the real and the imaginary MR images are obtained. The Fourier transform being a linear operator does not change the distribution of noise. Hence, the distribution of the noise in the spectral domain remains Gaussian. However, this is not the case in the magnitude MR images.

Let us denote a noisy magnitude MR image as $\widetilde{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and its vectorized version as $\tilde{\mathbf{y}} \in \mathbb{R}^{I_1 I_2 I_3}$. Also, assume a voxel at an arbitrary spatial location i from that volume is denoted by $\tilde{y}_i$ where $1 \leq i \leq I_l I_2 I_3$. Hence, $\tilde{y}_i$ follows the observation model

$$\tilde{y}_i = \sqrt{(y_i + \tau \bar{y}_i)^2 + (y_i + \tau \bar{y}_i)^2}. \tag{6.1}$$

where y_i denotes the magnitude of underlying noise-free MR image voxel and $\bar{y}_i$ is the zero mean unit variance Gaussian noise sample, i.e., $\bar{y}_i \sim \mathsf{N}(0,1)$ with an arbitrary constant $\tau > 0$. Clearly, $\tilde{y}_i$ follows the Rician distribution [119], i.e., $\tilde{y}_i \sim \mathsf{R}(y_i, \tau)$. Dropping location subscript i, we have the probability distribution function of $\tilde{y}$ given as

$$p(\tilde{y}|y,\tau) = \frac{\tilde{y}}{\tau^2} \exp\left\{-\frac{\tilde{y}^2 + y^2}{2\tau^2}\right\} B_0\left(\frac{\tilde{y}y}{\tau^2}\right) \quad \forall\ \tilde{y} \geq 0 \text{ and } y > 0, \tag{6.2}$$

where B_0 is the zeroth order modified Bessel function [144].

6.2.1 Variance stabilization

The mean and the variance of the Rician data y are, respectively, given as

$$E[\tilde{y}|y,\tau] \;=\; \tau\sqrt{\frac{\pi}{2}}\, L\left(-\frac{y^2}{2\tau^2}\right), \tag{6.3}$$

and

$$Var[\tilde{y}|y,\tau] \;=\; 2\tau^2 + y^2 - \frac{\pi\tau^2}{2}\, L^2\left(-\frac{y^2}{2\tau^2}\right), \tag{6.4}$$

where $L(z) = e^{z/2}\left[(1-z)B_0(-\frac{z}{2}) - zB_1(-\frac{z}{2})\right]$ where B_1 is the first order modified Bessel function. From (6.3) and (6.4), it is clear that neither the mean nor the variance is independent of underlying noise-free signal y. As a result, applying the denoising methods for the AWGN directly on the magnitude MR image introduces a bias and yields poor denoising performance.

To overcome this problem, a three-level denoising method is suggested in [143]. In the first level, the variance stabilization transform (VST) is applied on the noisy MR image to obtain a uniform variance in the transformed MR image. The noise in such variance stabilized image can be assumed to be approximately Gaussian. In the second level, the VST stabilized MR data is denoised by using an AWGN removal method. For example, in [143], the BM4D method is used. Further, the inverse VST is applied on this image to obtain the final denoised MR image. Mathematically, this three-level MR image denoising approach can be expressed as

$$\{\tilde{\mathbf{y}}_{\mathrm{VST}}, \tau_{\mathrm{VST}}\} = \mathrm{VST}(\tilde{\mathbf{y}}, \tau), \tag{6.5}$$

$$\hat{\mathbf{y}}_{\mathrm{VST}} = \mathrm{CF}(\tilde{\mathbf{y}}_{\mathrm{VST}}, \tau_{\mathrm{VST}}), \tag{6.6}$$

$$\text{and} \qquad \hat{\mathbf{y}} = \mathrm{VST}^{*}(\hat{\mathbf{y}}_{\mathrm{VST}}, \tau), \tag{6.7}$$

where $\mathrm{VST}(\cdot)$ and $\mathrm{VST}^{*}(\cdot)$, respectively, denote the forward and the inverse variance stabilization transform operator, τ_{VST} is the stabilized standard deviation, and the $\mathrm{CF}(\cdot)$ denotes the collaborative filtering operator. The block diagram of the three-level approach can be seen in Figure 6.1. The variance stabilized MR image obtained on applying the VST can be loosely treated as if it were corrupted with additive Gaussian noise. We can write $\tilde{\mathbf{y}}_{\mathrm{VST}} = \mathbf{y}_{\mathrm{VST}} + \bar{\mathbf{y}}_{\mathrm{VST}}$. Hence, we adopt this approach in both the proposed methods to be able to apply the SV shrinkage based matrix estimators for the denoising of MR image under Rician noise model.

6.3 Proposed MR image denoising method

The proposed denoising method uses the CF framework, which was originally proposed in [143] to exploit the intra- and inter-block correlations present in the MR image. Here, we modify the CF by making use of the adaptive SVSE instead of the HT of the 3D DWT coefficients. Moreover, unlike in the existing CF based MR image denoising approaches, we iterate the process a couple of times while updating the noise level estimate before obtaining the final denoised image. Figure 6.1 shows the block diagram of the CF for MR image denoising where iterative CF block is identical to the one shown in Figure 5.3 for 2D image. In the following, we discuss the stages involved in its first step of the proposed MR image denoising methods.

Figure 6.1: Block diagram of collaborative filtering for MR image denoising. The iterative CF block in the above is same as the one in Figure 5.3.

6.3.1 Block grouping

To exploit the non-local correlation of small 3D blocks of the MR image, these are grouped together. For that, the entire MR image volume of size $I_1 \times I_2 \times I_3$ is divided into several smaller blocks of size $n_1 \times n_2 \times n_3$ with an overlap of p voxels. Thus, the total number of such blocks turns out to be $\left(\frac{I_1-n_1}{n_1-p}+1\right)\left(\frac{I_2-n_2}{n_2-p}+1\right)\left(\frac{I_3-n_3}{n_3-p}+1\right)$. These blocks are then vectorized in the form $n_1 n_2 n_3 \times 1$ vectors and are referred to as the *reference-blocks*. A group of similar blocks are formed for each of the reference block using block-matching. For the same, the blocks similar to a reference block are searched in its neighborhood of size $N_1 \times N_2 \times N_3$. Thus, for an ith reference block we have $(N_1 - n_1 + 1)(N_2 - n_2 + 1)(N_3 - n_3 + 1) - 1$ number of blocks in its neighborhood for finding the similarity. The Euclidean distance between block vectors is taken as the measure of similarity. For each of the reference block, the m most similar vectorized blocks including itself are selected to form an SDM of size $n_1 n_2 n_3 \times m$. We denote this grouping process by an operator $\mathcal{G}_i$ which is applied on variance stabilized MR image vector $\tilde{\mathbf{y}}_{\text{VST}}$ to extract an SDM corresponding to the ith reference-block. Thus, $\widetilde{\mathbf{X}}_i = \mathcal{G}_i \tilde{\mathbf{y}}_{\text{VST}}$ and similarly $\mathbf{X}_i = \mathcal{G}_i \mathbf{y}_{\text{VST}}$ where $\mathbf{X}_i$ is the corresponding noise-free SDM. Thus, we have $\widetilde{\mathbf{X}}_i = \mathbf{X}_i + \overline{\mathbf{X}}_i$ where $\overline{\mathbf{X}}_i$ is the noise matrix. We can consider each element of the noise matrix as zero mean Gaussian distributed random variable with the standard deviation τ_{VST}. The MR image denoising problem is now turned into matrix estimation problem that can be solved by the SVSE methods.

6.3.2 Noise filtering

In this stage, $\widetilde{\mathbf{X}}_i$ are denoised by a suitable SVSE method. If the blocks grouped together were noisy versions of the exact copies of noise-free blocks, then the rank one approximation of this matrix would have resulted in an effective denoising of the group. In practice, all columns of the SDMs are not be identical but they are highly correlated. This correlation can be exploited by solving a rank regularized problem to find the denoised estimate

$$\widehat{\mathbf{X}}_i = \min_{\mathbf{X}_i} \|\mathbf{X}_i - \widetilde{\mathbf{X}}_i\|_{\text{Frob}}^2 + \mu \ \text{Rank}(\mathbf{X}_i), \tag{6.8}$$

where μ is the coefficient of regularization. An approximate solution to the above non-convex problem, popularly obtained by the SVHT [57], is given as

$$\widehat{\mathbf{X}}_{i,\,\mu}^{\text{SVHT}} = \sum_{j=1}^{L} \tilde{\sigma}_j \mathbf{1}_{(\sigma_j > \mu)} \tilde{\mathbf{u}}_j \tilde{\mathbf{v}}_j^T, \tag{6.9}$$

where $\mathbf{1}_{(\cdot)}$ is the indicator function $L = \min(n_1 n_2 n_3, m)$ and $\mathbf{X}_i = \sum_{j=1}^{L} \tilde{\sigma}_j \tilde{\mathbf{u}}_j \tilde{\mathbf{v}}_j^T$. As discussed earlier, the SVHT has several limitations and to overcome those, various SVSE methods have been discussed in Chapter 3. In a typical MR image denoising task, hundreds of thousands of SDMs are formed. Therefore, to denoise them, the applied SVSE methods need not only be accurate but also fast. Towards this, in the following subsections, we propose to use two different SV shrinkage approaches.

