

POLYNOMIAL FITTING AND TOTAL VARIATION
BASED TECHNIQUES ON 1-D AND 2-D SIGNAL
DENOISING

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
ELECTRONICS ENGINEERING

AND THE INSTITUTE OF ENGINEERING AND SCIENCES
OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By

Aykut Yıldız

July 2010

I certify that I have read this thesis and that in my opinion it is fully adequate,
in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Orhan Arıkan (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate,
in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Sinan Gezici

I certify that I have read this thesis and that in my opinion it is fully adequate,
in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. İbrahim Körpeoğlu

Approved for the Institute of Engineering and Sciences:

Prof. Dr. Levent Onural
Director of Institute of Engineering and Sciences

ABSTRACT

POLYNOMIAL FITTING AND TOTAL VARIATION BASED TECHNIQUES ON 1-D AND 2-D SIGNAL DENOISING

Aykut Yıldız

M.S. in Electrical and Electronics Engineering

Supervisor: Prof. Dr. Orhan Arıkan

July 2010

New techniques are developed for signal denoising and texture recovery. Geometrical theory of total variation (TV) is explored, and an algorithm that uses quadratic programming is introduced for total variation reduction. To minimize the staircase effect associated with commonly used total variation based techniques, robust algorithms are proposed for accurate localization of transition boundaries. For this boundary detection problem, three techniques are proposed. In the first method, the $1 - D$ total variation is applied in first derivative domain. This technique is based on the fact that total variation forms piecewise constant parts and the constant parts in the derivative domain corresponds to lines in time domain. The boundaries of these constant parts are used as the transition boundaries for the line fitting. In the second technique proposed for boundary detection, a wavelet based technique is proposed. Since the mother wavelet can be used to detect local abrupt changes, the Haar wavelet function is used for the purpose of boundary detection. Convolution of a signal or its derivative family with this Haar mother wavelet gives responses at the edge locations, attaining

local maxima. A basic local maximization technique is used to find the boundary locations. The last technique proposed for boundary detection is the well known Particle Swarm Optimization (PSO). The locations of the boundaries are randomly perturbed yielding an error for each set of boundaries. Pursuing the personal and global best positions, the boundary locations converge to a set of boundaries. In all of the techniques, polynomial fitting is applied to the part of the signal between the edges.

A more complicated scenario for $1 - D$ signal denoising is texture recovery. In the technique proposed in this thesis, the periodicity of the texture is exploited. Periodic and non-periodic parts are distinguished by examining total variation of the autocorrelation of the signal. In the periodic parts, the period size was found by PSO evolution. All the periods were averaged to remove the noise, and the final signal was synthesized.

For the purpose of image denoising, optimum one dimensional total variation minimization is carried to two dimensions by Radon transform and slicing method. In the proposed techniques, the stopping criterion for the procedures is chosen as the error norm. The processes are stopped when the residual norm is comparable to noise standard deviation. $1 - D$ and $2 - D$ noise statistics estimation methods based on Maximum Likelihood Estimation (MLE) are presented. The proposed denoising techniques are compared with principal curve projection technique, total variation by Rudin *et al*, total variation by Willsky *et al*, and curvelets. The simulations show that our techniques outperform these widely used techniques in the literature.

Keywords: Denoising, total variation, texture recovery, PSO, matched filtering, polynomial fitting

ÖZET

BİR BOYUTLU VE İKİ BOYUTLU SİNYALLERİN POLİNOM UYUMU VE TOPLAM DEĞİŞİME DAYALI GÜRÜLTÜ BASTIRMA TEKNİKLERİ

Aykut Yıldız

Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans

Tez Yöneticisi: Prof. Dr. Orhan Arıkan

Temmuz 2010

Tek boyutlu sinyallerin, dokuların ve imgelerin gürültülerinin bastırımı üzerine yeni teknikler geliştirilmiştir. Toplam değişim tekniğinin geometrik teorisi incelenmiş ve doğrusal programlamaya dayalı bir algoritma bulunmuştur. Bir boyut-taki optimal gürültü bastırım tekniği Willsky ve ekibi tarafından yayınlanan toplam değişim tekniğidir. Ancak bu teknik sinyaller üzerinde merdiven etkisi yaratmaktadır. Bu merdiven etkisini ortadan kaldırmak için kenarlar doğru olarak bulunmalıdır. Bunun için üç teknik önerilmiştir. Bunlardan birincisi, türev tanım kümesinde toplam değişim enküçültmesinin yapılmasıdır. İkincisi, Haar ana dalgacık sinyaliyle uyumlu süzgeçten geçirmektir. Son teknik ise Parçacık Sürüsü Optimizasyon tekniğidir. Her üç teknikte de, sinyalin bulunan kenarlar arasındaki kısmına polinom oturtulur. Bu tezde daha karmaşık bir senaryo olan doku onarımı üzerine de çalışılmıştır. Geliştirilen teknikte dokunun periyodikliğinden faydalanılmıştır. Periyotların hepsinin ortalaması alınarak gürültü ortadan kaldırılmıştır. Bulunan imge gürültü bastırım tekniklerinde bir boyutlu optimal gürültü azaltma tekniği Radon dönüşümü ve dilimleme teknikleriyle iki boyuta taşınmıştır. Bu tüm gürültü bastırım teknikleri için

durma kriteri hata normudur. Hata normu gürültü standart sapmasına eşit olarak kabul edilmiştir. Bir ve iki boyutlu sinyaller için enbüyük olabilirlik tekniğine dayalı gürültü istatistiği kestirim teknikleri geliştirilmiştir. Önerilen gürültü bastırım teknikleri ana eğriye izdüşüm, Rudin'in ve Willsky'nin TV metotlarıyla ve eğricik yöntemiyle karşılaştırılmıştır.

Anahtar Kelimeler: Gürültü bastırımı, toplam değişim, doku onarımı, Parçacık Sürüsü Optimizasyonu, uyumlu süzgeç, polinom oturtma

ACKNOWLEDGMENTS

My greatest gratitude goes to my supervisor Prof. Dr. Orhan Arikan. I thank him for encouraging me in my good times, and tolerating my bad times.

I wish to thank D. H. Donoho, G. Gilboa and U.Özertem for the MATLAB codes they provided for me.

I would like to thank Asst. Prof. Dr. Sinan Gezici for sharing his ideas when I consulted him. Also special thanks to my friends in the office EE514 for their cooperation on my research.

Finally, I am extremely appreciated with my father Aşır's love, support and encouragement during all my life.

Contents

1	INTRODUCTION	1
1.1	1 – D Signal Denoising	1
1.1.1	1 – D Noise Standard Deviation Estimation	6
1.1.2	Piecewise Linear and Higher Order Polynomial Denoising	6
1.1.3	Signal Denoising by Piecewise Continuous Polynomial Fitting	7
1.1.4	Denoising via Matched Filtering	7
1.1.5	1 – D Texture Recovery by Iterative Pasting	8
1.2	Image Denoising	9
1.2.1	Introduction of Viable Techniques in the Literature	9
1.2.2	2 – D Noise Standard Deviation Estimation	11
1.2.3	Slice-TV Technique	11
1.2.4	Radon-TV Technique	12
1.2.5	2 – D Texture Recovery by Total Variation	12

2	1 – D SIGNAL DENOISING	14
2.1	1 – D Noise Standard Deviation Estimation	15
2.2	Particle Swarm Optimization	16
2.2.1	Parameter Selection for Stability and Convergence	18
2.3	Theory of Total Variation Minimization	19
2.3.1	Geometry of Total Variation Reduction	19
2.3.2	Total Variation Reduction by Quadratic Programming	24
2.3.3	Finding the Points that have Constant Total Variation in Three Dimensions	28
2.3.4	An Algorithm to Reduce Total Variation of a Vector	32
2.4	Optimal Total Variation Reduction using Nonlinear Diffusion Equation	37
2.5	Piecewise Smooth Signal Denoising via Principal Curve Projections	40
2.6	Piecewise Linear Denoising by Total Variation	42
2.6.1	Formulation and Algorithm	43
2.6.2	Comparison of the Techniques	47
2.7	Piecewise Parabolic and Cubic Denoising by Total Variation	51
2.7.1	Formulation and Algorithm	53
2.7.2	Comparison of the Techniques	61
2.8	Signal Denoising by Piecewise Continuous Polynomial Fitting	65

2.8.1	Formulation and Algorithm	66
2.8.2	Comparison of the Techniques	71
2.9	Denoising Piecewise Polynomial Signals by Technique based on Convolution with Wavelets	74
2.9.1	Formulation and Algorithm	76
2.9.2	Comparison of the Techniques	79
2.10	Texture Recovery by Adaptive Patch Averaging	83
2.10.1	Formulation and Algorithm	84
2.10.2	Denoising Process	94
3	IMAGE DENOISING	107
3.1	$2 - D$ Noise Standard Deviation Estimation	108
3.2	Nonlinear Total Variation Based Noise Removal Algorithms . . .	109
3.2.1	Formulation	111
3.3	Curvelet Transform Based Denoising	112
3.4	Comparison of Viable Techniques	112
3.4.1	Examples Concerning the Comparison of the Output Im- ages of Five Different Algorithms	112
3.4.2	Comparison of Curvelet and other benchmark denoising techniques	113
3.5	$2 - D$ Denoising by Averaging Row-Wise and Column-Wise $1 - D$ Total Variation Outputs	114

3.5.1	Formulation and Algorithm	116
3.5.2	Comparison of Experiments	118
3.6	Image Denoising by Total Variation Minimization in Radon Transform Domain	125
3.6.1	Formulation and Algorithm	126
3.6.2	Comparison of the Techniques	130
3.7	Image Texture Recovery by Separable Implementation of $2 - D$ Total Variation	148
4	CONCLUSIONS AND FUTURE WORK	153
4.1	$1 - D$ Signal Denoising	153
4.2	Image Denoising	155
	APPENDIX	158
	A PROCESS AND OPTIMALITY OF PIECEWISE LINEAR POLYNOMIAL FITTING	158

List of Figures

1.1	The random electron motions that cause additive noise in a channel.	2
2.1	The table which compares the proposed polynomial fitting based algorithms.	15
2.2	Original piecewise linear signal with one jump and the noisy version.	16
2.3	The geometry of the solution for two dimensions.	21
2.4	The geometry of the solution for three dimensions.	22
2.5	The case if the variance sphere and the line $y_1 = y_2 = y_3$ do not intersect.	23
2.6	The points that have minimum total variation in two dimensions for different spheres.	24
2.7	The graph of the total variation of the denoised vector versus σ . .	25
2.8	Logarithmic plot of the total variation of the denoised vector versus σ	26
2.9	The points that have constant total variation for $x_2 = 0$	29
2.10	The points that have constant total variation for $x_2 = C/2$	30

2.11	The projection of the $3D$ shape that represents the intersection by a sphere in two dimensions.	31
2.12	The graph of constant total variation in three dimensions.	32
2.13	The convex sets with different total variation.	33
2.14	Finding the optimal μ	34
2.15	Piecewise constant behavior of total variation process.	35
2.16	The final signal for $\mu = 0$	36
2.17	The process of total variation reduction.	37
2.18	The same resulting vector for two steps.	38
2.19	Different resulting vectors.	39
2.20	Flow diagram of the proposed Piecewise Linear Denoising Technique.	43
2.21	Two lines to be merged.	46
2.22	Denoising multi-line signal with no jump-discontinuities by Total Variation by Rudin <i>et al.</i>	47
2.23	Denoising multi-line signal with no jump-discontinuities by Total Variation by Willsky <i>et al.</i>	48
2.24	Denoising multi-line signal with no jump-discontinuities by the proposed PWL technique.	49
2.25	First difference of a noisy signal consisting of two paralel lines separated by a jump.	50
2.26	Denoising the synthetic signal composed of two lines by the tech- nique of Total Variation by Rudin <i>et al.</i>	51

2.27	Denoising the synthetic signal composed of two lines by the technique of Total Variation by Willsky <i>et al.</i>	52
2.28	Denoising the synthetic signal composed of two lines by the proposed PWL technique.	53
2.29	Estimation error of Total Variation by Rudin <i>et al.</i> for the signal composed of two lines.	54
2.30	Estimation error of Total Variation by Willsky <i>et al.</i> for the signal composed of two lines.	55
2.31	Estimation error of the proposed PWL technique for the signal composed of two lines.	56
2.32	Normalized fit error of Total Variation by Rudin <i>et al.</i> for the signal composed of two lines at various noise standard deviations.	57
2.33	Normalized fit error of Total Variation by Willsky <i>et al.</i> for the signal composed of two lines at various noise standard deviations.	58
2.34	Normalized fit error of the proposed PWL technique for the signal composed of two lines at various noise standard deviations.	59
2.35	Denoising an image slice by Total Variation of Rudin <i>et al.</i>	62
2.36	Denoising an image slice by Total Variation of Willsky <i>et al.</i>	63
2.37	Denoising an image slice by the proposed piecewise parabolic technique.	64
2.38	Denoising an image slice by the proposed piecewise cubic technique.	65
2.39	Denoising a parabolic curve by the technique of Rudin <i>et al.</i>	66
2.40	Denoising a parabolic curve by the technique of Willsky <i>et al.</i>	67

2.41	Denoising a parabolic curve by our technique of piecewise parabolas.	68
2.42	Denoising a parabolic curve by our technique of piecewise cubics.	69
2.43	Estimation error of Total Variation by Rudin for the parabolic curve.	70
2.44	Estimation error of Total Variation by Willsky for the parabolic curve.	71
2.45	Estimation error of our Parabolic Denoising Technique for the parabolic curve.	72
2.46	Estimation error of our Cubic Denoising Technique for the parabolic curve.	73
2.47	Flow diagram of the proposed Piecewise Continuous Polynomial Fitting Technique.	74
2.48	The geometry of polynomial fitting problem and the solution based on orthogonal projection [1].	75
2.49	Piecewise constant signal with noise [2].	76
2.50	Error in estimation of the original signal by the proposed PSO based polynomial fit technique at five different noise levels [2]. . .	76
2.51	Error in estimation of the original signal by total variation technique at 5 different noise levels [2].	77
2.52	The denoising result obtained with the proposed technique [2]. . .	77
2.53	The graphs of total variation by Rudin <i>et al</i> for a signal with three jumps [2].	78
2.54	The graphs of Principal Curve Projections for a signal with three jumps [2].	78

2.55	The fit error of as a function of number of boundaries as D [2].	79
2.56	The fit error as a function of polynomial order k [2].	79
2.57	Flow diagram of the proposed Wavelet based Technique.	80
2.58	Original syntetic piecewise constant signal with one jump discontinuity.	81
2.59	The probability of detection for theoretical thresholds and for the actual stopping criterion about the case of piecewise constant signal.	82
2.60	The probability of false alarm for theoretical thresholds and for the actual stopping criterion about the case of piecewise constant signal.	83
2.61	Original synthetic triangular signal.	84
2.62	The probability of detection for theoretical thresholds and for the actual stopping criterion about the triangular signal.	85
2.63	The probability of false alarm for theoretical thresholds and for the actual stopping criterion about the triangular signal.	86
2.64	Original synthetic signal with piecewise constant parts.	87
2.65	Noisy piecewise constant signal with an additive Gaussian noise with a standard deviation of 5.	88
2.66	Piecewise constant signal denoised by the proposed technique.	89
2.67	Original synthetic signal with piecewise linear parts.	90
2.68	Noisy piecewise linear signal with an additive Gaussian noise with a standard deviation of 1.	91

2.69	Output of convolution at the first stage.	92
2.70	Output of convolution at the second stage.	93
2.71	Piecewise linear signal denoised by the proposed technique.	94
2.72	Flow diagram of the proposed Patch Averaging based Texture Recovery Technique.	95
2.73	Flow diagram of texture recovery by adaptive windowing.	96
2.74	The synthetic signal with three different texture steps.	97
2.75	Noisy signal with three different texture steps.	98
2.76	The signal with three smoothed steps.	99
2.77	The residual of the signal with three different texture steps.	99
2.78	Enlarged graph of the first periodic texture. Three periods are shown for convenience.	100
2.79	Enlarged graph of the second periodic texture. Three periods are shown for convenience.	100
2.80	Enlarged graph of the third periodic texture. Three periods are shown for convenience.	101
2.81	Signal recovered by the method of patch averaging.	101
2.82	A syntetic signal with a periodic and a non-periodic step.	102
2.83	The noisy version of the signal with a periodic and a non-periodic step.	102
2.84	The smoothed version of the signal with a periodic and a non- periodic step.	103

2.85	Autocorrelation of the first (periodic) part.	103
2.86	Autocorrelation of the second (non-periodic) part.	104
2.87	The signal recovered by patch averaging.	104
2.88	A synthetic signal consisting of four various parts.	105
2.89	Noisy version of the signal consisting of four parts.	105
2.90	The signal recovered by patch averaging.	106
3.1	Original synthetic image.	109
3.2	Noisy synthetic image whose noise σ is estimated.	110
3.3	Curvelet transform building blocks for image denoising [3].	113
3.4	The denoised outputs of five different algorithms for a synthetic image.	114
3.5	The denoised outputs of five different algorithms for the classical Barbara image.	115
3.6	Flow diagram of the proposed Slice TV Technique.	116
3.7	Original synthetic image.	119
3.8	Noisy synthetic image with a Gaussian noise of $\sigma=15$	120
3.9	Synthetic image denoised with total variation by Rudin <i>et al.</i>	121
3.10	Synthetic image denoised with enhanced total variation.	122
3.11	Synthetic image denoised with Slice-TV.	123

3.12	Images of difference between the original image and the denoised image at extreme parameters for total variation by Rudin <i>et al.</i>	124
3.13	Images of difference between the original image and the denoised image at extreme parameters for enhanced total variation.	124
3.14	Images of difference between the original image and the denoised image at various parameters for Slice-TV.	125
3.15	Frobenius norm of estimation error for total variation by Rudin <i>et al.</i>	126
3.16	Frobenius norm of estimation error for enhanced total variation.	127
3.17	Frobenius norm of estimation error for total variation by Slice-TV.	128
3.18	Original Barbara image.	129
3.19	Noisy Barbara image with Gaussian noise of $\sigma = 15$.	130
3.20	Image denoised with total variation by Rudin <i>et al.</i>	131
3.21	Image denoised with enhanced total variation.	132
3.22	Image denoised with with Slice-TV.	133
3.23	Flow diagram of the proposed Radon TV Technique.	134
3.24	Block diagram of the denoising process.	134
3.25	The geometry of Radon Transform (taken from MATLAB Radon Transform help).	135
3.26	Original synthetic image.	136
3.27	Noisy synthetic image with a Gaussian noise of $\sigma=15$.	137

3.28	Synthetic image denoised with total variation by Rudin <i>et al.</i> . . .	138
3.29	Synthetic image denoised with curvelet.	139
3.30	Synthetic image denoised with proposed R-TV.	140
3.31	Original Barbara image.	141
3.32	Noisy Barbara image with Gaussian noise of $\sigma = 15$	142
3.33	Barbara image denoised with total variation by Rudin <i>et al.</i> . . .	143
3.34	Barbara image denoised with curvelet transform.	144
3.35	Barbara image denoised by R-TV.	145
3.36	Images of difference between the original image and the denoised image at various parameters for total variation by Rudin <i>et al.</i> . .	145
3.37	Images of difference between the original image and the denoised image at various parameters for curvelet transform.	146
3.38	Images of difference between the original image and the denoised image at various parameters for R-TV.	146
3.39	Frobenius norm of estimation error for total variation by Rudin <i>et</i> <i>al.</i>	147
3.40	Frobenius norm of estimation error for curvelet transform.	148
3.41	Frobenius norm of estimation error for total variation by R-TV. .	149
3.42	The original Barbara image.	150
3.43	The noisy Barbara image with a noise σ of 8.	151
3.44	Barbara image denoised with the separable total variation technique.	152

A.1	The original signal $\check{u}[n]$ and the $\mathbf{u}[n]$ signal which is noisy.	160
A.2	$d[n, 0)$ which is equal to the first difference of $\mathbf{u}$	161
A.3	The original signal and $F_t(\mathbf{u}[n])$	162
A.4	$d[n, t)$ which is the result of total variation process on $d[n, 0)$	163
A.5	The original signal and final solution $F_T(\mathbf{u}[n])$	164
A.6	$d[n, T)$ which is the result of total variation process on $d[n, 0)$	165
A.7	Original signal, two lines to be merged and the merged line.	166

List of Tables

2.1	The total variation equations for four dimensional vector.	27
2.2	Error norms for a signal with three boundaries [2].	72
3.1	The comparison and comments for five renowned denoising techniques.	116

To the memories of my grandfathers . . .

Chapter 1

INTRODUCTION

1.1 1 – D Signal Denoising

Noise is a typical problem of remote sensing images, biomedical images, micro-imaging, graphics and telecommunication signals. This common problem is caused by the random electron motions as depicted in Figure 1.1. Whether these electron motions disturb a signal and create noise, human eye is able to perceive the object bodies and edges. On the other hand, removing the noise in an automated manner is an active research area and there have been many developments on this subject in recent years. Various techniques concerning different combinations of mathematical transforms and tools have been discovered that can reduce the noise of signals. In order to have a good performance, an important desired feature is the preservation of edges while removing the noise. Total variation reduction [4] is the most renowned technique for edge-preserving noise removal.

Total variation technique is based on gradually reducing the total variation of the pixel values of a signal without degrading the resolution. The definition of total variation of a one dimensional continuous space signal is straight forward.

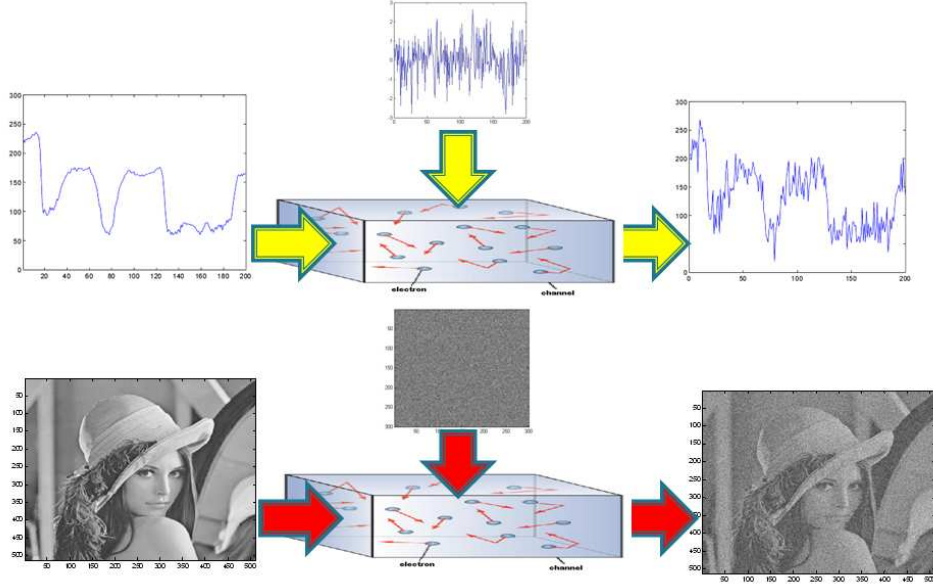


Figure 1.1: The random electron motions that cause additive noise in a channel.

$$TV(u(x)) = \int \left| \frac{du(x)}{dx} \right| dx.$$

The total variation based techniques are ideally suited for signals with piecewise constant regions. During the iterations, the total variation of the signal is reduced. If the iterations are not stopped, eventually, a constant valued signal is obtained. Therefore, total variation techniques make us of a stopping criterion to avoid oversmoothing of the noisy images.

A fast algorithm of total variation reduction is given in [4] for minimizing total variation by reducing the time complexity to $O(N \log N)$ and space complexity to $O(N)$ where N is the number of data samples. This algorithm solves the defined problem optimally. This problem takes total variation as a constraint and minimizes the difference L_2 norm.

To avoid a stopping criterion a non-parametric edge-preserving technique [5] was developed. In that technique so-called Principal Curve Projection, there are no artifacts such as stair-case effect of total variation, speckle noise of curvelet and Gibbs phenomenon of wavelet based techniques. The noise removal is achieved

by finding the maximum likelihood of the principal curve as maximum of the $2 - D$ pdf as the first stage. Then this ML signal is projected on the principal curve.

Matched filtering also finds a huge application area in telecommunications, signal processing and electronics. It is a crucial tool especially used in radar signal processing. References [6], [7], [8], [9], [10], [11] and [12] are decent examples in this field. Matched filtering is used for pulse compression especially as a more specific application in radar signal processing. Range profiles are found by the help of this technique. The techniques explained in [13], [14] and [15] are some examples. References [16], [17] and [18] provide application of matched filtering in denoising. Moreover, analog and digital electronics are areas that exploit matched filtering as well. References [19], [20] and [21] make use of matched filtering for these purposes. In this field, matched filtering is used for detecting where some predefined structure of signal resides. Matched filtering is achieved by convolution or cross-correlation of a signal with the sample signal kernel.

Polynomial fitting is a tool that is widely used in signal processing and numerical analysis. It is the act of finding the best polynomial of some order fitted to an arbitrary signal in least squares sense. The application of polynomial fitting on speech recognition is presented in [22]. An example of the application in image processing is [23]. Polynomial fitting is used on estimation problems in [24]. Least squares based polynomial fitting is benefited for the purpose of interpolation in [25]. It is told in the book that polynomial fitting is an invaluable tool to model smooth signals. Because, performing analytical operations such as derivative and integral on discrete signals is difficult and inaccurate. Instead, fitting an analytical function to the discrete signal is common practice. For example, if a polynomial is fitted to a discrete signals derivative and integral operations can be performed analytically in a straight-forward manner, and the resulting

function is again a polynomial. Evaluation of polynomials is also easy with so called *Horner's method*. In [25], it is also highlighted that polynomial fitting to a considerably long signal is impractical. Therefore, piecewise polynomial fitting was proposed for the purpose of interpolation [25]. Piecewise polynomial fitting is used in our algorithm as a means of denoising the piecewise smooth parts of the 1-D or 2-D signals. Because polynomials are most widely used type of signals to model so-called family of smooth signals.

Particle Swarm Optimization (PSO) is a recent optimization heuristic that has found a wide application area. Most of the optimal solutions in engineering and science require exhaustive search over some parameters. PSO is an extremely robust tool that reduces the computational complexity of the optimization procedure even for large dimensional parameter spaces. The analogy of the evolution of parameters to the flocking behavior of the birds has attracted the attention of psychologists and sociologists on PSO as well. However, besides the philosophy of PSO, the engineering application areas are more important to be mentioned here.

In electrical engineering, like in computer science, PSO is most widely used in machine learning. As soon as PSO algorithms became popular, it was started to be used in neural networks. There is an example of application of PSO in neural networks in [26]. In scheduling problems of networks which is more similar to industrial engineering counterpart fields, PSO was used as an optimizer of TDMA allocation [27]. In this thesis, we will use PSO to determine transition boundaries of piecewise smooth models.

Texture recovery is a new area of signal denoising which introduces sort of an artificial intelligence to the concept of denoising, because it somewhat requires to differentiate automatically between what texture and noise are. Texture is defined as a regular shape or structure in a signal or image. Excessive noise makes

it impossible to recover these details, but for moderate noise, there are techniques that can extract the original signal with an acceptable level of performance.

There are two main approaches to detect typical periodicity associated with a texture in a noisy signal. One of them is autocorrelation based and the other relies on frequency spectrum characteristics. In [28], it is proposed that an approximately periodic signal which contains dominantly high frequency components, its period can be estimated accurately according to its frequency distribution. A similar frequency based technique can be found in [29]. In [30], texture recovery is done by the method of autocorrelation. In [31], it is claimed that assigning the frequency component with the most power as the correct period is not a proper method, instead, it convolves observation signal with a kernel of autocorrelation.

There are techniques which use neither frequency spectrum nor autocorrelation. They solely modify generic denoising algorithms. In [32], an iterative algorithm is introduced which does not require edge detection, segmentation steps or any sophisticated constraints. It is just based on recovering a corrupted textured area by taking a patch with similar texture and pasting it on the corrupted one. However, if the textured signal is corrupted with homogeneous white Gaussian noise, then an uncorrupted patch can not be found and this technique fails. In [33], an adaptation of total variation reduction on texture recovery problem is introduced. The default implementations of total variation process stop when the residual norm reaches some value comparable to the noise standard deviation. However, in [33], the total variation stops when the distance to the signal to be recovered is minimum. This is achieved by tuning a penalty function that punishes the distance from the signal under noise so that denoising stops when the pure textured structure is reached. The drawback of this algorithm is that this tuning parameter is not a robust tool. It turned out to be an ad hoc approach since the tuning parameter depends on the textured image.