6.3.2.1 Noise filtering by SVLET

It is highlighted earlier that the SVLET has very low computational cost than the existing SVSE methods. For this reason, we propose to use the SVLET for the denoising of SDMs using (4.12). For the quick reference, the same is reproduced as

$$\widehat{\mathbf{X}}_i^{\text{SVLET}} = \sum_{j=1}^{L} f_{T,K}^{\text{SVLET}}(\tilde{\sigma}_j) \tilde{\mathbf{u}}_j \tilde{\mathbf{v}}_j^{\text{T}} \tag{6.10}$$

where $f_{T,K}^{\text{SVLET}}(\cdot)$ denotes the SVLET shrinkage operator and rest of the notations have usual meanings.

6.3.2.2 Noise filtering by OptShrink

The OptShrink is a matrix estimation method which does not involve any parameter tuning. As a result, it is suitable candidate for the MR image denoising. In our earlier explorations, it is also found to result in a competitive denoising performance when compared with that of the SVLET method. Thus, as a second approach, we propose to apply the OptShrink in place of the SVLET. The denoised SDMs in this case are obtained as

$$\widehat{\mathbf{X}}_i^{\text{OptShrink}} = \sum_{j=1}^{L} f^*_{\beta_+}(\tilde{\sigma}_j)\tilde{\mathbf{u}}_j\tilde{\mathbf{v}}_j^{\mathrm{T}} \tag{6.11}$$

where $f^*_{\beta_+}(\cdot)$ denotes the OptShrink operator and rest of the notations have usual meanings.

6.3.3 Aggregation

As the groups are formed in an overlapping manner, a voxel or a group of voxels can be a present in several SDMs. Thus, multiple estimates for a voxel are obtained. These estimates are then used to calculated the final estimate by the weighted averaging given as

$$\hat{\mathbf{y}}^{(l)}_{\text{VST}} = \sum_{\forall i} \frac{w_i \mathcal{G}_i^* \widehat{\mathbf{X}}_i}{\mathcal{G}_i^* \mathcal{G}_i} \tag{6.12}$$

where $\mathcal{G}_i^*$ denotes the inverse of the block-grouping operator $\mathcal{G}_i$. In the BM4D method, the weights w_i is set equal to the confidence in estimating the ith group. However, for simplicity, we take it to be unity, i.e., $w_i = 1$ for all i. Thus, in the proposed method, the aggregation is a simple averaging of all the available estimates for a voxel.

At the end of the first step, we get the basic estimate that is further refined in the second step. In the second step, the block-grouping is guided by this basic estimate and the formed SDMs are denoised using the 3D DWT domain Wiener filtering. The aggregation step is kept the same as in the first step. Thus, the second step of the proposed SVLET-CF is identical to the second step of the BM4D method. At the end of the second step, we obtain a better estimate of the MR image than the first step. Further enhancement in the denoised image can be achieved by iterating the entire process as follows.

Unlike the first iteration that takes the variance stabilized noisy image $\tilde{\mathbf{y}}_{\text{VST}}$ as the input, the subsequent iteration of the proposed SVLET-CF takes the partially denoised image obtained using back-projection as input. At lth iteration, the back-projected image is obtained using (6.13). For the ease of reference, the same can be reproduced as

$$\tilde{\mathbf{y}}^{(l)}_{\text{VST}} = \hat{\mathbf{y}}^{(l-1)}_{\text{VST}} + \delta(\tilde{\mathbf{y}}_{\text{VST}} - \hat{\mathbf{y}}^{(l-1)}_{\text{VST}}). \tag{6.13}$$

Also, the standard deviation of noise for the subsequent iterations is updated as

$$\hat{\tau}_{\mathrm{VST}}^{(l)} = \gamma\sqrt{\tau^2 - \frac{1}{N}\|\tilde{y}_{\mathrm{VST}} - \tilde{y}_{\mathrm{VST}}^{(l)}\|_2^2}. \tag{6.14}$$

This process is iterated a couple of times and finally the denoised MR image is obtained. For being free from any tunable parameter, the OptShrink is the fastest matrix denoising method.

6.4 Experimental set-up and parameter tuning

To evaluate the proposed MR image denoising methods, experimental studies are done on the simulated MR images from BrainWeb corpus [1]. It is well accepted and widely used database for validating the MR image restoration algorithms. It contains two anatomical model: normal brain simulator and multiple sclerosis (MS) lesion brain simulator. Each of them can simulate the MR images with three modalities: T1-, T2-, and PD-weighted and with various slice thicknesses and intensity non-uniformity levels. In our experimental studies, we consider one MR image volume from each of the anatomical models and modalities with the slice thickness of 1 mm and the non-uniformity level of 20%. The noisy MR images are obtained by adding the WGN according to the observation model in (6.1) with a chosen value of noise level τ. Various experiments on the test data are performed to establish the efficacy of the proposed methods.

6.4.1 Performance measures

To measure the denoising performance of the proposed and the contrast methods, we use the following quantitative performance metrics.

(i) **Peak Signal-to-Noise Ratio (PSNR)**: The PSNR defined for image in the last chapter is extended to the MR image considering its three dimensions. It measures the overall denoising performance in terms of mean squared error. For an MR image with the normalized voxel intensity level in $[0\ D]$, the PSNR is given as

$$\mathrm{PSNR}(\mathbf{y}, \widehat{\mathbf{y}}) = 10\,\log_{10}\left\{\frac{D^2}{MSE}\right\} \tag{6.15}$$

$$= 10\,\log_{10}\left\{\frac{D^2}{\frac{1}{I_1 I_2 I_3}\sum_{i=1}^{I_1 I_2 I_3}(y_i - \hat{y}_i)^2}\right\}. \tag{6.16}$$

---

[1] http://brainweb.bic.mni.mcgill.ca/brainweb/

Clearly, a better denoising method should result in a higher PSNR value.

(ii) **Structural Similarity Index Measure (SSIM) [145]**: In an image, the SSIM accounts the properties of human visual system and measures the structural similarity of the denoised image with the underlying noise-free image. For the MR image, it is given as

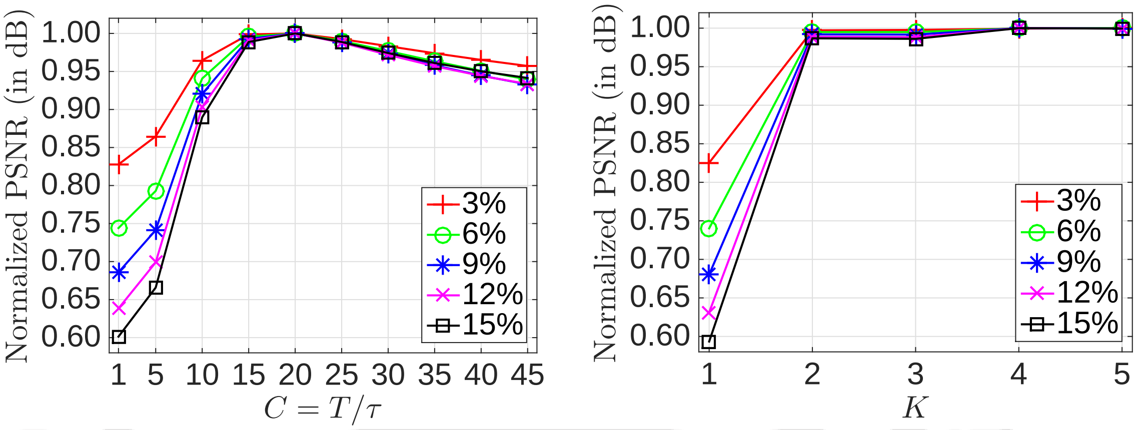
$$\text{SSIM}(\mathbf{y}, \widehat{\mathbf{y}}) = \frac{1}{I_1 I_2 I_3} \sum_{i=1}^{I_1 I_2 I_3} \frac{(2\mu_{y_i}\mu_{\hat{y}_i} + C_1)(2\tau_{y_i\hat{y}_i} + C_2)}{(\mu_{y_i}^2 + \mu_{\hat{y}_i}^2 + C_1)(\tau_{y_i}^2 + \tau_{\hat{y}_i}^2 + C_2)} \tag{6.17}$$

where μ_{y_i} and $\mu_{\hat{y}_i}$ are the local mean values of $\mathbf{y}$ and $\widehat{\mathbf{y}}$, respectively, whereas τ_{y_i} and $\tau_{\hat{y}_i}$ are corresponding local variances is used. The $\tau_{y_i\hat{y}_i}$ is the cross-covariance value whereas C_1 and C_2 are the small constants to provide stability in case of zero local means and variances. For a better denoising method the SSIM should be closer to unity. The SSIM is commonly assumed to be a better performance measure than the PSNR in terms of visual quality.

6.4.2 Selection of parameters

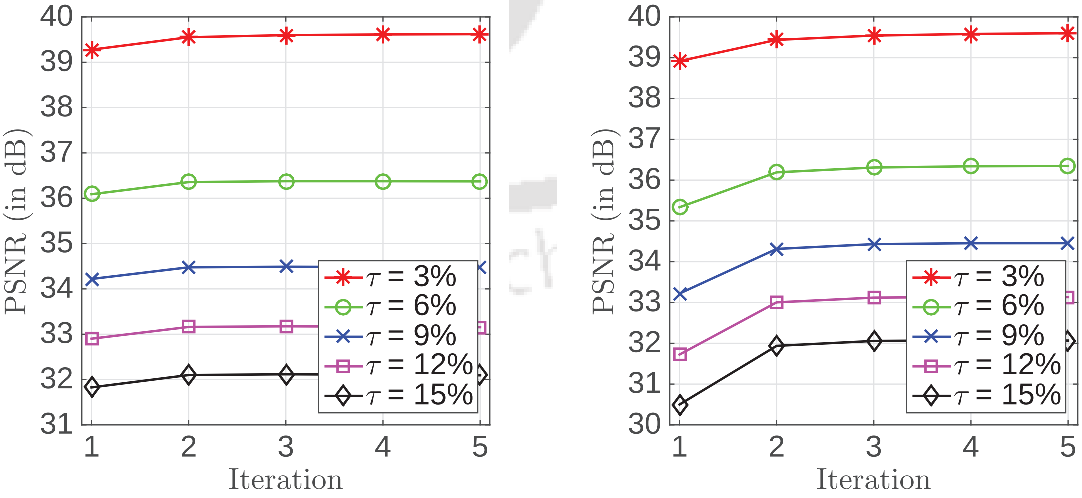
The proposed methods have various common parameters such as the block-size, the search-window size, and the overlap between successive reference blocks. As the similar 3D blocks of the noisy MR image are vectorised to be arranged as the columns of the data matrix, the block-size parameter determines the number of rows of the data matrix. With a larger block-size, it is difficult to find many similar blocks and hence the corresponding data matrix is not sufficiently low-rank. A very smaller blocks result in increased number of data matrices for a given MR volume, hence the denoising becomes slow. This is why a moderate choice of block-size is crucial to make. The search-window size is chosen small enough to reduce the search time and big enough to find a sufficiently large number of similar blocks. The overlap between successive reference blocks is required to get rid of blocking artifact in the denoised image. For a fair comparison with the BM4D method, the block-size, the search-window size, and the overlap between successive reference blocks are set at $4 \times 4 \times 4$, $11 \times 11 \times 11$, and 1, respectively. In the BM4D method, the group size (the number of similar blocks in a group) is decided by the matching threshold. However, in the proposed method, the group size m is set at 64 in order to keep the aspect ratio β of the SDM to 1.

Additionally, the proposed SVLET-CF method uses two specific parameters K and C, that are experimentally fixed. For that, we apply one step denoising (that includes block-grouping, group-denoising, and aggregation) on an MR image volume derived from the above database at different noise levels. In Figure 6.2(a), the normalized PSNRs are plotted against C, where $T = \tau C$, for denoising a noisy MR image at different noise levels. It can be seen that the best choice for C is approximately equal to 20. With this value of C, Figure 6.2(b) shows the plots of the normalized PSNR against K at different noise levels. From Figure 6.2(b), the value of K is chosen as 4.



(a) (b)

Figure 6.2: Tuning of parameter T and K of the SVLET-CF on an MR image volume cropped from the 'normal' SBD at different noise levels.



(a) SVLET-CF (b) OptShrink-CF

Figure 6.3: Effect of the iteration on the PSNR of denoised MR image corresponding to a volume cropped for the 'normal' case from BrainWeb corpus at different noise levels.

Table 6.1: Comparison of performances on 'normal' brain MR image extracted from BrainWeb corpus, in case of Gaussian noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB.

	3%		6%		9%		12%		15%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Modality: T1, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	38.56	0.9682	34.61	0.9183	32.27	0.8647	30.64	0.8116	29.36	0.7620
BM4D [134]:	39.64	**0.9780**	36.30	**0.9559**	34.35	0.9324	32.98	0.9091	31.88	0.8837
OptShrink-CF:	39.66	0.9762	36.39	0.9526	34.51	0.9287	33.20	0.9073	32.16	0.8845
SVLET-CF:	**39.69**	0.9766	**36.42**	0.9547	**34.53**	**0.9328**	**33.23**	**0.9136**	**32.18**	**0.8919**
Modality: T2, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	36.12	0.9680	32.86	0.9207	30.54	0.8602	28.74	0.8048	27.24	0.7487
BM4D [134]:	**37.56**	**0.9792**	**33.91**	**0.9585**	31.87	**0.9364**	30.43	**0.9137**	29.33	0.8901
OptShrink-CF:	37.50	0.9762	33.88	0.9512	31.88	0.9273	30.48	0.9040	29.42	0.8820
SVLET-CF:	37.42	0.9756	33.87	0.9528	**31.91**	0.9315	**30.52**	0.9111	**29.45**	**0.8930**
Modality: PD, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	37.69	0.9619	33.83	0.9017	31.66	0.8402	30.17	0.7786	28.98	0.7217
BM4D [134]:	**39.46**	**0.9764**	35.99	**0.9532**	33.97	**0.9287**	32.56	0.9039	31.46	0.8767
OptShrink-CF:	39.41	0.9742	36.04	0.9477	34.10	0.9227	32.74	0.8990	31.69	0.8737
SVLET-CF:	39.38	0.9742	**36.07**	0.9499	**34.14**	0.9268	**32.78**	**0.9064**	**31.71**	**0.8837**

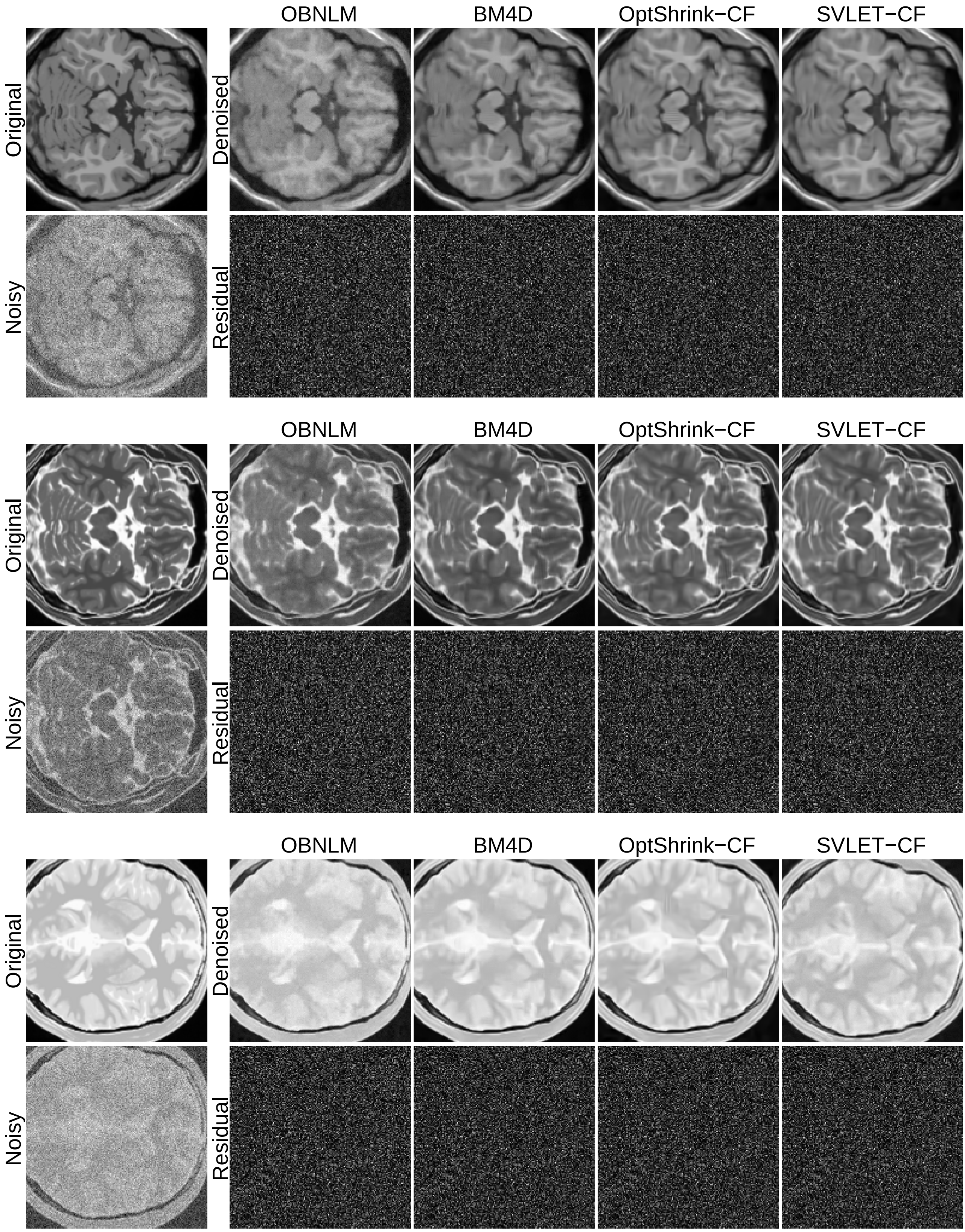


Figure 6.4: Qualitative performance comparison on ‘normal’ brain MR image extracted from Simulated BrainWeb Database, in case of Gaussian noise at 15% noise levels.

Table 6.2: Comparison of performances on 'MS Lesion' brain MR image extracted from BrainWeb corpus, in case of Gaussian noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB.