1.1.1 1 – D Noise Standard Deviation Estimation

Most of the denoising techniques smooth signals iteratively, hence they require a stopping criterion. The most natural constraint is the noise standard deviation. If estimated accurately, the iterations can be stopped when the distance from the noisy image is equal to the estimated noise standard deviation. If the removed residual is statistically similar to the noise, then the technique is successful and the error norm is small.

The technique developed to estimate the noise variance is based on coarse denoising. The signal is filtered with a running average filter. Then the distance between the noisy image and the denoised image is computed. This value is estimated as the noise variance. The computational complexity of this approach is $O(N)$.

1.1.2 Piecewise Linear and Higher Order Polynomial Denoising

First of all, the technique introduced here finds the first, second or third difference of the original signal. Total variation of this new signal is reduced and the region boundaries are obtained. (Region is the flat part of a signal). In each region, curve fitting is applied with least squares minimization. A polynomial of first, second or third degree is fit to the noisy part of the curve. When two regions merge in the total variation minimization process, two adjacent corresponding curves are merged by using the parameters of those two curves. As soon as a difference norm equal to the noise standard deviation is reached, the process is stopped. The noise standard deviation should be estimated before the process.

1.1.3 Signal Denoising by Piecewise Continuous Polynomial Fitting

Denoising signals is a constant requirement in signal analysis. There are a multitude of signal models that can be exploited for efficient denoising of signals. One of the commonly used model is piecewise polynomial model where the signal can be segmented to polynomial sections whose parameters can abruptly change at the section boundaries. This powerful model contains piecewise constant and piecewise linear models as its special case. When the number of sections and their boundaries are known, this model can perform highly effective signal denoising. However, typically, neither the number of these sections nor their boundaries are known *a priori*, the optimal choice of the section boundaries requires solution to an optimization problem with a complicated minima structure. To alleviate this problem, we propose to use particle swarm optimization technique [34], which efficiently provides the transition boundaries for a given number of sections. To determine the number of sections, we propose to monitor fit error for a sequentially increased number of sections until the fit error is comparable to the standard deviation of the noise on the signal.

1.1.4 Denoising via Matched Filtering

In the denoising tool introduced here, the discontinuities are searched for in the derivatives of a signal to segment it into piecewise polynomial parts so that we can apply polynomial fitting between those boundaries of discontinuities. Our approach is based on the fact that if the signal has piecewise polynomial segments, then it is strongly probable that its derivatives of some order have discontinuities. The steps that we take can be summarized as follows.

The derivative of the signal is taken iteratively and at each iteration matched filtering is applied with a Haar wavelet of fixed size to find the discontinuities

which form the boundaries of polynomials. A threshold is chosen and the maximum of the output of the convolution is found which is within the part of the signal that exceeds the threshold. The maximums of each derivative are merged to find all the boundaries of piecewise polynomials. At the end, the difference norm with respect to the original signal is found and the search for boundaries is stopped if this norm is smaller than the noise standard deviation. Otherwise, the thresholds are decreased so that more discontinuities are found and again the stopping criterion is in charge. If the stopping criterion is confirmed, then polynomial fitting is done within the boundaries. Noise standard deviation should be estimated before the process.

1.1.5 $1 - D$ Texture Recovery by Iterative Pasting

In the research conducted for this thesis, it was found out that autocorrelation is a viable tool to detect periodicity in a $1 - D$ signal. Autocorrelation is used in this algorithm to detect periodicity instead of frequency spectrum techniques. It is claimed here that edge detection is crucial to differentiate the textured and non-textured parts of an image, since these parts should be treated separately.

The method introduced here on texture recovery is based on the detection of periodic structures and recovering them by the method of iterative pasting. The first step for this purpose is to detect the boundaries of different periodic structures and to obtain the residual after a process of total variation. In this windowed signal, existence of periodicity is searched. If periodicity is found, the period of the periodic signal is found, and a number of these periods are averaged so that noise is removed. As a final step, these periods are pasted to their places iteratively. If no periodicity is found out, total variation turns back to an optimal estimated noise coefficient.

1.2 Image Denoising

1.2.1 Introduction of Viable Techniques in the Literature

Denoising of images have found diverse application areas including remote sensing, biomedical imaging and micro imaging. Recent work on this area is focused in two alternative techniques: $2-D$ total variation minimization [35] and curvelet domain denoising [36].

Curvelet transform is the application of Radon transform which takes the projection of a $2-D$ image to $1-D$ and the application of wavelet transform thereafter. It can be perceived as taking the average of points in the projection domain. Using Radon transform produces rotation independent results. In curvelet transform, the image is partitioned into dyadic squares in order to reduce the computational complexity. It yields surprisingly good results on images with high noise, however, it has not been proven that it is the optimal technique. For details, mathematical equations and algorithms, see [3], [36] and [37].

In two dimensions it involves the use of magnitude of gradient.

$$TV(u(x, y)) = \int |\nabla u(x, y)| \, dx \, dy.$$

The total variation technique explained in [35] was revolutionary for image denoising community. It gives optimal result by solving a partial differential equation. There are two constraints for the optimization problem which stem from the structure of the noise. One of them is zero mean residual constraint. The other one states that residual L_2 norm should be equal to noise standard deviation. The cost to minimize is the $2-D$ total variation. These costs and constraints mean that a sequence of signals in decreasing total variation are obtained such that each signal in the sequence is the best fit to the noisy signal

among the set of signals with the same total variation. The denoised signal is chosen from the obtained sequence of signals based on a statistical check on their corresponding fit error and noise standard deviation on the original signal. This optimization problem is solved by Euler-Lagrange equation which is a PDE technique. This method yields accurate and efficient results.

Staircase effect is a crucial problem in denoising based on total variation. This means that flat steps are observed at the outputs. However, the original signal without noise may not have flat steps, e.g. a linear approximation can be a better fit. The staircase effect (the flat parts in the output) was overcome in [38] by a technique at which a nonlinear fourth order term was added to the Euler-Lagrange equations of the total variation problem given in [35]. In [39], noise is reduced with total variation approximations combined with natural boundary conditions and no staircase-effect is observed. The stair-case effect was overcome by penalizing error to the original signal with a high order functional in order to regularize the output.

Generally, stair-case effect is overcome by adding a penalty term for regularization. Weighted combination of the total variation of the image and its L_2 norm distance to the original image is defined as the cost to be minimized. During the total variation iterations, this cost is monitored and the iterations are stopped at the observed minima of the cost function. The choice of the linear mixing weights determine the denoising performance. Several research has been done based on this idea. See [35], [40], [41], [42], [43], [44] and [45] for theory and examples. However, there is a problem with this method. It remarkably increases the computation burden. Actually, the algorithm given in [43] was implemented and only extremely small images could be denoised.

There are several different ways to solve total variation problems. In [46], total variation reduction is achieved by convex optimization. In [47], total variation reduction is accomplished by quadratic programming. In that method,

total variation of the two dimensional signal is the constraint of the optimization problem. In [48], total variation and wavelet transform approaches are combined. In [49] a Partial Differential Technique (PDE) is introduced to reduce edge artifacts that wavelet thresholding yields.

1.2.2 2 – D Noise Standard Deviation Estimation

The benchmark techniques, i.e. curvelet transform, total variation by Rudin *et al* and the proposed techniques Slice-TV and R-TV require *a priori* estimation of noise standard deviation. In this part of the thesis, a new technique is proposed. In this technique, the image is filtered with a median filter. The distance of noisy observation from the output of this filtering is taken as the noise σ estimate.

1.2.3 Slice-TV Technique

In this part of the thesis, a viable automated technique is introduced. This method achieves the denoising task without degrading the resolution of the images. Moreover, it has minimal undesired effects such as edge smoothing. In the method proposed, the efficiency problem of regularization in two dimensions was resolved by processing one dimensional slices of an image. In this method, the optimal 1 – D denoising technique given in [4] is used to denoise 1 – D row vectors of an image. However, the cost and constraint of the problem of that denoising tool was interchanged. In the original implementation, signal is smoothed and the signal with the best fit to the noisy signal is chosen among the signals with a given total variation. In the interchanged version, the signal with the least total variation is selected among the signals with a given distance to the noisy observation. This means that total variation of a vector is minimized with a known difference norm. In the image denoising technique proposed here, an image is denoised by decreasing the noise of its row-wise and column-wise slices.

The average of those two outputs is taken to reduce the discontinuities between neighboring $1 - D$ vectors.

1.2.4 Radon-TV Technique

Curvelet has an ambiguity which can be perceived as a drawback. It is not clear which components of the signals are kept via curvelet transform and which components are blocked. This means that the edges are not necessarily kept through curvelet transform. It shows low pass filtering effect on the edges which is not desired. Moreover, it has no optimality for the defined denoising problem. On the other hand, it has been proven that total variation is the optimal for the denoising of one dimensional signals. Though, the problem is that the optimality has not been carried to two dimensions. Hence, it is seen that both of these techniques have some drawbacks. It is reasonable to claim a method that combines the good sides of both of these denoising techniques. Curvelet uses radon projections as a first stage to carry the image to one dimension, so this part of curvelet can be kept at the new technique and total variation performs optimally in one dimension. After that, the image can be back projected and recovered to two dimensions.

In the approach named as R-TV which is introduced in this thesis, the total variation iterations for each projection is stopped when the L_2 norm distance between the denoised projection and the original projection reaches to a threshold determined by the estimated noise variance in the Radon projection domain.

1.2.5 $2 - D$ Texture Recovery by Total Variation

The fact that the optimal total variation can not be carried to two dimensions optimally is a challenging problem of image denoising. In this part of the thesis,

the total variation algorithm by [4] was carried to two dimensions by the separable implementation. The total variation reduction was carried out in vertical direction and this output was given as an input to the horizontal total variation reduction. In these two steps of $1-D$ total variation implementation, a relatively smaller stopping parameter was used in the process. Since total variation does not smooth out the edges; the texture composed of abrupt changes was observed to be kept successfully by this separable implementation. The simulations have shown that the performance of this technique is outstanding when it is applied on textured images such as classical Barbara image.

Chapter 2

1 – D SIGNAL DENOISING

The one dimensional signal denoising is an important area of signal processing. It comprises the denoising of speech signals, digital and analog telecommunication signals. Moreover, it can be used as a building block of image denoising. Since the algorithms such as total variation, polynomial fitting, matched filtering and dynamic programming provide optimal solution for one dimensional case, the $1 - D$ signal denoising can be accomplished relatively more reliably. Although these algorithms provide optimal solutions, there are still some challenging problems in $1 - D$ denoising. Since the modern denoising algorithms focus on high resolution techniques, the main objective in this chapter is to develop edge preserving denoising techniques. For this edge detection problem, three main approaches have been proposed, and for the purpose of smoothing in between the detected edges, optimal polynomial fitting is exploited. These least squares fitting based techniques and their comparison is introduced in Figure 2.1.

TECHNIQUE	Edge Detection Method	Smoothing Method	Input Parameters	Computational Complexity
<i>Piecewise Linear Fitting</i>	Total Variation	Line Merging	Noise σ	$O(N \log N)$
<i>Convolution with Wavelet</i>	Convolution Maximization	Adaptive Polynomial Fitting	Noise σ	$O(N^2)$
<i>PSO based polynomial fitting</i>	PSO	Polynomial Fitting	Polynomial orders, Number of transition boundaries	Extremely Fast

Figure 2.1: The table which compares the proposed polynomial fitting based algorithms.

2.1 1 – D Noise Standard Deviation Estimation

Denoising is the act of smoothing a signal. In discrete iterative implementations, the noisy signal is smoothed step by step and the stopping criterion is crucial to conclude the iterations. Noise standard deviation was used for this purpose in the standard algorithms [4], [35] and [3].

Actually, noise standard deviation can be achieved by crude non-parametric filtering. In this part, the running average filter is exploited for one dimensional noise standard deviation estimation. It gives nearly-optimal results for linear signals and piecewise constant signals. Therefore, it is good practice to perform this filtering in the parts of the signal with small standard deviation. In this estimation process, the difference of intensity of each sample from the average of the samples in the neighborhood of that sample is taken. The standard deviation of this difference yields the following estimate to the standard deviation of noise, σ estimate. The formula is:

$$\hat{\sigma} = \sqrt{\frac{1}{N_w} \sum_{n=N-\frac{N_w}{2}}^{n+\frac{N_w}{2}} \left(p_i[n] - \frac{1}{N_w} \sum_{k=n-\frac{N_w}{2}}^{n+\frac{N_w}{2}} p_i[k] \right)^2}. \quad (2.1)$$

This estimator was tested with a piecewise linear signal with a single jump discontinuity. The signal is shown in Figure 2.2. In that figure, the noise standard deviation is 1. This value was estimated as 0.9895 by the equation given in (2.1). The window length N_w was chosen as 8. A too short window overfits the signal giving a lower estimation. A too long window oversmooths the signal giving a value higher than the actual one.

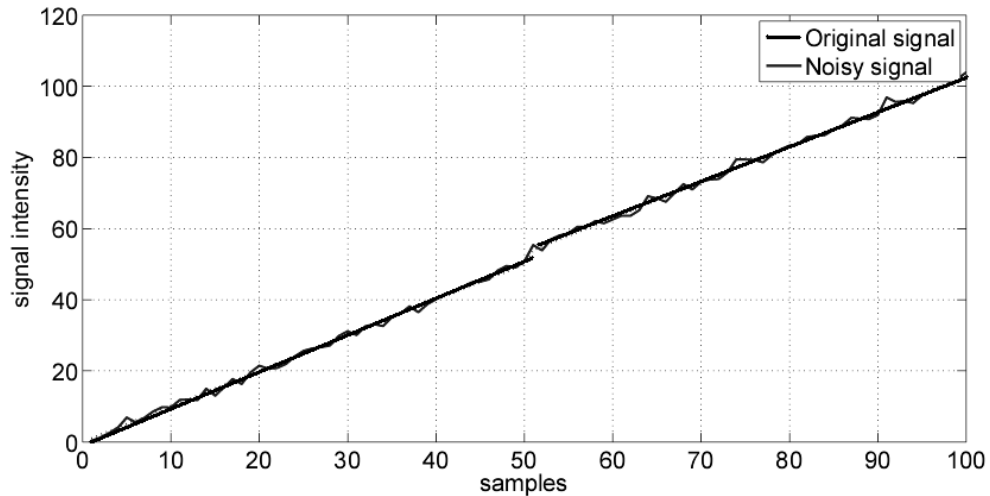


Figure 2.2: Original piecewise linear signal with one jump and the noisy version.

2.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a technique whose application to global optimization problems became more and more common in the recent years. It is similar to Differential Evolution [50] and Genetic Algorithms [51].

Particle Swarm Optimization was inspired by the flocking behavior of the birds. In the movement model of the bird flocks, the birds are assumed to exchange the information of best position to each other. With this communication,

their position converges to some point. The velocity evolution of the birds are influenced by their individual best position and global best position.

The particles have two attributes which evolve in time. One of them is their position, and the other is their velocity. These attributes are randomly initialized. Also, the particles keep two parameters in their memory. They are their individual best position and global best position. Packed into a mathematical model, the next velocity of the particles is determined by the linear combination of their previous velocity, the difference of the personal best position and the previous position, and the difference of global best position and previous position. This method so called Standard PSO is described in [34] in a detailed fashion. The summary given here was taken from that source. The random initialization of attributes and the movement towards the individual as well as global best position increases the diversity of the solution. These aspects name PSO as a global optimization method.

The evolution of the parameters is achieved by updating the attributes and parameters. The update equations are given as follows.

$$v_{ij} = \alpha(v_{ij} + w_1\zeta_1(I_{ij} - p_{ij}) + w_2\zeta_2(G_j - p_{ij})), \quad (2.2)$$

$$p_{ij} = p_{ij} + v_{ij}. \quad (2.3)$$

In these equations, v_{ij} is the j^{th} dimension of the velocity of the i^{th} particle. p_{ij} is the j^{th} dimension of the position of the i^{th} particle. I_{ij} is the j^{th} dimension of the individual best position of the i^{th} particle. G_j is the j^{th} dimension of the global best position attained until that time. ζ_1 and ζ_2 are chosen as independent uniform random variables in the interval $(0, 1)$.

It would make the process clearer to write these update equations as an algorithm.

Algorithm 1 The PSO procedure

```

for each time sample  $t$  until convergence do
  for each particle  $i$  in the swarm do
    update velocity by (2.2)
    update position by (2.3)
    update  $I_{ij}$  and  $G_j$  if necessary
  end for
end for

```

2.2.1 Parameter Selection for Stability and Convergence

The final output which is the position of convergence may get stuck to a local minimum, and this makes the PSO algorithm an heuristic approach. However, the parameters can be found analytically in an optimal manner [52]. Due to the theoretical analysis of the standard PSO algorithm, the relation between the parameters α , ζ_1 , and ζ_2 were derived as follows.

$$\alpha = \frac{2}{|2 - \xi - \sqrt{\xi^2 - 4\xi}|}, \quad (2.4)$$

$$\xi = \zeta_1 + \zeta_2. \quad (2.5)$$

It was proven that the convergence and stability of the standard PSO algorithm only depends on ξ . If $\xi < 4$ the position will oscillate which means that the position may not converge. If $\xi > 4$ convergence is certain. For ensured convergence but a more accurate result ξ should not be too high and must be slightly higher than 4. For this purpose, ξ was chosen as 4.1. This means that α is equal to 0.72984, the so-called magic number in the literature. ξ is constituted by the sum of ζ_1 and ζ_2 . If ζ_1 is increased, the diversity increases. On the other hand, if ζ_2 is increased, the exploration around the global minima is increased. To balance the compromise between these two effects, ζ_1 and ζ_2 are chosen to be equal to 2.05.

2.3 Theory of Total Variation Minimization

2.3.1 Geometry of Total Variation Reduction

Let $\mathbf{x}$ be an arbitrary one dimensional signal and $\mathbf{y}$ be a function which is the superposition of noise and the original function $\mathbf{x}$. Denoising is the act of estimating the original signal whose noisy version is observed:

$$\mathbf{y} = \mathbf{x} + \mathbf{u}, \quad (2.6)$$

where $\mathbf{u}$ is the noise on $\mathbf{x}$. Normally, noise is zero mean, that is, $E[\mathbf{u}] = 0$, Moreover, the noise at a point of $\mathbf{x}$ is uncorrelated with respect to a noise at any other arbitrary point and the variance of the noise is σ^2 . To represent mathematically,

$$E[u_i \cdot u_j^*] = \delta_{ij} \cdot \sigma^2. \quad (2.7)$$

The objective is to recover $\mathbf{x}$ from the noisy observations. We have to recover the original function $\mathbf{x}$ from the observable function $\mathbf{y}$. We can denote the estimated function as $\mathbf{x}_R$. We are throwing some part of the observed signal away while smoothing it. This part can be defined as an error and can be denoted as $\mathbf{e}$.

$$\mathbf{y} = \mathbf{x}_R + \mathbf{e}. \quad (2.8)$$

Ideally, the error should have the same characteristics as the noise. Therefore, it should have zero mean which implies zero average because of ergodicity:

$$\sum_{i=1}^N e_i = 0. \quad (2.9)$$

where N is the number of samples taken from the function. Summing (2.8) for this constraint:

$$\sum_{i=1}^N y_i = \sum_{i=1}^N x_{Ri}, \quad (2.10)$$

which constitutes the first constraint based on zero mean noise. If the number of samples is increased, the average L_2 norm of the error should converge to the variance of the noise.

$$\frac{1}{N} \|\mathbf{y} - \mathbf{x}_R\|_2^2 = \sigma^2, \quad (2.11)$$

or we can express this:

$$\|\mathbf{y} - \mathbf{x}_R\| = \sqrt{N}\sigma. \quad (2.12)$$

We can assume (2.12) as a second constraint. Expressing this as a summation:

$$\frac{1}{N} \sum_{i=1}^N |y_i - x_{Ri}|^2 = \sigma^2. \quad (2.13)$$

(2.12) represents a circle with a radius $r_0 = \sqrt{N} \cdot \sigma$ for two samples of the signal $\mathbf{y}$. The line $x_2 + x_1 = y_1 + y_2$ is a specific case of (2.10) for two samples. It is a line that passes through the point (y_1, y_2) and it is perpendicular to the line $x_2 = x_1$. The intersection of this line with the circle defined by (2.11) satisfies the two constraints about characteristics of noise. Naturally, solving these two equations yield two solutions. We can engage a cost based on total variation $((x_{R1}, x_{R2}) = \underset{(x_1, x_2)}{\operatorname{argmin}}\{|x_2 - x_1|\})$ s.t. $(x_1, x_2) \in \{(a_1, a_2), (b_1, b_2)\}$. The situation is depicted in Figure 3.5.

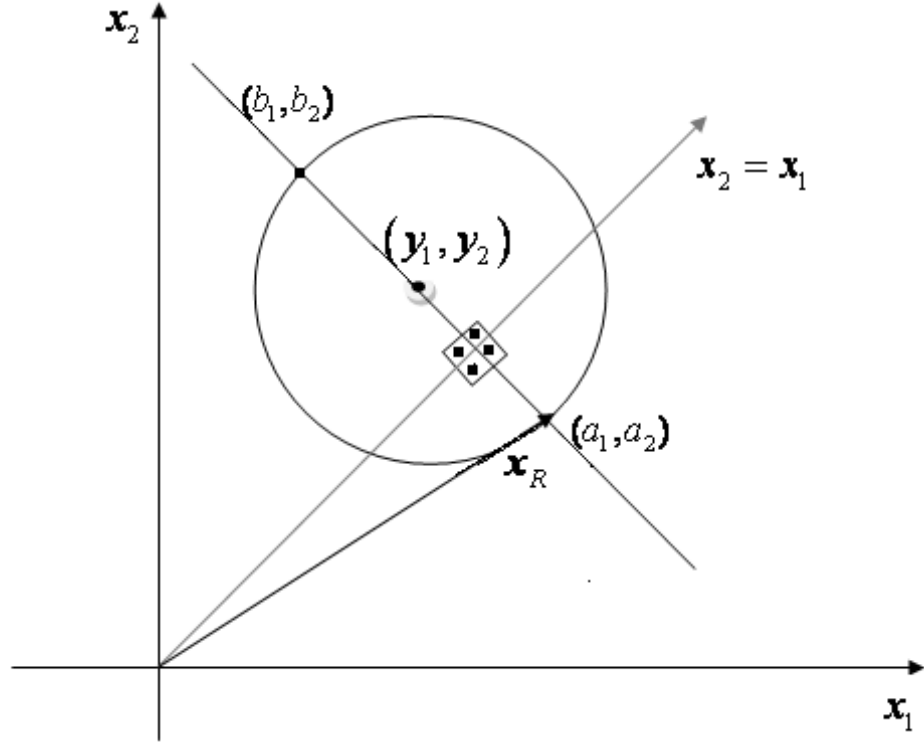


Figure 2.3: The geometry of the solution for two dimensions.

This geometry can be interpreted as follows as well. (2.10) for $N = 2$ implies that the projection of the vector $[y_1, y_2]$ on the line $x_1 = x_2$ should be the same as the projection of the recovery vector $\mathbf{x}_R$ on the same line. It is known that the recovery vector should be on the circle with a radius $r_0 = \sqrt{2} \cdot \sigma$ as imposed by (2.12). So the analytic solution is the set of vectors composed of (a_1, b_1) and (a_2, b_2) . The point with the least total variation is the unique solution. The total variation of a $1 - D$ signal is defined as:

$$TV(\mathbf{x}) = \sum_{i=1}^{N-1} |x_{i+1} - x_i|. \quad (2.14)$$

The situation can be generalized for three dimensions. The geometric shape which represents the vectors whose projections on $[1 \ 1 \ 1]$ have the same length as the observation vector $\mathbf{y}$ is a plane now. That plane passes through the origin of the sphere and is perpendicular to $[1 \ 1 \ 1]$. The intersection of the plane with the sphere is a circle now. The point of that circle that has the smallest total variation is the solution with the minimum cost.

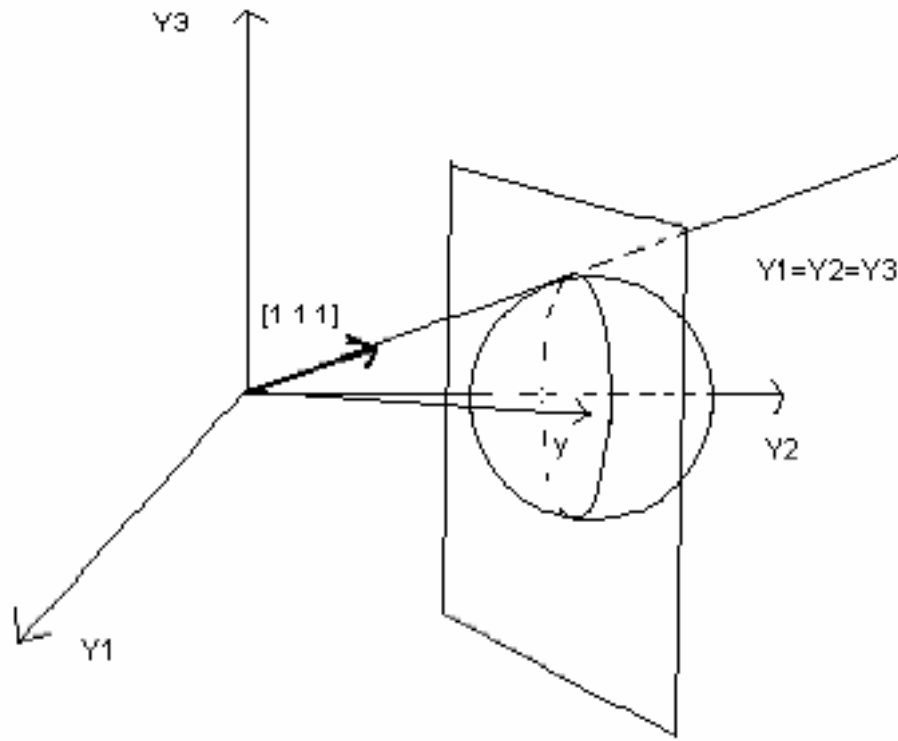


Figure 2.4: The geometry of the solution for three dimensions.

Since the minimum value of $|Y_2 - Y_1|$ is 0, the solution is the intersection of the constant variance sphere with the line $Y_2 = Y_1$. In case there is no intersection with the sphere, then the first line starting from $Y_2 = Y_1$ that intersects the sphere at a point is taken.

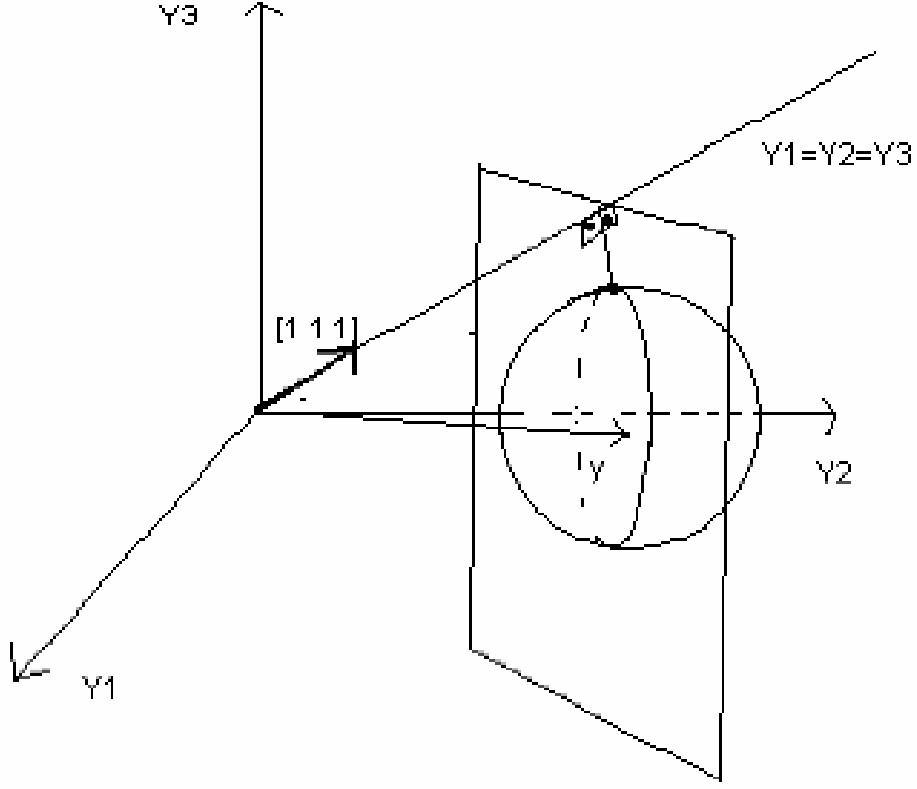


Figure 2.5: The case if the variance sphere and the line $y_1 = y_2 = y_3$ do not intersect.

In the techniques based on the known variance, if σ is supposed to be nearly zero, no noise removal is performed and the $\mathbf{x}_R$ is equal to $\mathbf{y}$. In $1 - D$, the graph of $TV(\mathbf{x})$ versus σ is like in Figure 2.7.

It is beneficial to plot the logarithmic graph of X_1 since it is sometimes easier to see the critical points at which $\log(\cdot)$ makes a jump. Logarithmic plot and the critical points usually look like in Figure 2.8.

Depending on the level of the estimated noise it may give more accurate results to choose the first jump or the second jump. This heuristic approach was introduced in a comprehensive paper of a total variation algorithm [4].

An elegantly chosen σ smooths the function properly making a good recovery of the original signal.

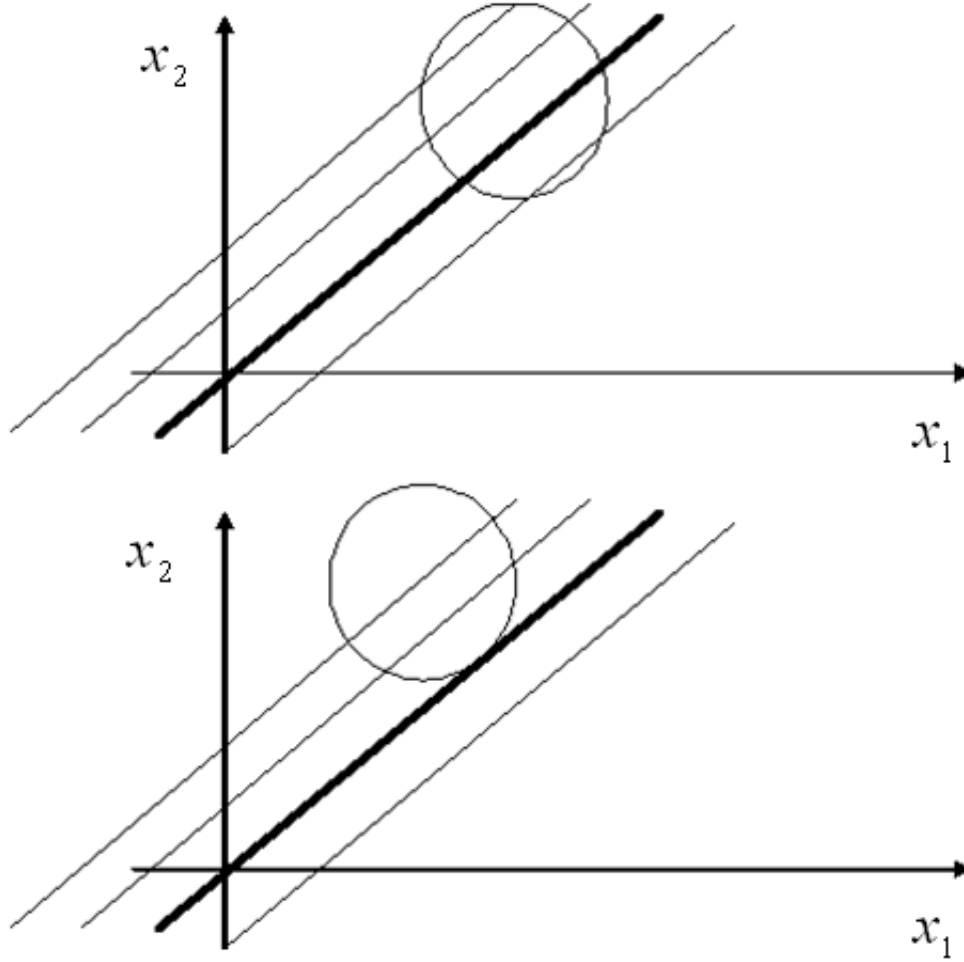


Figure 2.6: The points that have minimum total variation in two dimensions for different spheres.

2.3.2 Total Variation Reduction by Quadratic Programming

Let us try to find the analytic solution of the point that has the minimum total variation on an N dimension constant variance sphere. It will be proper to start with three dimensions. In $3D$, the problem is reduced to this constraint and cost.

$$\text{Given } (x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 = 3\sigma^2;$$

we should minimize $|x_2 - x_1| + |x_3 - x_2|$

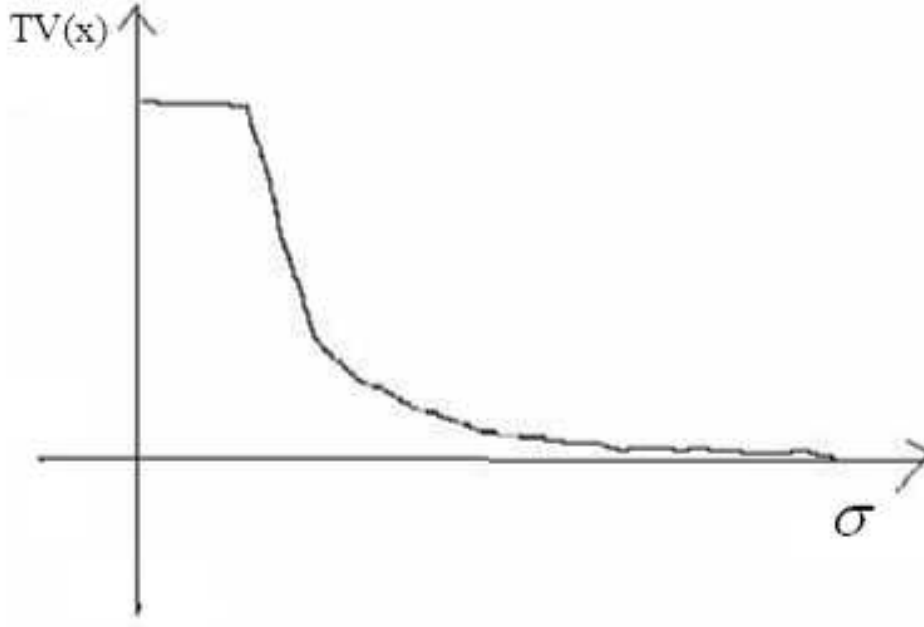


Figure 2.7: The graph of the total variation of the denoised vector versus σ .

The overall solution is the intersection of the sphere with the line $x_1 = x_2 = x_3$ if they intersect. If there is no intersection of the sphere with the line $x_1 = x_2 = x_3$, these cases should be considered:

- If the sphere is in the space of $x_2 \geq x_1$ and $x_3 \geq x_2$ we should minimize $x_3 - x_1$ for $x_3 - x_1 = C_1$ with minimum positive C_1 to intersect with the sphere at a point, and that intersection point is the solution.
- If the sphere is in the space of $x_2 \leq x_1$ and $x_3 \leq x_2$ we should minimize $x_1 - x_3$ for $x_1 - x_3 = C_2$ with minimum positive C_2 to intersect with the sphere at a point, and that intersection point is the solution.
- If the sphere is in the space of $x_2 \geq x_1$ and $x_3 \leq x_2$ we should minimize $2x_2 - x_3 - x_1$ for $2x_2 - x_3 - x_1 = C_3$ with minimum positive C_3 to intersect with the sphere at a point, and that intersection point is the solution.
- If the sphere is in the space of $x_2 \leq x_1$ and $x_3 \geq x_2$ we should minimize $-2x_2 + x_3 + x_1$ for $-2x_2 + x_3 + x_1 = C_4$ with minimum positive C_4 to intersect with the sphere at a point, and that intersection point is the solution.

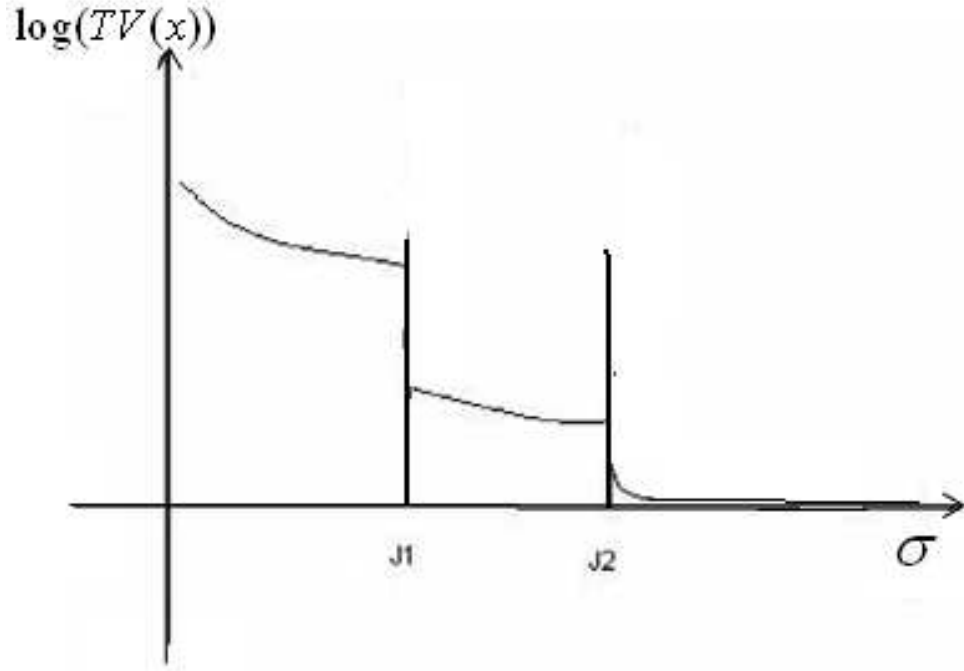


Figure 2.8: Logarithmic plot of the total variation of the denoised vector versus σ .

Note: If the sphere is in more than one space, more than one solution can occur. We should take the point for the minimum C .

Let us consider the situation in four dimensions.

1. The overall solution is the intersection of the sphere by the line $x_1 = x_2 = x_3 = x_4$ if they intersect.

2. If there is no intersection of the sphere with the line $x_1 = x_2 = x_3 = x_4$, the smallest C_i is taken for intersection to occur at the points specified below.

- $x_4 \geq x_3, x_3 \geq x_2, x_2 \geq x_1 \Rightarrow x_4 - x_1 = C_1$ is solution
- $x_4 \geq x_3, x_3 \geq x_2, x_2 \leq x_1 \Rightarrow 2.(x_1 - x_2) + x_4 - x_1 = C_2$ is solution
- $x_4 \geq x_3, x_3 \leq x_2, x_2 \geq x_1 \Rightarrow 2.(x_2 - x_3) + x_4 - x_1 = C_3$ is solution
- $x_4 \geq x_3, x_3 \leq x_2, x_2 \leq x_1 \Rightarrow 2.(x_2 - x_3) + 2.(x_1 - x_2) + x_4 - x_1 = C_4$ is solution

0+	0+	0+	$x_4 - x_1 = C_1$
0+	0+	$2.(x_1 - x_2) +$	$x_4 - x_1 = C_2$
0+	$2.(x_2 - x_3) +$	0+	$x_4 - x_1 = C_3$
0+	$2.(x_2 - x_3) +$	$2.(x_1 - x_2) +$	$x_4 - x_1 = C_4$
$2.(x_3 - x_4) +$	0+	0+	$x_4 - x_1 = C_5$
$2.(x_3 - x_4) +$	0+	$2.(x_1 - x_2) +$	$x_4 - x_1 = C_6$
$2.(x_3 - x_4) +$	$2.(x_2 - x_3) +$	0+	$x_4 - x_1 = C_7$
$2.(x_3 - x_4) +$	$2.(x_2 - x_3) +$	$2.(x_1 - x_2) +$	$x_4 - x_1 = C_8$

Table 2.1: The total variation equations for four dimensional vector.

- $x_4 \leq x_3, x_3 \geq x_2, x_2 \geq x_1 \Rightarrow 2.(x_3 - x_4) + x_4 - x_1 = C_5$ is solution
- $x_4 \leq x_3, x_3 \geq x_2, x_2 \leq x_1 \Rightarrow 2.(x_3 - x_4) + 2.(x_1 - x_2) + x_4 - x_1 = C_6$ is solution
- $x_4 \leq x_3, x_3 \leq x_2, x_2 \geq x_1 \Rightarrow 2.(x_3 - x_4) + 2.(x_2 - x_3) + x_4 - x_1 = C_7$ is solution
- $x_4 \leq x_3, x_3 \leq x_2, x_2 \leq x_1 \Rightarrow 2.(x_3 - x_4) + 2.(x_2 - x_3) + 2.(x_1 - x_2) + x_4 - x_1 = C_8$ is solution

If the matrix of the coefficients of the unknowns of eight above equations is formed, the matrix looks like in the Table 2.1.

The overall solution for N dimensions is the solution of the quadratic (2.15) for if there exists a real solution.

$$\frac{1}{N} \sum_{i=1}^N |y_i - x_R|^2 = \sigma^2. \quad (2.15)$$

If there exists no solution, the table is used. The table is analogous to a binary number which provides a means to write an algorithm for a general N dimension image vector. The algorithm is like in Algorithm ??:

write (2.15) as a constraint.

initialize second equation as $C = +x(N) - x(1)$

set l to 1

for $i = 1$ to 2^{N-1} **do**

 set k to i

while k not equal to 0 **do**

if $\text{mod}(k, 2) = 0$ **then**

 concatenate $+2.(x(l) - x(l + 1))$ to the second equation and take this
 as i^{th} equation.

end if

 set k to $\text{floor}(k/2)$

$l = l + 1$

end while

 set l to 1 and go to step (4)

end for

Solve the first equation as a pair with the equations given by Step (8). and
check the inequality for that equation to see if the point is in the right space.
Then, take the equation with minimum C for intersection at a point. That
point is the solution.

2.3.3 Finding the Points that have Constant Total Variation in Three Dimensions

The aim is to find the points that have a constant total variation C for three dimensions. The analytic solution is like in Figure 2.9.

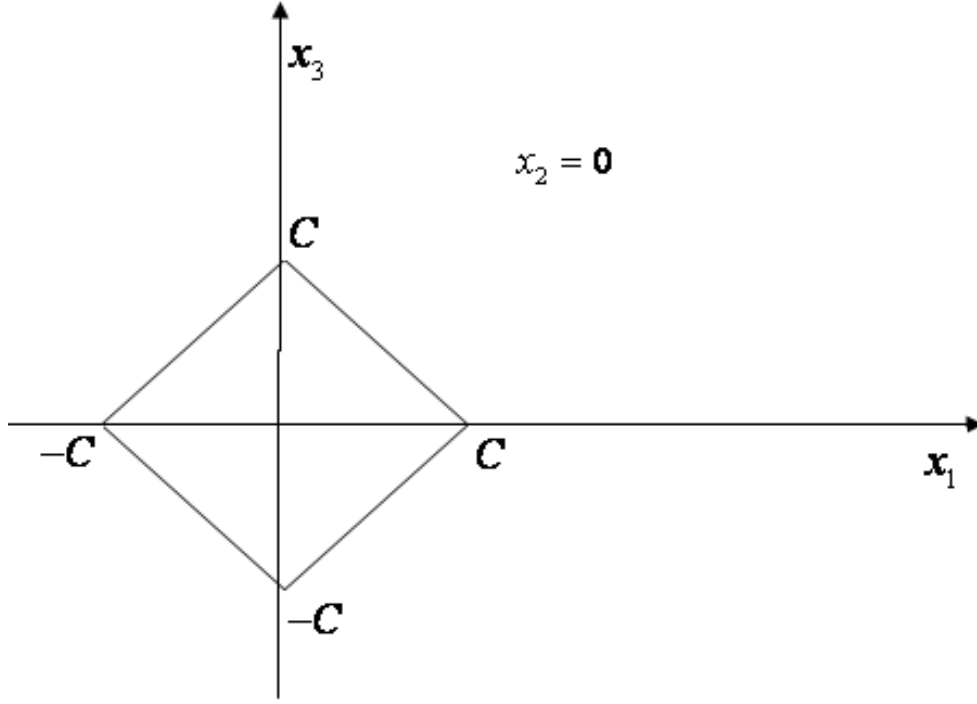


Figure 2.9: The points that have constant total variation for $x_2 = 0$.

If $x_2 = C/2$

1. $C/2 \geq x_1$ and $x_3 \geq C/2 \Rightarrow x_3 - x_1 = C$
2. $C/2 \leq x_1$ and $x_3 \leq C/2 \Rightarrow -x_3 + x_1 = C$
3. $C/2 \geq x_1$ and $x_3 \leq C/2 \Rightarrow -x_3 - x_1 = 0$
4. $C/2 \leq x_1$ and $x_3 \geq C/2 \Rightarrow x_3 + x_1 = 2C$

These four partial functions represent the shape in Figure 2.10.

Knowing that this is a linear system, we can come up with the conclusion that the 3D shape is a square prism whose center is shifted on the line $x_1 = x_2 = x_3$ and whose corners are C units away from the center in the directions of x_1 and x_3 . Returning back to the solution in three dimensions, this prism can be used to find the solution set. The overall solution is the point that intersects the variance sphere. If this prism does not intersect the sphere for C , we should increase C until it cuts the sphere at a point. Let us examine the case of the intersection

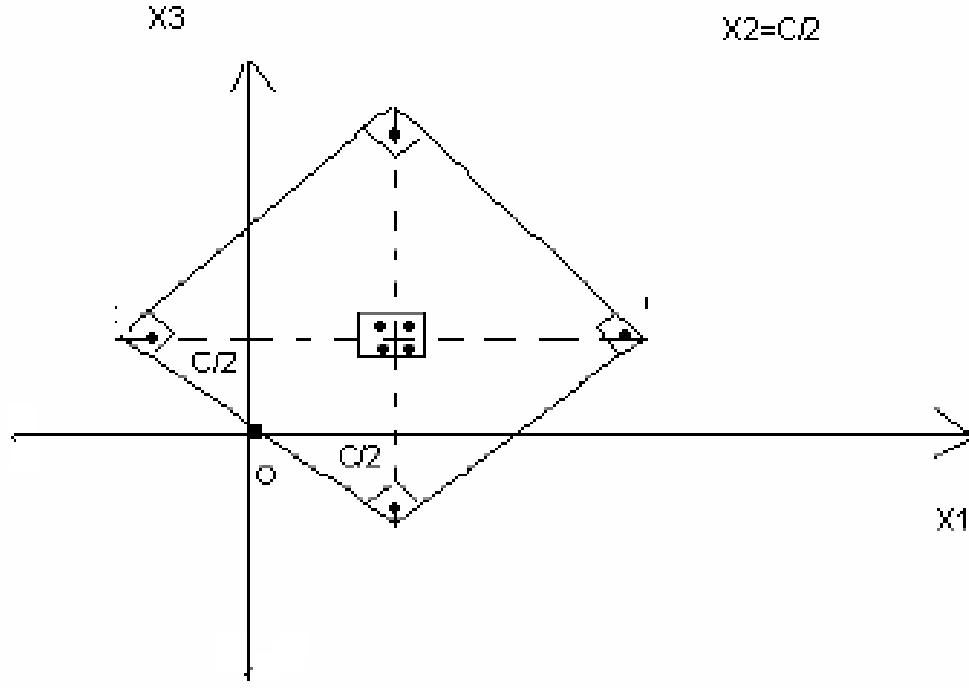


Figure 2.10: The points that have constant total variation for $x_2 = C/2$.

of the square prism characterized as above with a constant variance sphere. If the intersection were in two dimensions, it would be a circle, since the slice of a sphere is a circle. However, it is known that the slice of a square prism is a square, therefore, the intersection is forced to take place in $3D$. The intersection consists of two of the shapes below one of which occurs when the prism enters the sphere and the other occurs when the prism leaves the sphere. The projection of the $3D$ intersection to the planes of x_1 and x_3 is as shown in Figure 2.11.

In three dimensions, the points that have a constant total variation of 50 are as shown in the Figure 2.12.

The algorithm to draw these points is as follows:

Algorithm 3 Constructing points with constant total variation in $3 - D$

for $j = -L$ to L **do**

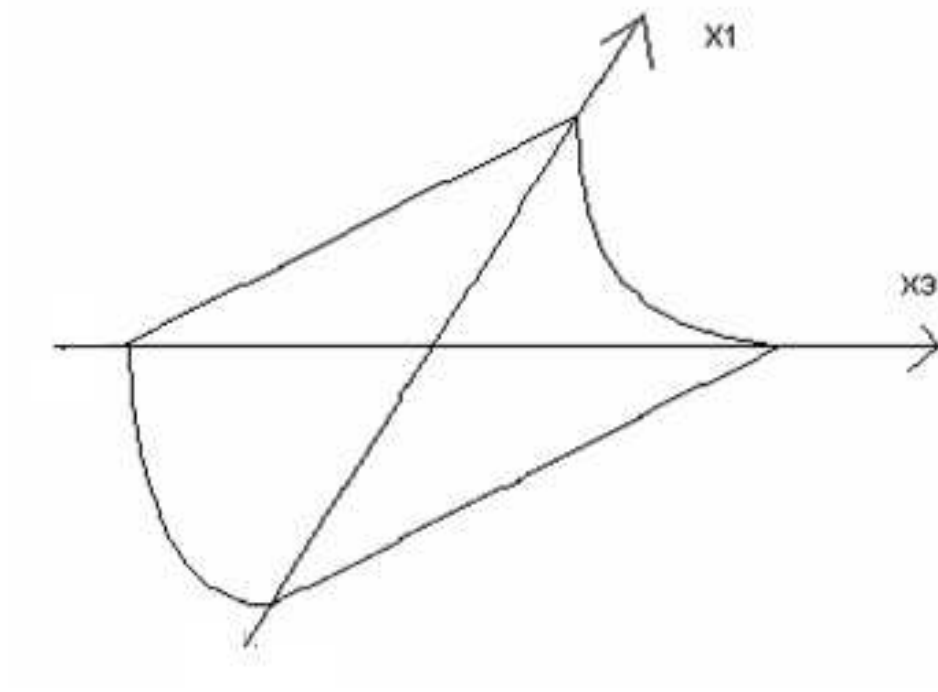


Figure 2.11: The projection of the 3D shape that represents the intersection by a sphere in two dimensions.

```

for  $i = -C + j$  to  $C + j$  do
     $x_1(i) = i$ ;
     $x_2(i) = j$ ;
    if  $j > i$  then
         $x_{3a} = C + i$ 
         $x_{3b} = 2j - i - C$ 
    else
         $x_{3a} = C + 2j - i$ 
         $x_{3b} = i - C$ 
    end if
end for
end for

```

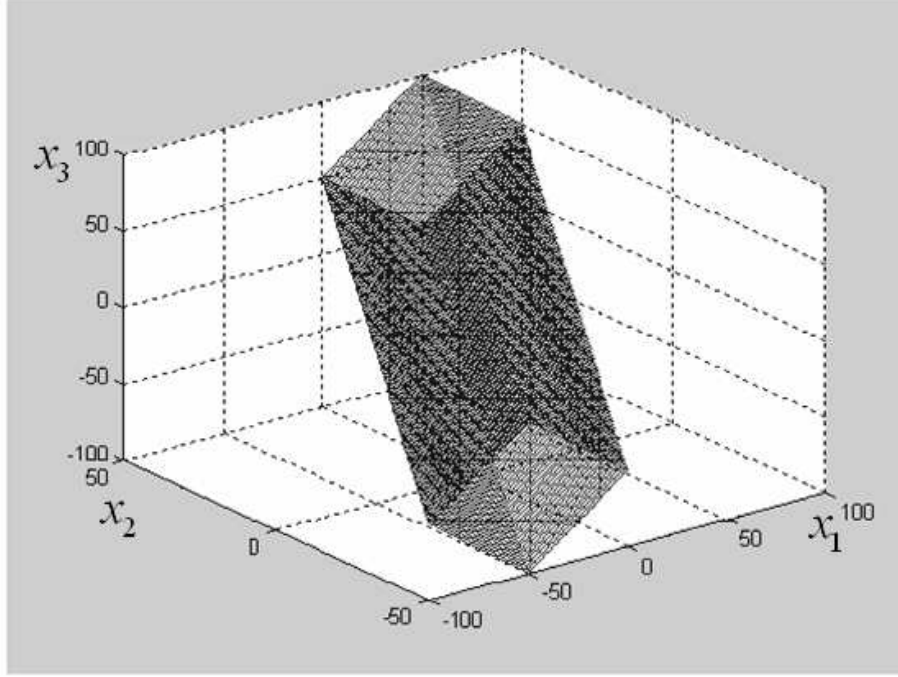


Figure 2.12: The graph of constant total variation in three dimensions.

2.3.4 An Algorithm to Reduce Total Variation of a Vector

An iterative optimal algorithm was published and the algorithm introduced here is a simpler version that achieves the same task. Let us assume that we should minimize $\|\mathbf{x} - \mathbf{x}_0\|_2^2$ such that $TV(\mathbf{x}) = \mu$.

It is shown in [4] that $TV(\mathbf{x}) \leq \mu$ is convex. Because, all smooth differentiable signals consist of constant, increasing or decreasing segments. Therefore, their linear combination will have total variation less than the part with a highest total variation and greater than the part with least total variation.

As shown above, there are different convex sets with different total variation. To find the optimal total variation, the plot of $\|\mathbf{x} - \mathbf{x}_0\|_2^2$ shown in figure can be used. The critical total variation is the μ_{opt} shown in the plot. Here, the objective is to minimize $J(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_0\|_2^2$ s.t. $\mu = \mu_{opt}$. If μ is increased beyond μ_{opt} , the cost does not decrease much.

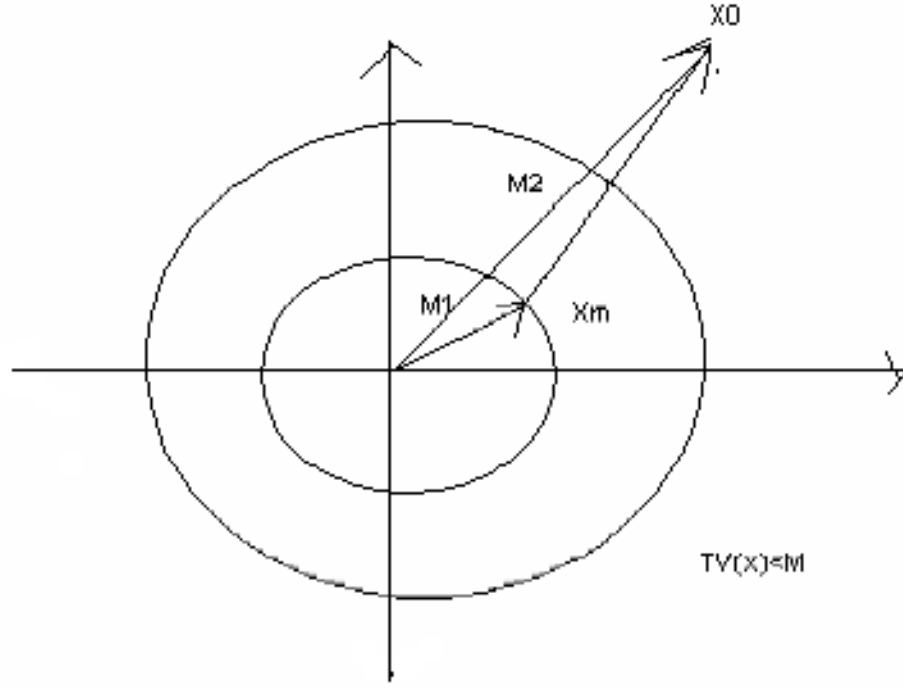


Figure 2.13: The convex sets with different total variation.

The effect of total variation reduction is like taking the piecewise average of the signal as in Figure 2.15. The final signal is a constant signal if the total variation is reduced to 0. This is depicted in Figure 2.16.