	3%		6%		9%		12%		15%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Modality: T1, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	38.55	0.9674	34.61	0.9202	32.29	0.8646	30.64	0.8104	29.36	0.7599
BM4D [134]:	39.64	**0.9778**	36.29	**0.9557**	34.37	0.9328	32.97	0.9087	31.89	0.8850
OptShrink-CF:	39.66	0.9762	36.39	0.9520	34.53	0.9291	33.19	0.9064	32.16	0.8848
SVLET-CF:	**39.69**	0.9766	**36.41**	0.9547	**34.56**	**0.9328**	**33.22**	**0.9136**	**32.18**	**0.8919**
Modality: T2, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	36.63	0.9662	33.06	0.9156	30.66	0.8576	28.88	0.7923	27.47	0.7369
BM4D [134]:	**37.97**	**0.9779**	34.40	**0.9561**	32.38	**0.9323**	30.97	0.9082	29.88	0.8833
OptShrink-CF:	37.93	0.9754	34.40	0.9498	32.45	0.9250	31.09	0.9010	30.05	0.8777
SVLET-CF:	37.87	0.9751	**34.41**	0.9518	**32.48**	0.9291	**31.12**	**0.9086**	**30.07**	**0.8892**
Modality: PD, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	38.07	0.9597	34.36	0.9037	32.30	0.8411	30.80	0.7811	29.66	0.7436
BM4D [134]:	**40.01**	**0.9765**	36.52	**0.9541**	34.52	**0.9305**	33.11	0.9060	32.03	0.8814
OptShrink-CF:	39.95	0.9742	36.59	0.9489	34.65	0.9242	33.26	0.8996	32.21	0.8773
SVLET-CF:	39.93	0.9743	**36.61**	0.9511	**34.68**	0.9285	**33.30**	**0.9074**	**32.24**	**0.8888**

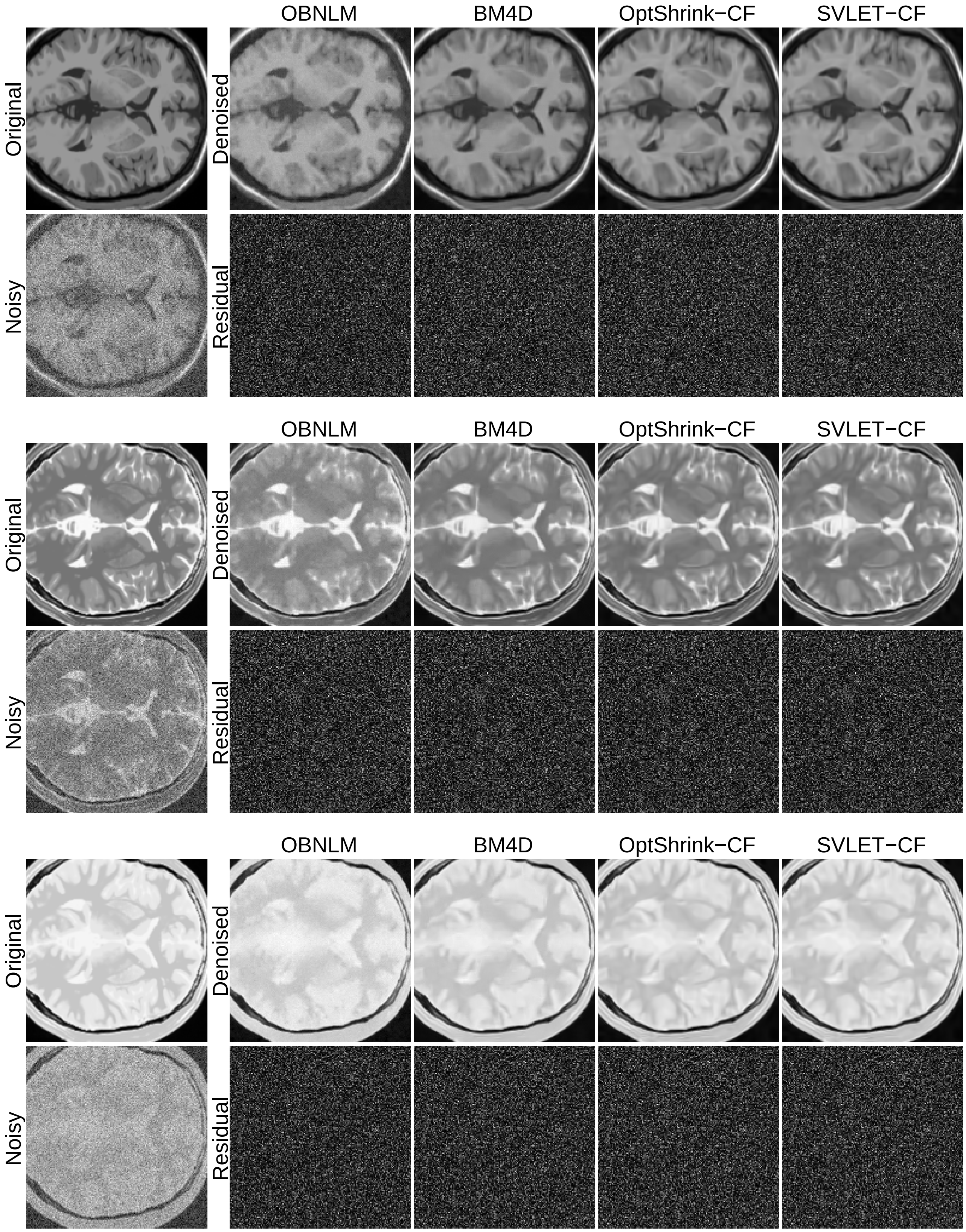


Figure 6.5: Qualitative performance comparison on ‘MS Lesion’ brain MR image extracted from Simulated BrainWeb Database, in case of Gaussian noise at 15% noise levels.

Table 6.3: Comparison of performances on 'normal' brain MR image extracted from BrainWeb corpus, in case of Rician noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB.

	3%		6%		9%		12%		15%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Modality: T1, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	37.94	0.9376	33.51	0.8599	30.67	0.7856	28.57	0.7192	26.87	0.6575
BM4D [134]:	38.60	0.9341	34.25	0.8949	30.83	0.8479	27.78	0.7921	25.12	0.7347
OptShrink-CF:	38.59	0.9383	34.36	0.8954	31.07	0.8514	28.63	0.8014	26.86	0.7470
SVLET-CF:	**38.69**	**0.9451**	**34.59**	**0.8978**	**32.06**	**0.8564**	**30.20**	**0.8092**	**28.67**	**0.7574**
Modality: T2, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	35.60	0.9268	31.86	0.8388	29.20	0.7624	27.06	0.6929	25.33	0.6326
BM4D [134]:	**36.64**	0.8862	32.84	0.8586	**30.38**	**0.8370**	**28.17**	0.8101	25.9588	0.7691
OptShrink-CF:	36.55	0.9278	**32.76**	0.8555	30.30	0.8323	28.09	0.8042	25.95	0.7698
SVLET-CF:	36.41	**0.9323**	32.71	**0.8628**	30.30	0.8358	28.13	**0.8126**	**26.02**	**0.7853**
Modality: PD, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	37.03	0.9218	32.86	0.8245	30.29	0.7439	28.37	0.6799	26.76	0.6229
BM4D [134]:	37.98	0.8793	34.45	0.8476	32.24	0.8254	30.45	0.8047	28.92	0.7859
OptShrink-CF:	37.91	0.9255	34.49	0.8475	32.38	0.8264	30.67	0.8072	29.22	0.7903
SVLET-CF:	**38.16**	**0.9332**	**34.57**	**0.8615**	**32.49**	**0.8286**	**30.78**	**0.8103**	**29.34**	**0.7940**

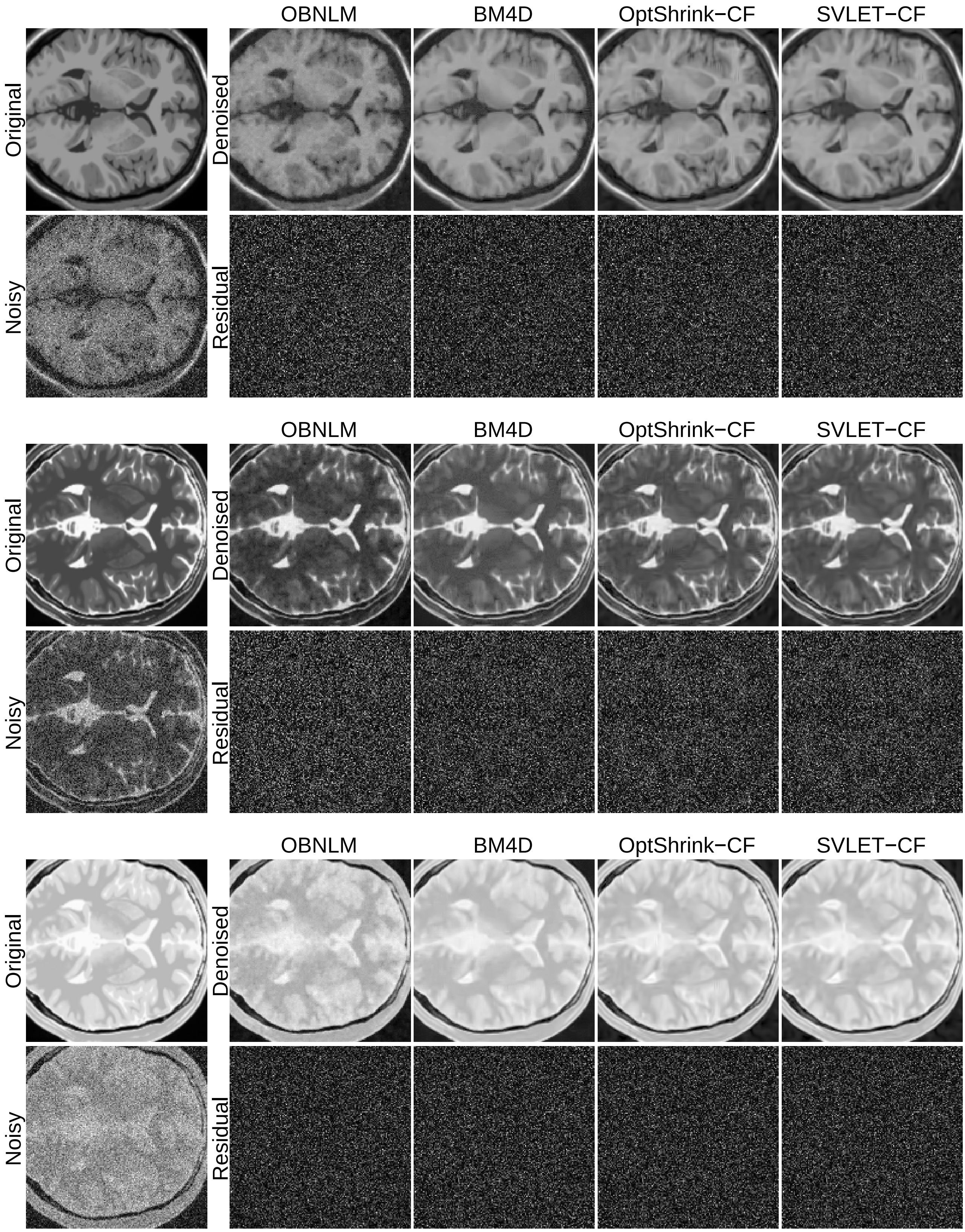


Figure 6.6: Qualitative performance comparison on ‘normal’ brain MR image extracted from Simulated BrainWeb Database, in case of Rician noise at 15% noise levels.