Let us take a vector to show the process of reducing total variation explained in [4].

$$\mathbf{x} = \begin{pmatrix} 1 & 3 & 2 & 4 \end{pmatrix}. \quad (2.16)$$

Let us perturb this vector so that we can reduce total variation.

$$\mathbf{x}_\delta = \begin{pmatrix} 1 + \delta_1 & 3 + \delta_2 & 2 + \delta_3 & 4 + \delta_4 \end{pmatrix}. \quad (2.17)$$

Now the total variation becomes:

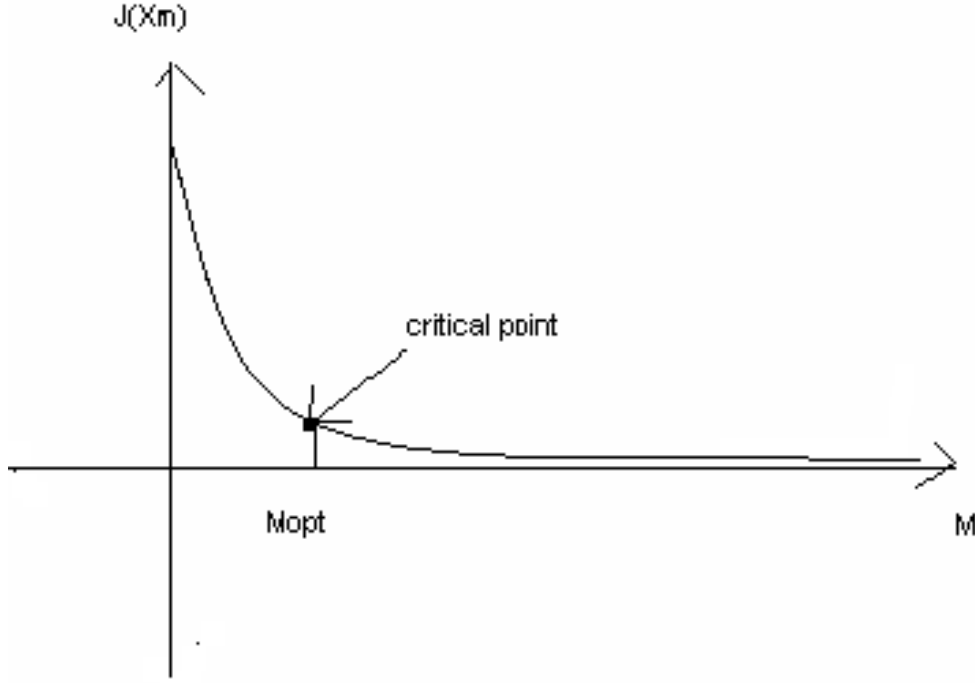


Figure 2.14: Finding the optimal μ .

$$\begin{aligned}
 TV(\mathbf{x}_\delta) &= (2 + \delta_2 - \delta_1) + (1 + \delta_2 - \delta_3) + (2 + \delta_4 - \delta_3) \\
 &= 5 - \delta_1 + 2\delta_2 - 2\delta_3 + \delta_4 \\
 &= TV(\mathbf{x}) + \begin{pmatrix} -1 & 2 & -2 & 1 \end{pmatrix} \cdot \delta^T
 \end{aligned}$$

So, a vector $\mathbf{a} = \begin{pmatrix} -1 & 2 & -2 & 1 \end{pmatrix}$ can be defined.

Then, the minimization cost becomes $\|\mathbf{x} - \mathbf{x}_\delta\|_2^2 = \|\underline{\delta}\|_2^2$.

Let us try to minimize $TV(\mathbf{x}_\delta)$ without changing the sorting of the entries. This means maximizing $\mathbf{a}^T \delta$. The vector $\mathbf{a}$ can be written by inspection. If the number of adjacent elements of an entry that are larger than it in signed manner (-1 for lower ones) is the number of the corresponding entry of $\mathbf{a}$. For example, if we consider the vector $\begin{pmatrix} 1 & 3 & 2 & 2 & 5 \end{pmatrix}$, the corresponding vector $\mathbf{a}$ is $\begin{pmatrix} 1 & -2 & 1 & 1 & -1 \end{pmatrix}$.

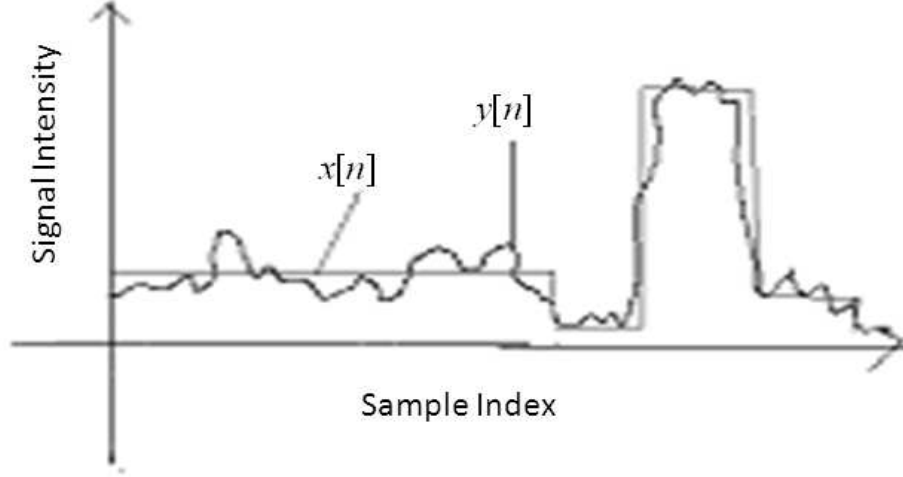


Figure 2.15: Piecewise constant behavior of total variation process.

Let us take the vector δ with the same entries δ . The perturbed vector becomes: $\begin{pmatrix} 1 + \delta & 3 - 2\delta & 2 + 2\delta & 4 - \delta \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 & 4 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 2 & -1 \end{pmatrix} \cdot \delta$

The maximum $\delta = 1/4$ for the sorting of the adjacent entries not to change. Because, $2 + 2\delta \leq 3 - 2\delta \Rightarrow 4\delta \leq 1 \Rightarrow \delta \leq 1/4$. After the perturbation of the vector with $\delta = 1/4$, the vector becomes: $\begin{pmatrix} 5/4 & 5/2 & 5/2 & 15/4 \end{pmatrix}$.

The total variation reduces from 5 to $-5/4 + 5/2 + 15/4 - 5/2 = 5/2$. If one of the adjacent different entries to the equal entries is larger and the other is smaller, those same entries should not be perturbed as is the case for the vector above. Because, if they are perturbed, the same entries approach one of the entries and go away from the other one, keeping the total variation unchanged. Mathematically, $\begin{pmatrix} 5/4 + \delta_1 & 5/2 + \delta_2 & 5/2 + \delta_3 & 15/4 + \delta_4 \end{pmatrix} \Rightarrow TV(\mathbf{x}_\delta) = 5/2 - 5/4 + \delta_2 - \delta_1 + |\delta_3 - \delta_2| + 15/4 - 5/2 + \delta_4 - \delta_3$.

$\mathbf{x}_\delta = \begin{pmatrix} 5/4 + \delta & 5/2 & 5/2 & 15/4 + \delta \end{pmatrix}$ If we go on perturbation, the total variation becomes zero for $\delta = 5/4$. The process by means of vector geometry is like in Figure 2.17:

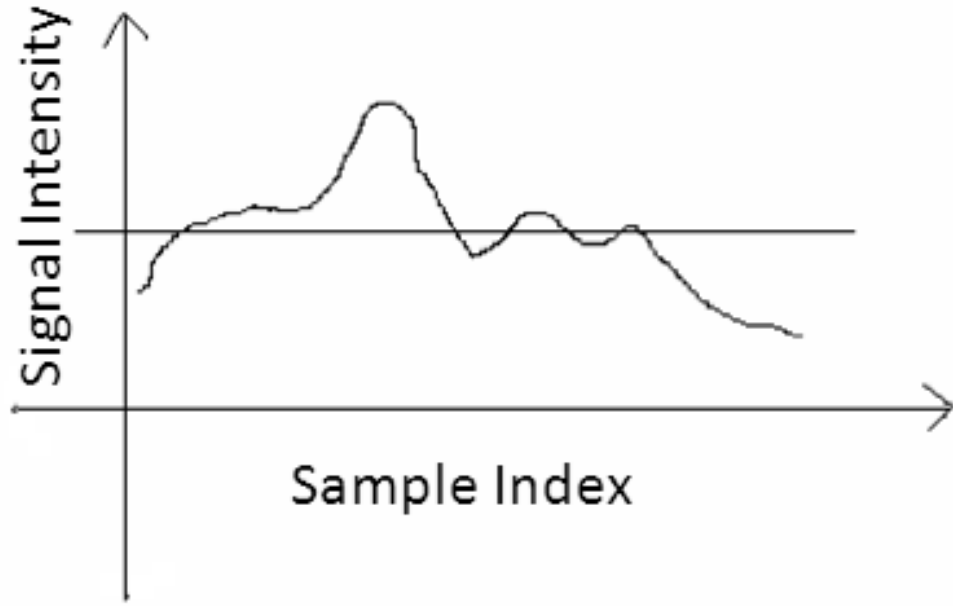


Figure 2.16: The final signal for $\mu = 0$.

With this process, we may not go to the vector with minimum total variation with one step and with more than one step we may end up with different minimization of L_2 norms. For the same resulting vectors, picture is like in Figure 2.18.

For different resulting vectors to be possible, the proof is geometric as shown in Figure 2.19:

The introduced algorithm should preserve the mean of the vector. This is because the noise is assumed to be zero mean. This mean constraint is achieved by taking the average of $\mathbf{a}$ for the flat parts.

$$\begin{pmatrix} 3 - \delta \\ 2 \\ 1 + 2\delta \\ 4 - \delta \end{pmatrix} \rightarrow \begin{pmatrix} 5/2 - \delta \\ 2 + \delta \\ 2 + \delta \\ 7/2 - \delta \end{pmatrix} \rightarrow \begin{pmatrix} 9/4 + \delta/3 \\ 9/4 + \delta/3 \\ 9/4 + \delta/3 \\ 13/4 - \delta \end{pmatrix} \rightarrow \begin{pmatrix} 5/2 \\ 5/2 \\ 5/2 \\ 5/2 \end{pmatrix}$$

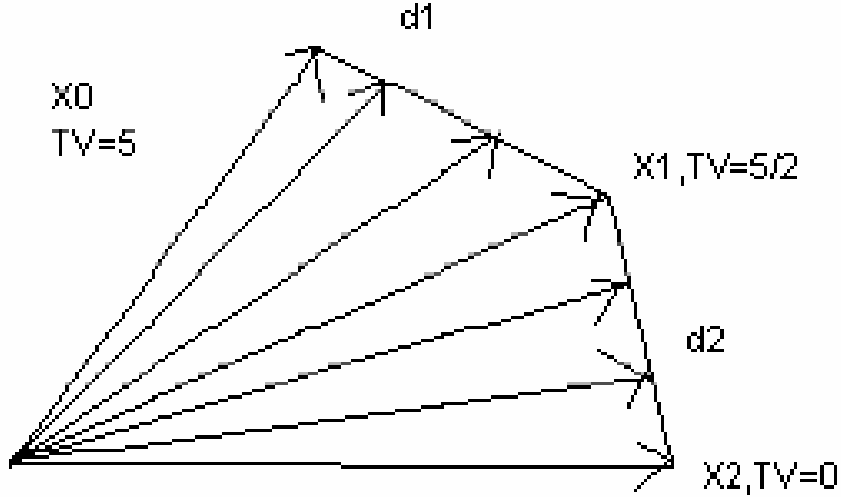


Figure 2.17: The process of total variation reduction.

2.4 Optimal Total Variation Reduction using Nonlinear Diffusion Equation

For one dimensional total variation minimization, the algorithm proposed by Willsky, *et al.* presents an optimal algorithm. The algorithm was compared by the method proposed by Rudin, *et al.* It is given with a constraint of total variation and a cost of difference norm:

$$\max\{\|\mathbf{u}(t_f) - \mathbf{u}(o)\|\} \quad s.t. \quad TV(\mathbf{u}) = t_v. \quad (2.18)$$

However, it is proposed in the same paper that the cost and constraint can be interchanged. However, this breaks the optimality of the algorithm in [4]. After this modification, the problem definition becomes:

$$\min\{TV(\mathbf{u})\} \quad s.t. \|\mathbf{u}(t_f) - \mathbf{u}(o)\| = N.\sigma^2, \quad (2.19)$$

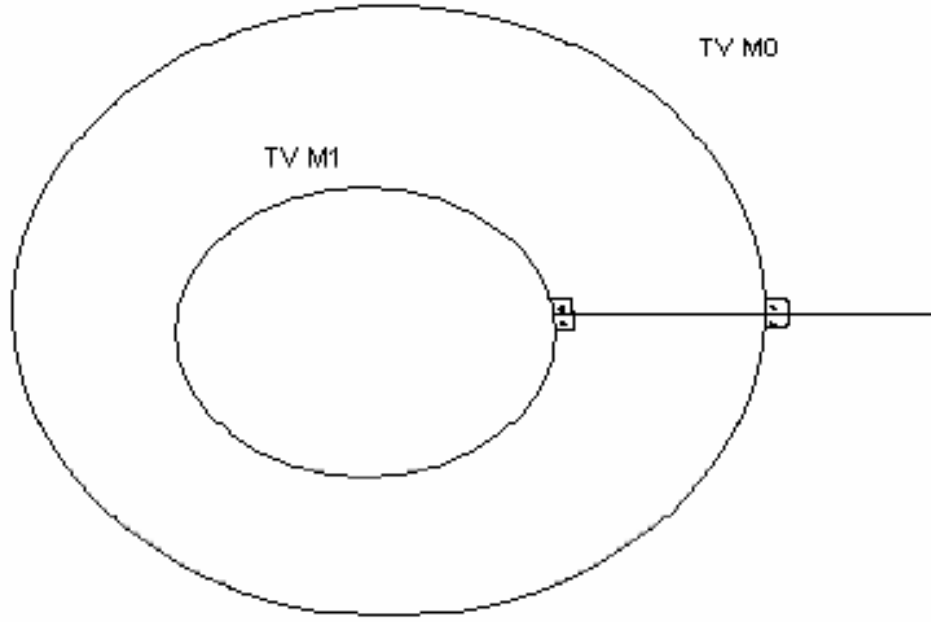


Figure 2.18: The same resulting vector for two steps.

here, N is the number of signal samples. We use this modified version in our algorithm, because we can estimate the noise standard deviation of the noise as proposed previously in above lines .

The total variation minimization algorithm in [4] only keeps tracks of region averages, that are updated efficiently when two regions are merged. When its total variation decreases, the number of regions in the image decreases and the regions enlarge. When a region has equal intensity with an adjacent one, that time is called hitting time. The parameters are evaluated only in the hitting times. The parameters that are kept in memory for each region are called segmentation parameters. The definitions of them are as follows. The following segmentation parameters are recomputed only at the hitting times:

Number of regions: $p(\mathbf{u})$

Index of left endpoint of region i: $n_i(\mathbf{u})$

Size of region i: $m_i(\mathbf{u})$

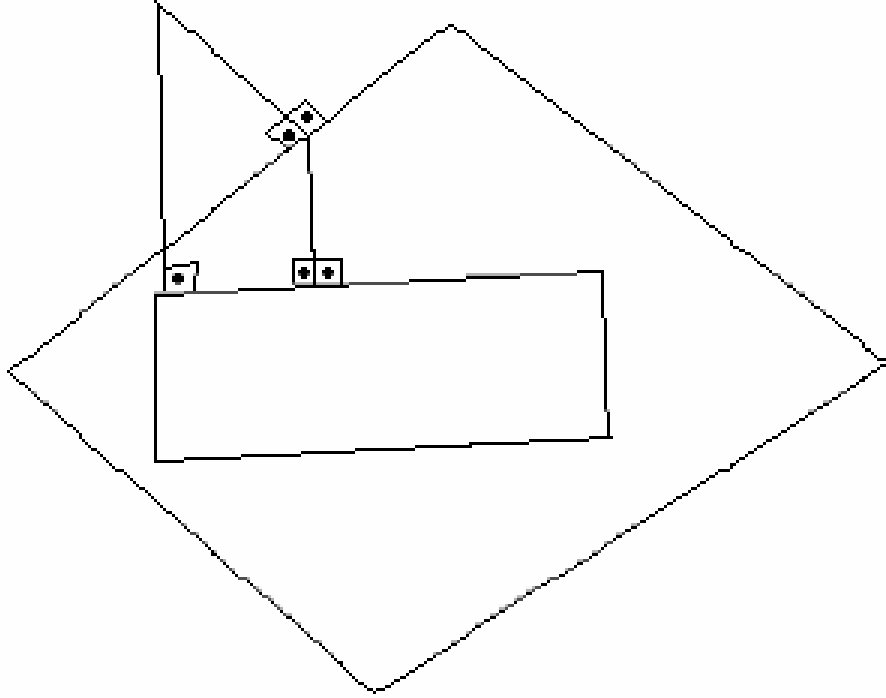


Figure 2.19: Different resulting vectors.

Number of neighbors of region i : $g_i(\mathbf{u})$

Is region i a local max, a local min, or neither? $B_i(\mathbf{u})$

The steps of the algorithm [4] is summarized as below:

1. Assign $l = 0$; $t_l = 0$; $\mathbf{u}^*(t_l) = \mathbf{u}_0$; initialize the segmentation parameters.
2. For each $i = 1, \dots, p - 1$ find $v_i(t_l) = |M_{i+1}(\mathbf{u}(t_l)) - M_i(\mathbf{u}(t_l))|$ from $\mathbf{u}^*(t_l)$ using the (2.20).

$$M_i(\mathbf{u}(t)) = M_i(\mathbf{u}^*(t)) - B(i)g(i)t/m(i). \quad (2.20)$$

Here, M_i is the intensity of region i . Calculate $r_i(t_l) = \dot{v}_i(t_l)$ from the segmentation parameters of $\mathbf{u}^*(t_l)$. Here r_i is the rate of change of an

intensity and it is equal to $-B(i)g(i)/m(i)$. Use $E_i(T) = -v_i(T)/r_i(T)$ to find all the candidates for the next hitting time: $t_l + E_l(t_l) \dots t_l + E_{p-1}(t_l)$

3. Find

$$j = \underset{i}{\operatorname{argmin}}[t_l + E_i(t_l)]$$

find the next hitting time $t_{l+1} = t_l + E_j(t_l)$

4. Find the difference norm to check if the target variance is reached.

$$a(t) = a(t_l) + (t^2 - t_l^2) \sum_{i=1}^p B(i)^2 g(i)^2 / m(i);$$

Here $a(t) = \|u(t) - u(0)\|^2$;

if this is greater than $N.\sigma^2$, stop. If it is smaller,

i)update $\mathbf{u}^*$;

$$M_j(\mathbf{u}^*(t_{l+1})) = \frac{m_j(\mathbf{u}^*(t_l)).M_j(\mathbf{u}^*(t_l)) + m_{j+1}(\mathbf{u}^*(t_l)).M_{j+1}(\mathbf{u}^*(t_l))}{m_j(\mathbf{u}^*(t_l)) + m_{j+1}(\mathbf{u}^*(t_l))};$$

ii)update the segmentation parameters.

iii)increment l by 1.

iv)go to step four.

All these are achieved by a time complexity of $O(N \log N)$ and space complexity of $O(N)$ where N is the number of data samples by the algorithm above in [4].

2.5 Piecewise Smooth Signal Denoising via Principal Curve Projections

Curvelet decomposition and the total variation are two widely used techniques in signal denoising. Curvelets which is a wavelet domain technique suffer from Gibbs phenomenon near the edges. Furthermore, its output has speckle noise characteristics. On the other hand, total variation based techniques suffer from

stair-case effect. Principal Curve Projection method is a recent technique which does not yield those artifacts of curvelets and total variation [5]. It is developed for the signals which are smooth between the edges. The edges are not disturbed through this technique.

The Principal Curve Projection technique first transforms a $1 - D$ noisy signal to $2 - D$:

$$\mathbf{u}_n = [u[n] \quad n]^T, \quad (2.21)$$

where $u[n]$ is the $1 - D$ discrete signal. To generate a $2 - D$ signal from these samples, interpolation with $2 - D$ Gaussian kernels is used:

$$F(\mathbf{u}) = \sum_{n=1}^N G_{\Sigma_n}(\mathbf{u} - \mathbf{u}_n), \quad (2.22)$$

where $G_{\Sigma_n}(\mathbf{u})$ is a $2 - D$ Gaussian curve with a mean of $\mu = 0$ and a covariance of $\Sigma = \Sigma_n$

The algorithm of principal curve projections relies on the fact that the principal curves are approximately equal to the peak of the pdf. Therefore, at each iteration, $\mathbf{u}$ that maximizes the pdf is found. That is,

$$\mathbf{u} := \left(\sum_{n=1}^N \Sigma_n^{-1} G_{\Sigma_n}(\mathbf{u} - \mathbf{u}_n) \right)^{-1} \sum_{n=1}^N \Sigma_n^{-1} G_{\Sigma_n}(\mathbf{u} - \mathbf{u}_n) \mathbf{u}_n. \quad (2.23)$$

This is called mean shift update in the literature [53].

The denoising algorithm can be summarized as follows.

1. Construct the $2 - D$ pdf by selecting the Gaussian covariance. The most reliable method for this selection is the one that is based on maximum likelihood estimation adaptive to the noise [25].
2. Calculate mean shift update in the (2.23).
3. Check the first entry of the output of mean shift update to see whether it is within correct boundaries.
4. If the result converged to a value, go to next step, otherwise go to Step (2).
5. Calculate the mean shift update in the (2.23) for each time sample n .
6. $\tilde{\mathbf{u}} = \mathbf{v}\mathbf{v}^T\mathbf{m}(\mathbf{u}(n))$ where $\mathbf{v}$ is the eigenvector of Σ^{-1} corresponding to the eigenvalue with the greatest magnitude and $\mathbf{m}$ is the mean shift update.
7. If $\frac{|\nabla^T\mathbf{H}\nabla|}{\|\nabla\|\|\mathbf{H}\nabla\|}$ is greater than a threshold, stop the algorithm, otherwise assign $\mathbf{u}(n+1)$ as $\tilde{\mathbf{u}}$. Increase n and go to Step (5). In this inequality, ∇ is the gradient and $\mathbf{H}$ is the Hessian of the $2 - D$ pdf.

The second part of the algorithm, starting from Step (5) and ending at Step (7) is called SCMS algorithm in the literature [54]. It is used in order to shift the maximum of the pdf to the principal curve.

2.6 Piecewise Linear Denoising by Total Variation

The superior performance of the total variation in denoising of piecewise constant signals is investigated in detail in [4] and [35]. Here, we propose a generalization of the original total variation technique for denoising of signals with piecewise linear sections. For this purpose, the total variation denoising is applied to the

first difference of the signal to be denoised to obtain the hit points. Based on the obtained hit points, the original signal is partitioned into segments and in each segment the signal is approximated with a line segment, when two regions are merged by the total variation iterations on the first difference signal, the respective line segments are merged with an efficient technique. In this technique, the parameters of the adjacent lines are used. The flow diagram which illustrates this proposed technique is shown in Figure 2.20. Based on extensive set of comparisons, it is demonstrated that the proposed technique is a very effective denoising technique for piecewise linear signals.

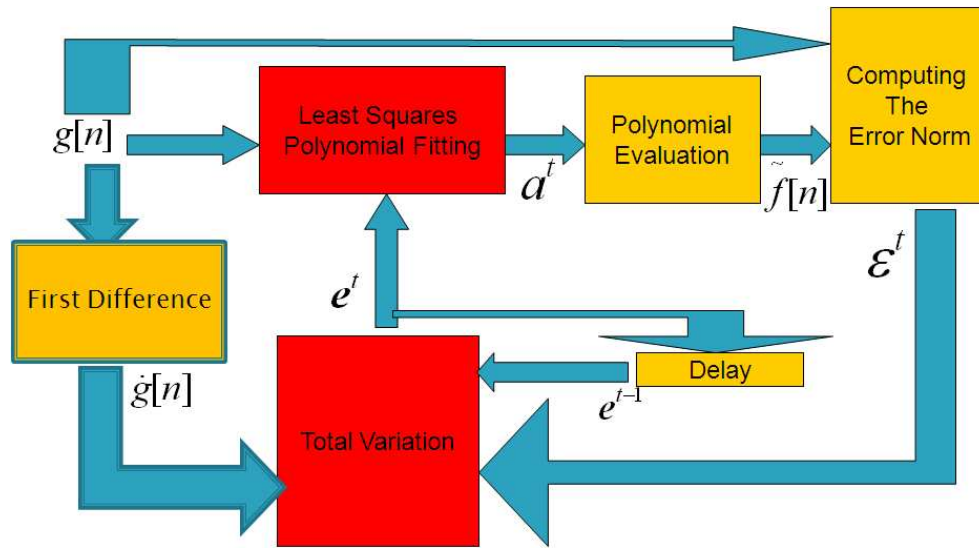


Figure 2.20: Flow diagram of the proposed Piecewise Linear Denoising Technique.

2.6.1 Formulation and Algorithm

Let us find the best line fitted to a signal. Let that signal be defined in the interval $\mathbf{x} = (1 \dots N)^T$. Let us denote a line $\mathbf{f}_{\alpha, \beta}$.

$$\mathbf{f}_{\alpha, \beta} = \beta + \alpha \mathbf{x}. \quad (2.24)$$

Here, the issue is to find the least squares estimate of $\mathbf{y}$ which is the observed noisy signal. That is,

$$\begin{aligned}
\mathbf{w}_{LS} &= \underset{[\alpha, \beta]}{\operatorname{argmin}} \{ \|\mathbf{y} - \mathbf{f}_{\alpha, \beta}\|_2 \} \\
&= \underset{[\alpha, \beta]}{\operatorname{argmin}} \{ \|\mathbf{y} - (\alpha \mathbf{x} + \beta \mathbf{1})\|_2 \} \\
&= \underset{[\alpha, \beta]}{\operatorname{argmin}} \left\{ \left\| \mathbf{y} - \begin{pmatrix} \mathbf{x} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\|_2 \right\}.
\end{aligned}$$

This is in the format of;

$$\mathbf{w}_{LS} = \underset{\mathbf{w}}{\operatorname{argmin}} \{ \|\mathbf{y} - \mathbf{A} \cdot \mathbf{w}\|_2 \}, \quad (2.25)$$

and the solution is;

$$\mathbf{w}_{LS} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (2.26)$$

Therefore,

$$\begin{pmatrix} \alpha_{LS} \\ \beta_{LS} \end{pmatrix} = \left(\begin{pmatrix} \mathbf{x}^T \\ \mathbf{1}^T \end{pmatrix} \begin{pmatrix} \mathbf{x} & \mathbf{1} \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathbf{x}^T \\ \mathbf{1}^T \end{pmatrix} \mathbf{y}, \quad (2.27)$$

$$\begin{pmatrix} \alpha_{LS} \\ \beta_{LS} \end{pmatrix} = \begin{pmatrix} \mathbf{x}^T \mathbf{x} & \mathbf{x}^T \mathbf{1} \\ \mathbf{1}^T \mathbf{x} & \mathbf{1}^T \mathbf{1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x}^T \mathbf{y} \\ \mathbf{1}^T \mathbf{y} \end{pmatrix}. \quad (2.28)$$

In the (2.28);

$$\mathbf{x}_1^T \mathbf{x}_1 = \frac{N_1(2N_1 + 1)(N_1 + 1)}{6}, \mathbf{x}_1^T \mathbf{1}_1 = \frac{N_1(N_1 + 1)}{2}, \mathbf{1}_1^T \mathbf{1}_1 = N_1, \quad (2.29)$$

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \frac{\begin{pmatrix} N_1 & \frac{-N_1(N_1+1)}{2} \\ \frac{-N_1(N_1+1)}{2} & \frac{-N_1(N_1+1)(2N_1+1)}{6} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{N_1} i y_1(i) \\ \sum_{i=1}^{N_1} y_1(i) \end{pmatrix}}{\frac{N_1^2(N_1+1)(N_1-1)}{12}}, \quad (2.30)$$

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \frac{12}{N_1(N_1-1)(N_1+1)} & \frac{-6}{N_1(N_1-1)} \\ \frac{-6}{N_1(N_1-1)} & \frac{2 \cdot (2N_1+1)}{N_1(N_1-1)} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{N_1} i y_1(i) \\ \sum_{i=1}^{N_1} y_1(i) \end{pmatrix} = \mathbf{A}_1^{-1} \cdot \begin{pmatrix} s_1 \\ r_1 \end{pmatrix}. \quad (2.31)$$

Here, counting the multiplications is an important issue. Computation of $\frac{2}{N_1(N_1-1)}$ requires two multiplications and it is computed once. There are four more multiplications in that matrix. Computation of the first term in the second array requires N_1 multiplications. Computation of the multiplication of the vector and the matrix requires four more multiplications. In total, the number of multiplications required is N_1+10 .

Let there be two lines to be merged one of which is $y_1[n]$ and the other is $y_2[n]$. The first is defined in the interval from 1 to N_1 and the other is defined from $N_1 + 1$ to $N_2 + N_1$. That is;

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ N_1 \end{pmatrix} \quad \mathbf{x}_2 = \begin{pmatrix} N_1 + 1 \\ \cdot \\ \cdot \\ N_1 + N_2 \end{pmatrix} \quad \mathbf{y}_1 = \begin{pmatrix} y(1) \\ \cdot \\ \cdot \\ y(N_1) \end{pmatrix} \quad \mathbf{y}_2 = \begin{pmatrix} y(N_1 + 1) \\ \cdot \\ \cdot \\ y(N_1 + N_2) \end{pmatrix}. \quad (2.32)$$

Having found how to efficiently calculate the line parameters, now the issue is to merge the lines. The optimal best line defined between 1 and $N_1 + N_2$ might be found directly without merging the lines. However, if the parameters of the two best-fit lines defined in the interval are used, the best fitting line to combine the interval can be found with less computation. If the overall best line is computed directly $N_1 + N_2+10$ multiplications are performed. On the other hand, as it is shown below, if the parameters of the adjacent lines are used, we only 11 multiplications are computed.

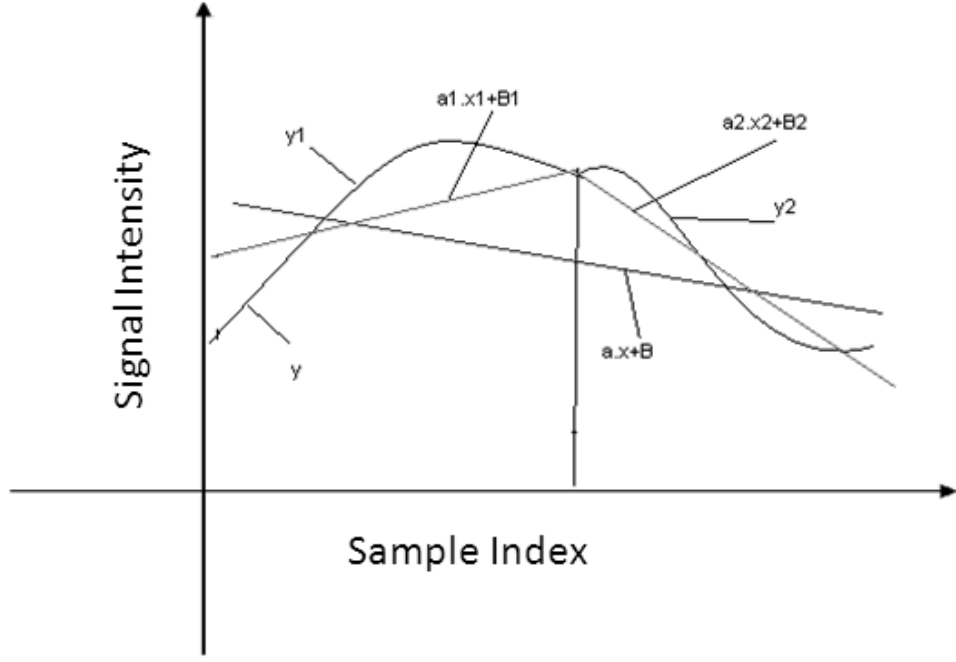


Figure 2.21: Two lines to be merged.

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \mathbf{A}_1^{-1} \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{1}_1^T \end{pmatrix} \mathbf{y}_1 = \mathbf{A}_1^{-1} \begin{pmatrix} s_1 \\ r_1 \end{pmatrix}. \quad (2.33)$$

$$\begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix} = \mathbf{A}_2^{-1} \begin{pmatrix} \mathbf{x}_2^T \\ \mathbf{1}_2^T \end{pmatrix} \mathbf{y}_2 = \mathbf{A}_2^{-1} \begin{pmatrix} s_2 \\ r_2 \end{pmatrix}, \quad (2.34)$$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} \mathbf{x}_1^T & \mathbf{x}_2^T + N_1 \mathbf{1}_2^T \\ \mathbf{1}_1^T & \mathbf{1}_2^T \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} s \\ r \end{pmatrix}, \quad (2.35)$$

$$s = s_1 + s_2 + N_1 \cdot r_2, \quad (2.36)$$

$$r = r_1 + r_2. \quad (2.37)$$

By using the updates of s and r parameters given in (2.36), there is no need to compute the s for the entire interval which requires $N_1 + N_2$ multiplications,

but the already computed s_1 , s_2 and r can be used by performing just a single multiplication.

2.6.2 Comparison of the Techniques

The simulations of the techniques of Rudin, Willsky and our technique were compared.

In Figure 2.22, we illustrate a denoising result that can be obtained by using the technique by Rudin, et. al. In this figure, the black curve is the original signal; the blue curve is the noisy signal and the red one is the denoised signal. As seen from the obtained result, the signal could not be denoised at an acceptable level. Staircase effect is visible in the reconstruction.

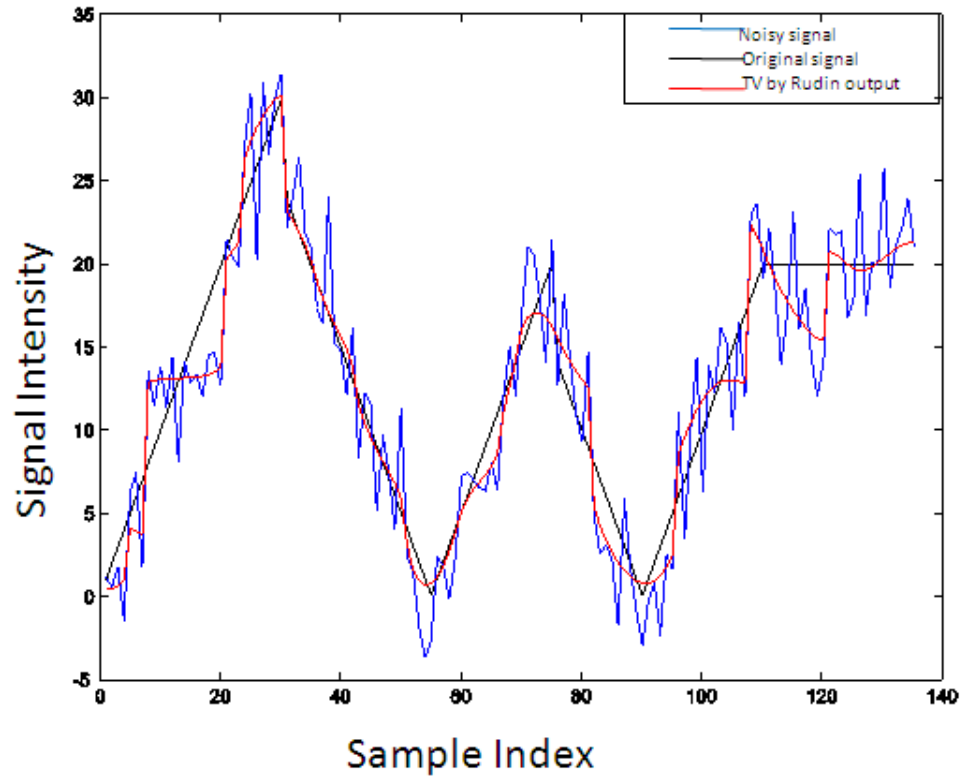


Figure 2.22: Denoising multi-line signal with no jump-discontinuities by Total Variation by Rudin *et al.*

In Figure 2.23, we illustrate the performance of the technique by Willsky, et. al. In this figure, the black curve is the original signal; the blue curve is the noisy signal and the red one is the denoised signal. As seen from this result, the signal could not be denoised because of significant staircase effect.

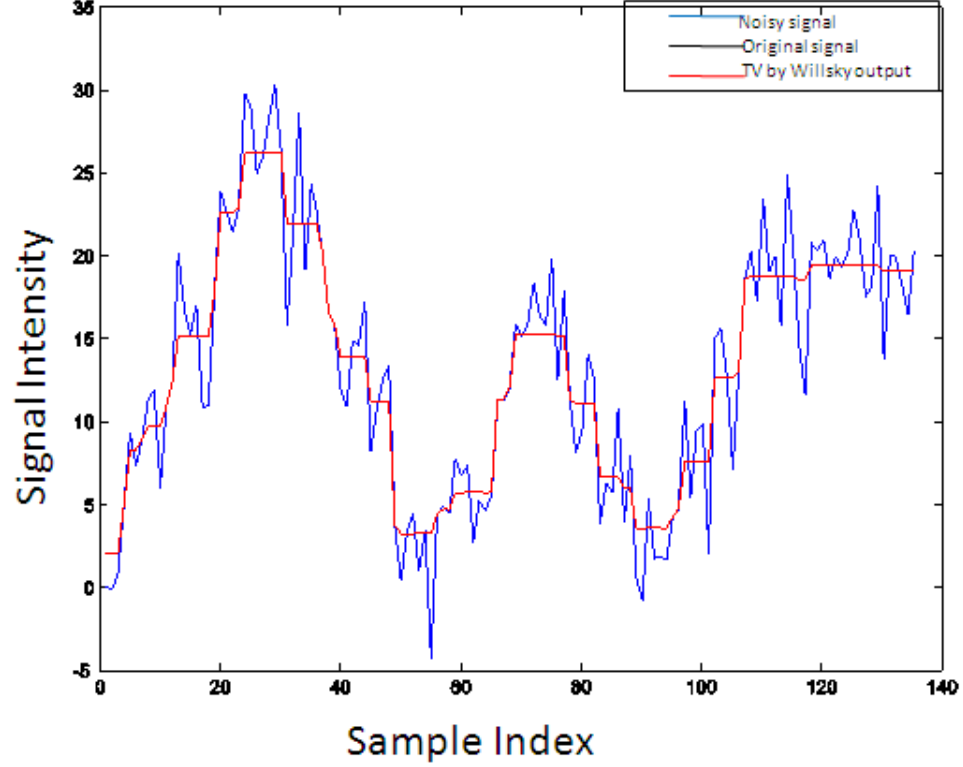


Figure 2.23: Denoising multi-line signal with no jump-discontinuities by Total Variation by Willsky *et al.*

Finally, in Figure 2.24, we illustrate the performance of the proposed technique on the same data. In this figure, the black curve is the original signal; the blue curve is the noisy signal and the red one is the denoised signal. As seen from this figure, the proposed technique provides significantly better denoising than the alternatives. Note that the proposed technique has the same computational complexity with the technique by Willsky, et. al.

Having studied the multi-boundary case, here, a challenging two segment case will be discussed. The challenging part of this signal type is its first difference structure shown in Figure 2.25. Except the one-sample-spike in the middle, the

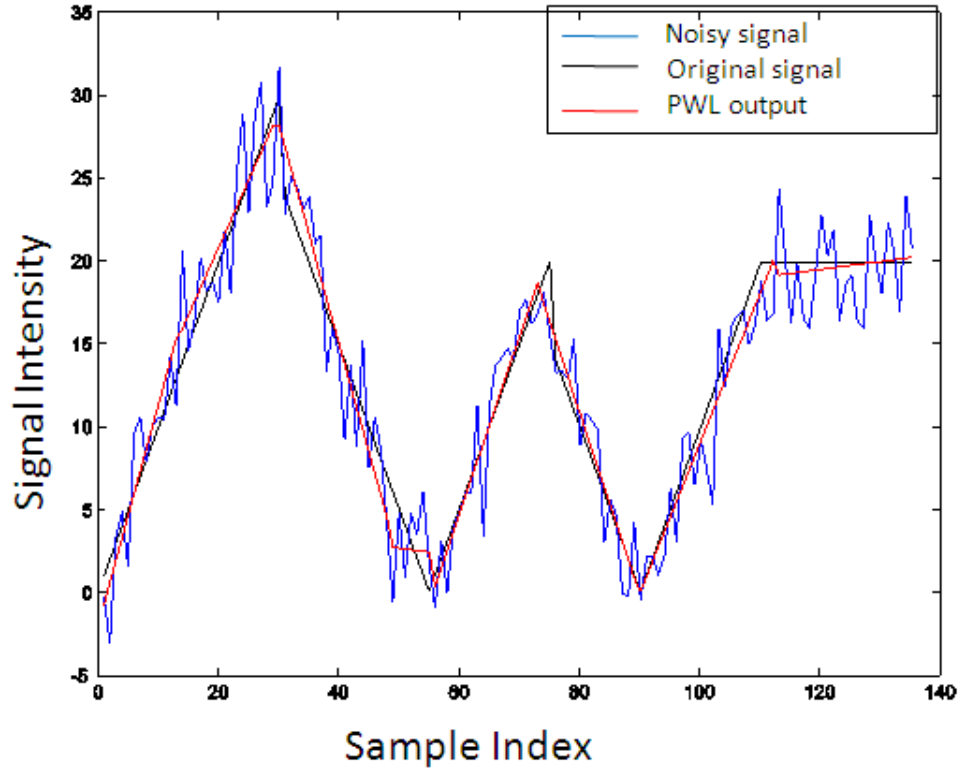


Figure 2.24: Denoising multi-line signal with no jump-discontinuities by the proposed PWL technique.

first difference signal is like a noisy constant signal. Here, it will be observed that the spike between the segments will be sufficient to detect the edge.

Similar conclusions as the first experiment can be drawn out of this new signal type. The most crucial difference about the performance is again about the stair-case effect which disturbs the output of Total Variation based techniques and which is not observed in the denoising results of the proposed technique. Another remarkable point is the collaboration of the only one spike to detect the boundary.

It would be a if the plots of the error norms of the signals versus the stopping parameters are given.

In Figure 2.29, the error norm of Rudin's technique is given. It is seen that the technique is highly robust against deviations of the iterations. However,

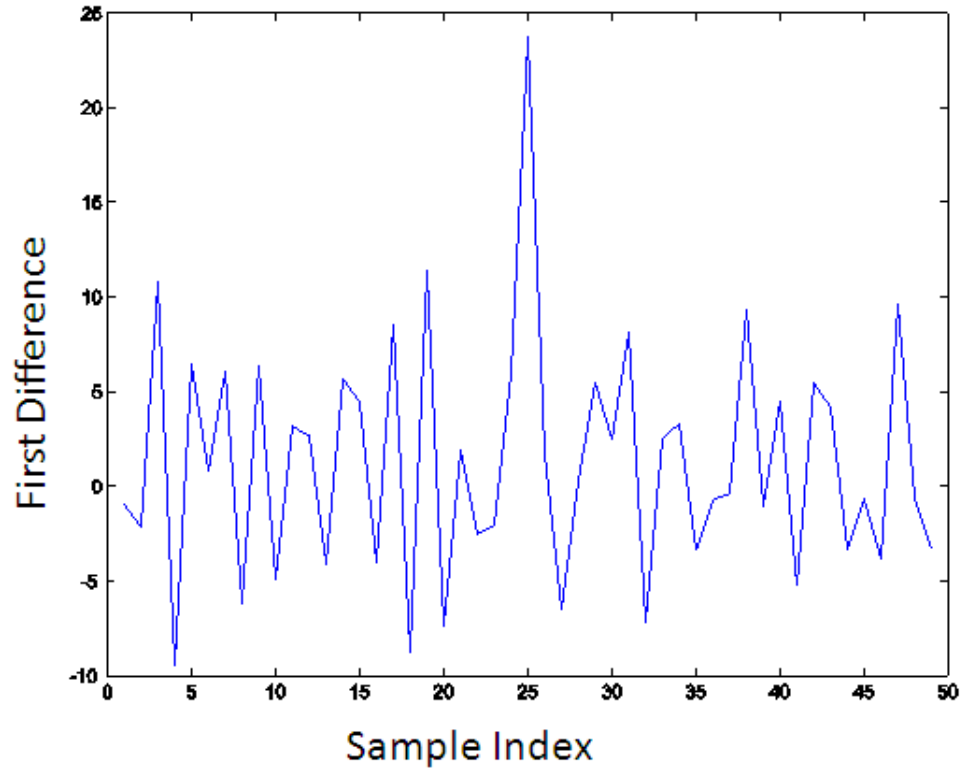


Figure 2.25: First difference of a noisy signal consisting of two parallel lines separated by a jump.

the smallest value of the error norm which is 15.71 is the worse of the three techniques.

In Figure 2.30, the error norm of Willsky's technique is given. It is interesting that the technique is not robust against deviations of the estimated noise standard deviation. But, the smallest value of the error norm which is 14.93 is the moderate one.

In Figure 2.31 the error norm of our technique is given. It is obvious that the technique is moderately robust against deviations of the estimated noise standard deviation. But, the smallest value of the error norm which is 8.63 is the best one. Here, we can draw the conclusion that our technique is the best one for this setting.

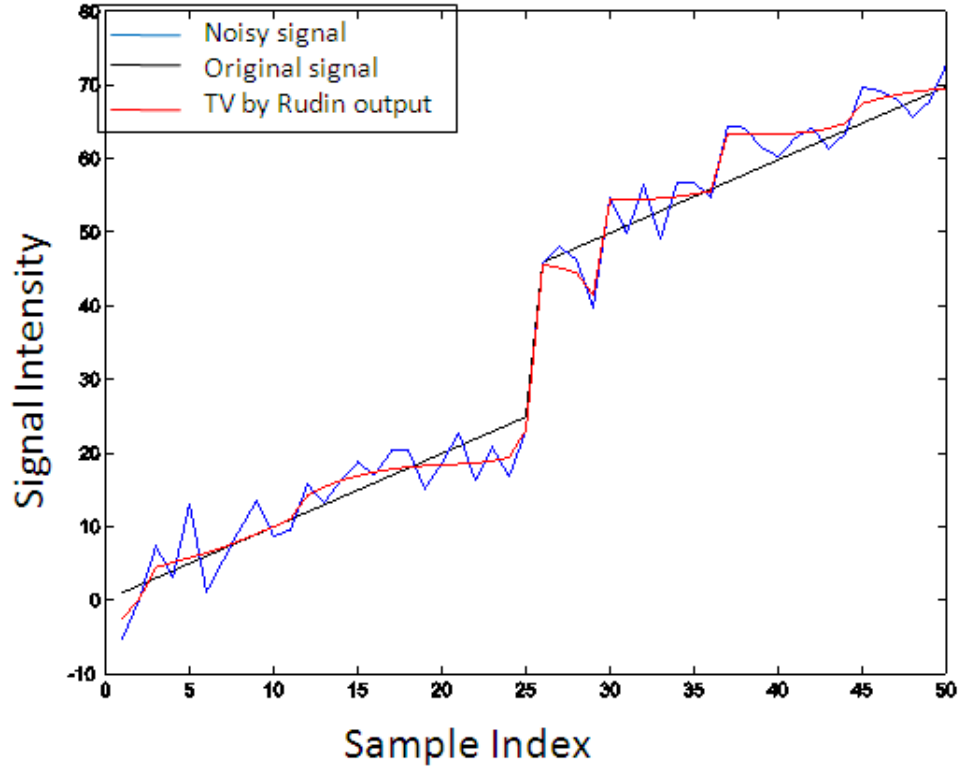


Figure 2.26: Denoising the synthetic signal composed of two lines by the technique of Total Variation by Rudin *et al.*

For the sake of convenience, another crucial metric, the fit error norms are also given in the successive plots. The cross-overs among different SNR values is interesting in these plots.

2.7 Piecewise Parabolic and Cubic Denoising by Total Variation

Signal denoising is an important area that is at the intersection of telecommunications and signal processing. Two dimensional signal processing is used in image processing and one dimensional signal processing is used in telecommunications. In this thesis, we are going to focus on one dimensional signal denoising. In this area, there are two well known approaches proposed in [4] and [35]. In [4], the total variation of the image is minimized by using a set of parameters. In this

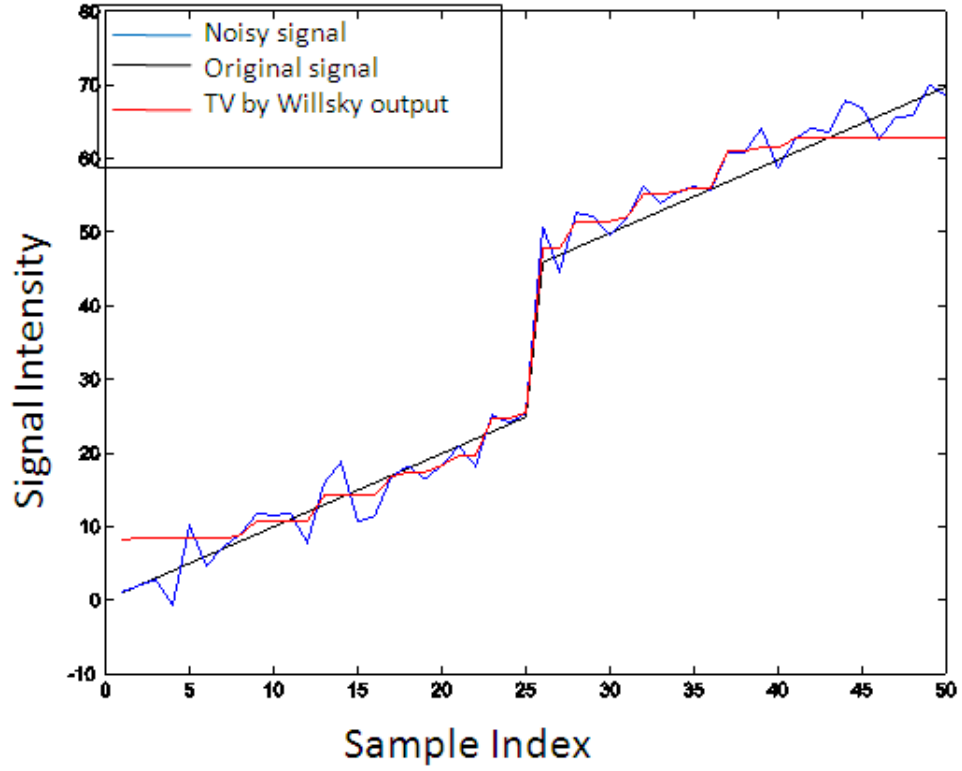


Figure 2.27: Denoising the synthetic signal composed of two lines by the technique of Total Variation by Willsky *et al.*

process, some regions at which signal takes a constant value are formed. These regions are merged to form larger regions. This forms some undesired flat steps as a result of so-called staircase effect. In [35], total variation integral equation is solved by using Euler-Lagrange equations. Noise standard deviation is used as an L2 norm constraint. Staircase effect is again an undesired result for initial signals that contain arbitrary curvature. The flat steps could only be desired for piecewise constant signals. Our technique at first finds the second or third difference of the original signal. Total variation of this new signal is reduced and the region boundaries are obtained. In each region, curve fitting is applied with least squares minimization. A polynomial of second or third order is fit to the noisy part of the signal. When two regions merge in the total variation minimization process, two adjacent corresponding segments are merged by using the parameters of those two segments. When the noise standard deviation is

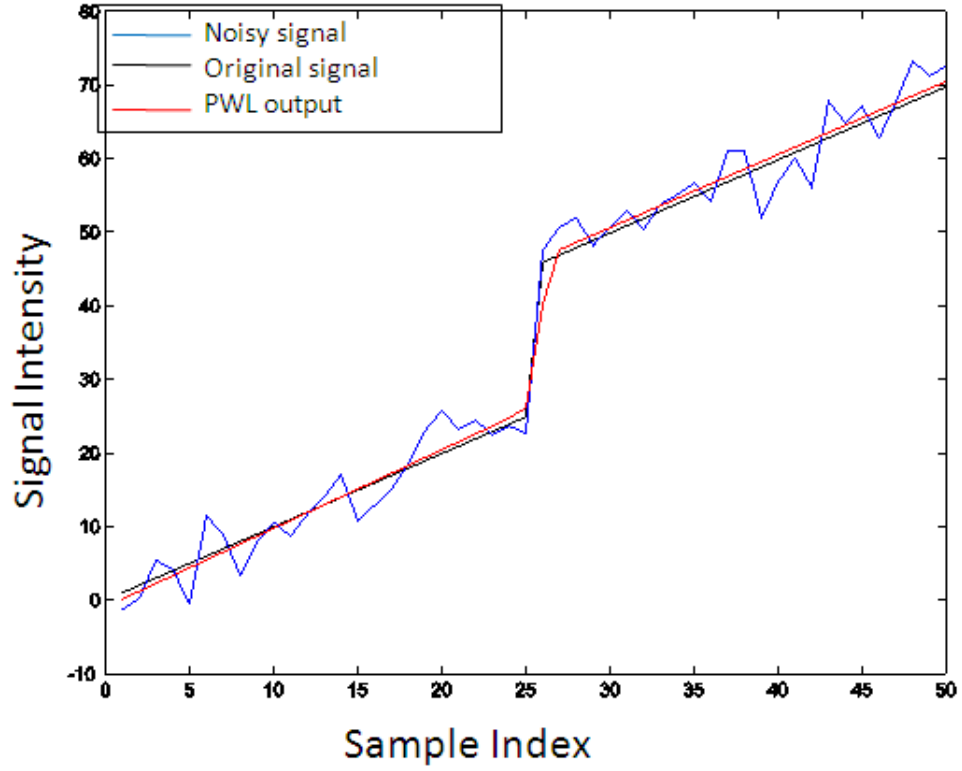


Figure 2.28: Denoising the synthetic signal composed of two lines by the proposed PWL technique.

reached, the process is terminated. Hence, it is required to know or obtain an estimate to the noise standard deviation.

2.7.1 Formulation and Algorithm

Let us find the parabola fitting for a signal. Let that signal be defined in the interval $\mathbf{x} = (1 \dots N)^T$. Let us denote the best curve by $\mathbf{f}$.

$$\mathbf{f}_{a,b,c} = a\mathbf{x}^2 + b\mathbf{x} + c. \quad (2.38)$$

Here, the issue is to find the least squares estimate of $\mathbf{y}$ which is the observed noisy signal. That is,

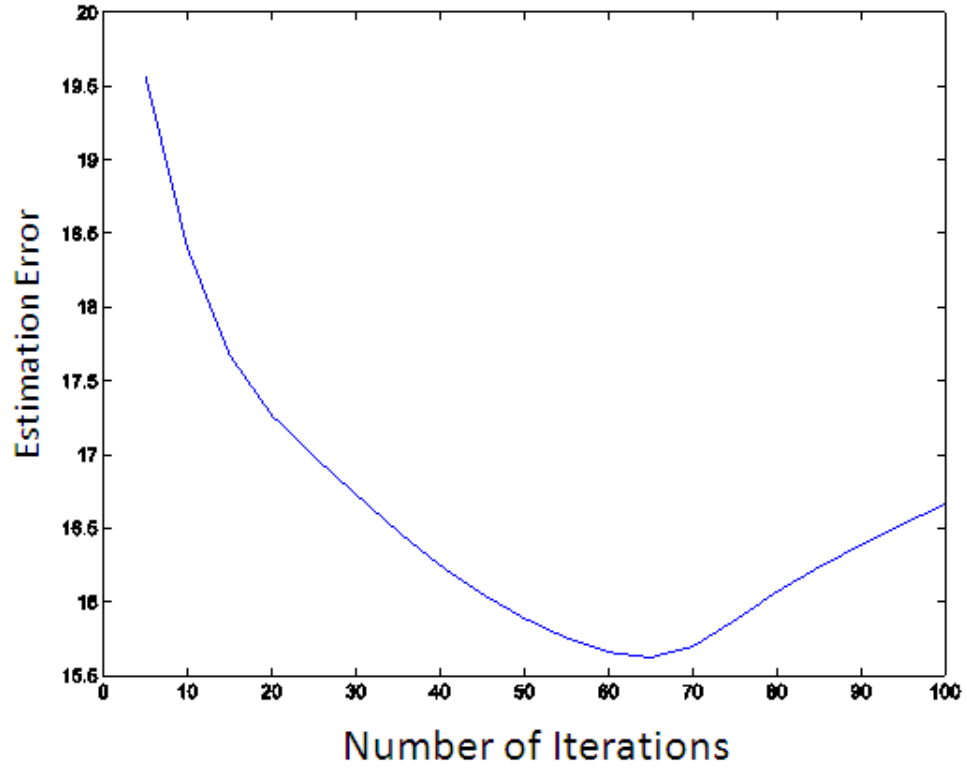


Figure 2.29: Estimation error of Total Variation by Rudin *et al.* for the signal composed of two lines.

$$\begin{aligned}
\mathbf{w}_{LS} &= \underset{[a,b,c]}{\operatorname{argmin}} \{ \|\mathbf{y} - \mathbf{f}_{a,b,c}\|_2 \} \\
&= \underset{[a,b,c]}{\operatorname{argmin}} \{ \|\mathbf{y} - (a\mathbf{x}^2 + b\mathbf{x} + c\mathbf{1})\|_2 \} \\
&= \underset{[a,b,c]}{\operatorname{argmin}} \left\{ \left\| \mathbf{y} - \begin{pmatrix} \mathbf{x}^2 & \mathbf{x} & \mathbf{1} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right\|_2 \right\}.
\end{aligned}$$

This is in the format of;

$$\mathbf{w}_{LS} = \underset{\mathbf{w}}{\operatorname{argmin}} \{ \|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2 \}, \quad (2.39)$$

and the solution is;

$$\mathbf{w}_{LS} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (2.40)$$

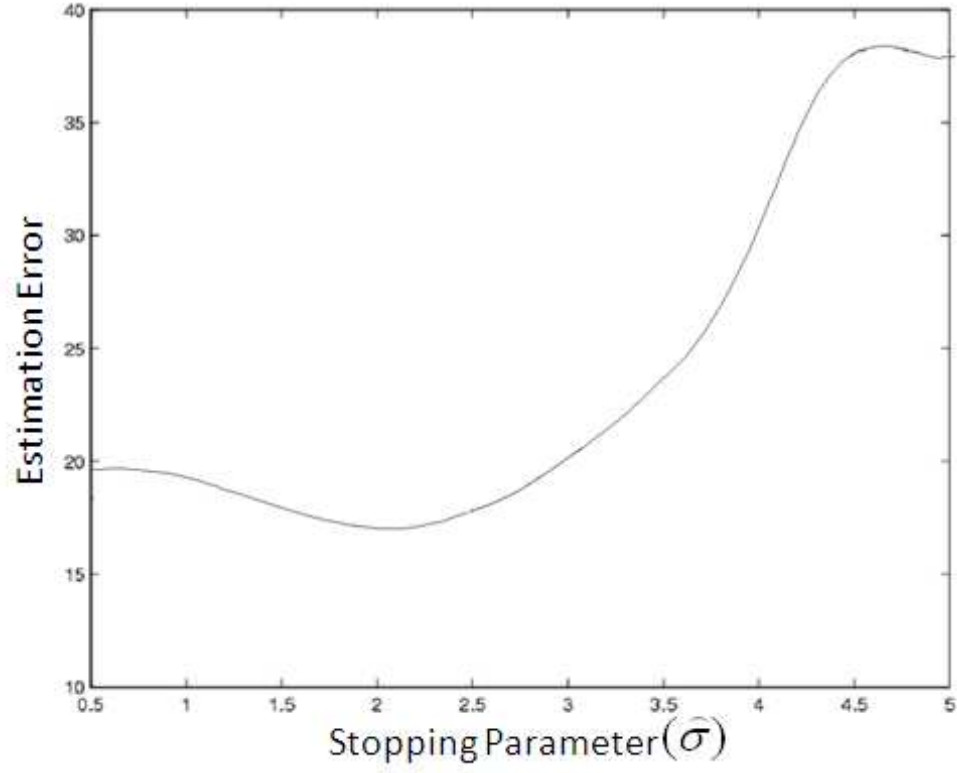


Figure 2.30: Estimation error of Total Variation by Willsky *et al.* for the signal composed of two lines.