Table 6.4: Comparison of performances on 'MS Lesion' brain MR image extracted from BrainWeb corpus, in case of Rician noise at different noise levels. The best performances in each case are highlighted. All PSNR values are in dB.

	3%		6%		9%		12%		15%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Modality: T1, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	37.94	0.9371	33.49	0.8587	30.66	0.7844	28.60	0.7189	26.87	0.6557
BM4D [134]:	38.60	0.9346	34.20	0.8941	30.76	0.8454	27.61	0.7868	24.96	0.7280
OptShrink-CF:	38.59	0.9377	34.32	0.8947	31.02	0.8493	28.65	0.7969	26.83	0.7395
SVLET-CF:	**38.69**	**0.9444**	**34.57**	**0.8972**	**32.05**	**0.8545**	**30.24**	**0.8048**	**28.66**	**0.7502**
Modality: T2, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	36.09	0.9259	32.14	0.8354	29.42	0.7572	27.36	0.6913	25.68	0.6332
BM4D [134]:	36.89	0.8822	33.28	0.8535	31.03	0.8324	29.26	0.8134	27.67	0.7938
OptShrink-CF:	36.85	0.9289	33.26	0.8522	31.09	0.8313	29.38	0.8130	27.83	0.7936
SVLET-CF:	**36.99**	**0.9344**	**33.32**	**0.8640**	**31.18**	**0.8337**	**29.50**	**0.8162**	**27.96**	**0.7986**
Modality: PD, Slice thickness: 1mm, Intensity non-uniformity: 20%										
OBNLM [125]:	37.34	0.9209	33.26	0.8247	30.72	0.7487	28.76	0.6839	27.12	0.6277
BM4D [134]:	38.34	0.8800	34.84	0.8496	32.60	0.8280	30.82	0.8087	29.27	0.7903
OptShrink-CF:	38.27	0.9241	34.88	0.8492	32.74	0.8283	31.04	0.8104	29.54	0.7928
SVLET-CF:	**38.53**	**0.9331**	**34.97**	**0.8597**	**32.85**	**0.8309**	**31.15**	**0.8137**	**29.65**	**0.7974**

6.4.3 Effect of iterations

In addition to the parameters discussed above, the proposed CF requires another two parameters. These are the back-projection parameters δ and noise update parameter γ. We have experimentally fixed these parameters at 0.05 and 0.4, respectively, that maximize the PSNR. Using these values, Figure 6.3(a) and (b) show the PSNR of the estimated MR image with varying number of iterations for the proposed SVLET-CF and OpriShrink-CF methods, respectively. It can be seen that the PSNR converges around 4th iteration. Hence, in the rest of the experiments we take the denoised image after 4th iteration as the final MR image estimate in both the methods.

6.5 Results and discussion

In this subsection, we present the MR image denoising performance of the proposed and existing methods. The parameters of the proposed methods have been selected as discussed above. For the existing methods, the parameters are chosen as suggested by their authors. The experiments are conducted on the simulated MR images derived from the BrainWeb at different noise levels. Recently, a few competitive MR image denoising methods are reported [125, 134, 141, 142]. We consider the OBNLM [125] and the BM4D [134] methods for comparison. The availability of the executable MATLAB codes enables quick and accurate performance comparison. The parameters of these methods are chosen as suggested by their authors.

In Table 6.1, we show the quantitative performance of the proposed and the chosen contrast methods on the normal brain MR image consisting of all three different modalities at varying Gaussian noise levels. For the Gaussian noise removal, the VST is not needed, hence only the iterative CF block of the block diagram in Figure 6.1 is applicable. The standard deviation τ corresponding to a given percentage of noise level p can be computed as $\tau = D \times p/100$. In addition, we also show the corresponding visual performance comparison at a chosen noise level in Figure 6.4. The residual images are obtained by subtracting the denoised image from its noisy counterpart. In the best case, the voxels of the residual image should have no correlations.

The denoising performances on the MS lesion MR images are listed in Table 6.2 while the visual comparison is shown in Figure 6.5. On comparing the quantitative performances, we

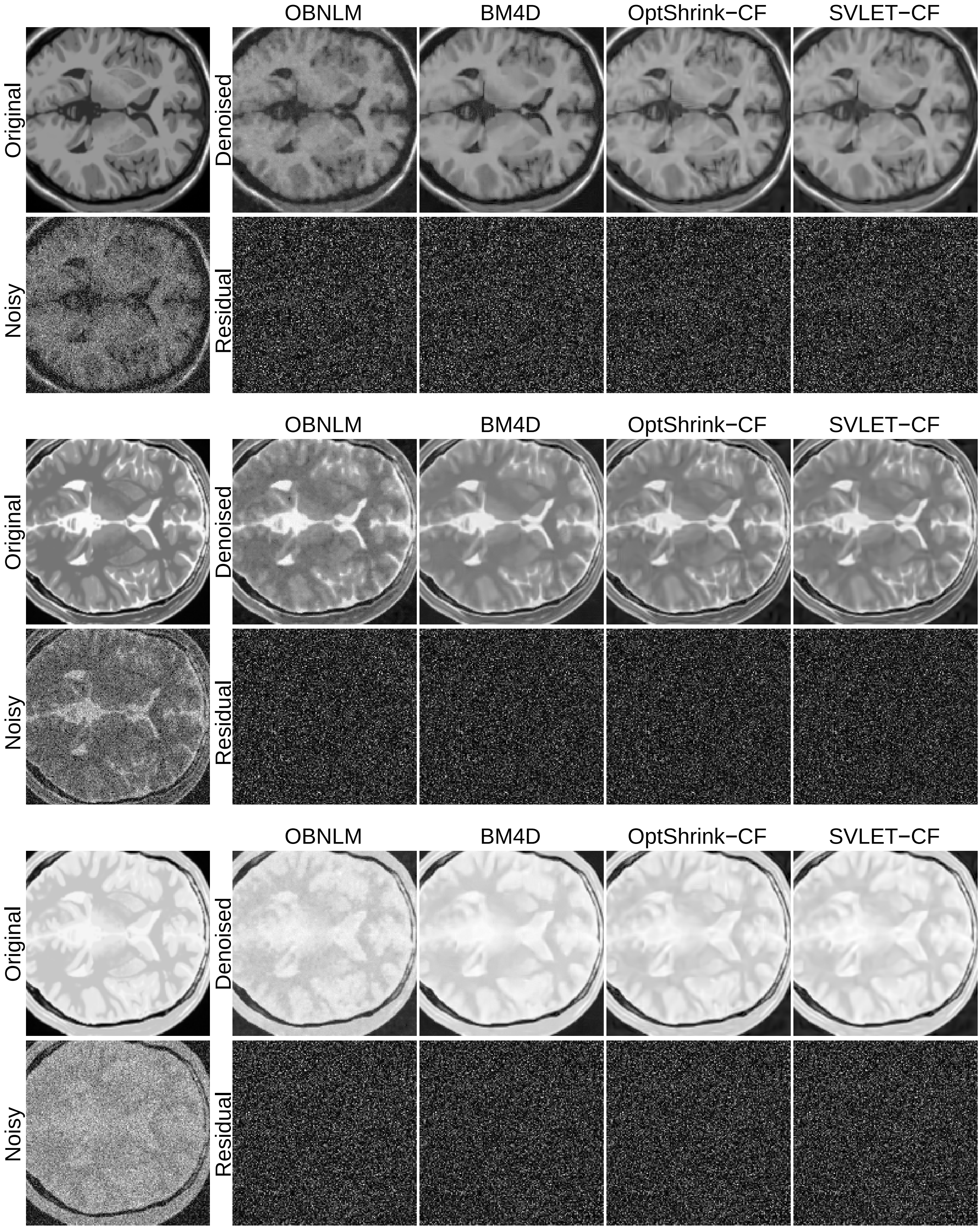


Figure 6.7: Qualitative performance comparison on ‘MS Lesion’ brain MR IMAGE data extracted from Simulated BrainWeb Database, in case of Rician noise at 15% noise levels.

note that the proposed methods outperform both the chosen contrast methods. At low noise level, the proposed OptShrink-CF and the existing BM4D method yield the best performances. However, at high noise level, the SVLET-CF method yields the best results. Almost similar observation can be made in visual performance comparison.