Therefore,

$$\begin{pmatrix} a_{LS} \\ b_{LS} \\ c_{LS} \end{pmatrix} = \left(\begin{pmatrix} \mathbf{x}^{.2T} \\ \mathbf{x}^T \\ \mathbf{1}^T \end{pmatrix} \begin{pmatrix} \mathbf{x}^{.2} & \mathbf{x} & \mathbf{1} \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathbf{x}^{.2T} \\ \mathbf{x}^T \\ \mathbf{1}^T \end{pmatrix} \mathbf{y}. \quad (2.41)$$

Let there be two curves to be merged one of which is $y_1[n]$ and the other is $y_2[n]$. The first is defined in the interval from 1 to N_1 and the other is defined from $N_1 + 1$ to $N_2 + N_1$. That is;

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ N_1 \end{pmatrix} \quad \mathbf{x}_2 = \begin{pmatrix} N_1 + 1 \\ \cdot \\ \cdot \\ N_1 + N_2 \end{pmatrix} \quad \mathbf{y}_1 = \begin{pmatrix} y(1) \\ \cdot \\ \cdot \\ y(N_1) \end{pmatrix} \quad \mathbf{y}_2 = \begin{pmatrix} y(N_1 + 1) \\ \cdot \\ \cdot \\ y(N_1 + N_2) \end{pmatrix}. \quad (2.42)$$

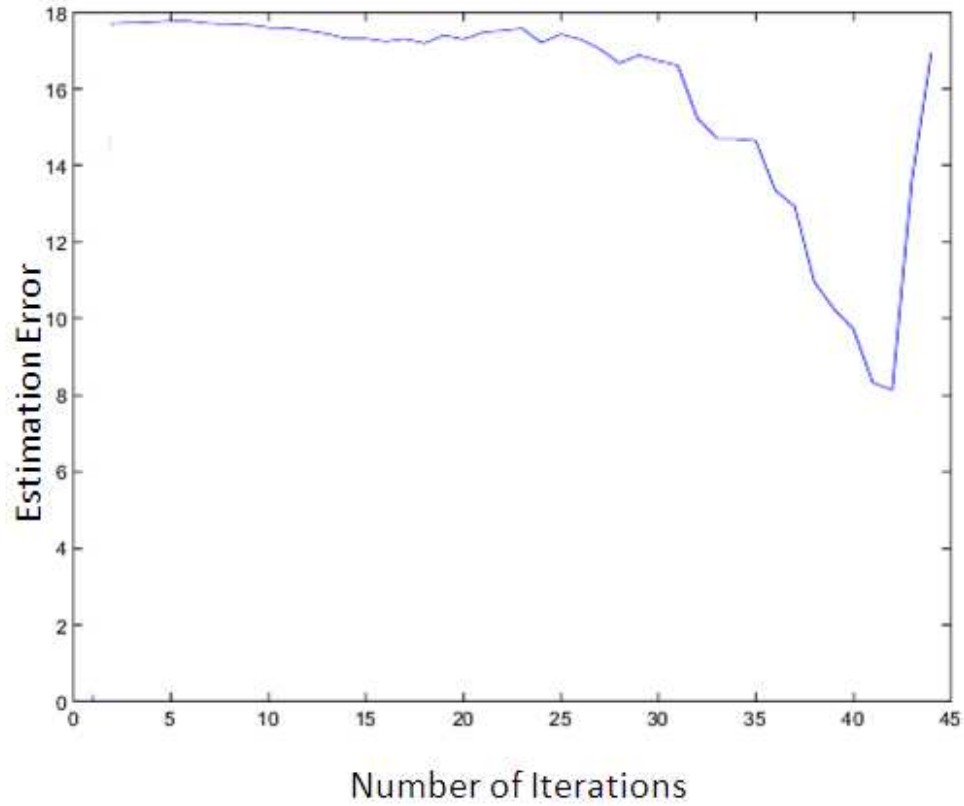


Figure 2.31: Estimation error of the proposed PWL technique for the signal composed of two lines.

Having found how to efficiently calculate the curve parameters, now the issue is to merge the curves. We might find the optimal best line defined between 1 and $N_1 + N_2$ directly without merging the lines. However, if we make use of the parameters of the two best-fit curves defined in the interval, we can obtain the best fitting curve to the combine of interval with less computation. This way we no longer have to compute the s for the entire interval which requires $N_1 + N_2$ multiplications, but can use already computed s_1 and s_2 and r as given in Figure 2.47 by performing just a single multiplication.

$$\begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} = \mathbf{A}_1^{-1} \begin{pmatrix} \mathbf{x}_1^{2T} \\ \mathbf{x}_1^T \\ \mathbf{1}_1^T \end{pmatrix} \mathbf{y}_1 = \mathbf{A}_1^{-1} \begin{pmatrix} t_1 \\ s_1 \\ r_1 \end{pmatrix} \quad (2.43)$$

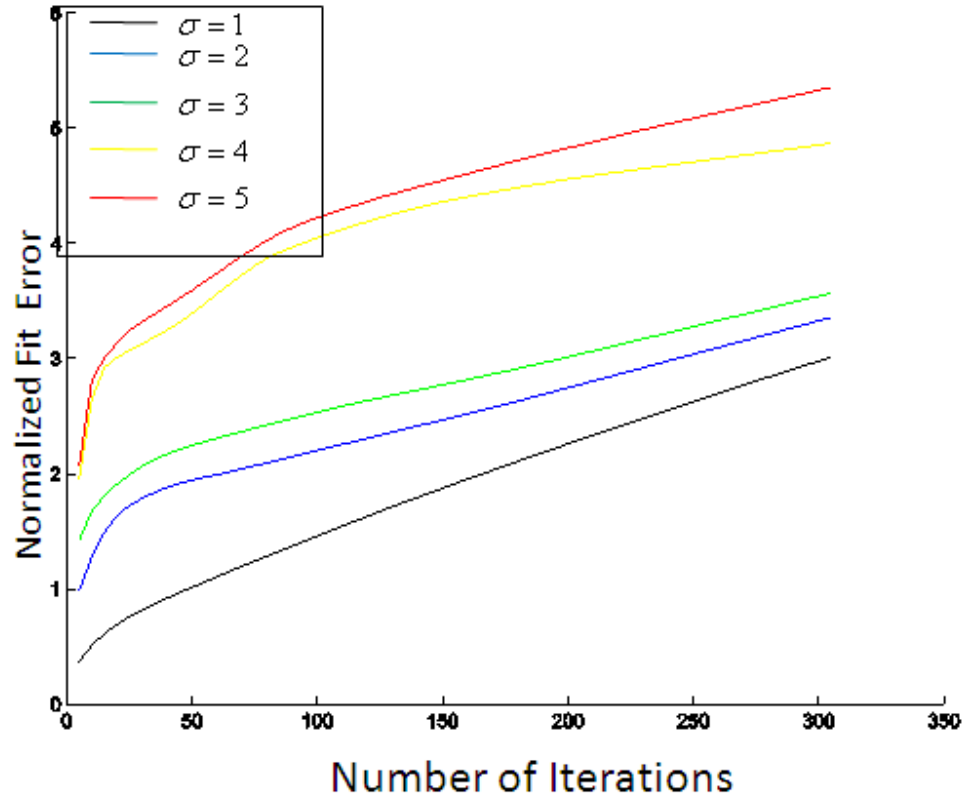


Figure 2.32: Normalized fit error of Total Variation by Rudin *et al.* for the signal composed of two lines at various noise standard deviations.

$$\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = \mathbf{A}_2^{-1} \begin{pmatrix} \mathbf{x}_2^{2T} \\ \mathbf{x}_2^T \\ \mathbf{1}_2^T \end{pmatrix} \mathbf{y}_2 = \mathbf{A}_2^{-1} \begin{pmatrix} t_2 \\ s_2 \\ r_2 \end{pmatrix}, \quad (2.44)$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} \mathbf{x}_1^{2T} & \mathbf{x}_2^{2T} + 2.N_1.\mathbf{x}_2^T + N_1^2.\mathbf{1}_2^T \\ \mathbf{x}_1^T & \mathbf{x}_2^T + N_1.\mathbf{1}_2^T \\ \mathbf{1}_1^T & \mathbf{1}_2^T \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} t \\ s \\ r \end{pmatrix}, \quad (2.45)$$

$$t = t_1 + t_2 + t_2^2 + 2.N_1.s_2 + N_1^2.r_2, \quad (2.46)$$

$$s = s_1 + s_2 + N_1.r_2, \quad (2.47)$$

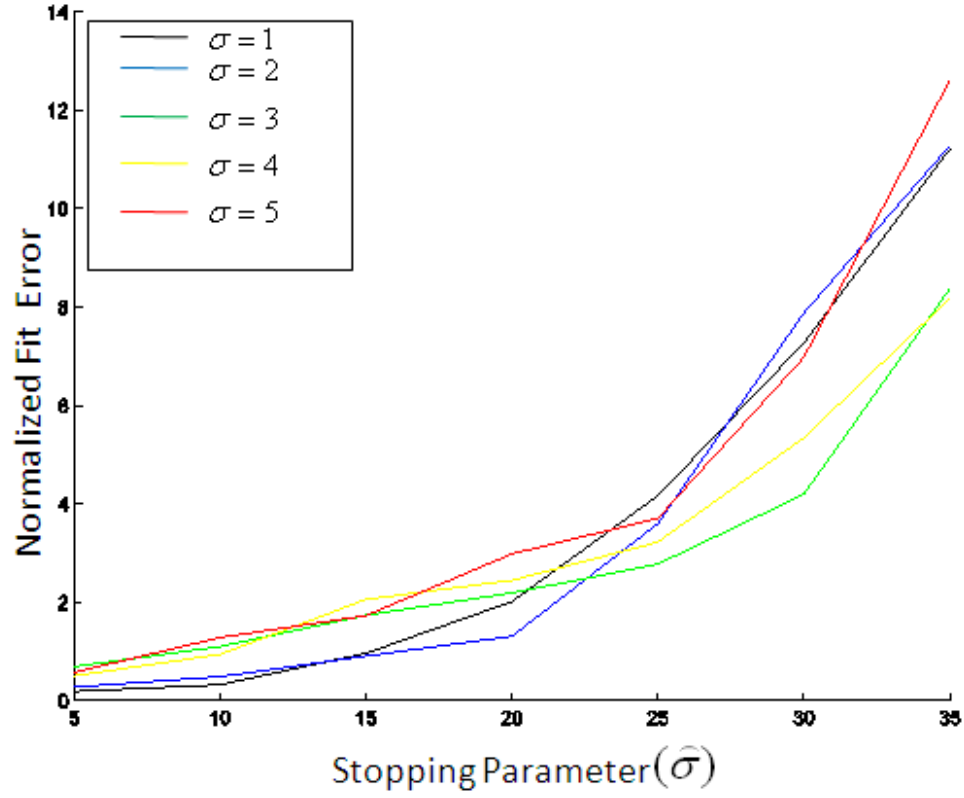


Figure 2.33: Normalized fit error of Total Variation by Willsky *et al.* for the signal composed of two lines at various noise standard deviations.

$$r = r_1 + r_2. \quad (2.48)$$

Let us find the cubic fitting for a signal. Let that signal be defined in the interval $\mathbf{x} = [1 \dots N]^T$. Let us denote the best curve by $\mathbf{f}$.

$$\mathbf{f} = a.\mathbf{x}.^3 + b.\mathbf{x}.^2 + c.\mathbf{x} + d, \quad (2.49)$$

where $\mathbf{x}.^2 = (1 \dots N^2)^T$.

Here, the issue is to find the least squares estimate of $\mathbf{y} - \mathbf{f}$ where $\mathbf{y}$ is the given signal. That is,

$$\|\mathbf{y} - (a.\mathbf{x}.^3 + b.\mathbf{x}.^2 + c.\mathbf{x} + d)\|_{LS}, \quad (2.50)$$

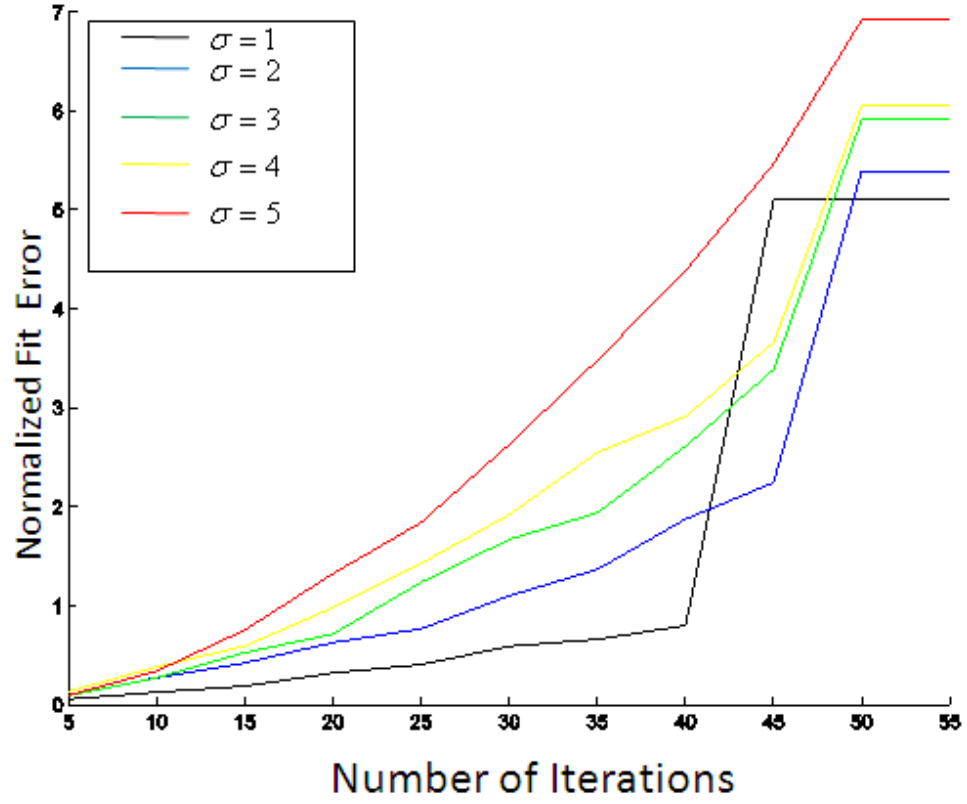


Figure 2.34: Normalized fit error of the proposed PWL technique for the signal composed of two lines at various noise standard deviations.

where $\mathbf{1} = (1 \dots 1)^T$

$$\left\| \mathbf{y} - \begin{pmatrix} \mathbf{x} \cdot^3 & \mathbf{x} \cdot^2 & \mathbf{x} & \mathbf{1} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \right\|_{LS}. \quad (2.51)$$

This is in the format of;

$$\|\mathbf{y} - \mathbf{A} \cdot \mathbf{x}\|_{LS}, \quad (2.52)$$

and the solution is;

$$\mathbf{x}_{LS} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (2.53)$$

Therefore,

$$\begin{pmatrix} a_{LS} \\ b_{LS} \\ c_{LS} \\ d_{LS} \end{pmatrix} = \left(\begin{pmatrix} \mathbf{x}^{.3T} \\ \mathbf{x}^{.2T} \\ \mathbf{x}^T \\ \mathbf{1}^T \end{pmatrix} \begin{pmatrix} \mathbf{x}^{.3} & \mathbf{x}^{.2} & \mathbf{x} & \mathbf{1} \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathbf{x}^{.3T} \mathbf{x}^{.2T} \\ \mathbf{x}^T \\ \mathbf{1}^T \end{pmatrix} \mathbf{y}. \quad (2.54)$$

Let there be two segments to be merged one of which is $y_1[n]$ and the other is $y_2[n]$. The first is defined in the interval from 1 to N_1 and the other is defined from $N_1 + 1$ to $N_2 + N_1$. That is;

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ N_1 \end{pmatrix} \quad \mathbf{x}_2 = \begin{pmatrix} N_1 + 1 \\ \cdot \\ \cdot \\ N_1 + N_2 \end{pmatrix} \quad \mathbf{y}_1 = \begin{pmatrix} y(1) \\ \cdot \\ \cdot \\ y(N_1) \end{pmatrix} \quad \mathbf{y}_2 = \begin{pmatrix} y(N_1 + 1) \\ \cdot \\ \cdot \\ y(N_1 + N_2) \end{pmatrix}. \quad (2.55)$$

Having found how to efficiently calculate the section parameters, now the issue is to merge the corresponding sections. We might find the optimal best section defined between 1 and $N_1 + N_2$ directly without merging the corresponding sections. However, if we make use of the parameters of the two best-fit sections defined in the interval, we can obtain the best fitting curve to the combine of interval with less computation. This way we no longer have to compute the s for the entire interval which requires $N_1 + N_2$ multiplications, but can use already computed s_1 and s_2 and r as given in the (2.60) by performing just a single multiplication.

$$\begin{pmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{pmatrix} = \mathbf{A}_1^{-1} \begin{pmatrix} \mathbf{x}_1^{.3T} \\ \mathbf{x}_1^{.2T} \\ \mathbf{x}_1^T \\ \mathbf{1}_1^T \end{pmatrix} \quad \mathbf{y}_1 = \mathbf{A}_1^{-1} \begin{pmatrix} v_1 \\ t_1 \\ s_1 \\ r_1 \end{pmatrix}, \quad (2.56)$$

$$\begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} = \mathbf{A}_2^{-1} \begin{pmatrix} \mathbf{x}_2^{3^T} \\ \mathbf{x}_2^{2^T} \\ \mathbf{x}_2^T \\ \mathbf{1}_2^T \end{pmatrix} \mathbf{y}_2 = \mathbf{A}_2^{-1} \begin{pmatrix} v_2 t_2 \\ s_2 \\ r_2 \end{pmatrix} \quad (2.57)$$

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} \mathbf{x}_1^{3^T} & \mathbf{x}_2^{3^T} + 3.N_1.\mathbf{x}_2^{2^T} + 3.N_1^2.\mathbf{x}_2^T + N_1^3.\mathbf{1}_2^T \\ \mathbf{x}_1^{2^T} & \mathbf{x}_2^{2^T} + 2.N_1.\mathbf{x}_2^T + N_1^2.\mathbf{1}_2^T \\ \mathbf{x}_1^T & \mathbf{x}_2^T + N_1.\mathbf{1}_2^T \\ \mathbf{1}_1^T & \mathbf{1}_2^T \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$$

$$= \mathbf{A}^{-1} \begin{pmatrix} v \\ t \\ s \\ r \end{pmatrix}$$

$$v = v_1 + v_2 + v_2^3 + 3.N_1.t_2^2 + 3.N_1^2.s_2 + N_1^3.r_2, \quad (2.58)$$

$$t = t_1 + t_2 + t_2^2 + 2.N_1.s_2 + N_1^2.r_2, \quad (2.59)$$

$$s = s_1 + s_2 + N_1.r_2, \quad (2.60)$$

$$r = r_1 + r_2. \quad (2.61)$$

2.7.2 Comparison of the Techniques

In this section, we provide comparison results between the techniques of Rudin, Willsky and the technique proposed in this thesis.

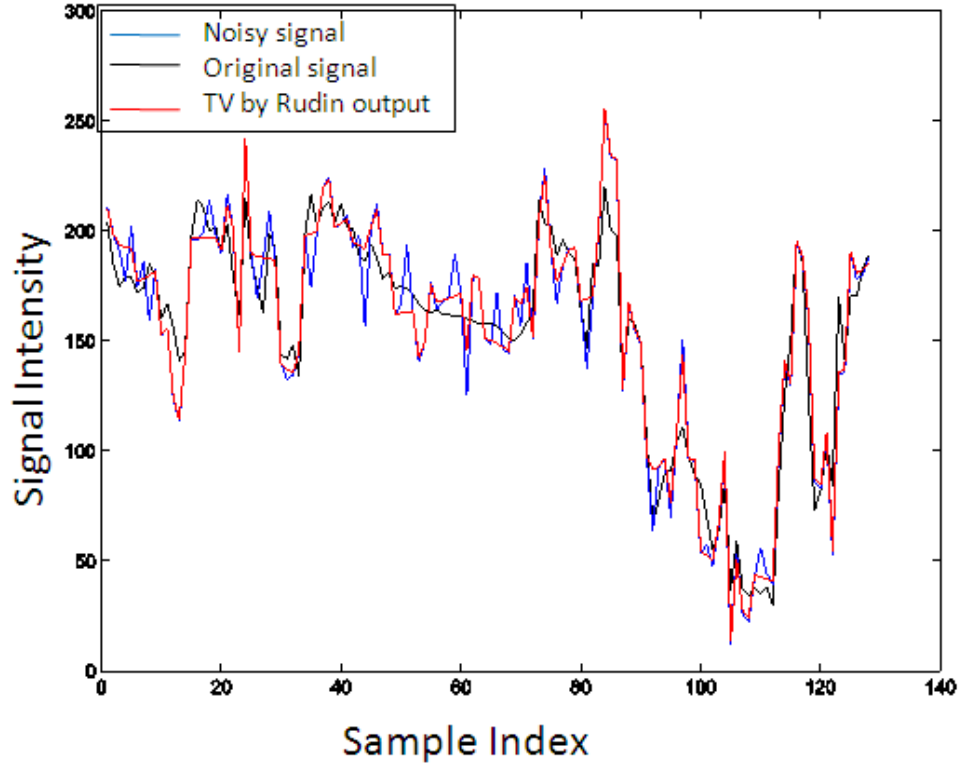


Figure 2.35: Denoising an image slice by Total Variation of Rudin *et al*.

In Figure 2.35, the denoising performance of Total Variation by Rudin *et al* is illustrated. It can be concluded that the signal could not be denoised at an acceptable level, since staircase effect is observed in the output.

In Figure 2.36, the performance of the Total Variation by Willsky *et al*. is depicted. Here, it is seen that the denoising performance is not acceptable because of an excessive staircase effect.

In Figure 2.37, the simulation based performance of the proposed technique is illustrated. Here, it can be concluded that our technique provides significant improvement over the alternatives. Because, the well-known staircase effect problem of total variation is alleviated by the proposed technique.

It will be investigated if the performance comparison will confirm the superiority of the proposed technique against the alternatives. For this purpose, a simple signal which is composed of a parabolic curve will be used.

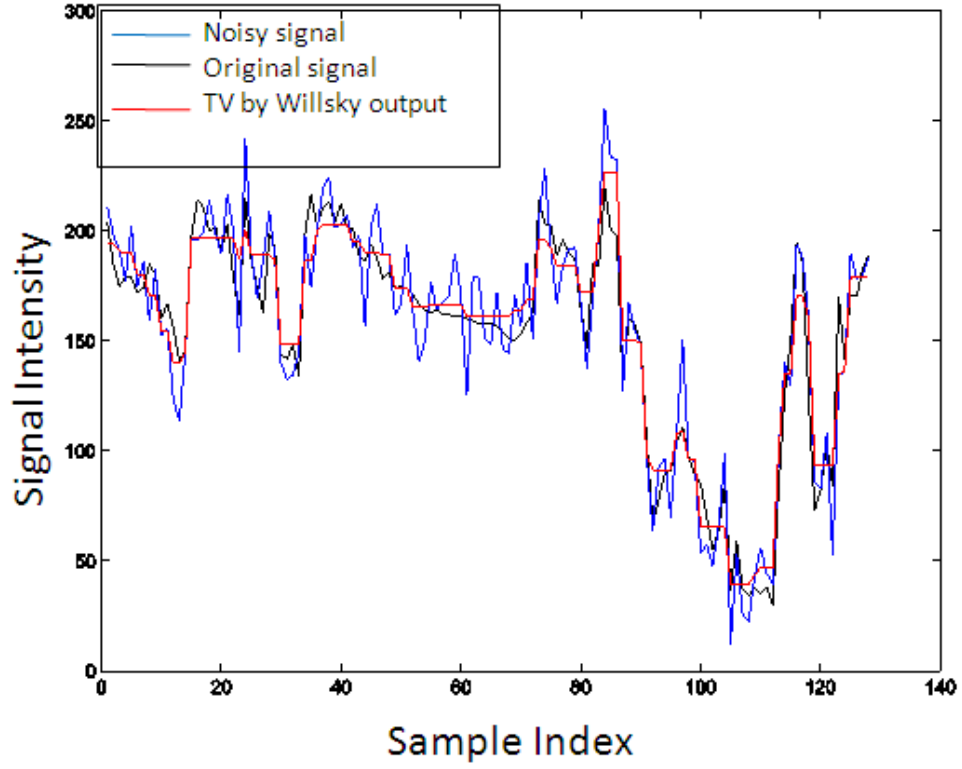


Figure 2.36: Denoising an image slice by Total Variation of Willsky *et al.*

The comparison of these simulation results reveals that our proposed techniques outperform the alternatives in the literature as expected. The stair-case effect prevents to catch parabolic behaviour which was observed for the linear behaviour as well.

In the next plots, the estimation error norms of the techniques versus the stopping parameters will be given.

In Figure 2.43 the error norm of Rudin's technique is given. We see that the technique is highly robust against deviations of the iterations. However, the smallest value of the error norm is moderate of the three techniques.

In Figure 2.44 the error norm of Willsky's technique is given. We see that the technique is not robust against deviations of the estimated noise standard deviation. But, the smallest value of the error norm which is the worst of all.