As discussed earlier, the noise in magnitude MR images is Rician distributed. Hence, the MR image denoising methods need to be evaluated on the Rician noise as well. For this purpose, the three level approach given in Figure 6.1 employing VST is applied on the noisy MR image. In Table 6.3 and Table 6.4, we compare the denoising performances on the normal and the MS lesion MR data generated using Rician noise model, respectively. In Figure 6.6 and Figure 6.7, the corresponding visual performances are shown. Again, the similar conclusions can be drawn on comparing the denoising performances.

6.6 Summary

In this chapter, we evaluated the OptShrink and the SVLET matrix estimation methods for the denoising of MR images. The proposed approaches use the CF framework that exploits the local as well as nonlocal correlation present in the MR image with the help of these SV shrinkage methods. They are found to outperform the OBNLM and the BM4D method in case of the Gaussian as well as the Rician noise. Moreover, the SVLET-CF is found to be more effective as compared to the OptShrink-CF, especially in case of high noise levels. Hence, we suggest to use the SVLET in CF for the denoising of the MR images.

7

Conclusions and Future Directions

Contents

7.1 Conclusions **118**

7.2 Future directions **119**

7.1 Conclusions

This thesis presented a few approaches in the direction of improving the spectral shrinkage estimators in terms of the estimation accuracy as well as the efficiency in implementation. Essentially, it proposes two shrinkage estimators: the SVLT and the SVLET. The motivation behind the first one is to improve the estimation accuracy of the existing shrinkage estimators by decoupling of the truncation and the shrinkage of SVs. This is done by means of a logistic function based thresholding that uses three parameters. The proposed SVLT estimator though has more number of parameters than the existing estimators, it allows a better control over the shrinkage and the truncation. The improvement in estimation accuracy is achieved at the cost of increased computational complexity in parameter tuning. Thus, it does not have a wide practical applicability.

The second proposed estimator is motivated towards addressing this constraint. In this approach, the thresholding function is expressed as the linear combination of the derivative of Gaussian functions. This linearization helped deriving the closed form expression of the optimal parameters of the shrinkage estimator, thus obviating the need of the costly grid search. In addition, the linearization also allowed us to increase the number of parameters arbitrarily, therefore, it achieves the greater flexibility in shrinkage. Further, we have shown that the proposed SVLET estimator achieves the optimality (in mean-squared sense) in the large matrix limits. In finite matrix case, it performs as good as the existing OptShrink estimator derived from the random matrix theory.

On analysing the results of various experiments conducted on artificially generated noisy matrices, it is concluded that the SVLT performs better rank estimation and better matrix estimation than the existing methods. The experimental study of the SVLET showed that it is one of the fastest methods. It is also noted to outperform the existing parametric shrinkage estimators for varying noise levels and ranks. Hence, for the real time applications, where computational complexity is a huge concern, the SVLET method forms a compelling candidate.

Towards establishing the efficacy of the proposed spectral shrinkage estimators, they have been explored in various signal processing applications. Specifically, we chose the denoising of the signals such as ECG, photographic images, and MR images.

The salient concluding points of the work presented in this thesis can be listed as follows.

- The essence of better estimator of SVs for matrix rank estimation (model order selection) is recognized and explained.
- The fact that a better rank estimation lead to better matrix estimation is reaffirmed.
- A fast and adaptive shrinkage of SVs to remove Gaussian noise from the matrix elements is proposed and successfully demonstrated.
- It is shown that, it is possible to construct a non-linear SVSE, that reaches asymptotic optimality using SURE principle and random matrix theory, by linearizing the shrinkage function.
- The speed achieved by this linearization made it possible to apply the adaptive SV shrinkage methods in practical applications, e.g. MRI denoising.
- The proposed estimators have been shown to perform better than the contrast methods for rank estimation, matrix denoising, ECG signal denoising, and MRI denoising.

7.2 Future directions

The matrix denoising is a simpler category in the class of more challenging inverse problems that include matrix recovery from compressed measurements, the recovery from the data undergone through the motion artifacts, etc. Also, another open challenge in developing the shrinkage estimators is when the distribution of the corrupting noise is non-Gaussian. Mathematically, a more generalized data model than the one considered in the thesis can be given as

$$\tilde{\mathbf{x}} = \mathbf{A}\mathrm{vec}(\mathbf{X}) + \mathrm{vec}(\overline{\mathbf{X}}), \tag{7.1}$$

where $\mathbf{A} \in \mathbb{R}^{p\times mn}$ such that $p \leq mn$ is a matrix operator, the operator $\mathrm{vec}(\cdot)$ represents the vectorization operation, and the elements of noise matrix $\overline{\mathbf{X}}$ is random variables with known/unknown distribution. The focus of the presented works in this thesis was limited to the denoising problems (considering $\mathbf{A} = \mathbf{I}$) with Gaussian noise models (considering $\bar{x}_{j,i} \overset{\text{IID}}{\sim} \mathcal{N}(0, \tau^2)$), only.

The following issues can be explored as a future research direction.

- The proposed SVLT method for rank estimation can be applied to a wide range of machine learning problems for model order selection.

- The adaptive shrinkage strategies enable us to preserve rare-features in medical signals. Hence, the proposed SVLT and SVLET are still to be examined in various medical signal processing applications. For instance, denoising the multi-channel ECG can be a possible extension of the single channel ECG denoising using the SVLET.

- The proposed shrinkage estimators can be applied to various multidimensional signal denoising tasks such as hyper-spectral images, videos, etc. Examining the performance in such applications can be explored in future.

- The proposed methods use SURE for parameter estimation under Gaussian noise assumptions. SURE for exponential family of noise exists but adapting it for SVD domain is mathematically involved. Having done so, SVLET can be modified to incorporate a wide range of noise distribution.

- Notably, this thesis aimed only at matrix denoising problem where all entries are observed. Incorporating the case of missing entries, the proposed estimators need to be modified and can be explored in future.

References

[1] M. Gavish and D. L. Donoho, "Optimal shrinkage of singular values," *May 2016 [Online]. Available: https://arxiv.org/abs/1405.7511v3.*

[2] J. Josse and S. Sardy, "Adaptive shrinkage of singular values," *Statistics and Computing*, vol. 26, no. 3, pp. 715–724, 2016.

[3] S. K. Yadav, R. Sinha, and P. K. Bora, "An efficient SVD shrinkage for rank estimation," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2406–2410, Dec 2015.

[4] H. Andrews and C. Patterson, "Singular value decompositions and digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 1, pp. 26–53, Feb 1976.

[5] V. Klema and A. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164–176, 1980.

[6] L. L. Scharf, "The SVD and reduced rank signal processing," *Signal Processing*, vol. 25, no. 2, pp. 113–133, Nov. 1991.

[7] I. T. Jolliffe, *Principal Component Analysis.* Springer, 2002.

[8] R. Ma, N. Barzigar, A. Roozgard, and S. Cheng, "Decomposition approach for low-rank matrix completion and its applications," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1671–1683, April 2014.

[9] S. S. Vempala, *Spectral Algorithms for Learning and Clustering.* Springer Berlin Heidelberg, 2007.

[10] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *Journal of Machine Learning Research*, vol. 11, pp. 2287–2322, 2010.

[11] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition." *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 931–939, 1998.

[12] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero III, "Shrinkage algorithms for MMSE covariance estimation." *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5016–5029, 2010.

[13] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[14] J. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[15] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee, "Alternating least-squares for low-rank matrix reconstruction," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 231–234, April 2012.

[16] S. Chatterjee, “Matrix estimation by universal singular value thresholding,” *The Annals of Statistics*, vol. 43, no. 1, pp. 177–214, Feb 2015.

[17] C. Deledalle, S. Vaiter, and G. Peyré, “Risk estimation for matrix recovery with spectral regularization,” *May 2012 [Online]. Available: https://arxiv.org/abs/1205.1482v3.*

[18] E. J. Candès, C. A. Sing-Long, and J. D. Trzasko, “Unbiased risk estimates for singular value thresholding and spectral estimators,” *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4643–4657, Oct 2013.

[19] A. A. Shabalin and A. B. Nobel, “Reconstruction of a low-rank matrix in the presence of gaussian noise,” *Journal of Multivariate Analysis*, vol. 118, pp. 67–76, 2013.

[20] R. R. Nadakuditi, “OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage,” *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3002–3018, May 2014.

[21] M. Gavish and D. L. Donoho, “The optimal hard threshold for singular values is $4/\sqrt{3}$,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 5040–5053, 2014.

[22] J. Josse and S. Wager, “Bootstrap-based regularization for low-rank matrix estimation,” *Journal of Machine Learning Research*, vol. 17, no. 124, pp. 1–29, 2016.

[23] A. Parekh and I. W. Selesnick, “Enhanced low-rank matrix approximation,” *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 493–497, April 2016.

[24] J. Bigot, C. Deledalle, and D. Féral, “Generalized SURE for optimal shrinkage of singular values in low-rank matrix denoising,” *May 2016 [Online]. Available: https://arxiv.org/abs/1605.07412*, 2016.

[25] C. Eckart and G. Young, “The approximation of one matrix by another of lower rank,” *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[26] L. Mirsky, “Symmetric gauge functions and unitarily invariant norms,” *The Quarterly Journal of Mathematics*, vol. 11, no. 1, p. 50, 1960.

[27] M. Fazel, H. Hindi, and S. P. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” *in Proceedings of the American Control Conference*, vol. 6, pp. 4734–4739, 2001.