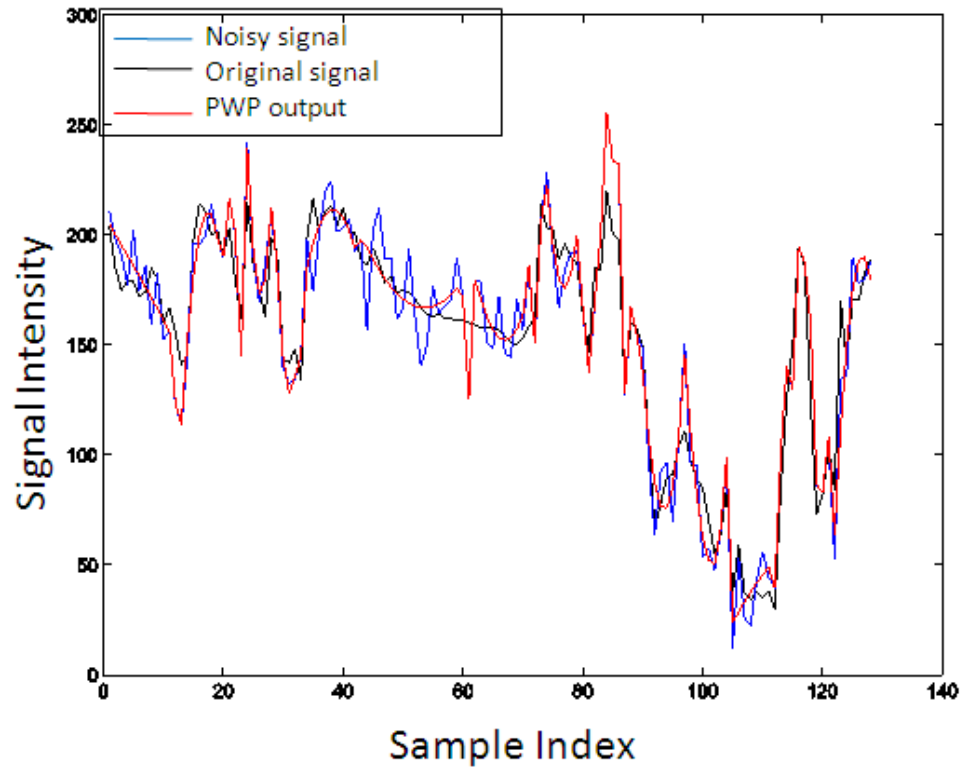


Figure 2.37: Denoising an image slice by the proposed piecewise parabolic technique.

In Figure 2.45 the error norm of our parabola technique is given. We see that the technique is moderately robust against deviations of the estimated noise standard deviation. But, the smallest value of the error norm is the best one. Here, we can draw the conclusion that our technique is the best one for this setting.

In Figure 2.46 the error norm of our cubic technique is given. We see that the technique is moderately robust against deviations of the estimated noise standard deviation. But, the smallest value of the error norm is the second one.

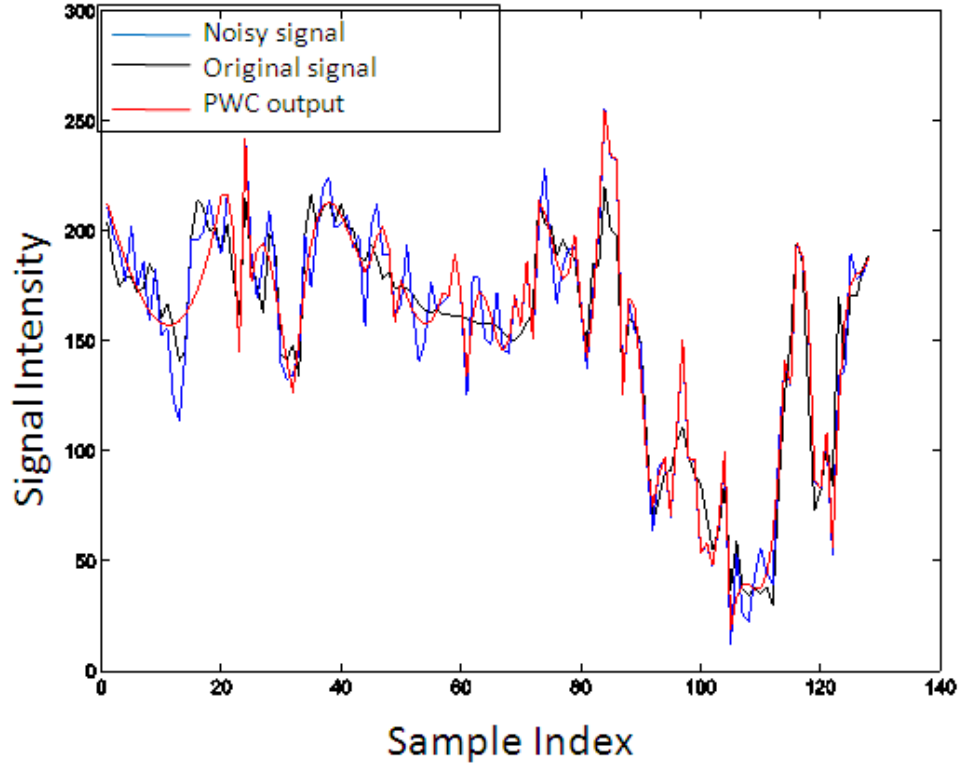


Figure 2.38: Denoising an image slice by the proposed piecewise cubic technique.

2.8 Signal Denoising by Piecewise Continuous Polynomial Fitting

Piecewise smooth signal denoising is cast as a non-linear optimization problem in terms of transition boundaries and a parametric smooth signal family. Optimal transition boundaries for a given number of transitions are obtained by using particle swarm optimization. The piecewise smooth section parameters are obtained as the maximum likelihood estimates conditioned on the optimal transition boundaries. The proposed algorithm is extended to the case where the number of transition boundaries are unknown by sequentially increasing number of sections until the residual error is at the level of noise standard deviation. This proposed technique is illustrated in Figure 2.47 Performance comparison

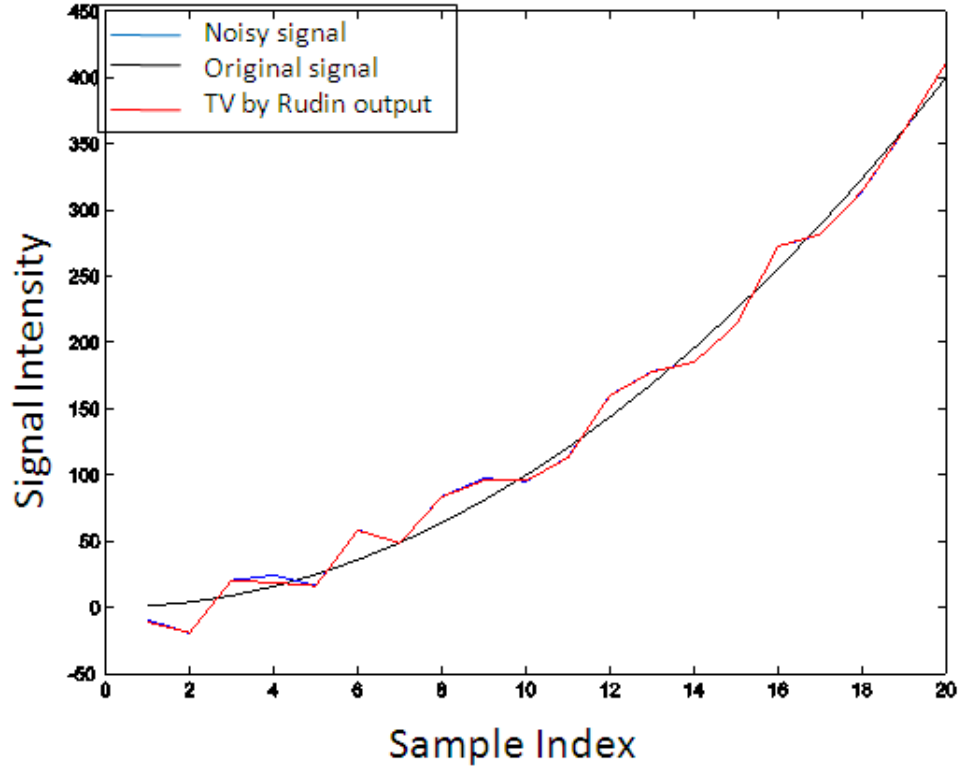


Figure 2.39: Denoising a parabolic curve by the technique of Rudin *et al.*

with the state of the art techniques reveals the important advantages of the proposed technique. The conference paper about this technique was accepted to SIU 2010.

2.8.1 Formulation and Algorithm

Let $f[n]$ be a noisy version of piecewise polynomial sequence of length N with transition boundaries at:

$$0 < e_1 < e_2 \dots < e_D < N - 1, \quad (2.62)$$

$$\hat{f}[n; \mathbf{a}, \mathbf{e}] = a_{0,i} + a_{1,i}n + \dots + a_{k,i}n^k \text{ if } e_i \leq n < e_{i+1}, \quad (2.63)$$

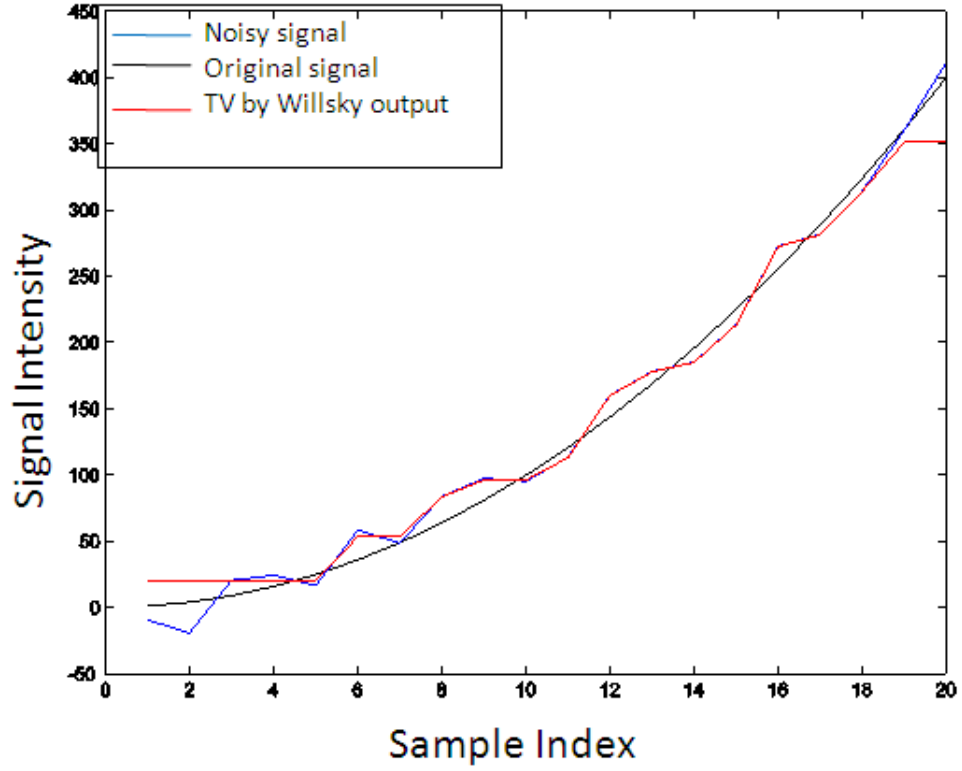


Figure 2.40: Denoising a parabolic curve by the technique of Willsky *et al.*

$$\min_{\mathbf{e}} \left[\min_{\mathbf{a}} \sum_{n=0}^{N-1} |f[n] - \hat{f}[n; \mathbf{a}, \mathbf{e}]|^2 \right]. \quad (2.64)$$

For a given estimate of transition boundaries, we can find the optimal $\mathbf{a}$ that provides the best fit between the boundaries easily by solving the following set of least squares polynomial fit problems for each $D + 1$ intervals between the boundaries where $e_0 = 0$, $e_{D+1} = N$:

$$\min_{\mathbf{a}_i} \sum_{n=e_i}^{e_{i+1}-1} |f[n] - (a_{0,i} + a_{1,i}n + \dots + a_{k,i}n^k)|^2, 0 \leq i \leq D, \quad (2.65)$$

which has the following closed form solution:

$$\mathbf{a}_i = (\mathbf{A}_i^T \mathbf{A}_i)^{-1} \mathbf{A}_i^T \mathbf{f}_i. \quad (2.66)$$

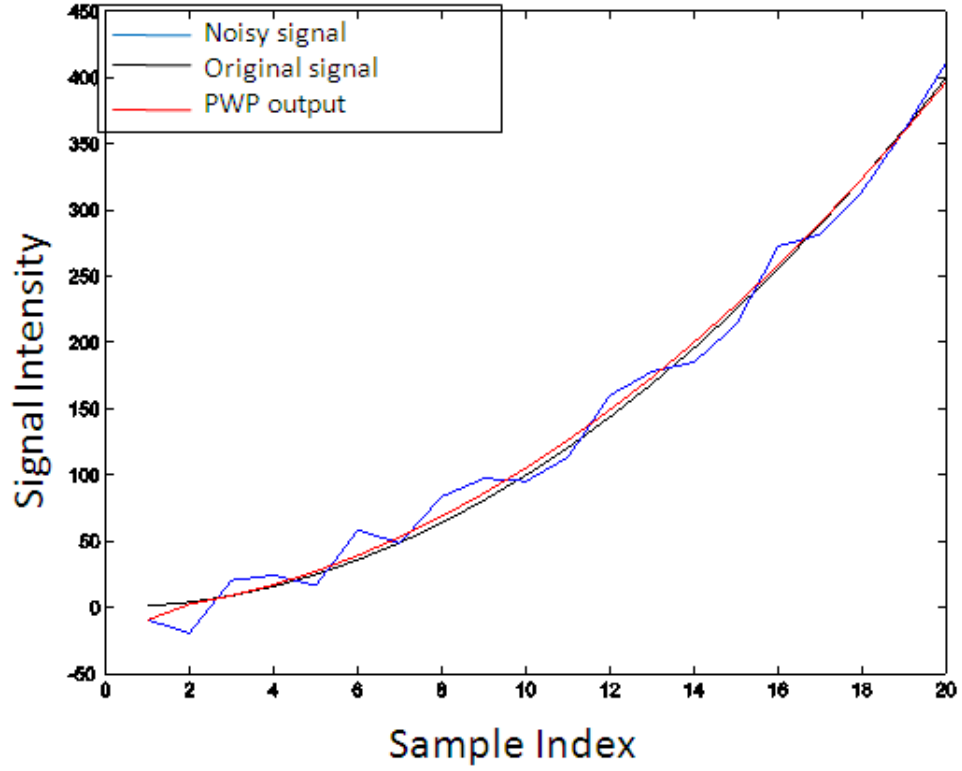


Figure 2.41: Denoising a parabolic curve by our technique of piecewise parabolas.

here, $\mathbf{A}_i = [\mathbf{1} \quad \mathbf{n} \quad \dots \mathbf{n}^k]$, where $\mathbf{n}$ is a vector of entries $e_i \leq n_j < e_{i+1}$.

This solution is achieved by the orthogonal projection method. Since the noisy observation vector $\mathbf{f}_i$ is not in the range space of polynomial model matrix $\mathbf{A}_i$ as illustrated in Figure 2.48, it is projected on that range space. This way the linear system becomes solvable and then the closed form solution in (2.66) is obtained.

To avoid matrix inversion, the $\mathbf{a}$ vector can be found by QR decomposition as well. For QR decomposition, the methods of Householder transform, Givens rotations and Gram Schmidt Orthonormalization were proposed in [25].

Let the minimum cost of the best polynomial fit in the i^{th} section be $J_i(\mathbf{f}, \mathbf{e})$. Then the original optimization problem can be cast as an optimization in terms of the transition boundaries alone as:

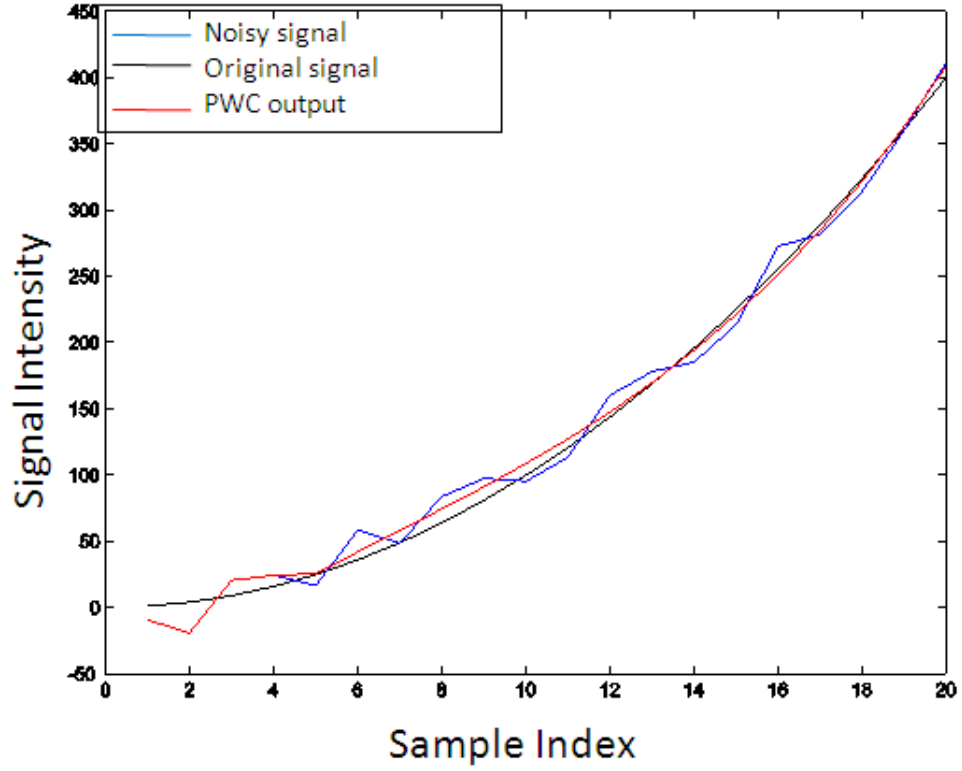


Figure 2.42: Denoising a parabolic curve by our technique of piecewise cubics.

$$\min_{\mathbf{e}} \sum_{i=0}^D J_i(\mathbf{f}, \mathbf{e}). \quad (2.67)$$

The reduced optimization has far few variables than the original optimization problem. Hence we have better chance of finding its solution. However, the cost function typically has a complicated minima structure that makes it difficult to search for the optimal solution through a local optimization technique. Therefore, we propose to use Particle Swarm Optimization [34] to increase our chances of obtaining the globally optimal solution for the transition boundaries.

To obtain the global minima of the cost function in (2.67) in terms of the boundary vector $\mathbf{e}$, we use a standard implementation of PSO algorithm as follows. Let N_p be the number of particles used in the PSO implementation. Note that each particle of the PSO is a solution candidate for the optimization problem, hence, they are length D vectors with entries as candidate positions for the

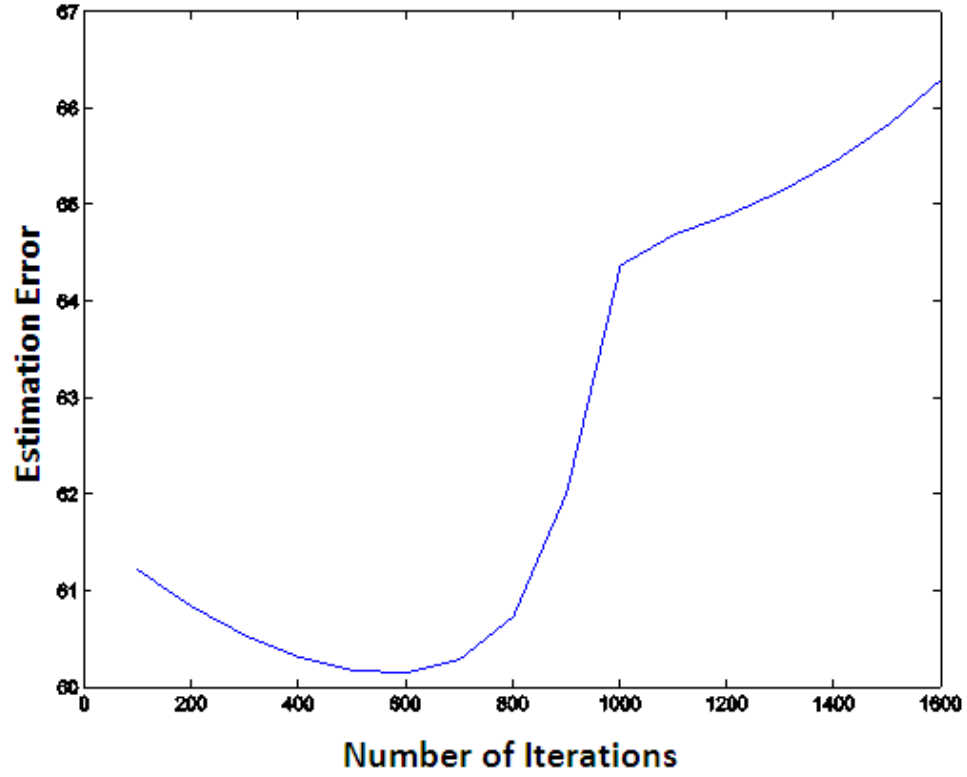


Figure 2.43: Estimation error of Total Variation by Rudin for the parabolic curve.

boundaries. We start with a randomly chosen population of N_p vectors. Then we use the following standard update equations of the PSO [34] to produce new populations:

$$\mathbf{v}^{p,t+1} = \alpha(\mathbf{v}^{p,t} + w_1\zeta_1(\mathbf{e}_{pb}^{p,t} - \mathbf{e}^{p,t}) + w_2\zeta_2(\mathbf{e}_{gb}^t - \mathbf{e}^{p,t})), \quad (2.68)$$

$$\mathbf{e}^{p,t+1} = \mathbf{e}^{p,t} + \mathbf{v}^{p,t}. \quad (2.69)$$

In these equations, $\mathbf{v}^{p,t}$ is the velocity and $\mathbf{e}^{p,t}$ is the position of the p^{th} particle respectively. $\mathbf{e}_{pb}^{p,t}$ is the historically best position of the p^{th} particle; $\mathbf{e}_{gb}^t$ is the historically best position among all particles; ζ_1 and ζ_2 are i.i.d. random variables uniformly distributed in $[0, 1]$.

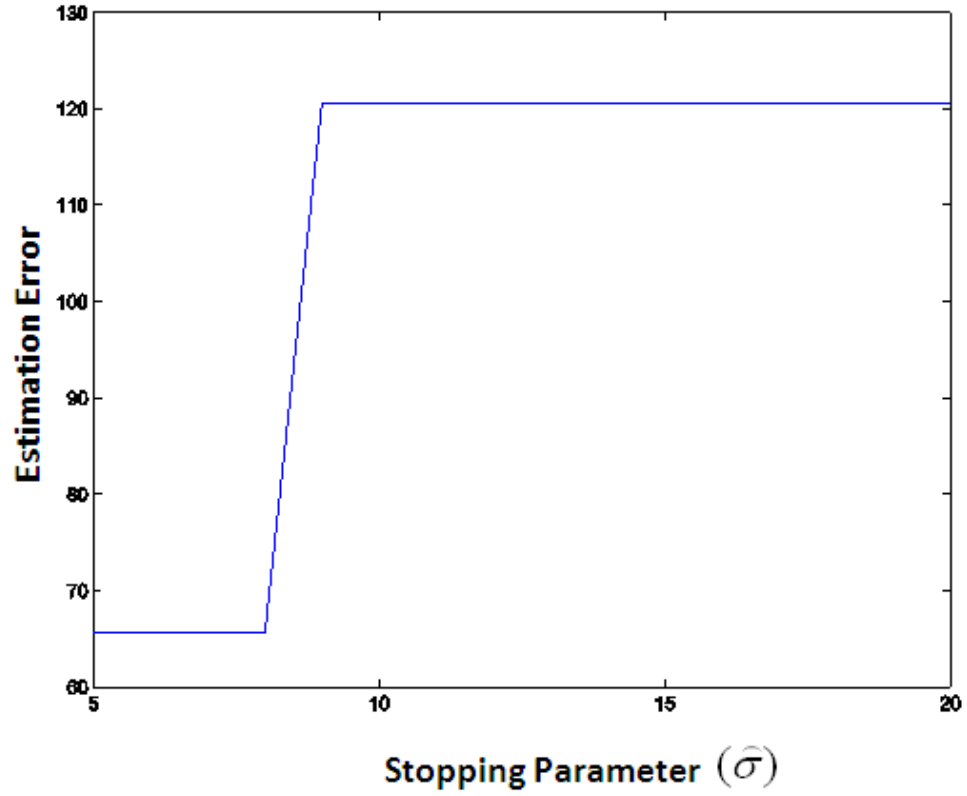


Figure 2.44: Estimation error of Total Variation by Willsky for the parabolic curve.

In the following section, we investigate the performance of the PSO based polynomial denoising technique over a set of simulations. We also address the issue of determining the number of intervals D , and the order of polynomials, k .

2.8.2 Comparison of the Techniques

First we investigate a simple case of piecewise constant signal with a discontinuity. This problem has also been addressed by the total variation technique [4]. Therefore, we will compare performance of the proposed technique with total variation approach. For the signal shown in Figure 2.8.2, the simulations were realized at five different values of the noise standard deviation. Comparison of the results shown in Figure 2.50 and Figure 2.51 reveals that, the proposed technique denoise the signal more effectively than the total variation technique. Note

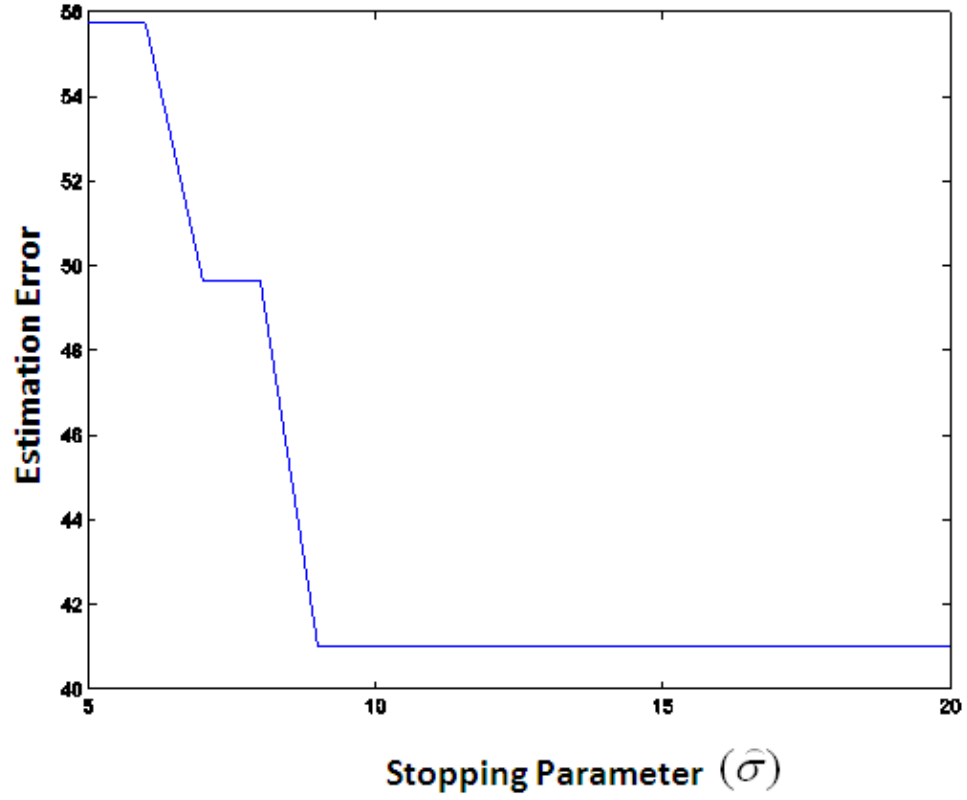


Figure 2.45: Estimation error of our Parabolic Denoising Technique for the parabolic curve.

that, for the comparison we used the lowest estimation error in the total variation curves given in Figure 2.51. Since the actual signal is unknown, the estimation error curves for the total variation as a function of iterations cannot be produced. Therefore, in practice, the performance of the total variation denoising cannot achieve the levels we used here for the comparison.

Technique	Error Norm
Total Variation by Rudin <i>et al</i>	1.3754
Principal Curve Projections	1.1866
Our technique	1.0859

Table 2.2: Error norms for a signal with three boundaries [2].