[28] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[29] M. L. Mehta, *Random Matrices*, 3rd ed. Academic Press, New York, 2004.

[30] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices.* Cambridge University Press, 2010, vol. 118.

[31] F. Benaych-Georges and R. R. Nadakuditi, “The singular values and vectors of low rank perturbations of large rectangular random matrices,” *Journal of Multivariate Analysis*, vol. 111, pp. 120–135, 2012.

[32] D. Paul, “Asymptotics of sample eigenstructure for a large dimensional spiked covariance model,” *Statistica Sinica*, vol. 17, pp. 1617–1642, 2007.

[33] M. Verbanck, J. Josse, and F. Husson, “Regularised PCA to denoise and visualise data,” *Statistics and Computing*, vol. 25, no. 2, pp. 471–486, 2015.

[34] C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, Nov 1981.

[35] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, April 2004.

[36] H. Zou, T. Hastie, and R. Tibshirani, “On the degrees of freedom of the LASSO,” *The Annals of Statistics*, vol. 35, no. 5, pp. 2173–2192, 10 2007.

[37] R. J. Tibshirani and J. Taylor, “Degrees of freedom in LASSO problems,” *The Annals of Statistics*, vol. 40, no. 2, pp. 1198–1232, April 2012.

[38] X. Xie, S. C. Kou, and L. D. Brown, “SURE estimates for a heteroscedastic hierarchical model,” *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1465–1479, 2012.

[39] A. Mukherjee, K. Chen, N. Wang, and J. Zhu, “On the degrees of freedom of reduced-rank estimators in multivariate regression,” *Biometrika*, vol. 102, no. 2, pp. 457–477, 2015.

[40] J. R. Magnus, “On differentiating eigenvalues and eigenvectors,” *Econometric Theory*, vol. 1, no. 2, pp. 179–191, 1985.

[41] D. Sun and J. Sun, “Nonsmooth matrix valued functions defined by singular values,” Department of Decision Sciences, National University of Singapore, Tech. Rep., 2003.

[42] A. S. Lewis and H. S. Sendov, “Twice differentiable spectral functions,” *SIAM Journals on Matrix Analysis Applications*, vol. 23, pp. 368–386, 2001.

[43] B. Efron and C. Morris, “Empirical Bayes on vector observations: An extension of Stein’s method,” *Biometrika*, vol. 59, no. 2, pp. 335–347, 1972.

[44] ——, “Multivariate empirical Bayes and estimation of covariance matrices,” *The Annals of Statistics*, vol. 4, no. 1, pp. 22–32, 1976.

[45] O. Kuybeda, D. Malah, and M. Barzohar, “Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors,” *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5579–5592, Dec 2007.

[46] J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, “Speckle reduction of SAR images using SURE-based adaptive sigmoid thresholding in the wavelet domain,” *in Proceedings of International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 462–465, 2009.

[47] P. O. Perry and P. J. Wolfe, “Minimax rank estimation for subspace tracking,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 504–513, Jun 2010.

[48] W. Dong, G. Shi, X. Hu, and Y. Ma, “Nonlocal sparse and low-rank regularization for optical flow estimation,” *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4527–4538, Oct 2014.

[49] B. Rasti, M. Ulfarsson, and J. Sveinsson, “SURE based model selection for hyperspectral imaging,” *in Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4636–4639, July 2014.

[50] T. Zemen and A. Molisch, “Adaptive reduced-rank estimation of nonstationary time-variant channels using subspace selection,” *IEEE Transactions on Vehicular Technology*, vol. 61, no. 9, pp. 4042–4056, Nov 2012.

[51] T. W. Anderson, “Asymptotic theory for principal component analysis,” *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, Mar 1963.

[52] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec 1974.

[53] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, Mar 1978.

[54] J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, no. 5, pp. 465 – 471, Sep 1978.

[55] M. Stone, “Cross-validatory choice and assessment of statistical predictions,” *Journal of Royal Statistic Society*, vol. 36, pp. 111–147, 1974.

[56] M. O. Ulfarsson and V. Solo, “Selecting the number of principal components with SURE,” *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 239–243, Feb 2015.

[57] Q. Guo, C. Zhang, Y. Zhang, and H. Liu, “An efficient SVD-based method for image denoising,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 868–880, May 2016.

[58] K. V. Mardia, J. T. Kent, and J. M. Bibby, “Multivariate analysis,” *Academic Press*, 1979.

[59] I. Jolliffe, “Principal Component Analysis,” *John Wiley & Sons, Ltd*, 2002.

[60] A. B. Owen and P. O. Perry, “Bi-cross-validation of the svd and the nonnegative matrix factorization,” *The Annals of Applied Statistics*, vol. 3, no. 2, pp. 564–594, 06 2009.

[61] M. O. Ulfarsson and V. Solo, “Tuning parameter selection for underdetermined reduced-rank regression,” *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 881–884, Sep 2013.

[62] C. Stein, “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution,” *in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 197–206, 1956.

[63] W. James and C. Stein, “Estimation with quadratic loss,” in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press, 1961, pp. 361–379.

[64] I. Markovsky, “Structured low-rank approximation and its applications,” *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[65] I. Markovsky and K. Usevich, “Structured low-rank approximation with missing data,” *SIAM Journals on Matrix Analysis Applications*, vol. 34, no. 2, pp. 814–830, 2013.

[66] R. Jenatton, G. Obozinski, and F. R. Bach, “Structured sparse principal component analysis,” *in Proceedings of Journal of Machine Learning Research*, vol. 9, pp. 366–373, 2010.

[67] G. Y. Li, K. J. R. Liu, and J. Razavilar, “A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation,” *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 481–486, 1997.

[68] W.-H. Fang and A. E. Yagle, “Two methods for Toeplitz-plus-Hankel approximation to a data covariance matrix,” *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1490–1498, 1992.

[69] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” *in Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2862–2869, June 2014.

[70] F. Luisier, T. Blu, and M. Unser, “A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding,” *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, March 2007.

[71] J. Bai, Y. Zhang, D. Shen, L. Wen, C. Ding, Z. Cui, F. Tian, B. Yu, B. Dai, and J. Zhang, “A portable ECG and blood pressure telemonitoring system,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 4, pp. 63–70, 1999.

[72] A. Alesanco and J. Garcı́a, “Clinical assessment of wireless ECG transmission in real-time cardiac telemonitoring,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1144–1152, 2010.

[73] M. Rahman, R. A. Shaik, and D. V. R. K. Reddy, “Efficient and simplified adaptive noise cancelers for ECG sensor based remote health monitoring,” *IEEE Sensors Journal*, vol. 12, no. 3, pp. 566–573, 2012.

[74] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *in Proceedings of the Royal Society London. Series A: Mathematics, Physics, Engineering, and Science*, vol. 454, no. 1971, pp. 903–995, 1998.

[75] P. M. Agante and J. de Sa, “ECG noise filtering using wavelets with soft-thresholding methods,” *Computers in Cardiology*, pp. 535–538, 1999.

[76] R. Sameni, M.-B. Shamsollahi, C. Jutten, and G. Clifford, “A nonlinear bayesian filtering framework for ECG denoising,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2172–2185, 2007.

[77] O. Sayadi and M.-B. Shamsollahi, “ECG denoising and compression using a modified extended Kalman filter structure,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2240–2248, 2008.

[78] M. Blanco-Velasco, B. Weng, and K. E. Barner, “ECG signal denoising and baseline wander correction based on the empirical mode decomposition,” *Computers in Biology and Medicine*, vol. 38, no. 1, pp. 1–13, 2008.

[79] J. Gao, H. Sultan, J. Hu, and W. wen Tung, “Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison,” *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 237–240, 2010.

[80] H. Kestler, M. Haschka, W. Kratz, F. Schwenker, G. Palm, V. Hombach, and M. Hoher, “Denoising of high-resolution ECG signals by combining the discrete wavelet transform with the Wiener filter,” *Computers in Cardiology*, pp. 233–236, 1998.

[81] B. Weng, M. Blanco-Velasco, and K. Barner, “ECG denoising based on the empirical mode decomposition,” *in Proceedings of IEEE International Conference of Engineering in Medicine and Biology Society*, pp. 1–4, Aug 2006.

[82] M. A. Kabir and C. Shahnaz, “Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains,” *Biomedical Signal Processing Control*, vol. 7, no. 5, pp. 481–489, 2012.

[83] A. Buades, B. Coll, and J. M. Morel, “A non-local algorithm for image denoising,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005.

[84] B. Tracey and E. Miller, “Nonlocal means denoising of ECG signals,” *IEEE Trans. Biomed. Engg.*, vol. 59, no. 9, pp. 2383–2386, 2012.

[85] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[86] P. Chatterjee and P. Milanfar, “Clustering-based denoising with locally learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1438–1451, 2009.

[87] A. Rajwade, A. Rangarajan, and A. Banerjee, “Image denoising using the higher order singular value decomposition,” *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 35, no. 4, pp. 849–862, 2013.

[88] X. Zhang, X. Feng, and W. Wang, “Two-direction nonlocal model for image denoising,” *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 408–412, 2013.

[89] D. Van De Ville and M. Kocher, “SURE-based non-local means,” *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 973–976, 2009.

[90] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[91] D. L. Donoho and J. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[92] G. Moody and R. Mark, “The impact of the MIT-BIH arrhythmia database,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, May 2001.

[93] R. Bousseljot, D. Kreiseler, and A. Schnabel, “Nutzung der EKG-signaldatenbank CARDIODAT der PTB ber das internet.” *Biomedizinische Technik. (Biomed. Tech.)*, vol. 40, no. 1, pp. 317–318, 1995.

[94] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, “Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals.” *Circulation*, vol. 101, no. 23, pp. 215–220, June 2000.

[95] J. Mateo, A. M. Torres, C. Soria, and J. L. Santos, “A method for removing noise from continuous brain signal recordings,” *Computers & Electrical Engineering*, vol. 39, no. 5, pp. 1561–1570, 2013.

[96] J. Mateo and J. J. Rieta, “Application of artificial neural networks for versatile preprocessing of electrocardiogram recordings,” *Journal of Medical Engineering and Technology*, vol. 36, no. 2, pp. 90–101, 2012.

[97] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, “Optimal weighted median filtering under structural constraints,” *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 591–604, Mar 1995.

[98] F. Catta, P.-L. Lions, J.-M. Morel, and T. Coll, “Image selective smoothing and edge detection by nonlinear diffusion,” *SIAM Journal on Numerical Analysis*, vol. 29, no. 1, pp. 182–193, 1992.

[99] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, Jul 1990.

[100] T. D. Bui and G. Chen, “Translation-invariant denoising using multiwavelets,” *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3414–3420, Dec 1998.

[101] H. Zhang, A. Nosratinia, and R. O. Wells, “Image denoising via wavelet-domain spatially adaptive fir Wiener filtering,” *in Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 2179–2182, 2000.

[102] S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, Sep 2000.

[103] ——, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, Sep 2000.

[104] B. Goossens, A. Pizurica, and W. Philips, “Image denoising using mixtures of projected gaussian scale mixtures,” *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1689–1702, Aug 2009.

[105] W. Dong, L. Zhang, G. Shi, and X. Li, “Nonlocally centralized sparse representation for image restoration,” *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.

[106] H. Takeda, S. Farsiu, and P. Milanfar, “Kernel regression for image processing and reconstruction,” *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.

[107] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Transactions on Image Processesing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[108] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, “A multilinear singular value decomposition,” *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, Mar 2000.

[109] G. Bergqvist and E. G. Larsson, “The higher-order singular value decomposition: Theory and an application [lecture notes],” *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 151–154, May 2010.

[110] J. Liang, Y. He, D. Liu, and X. Zeng, “Image fusion using higher order singular value decomposition,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2898–2909, May 2012.

[111] Q. Li, X. Shi, and D. Schonfeld, “Robust hosvd-based higher-order data indexing and retrieval,” *IEEE Signal Processing Letters*, vol. 20, no. 10, pp. 984–987, Oct 2013.

[112] N. D. Sidiropoulos, L. D. Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, “Tensor decomposition for signal processing and machine learning,” *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, July 2017.

[113] M. Nejati, S. Samavi, S. M. R. Soroushmehr, and K. Najarian, “Low-rank regularized collaborative filtering for image denoising,” *in Proceedings of IEEE International Conference on Image Processing (ICIP)*, pp. 730–734, Sept 2015.

[114] J.-L. Starck, E. Candès, and D. Donoho, “The curvelet transform for image denoising,” *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[115] E. J. Candès, “Ridgelets: Theory and application,” Ph.D. dissertation, Dept. of Stat., Stanford Univ., 1998.

[116] M. Do and M. Vetterli, “The finite ridgelet transform for image representation,” *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 16–28, 2003.

[117] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, “Nonlinear anisotropic filtering of MRI data,” *IEEE Transactions on Medical Imaging*, vol. 11, no. 2, pp. 221–232, Jun 1992.

[118] H. Gudbjartsson and S. Patz, “The Rician distribution of noisy MRI data,” *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, 1995.

[119] A. Macovski, “Noise in MRI,” *Magnetic Resonance in Medicine*, vol. 36, no. 3, pp. 494–497, 1996.

[120] J. Sijbers and A. den Dekker, "Maximum likelihood estimation of signal amplitude and noise variance from MR data," *Magnetic Resonance in Medicine*, vol. 51, no. 3, pp. 586–594, 2004.

[121] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1408–1419, Oct 1999.

[122] A. A. Samsonov and C. R. Johnson, "Noise-adaptive nonlinear diffusion filtering of MR images with spatially varying noise levels," *Magnetic Resonance in Medicine*, vol. 52, no. 4, pp. 798–806, 2004.

[123] K. Krissian and S. Aja-Fernandez, "Noise-driven anisotropic diffusion filtering of MRI," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2265–2274, Oct 2009.

[124] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, "MRI denoising using non-local means," *Medical Image Analysis*, vol. 12, no. 4, pp. 514 – 523, 2008.

[125] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 425–441, April 2008.

[126] L. He and I. R. Greenshields, "A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 2, pp. 165–172, Feb 2009.

[127] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "New methods for MRI denoising based on sparseness and self-similarity," *Medical Image Analysis*, vol. 16, no. 1, pp. 18 – 27, 2012.

[128] J. Rajan, J. Veraart, J. V. Audekerke, M. Verhoye, and J. Sijbers, "Nonlocal maximum likelihood estimation method for denoising multiple-coil magnetic resonance images," *Magnetic Resonance Imaging*, vol. 30, no. 10, pp. 1512 – 1518, 2012.

[129] J. Hu, J. Zhou, and X. Wu, "Non-local MRI denoising using random sampling," *Magnetic Resonance Imaging*, vol. 34, no. 7, pp. 990 – 999, 2016.

[130] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 323–331, March 2003.

[131] P. Bao and L. Zhang, "Noise reduction for magnetic resonance images via adaptive multiscale products thresholding," *IEEE Transactions on Medical Imaging*, vol. 22, no. 9, pp. 1089–1099, Sept 2003.

[132] X. Yang and B. Fei, "A wavelet multiscale denoising algorithm for magnetic resonance (MR) images," *Measurement Science and Technology*, vol. 22, no. 2.

[133] F. Luisier, T. Blu, and P. J. Wolfe, "A CURE for noisy magnetic resonance images: Chi-square unbiased risk estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3454–3466, Aug 2012.

[134] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119–133, Jan 2013.

[135] S. Li, L. Fang, and H. Yin, "An efficient dictionary learning algorithm and its application to 3-D medical image denoising," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 417–427, Feb 2012.

[136] A. Lorbert, J. S. Guntupalli, D. J. Eis, J. V. Haxby, and P. J. Ramadge, "Collaborative denoising of multi-subject fMRI data," *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1008–1012, May 2013.

[137] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X. P. Zhang, "Bayesian nonparametric dictionary learning for compressed sensing MRI," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5007–5019, Dec 2014.

[138] L. Xu, C. Wang, W. Chen, and X. Liu, "Denoising multi-channel images in parallel MRI by low rank matrix decomposition," *IEEE Transactions on Applied Superconductivity*, vol. 24, no. 5, pp. 1–5, Oct 2014.

[139] H. M. Nguyen, X. Peng, M. N. Do, and Z. P. Liang, "Denoising MR spectroscopic imaging data with low-rank approximations," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 78–89, Jan 2013.

[140] J. P. Haldar, "Low-rank modeling of local k -space neighborhoods (LORAKS) for constrained MRI," *IEEE Transactions on Medical Imaging*, vol. 33, no. 3, pp. 668–681, March 2014.

[141] X. Zhang, Z. Xu, N. Jia, W. Yang, Q. Feng, W. Chen, and Y. Feng, "Denoising of 3D magnetic resonance images by using higher-order singular value decomposition," *Medical Image Analysis*, vol. 19, no. 1, pp. 75 – 86, 2015.

[142] Y. Fu and W. Dong, "3D magnetic resonance image denoising using low-rank tensor approximation," *Neurocomputing*, vol. 195, pp. 30–39, 2016.

[143] A. Foi, "Noise estimation and removal in MR imaging: The variance-stabilization approach," *in Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1809–1814, March 2011.

[144] G. B. Arfken, H. J. Weber, and F. E. Harris, "Mathematical methods for physicists: A comprehensive guide," *Academic Press*, Jul 2005.

[145] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

LIST OF PUBLICATIONS

Refereed Journals:

1. S. K. Yadav, R. Sinha and P. K. Bora, "An Efficient SVD Shrinkage for Rank Estimation," in IEEE Signal Processing Letters, vol. 22, no. 12, pp. 2406-2410, Dec. 2015. doi: 10.1109/LSP.2015.2487600
2. S. K. Yadav, R. Sinha and P. K. Bora, "Electrocardiogram signal denoising using non-local wavelet transform domain filtering," in IET Signal Processing, vol. 9, no. 1, pp. 88-96, 2 2015. doi: 10.1049/iet-spr.2014.0005

Refereed Conferences:

1. S. K. Yadav, R. Sinha and P. K. Bora, "Image denoising using ridgelet transform in a collaborative filtering framework," 2014 International Conference on Signal Processing and Communications (SPCOM), Bangalore, 2014, pp. 1-5. doi: 10.1109/SPCOM.2014.6983964

Manuscripts under Review/Preparation

1. S. K. Yadav, R. Sinha, and P. K. Bora, "ECG Signal Denoising using Adaptive Shrinkage of Singular Values in a Collaborative Filtering Framework," (To be submitted to IET Signal Processing).
2. S. K. Yadav, R. Sinha, and P. K. Bora, "A Fast Date-Driven Shrinkage of Singular Values for Arbitrary Rank Signal Matrix Denoising," (Under review in IEEE Transaction on Signal Processing).
3. S. K. Yadav, R. Sinha, and P. K. Bora, "A Fast Nonlocal Method for MRI Denoising using Adaptive Shrinkage of Singular Values," (To be submitted to IEEE Transaction on Biomedical Engineering).