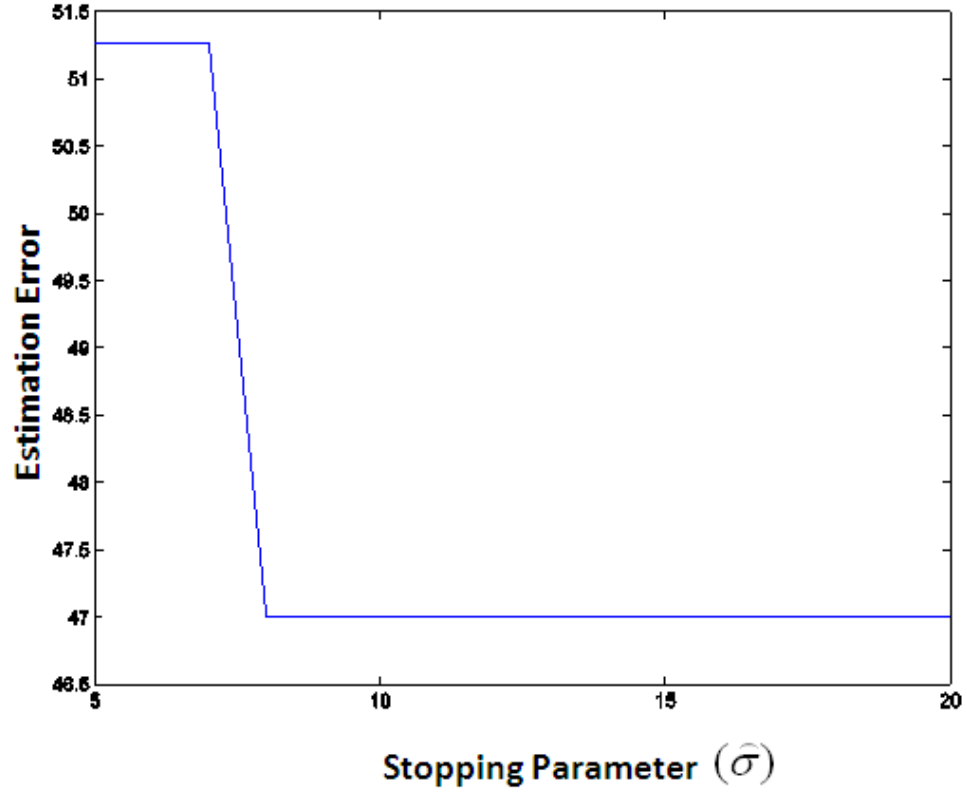


Figure 2.46: Estimation error of our Cubic Denoising Technique for the parabolic curve.

Now, we present results obtained in a more complicated situation where the original signal has four polynomial sections as shown in Figure 2.52. We compared the performance of the proposed technique with the Rudin's total variation technique, which is capable of denoising not necessarily piecewise constant signals, and, the principal curve projection technique, which is a non-parametric denoising technique [5].

The obtained denoising results are shown in Figures 2.52-2.54. As seen from these figures and estimation errors given in Table 2.2, the proposed technique provides a significantly improved estimate to the original signal.

To illustrate an effective method of determining D , we use the signal shown in Figure 2.52. As shown in Figure 2.55, the normalized fit error of the estimated polynomial signal decreases monotonically as a function of D . By monitoring

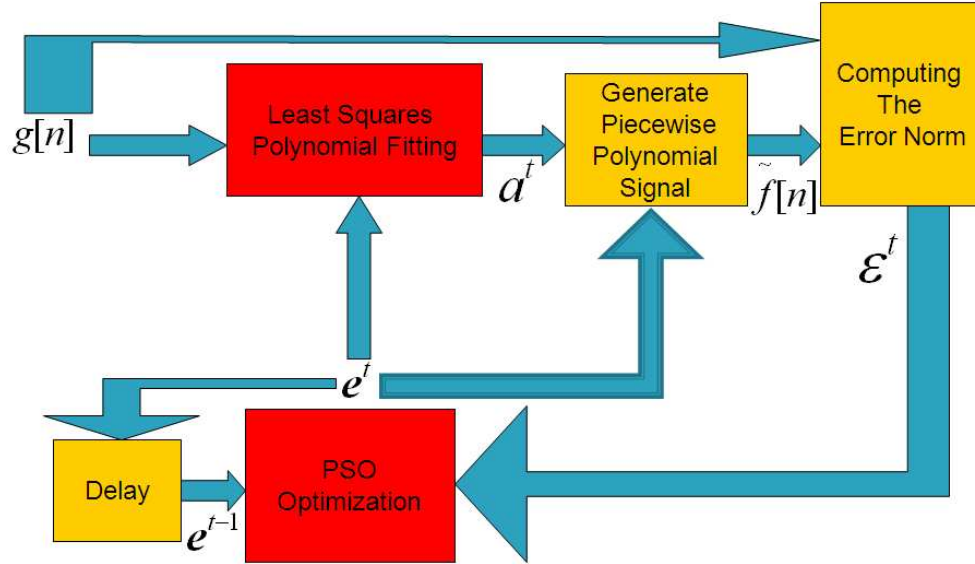


Figure 2.47: Flow diagram of the proposed Piecewise Continuous Polynomial Fitting Technique.

this fit error, we can increase D until the fit error is about the noise standard deviation.

2.9 Denoising Piecewise Polynomial Signals by Technique based on Convolution with Wavelets

Matched filters have various application fields from telecommunications to signal processing. Here, the application of matched filters on signal denoising is presented. One dimensional piecewise polynomial signals are denoised with a constraint of noise standard deviation. Signal's first difference is taken iteratively until a piecewise constant signal is obtained and at each step, the signal is convolved with a Haar wavelet of fixed size. If there is a discontinuity at a level, then the matched filtering results in impulses. The maximum of these points is the location of discontinuity. A threshold is chosen and false alarms and detection are decided that way. False alarm occurs if the convolved signal passes the

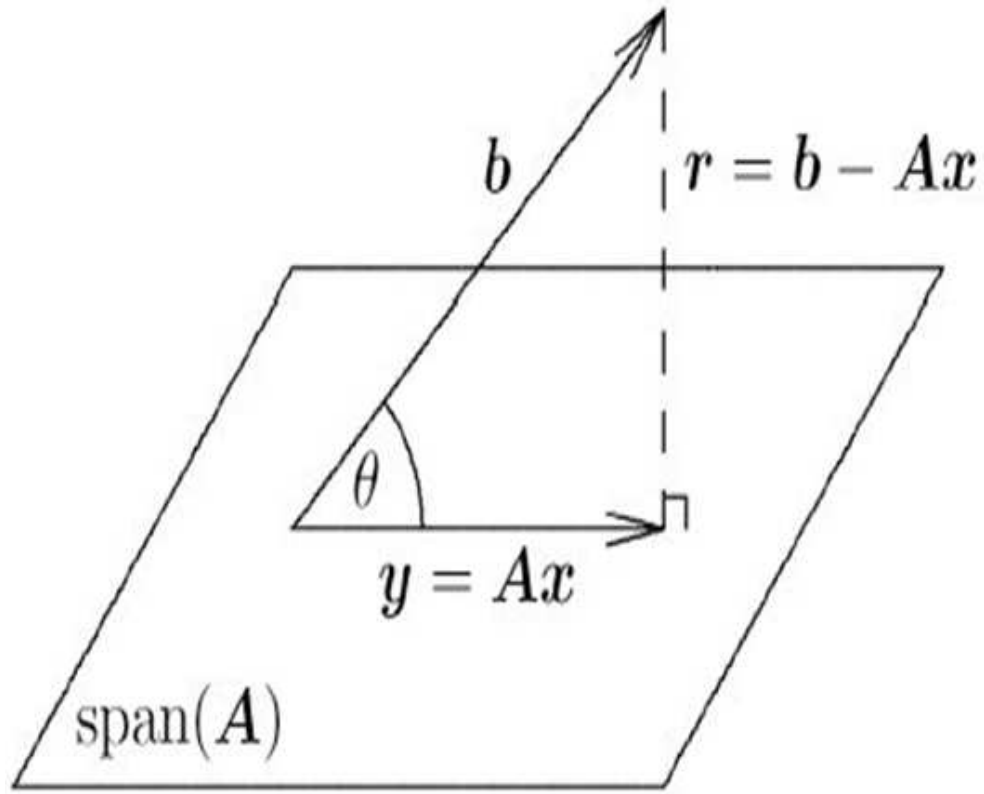


Figure 2.48: The geometry of polynomial fitting problem and the solution based on orthogonal projection [1].

threshold in a position where there is no discontinuity. After at each derivative level some edges are obtained, these edges are merged. From the final set of these edges, the region boundaries are determined and the curve fitting with a known polynomial degree is applied between the boundaries. The threshold is decreased and the process is repeated until the signal after the curve fitting is applied is within some difference norm smaller than the noise standard deviation. This proposed technique is depicted in 2.57 Simulations are presented at which it is shown that false alarms are in acceptable values.

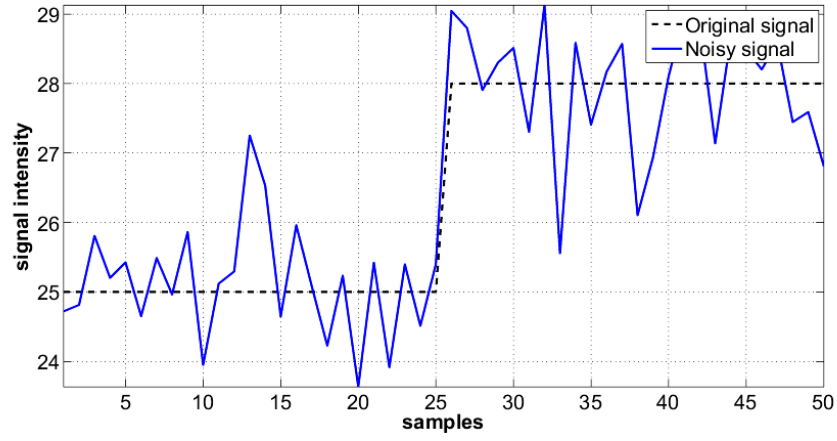


Figure 2.49: Piecewise constant signal with noise [2].

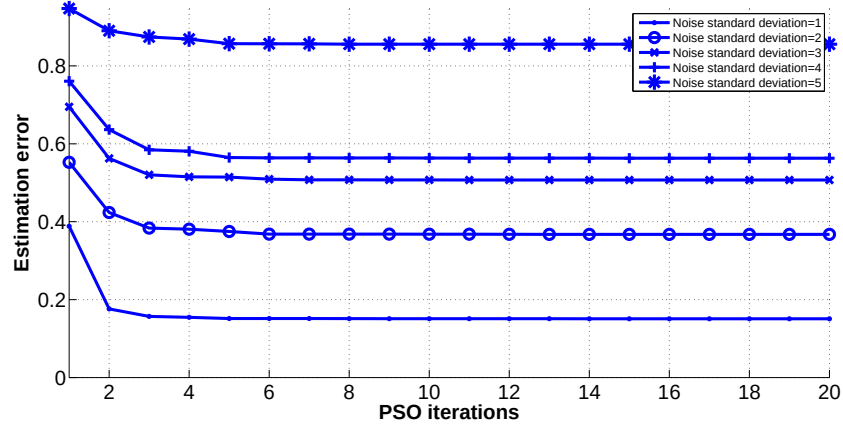


Figure 2.50: Error in estimation of the original signal by the proposed PSO based polynomial fit technique at five different noise levels [2].

2.9.1 Formulation and Algorithm

Let us assume a one dimensional noisy signal $f[n]$. The convolution kernel $h[n]$ is a Haar wavelet such that in its support, $h[n] = 1$ for $n = 1, \dots, 10$ and $h[n] = -1$ for $n = 11, \dots, 20$

1. Compute the matched filter output so that: $t = 1$; $g^t[n] = f[n]$;
2. $y^t[n] = g^t[n] * h[n]$

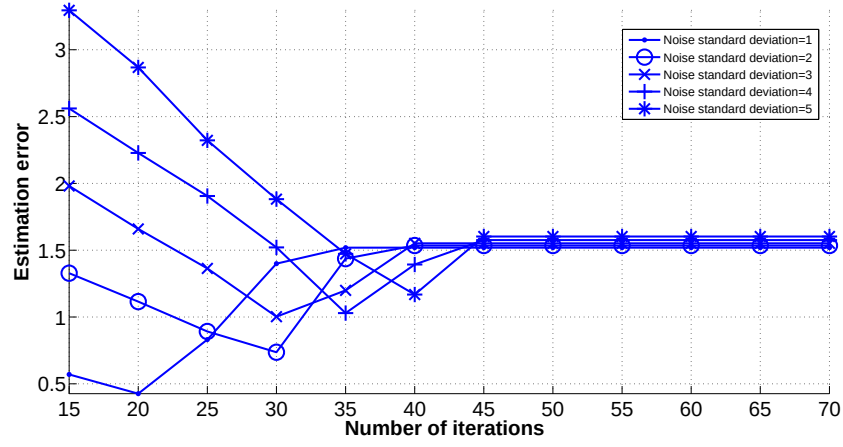


Figure 2.51: Error in estimation of the original signal by total variation technique at 5 different noise levels [2].

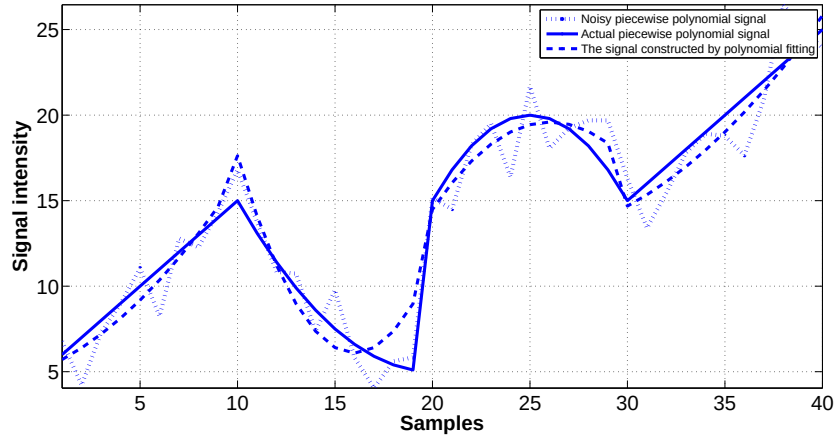


Figure 2.52: The denoising result obtained with the proposed technique [2].

3. Let us assume that the location at which $y^t[n]$ takes the highest value is point n_p such that $n_p = \underset{n}{\operatorname{argmax}}\{y^t[n]\}$. If $y[n_p] - y[n_p - 10] < N$ then go to Step (6). It means that there is no discontinuity at this level of derivative.

4. Take a threshold L and find the signal parts that exceed this threshold. (In this case we assume there are N parts)

$$w_1^t[n] = y^t[n] \text{ for } n = k_1, k_1 + 1, \dots, k_1 + M_1 \text{ if } y^t[n] > L \dots$$

$$w_N^t[n] = y^t[n] \text{ for } n = k_N, k_N + 1, \dots, k_N + M_N \text{ if } y^t[n] > L.$$

5. Then find $d^t[k] = \underset{n}{\operatorname{argmax}}\{w_k^t[n]\}$. These are the discontinuity points from the t^{th} derivative.

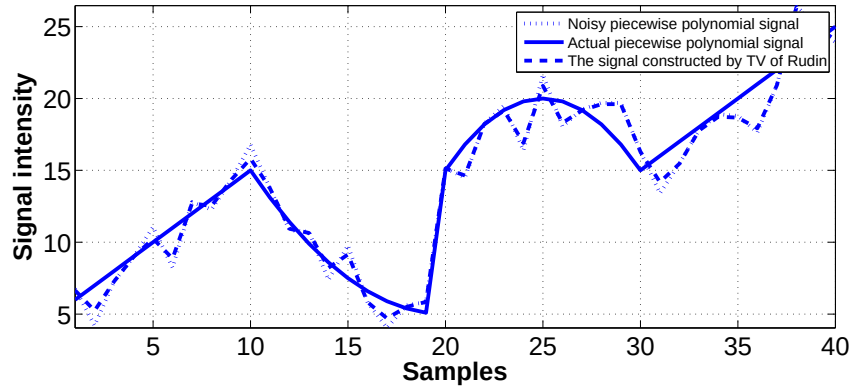


Figure 2.53: The graphs of total variation by Rudin *et al* for a signal with three jumps [2].

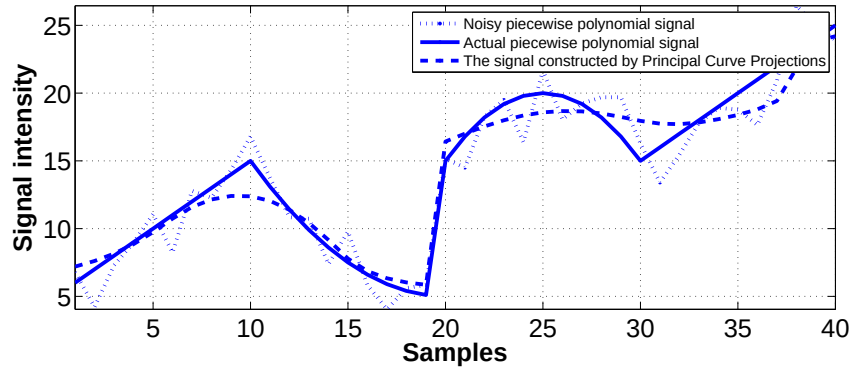


Figure 2.54: The graphs of Principal Curve Projections for a signal with three jumps [2].

6. Compute the first difference of $g^t[n]$ such that $g^{t+1}[n] = g^t[n] - g^t[n - 1]$ and go to Step (2) if $t < T$.
7. Merge all the $d^t[k]$ for all t .
8. Apply polynomial fitting with a least squares sense between each of the discontinuity boundaries. The polynomial degree D_k is found by finding minimum degree for which normalized residual norm falls below σ . Mathematically,

$$\min\{D_k\} \text{ such that } \|\mathbf{f}_k - LS_{D_k}\{\mathbf{f}_k\}\| < \sigma \cdot \sqrt{K},$$

where K is the length of the k^{th} segment.

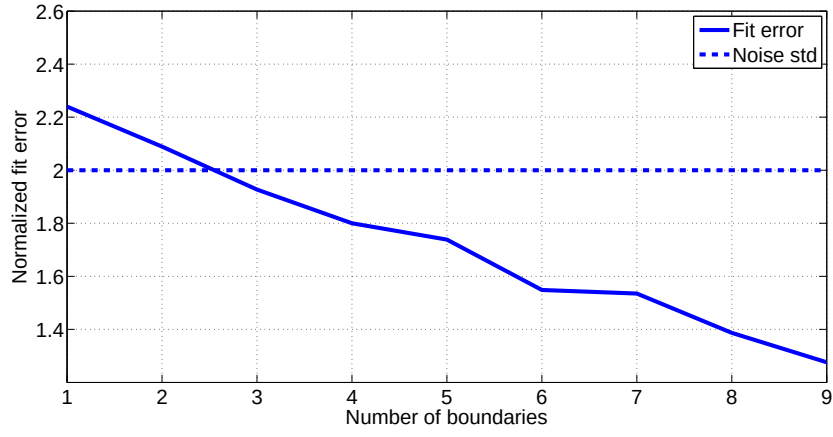


Figure 2.55: The fit error of as a function of number of boundaries as D [2].

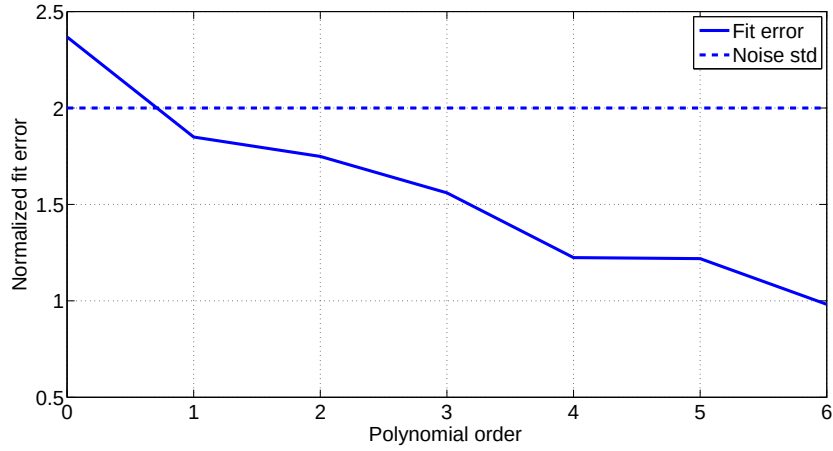


Figure 2.56: The fit error as a function of polynomial order k [2].

2.9.2 Comparison of the Techniques

It will be useful to begin the comparisons with relatively simple signals. The first case we will examine is for a signal that is a piecewise constant signal that includes one jump. The signal is shown in Figure 2.58.

The probability of detection (PD) is given in Figure 2.59. By detection, we mean the success of finding the discontinuity that actually resides in the signal. In the graph, X axis the SNR Y axis is the probability of detection. The parallel lines are theoretical values for thresholds of 100 to 500. Highest PD is obtained with a smaller threshold therefore the upmost curve is for threshold of 100. It is

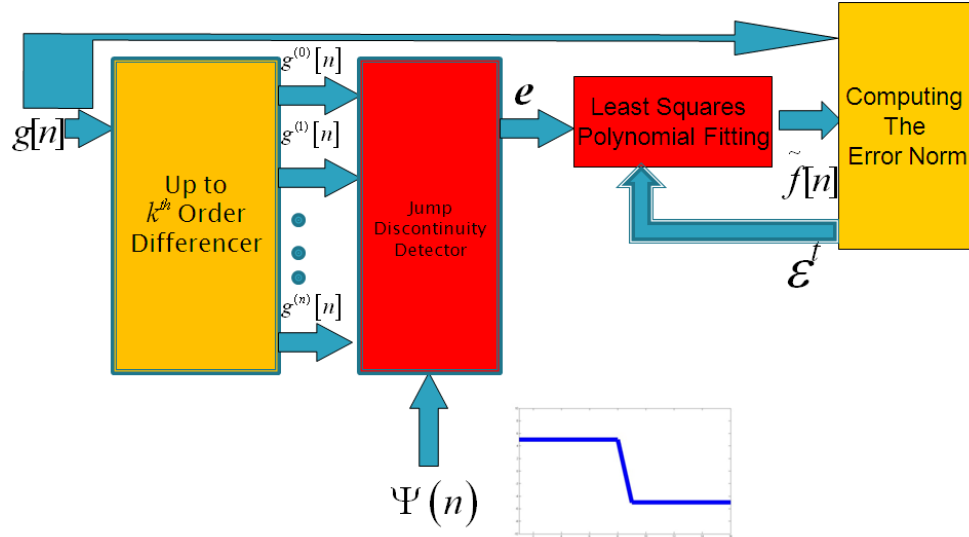


Figure 2.57: Flow diagram of the proposed Wavelet based Technique.

seen that the threshold stops at about 400, because the probabilities of detection are similar.

The probability of false alarm (PF) is shown in Figure 2.60. False alarm occurs when the output of the matched filter exceeds the threshold because of noise where there is no actual discontinuity. In the figure, the PF achieved is much smaller than the theoretical values which is a big plus for our technique.

The next simulation is based on a signal which is a basic triangle. The signal is shown in Figure 2.61. The comparison of probability of detection for theory and application is given in Figure 2.62. The parallel lines are for thresholds of 10 to 60. The process seems to stop at a threshold of 60 since the probability of detection is similar. The probability of false alarm is depicted in Figure 2.63. The probability of false alarm with a stopping criterion of noise standard deviation takes a moderate value compared to the theory which uses constant thresholds.

It is reasonable to go on with the simulations with a relatively more complicated signal. The signal is shown in Figure 2.64 .

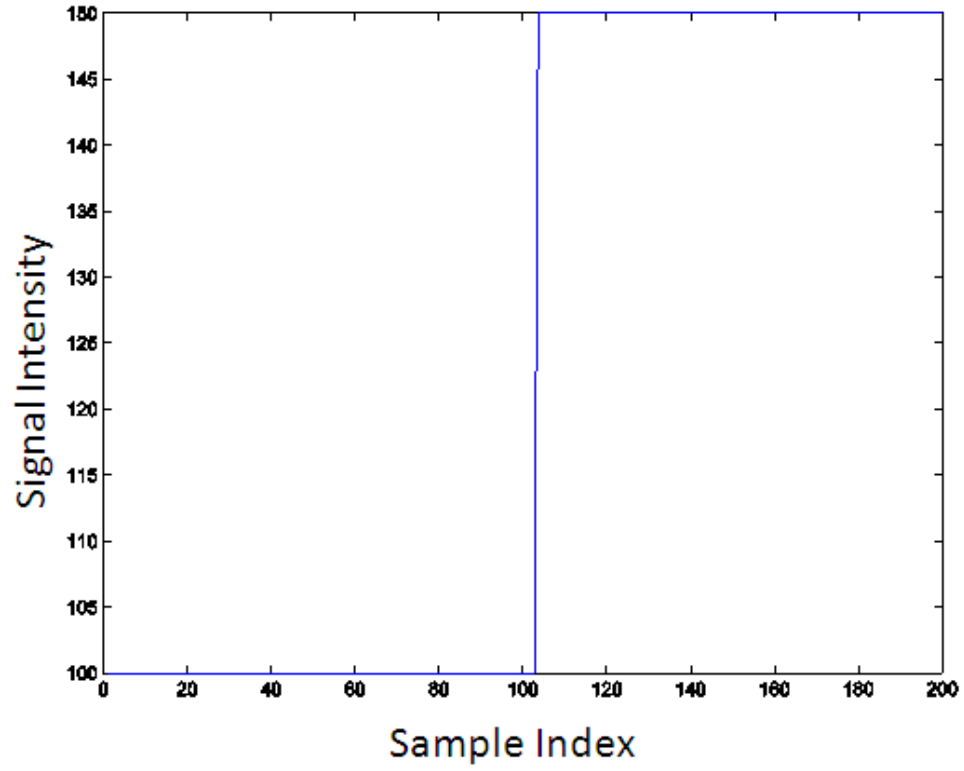


Figure 2.58: Original syntetic piecewise constant signal with one jump discontinuity.

Gaussian noise of standard deviation of 5 was added to the signal and it is depicted in Figure 2.65.

The signal was denoised with the algorithm presented in this thesis and it is shown in Figure 2.66. The error norm with respect to the original signal turned out to be only 10.22. There is no false alarm and all the discontinuities are detected.

The question arises whether this technique can denoise more difficult signals. For example, can a signal which contains two adjacent lines with the same slope and with a discontinuity in between be denoised. Such a signal is shown in Figure 2.67.

Normal noise of standard deviation of 1 was added on this signal and the noisy signal is given in Figure 2.68.

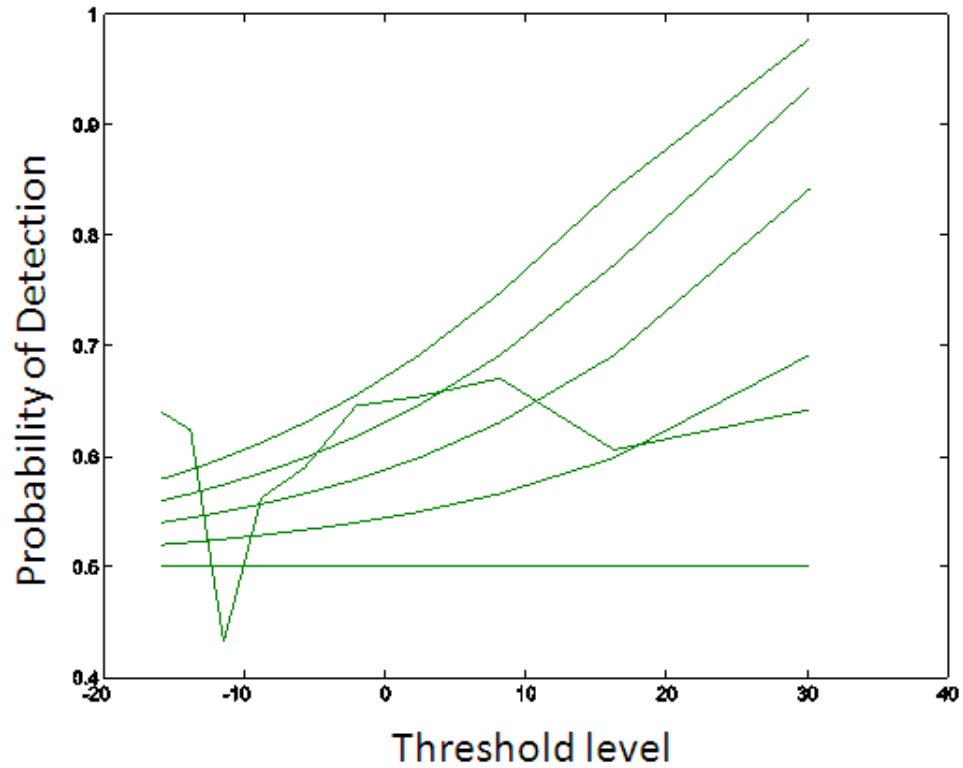


Figure 2.59: The probability of detection for theoretical thresholds and for the actual stopping criterion about the case of piecewise constant signal.

The signal is constituted by lines, therefore it has polynomials of order 1. Therefore, discontinuities are going to be searched for at two stages of matched filtering. The result of convolution with the signal itself is given in Figure 2.69. Only the rightmost end of the signal can be detected as an edge from this stage. The output of convolution with the signal's first difference is given in Figure 2.70. All the discontinuity boundaries are detected at the derivative. The signal after line fitting between these boundaries is shown in Figure 2.71. It is clearly observed that the challenging discontinuity was detected.

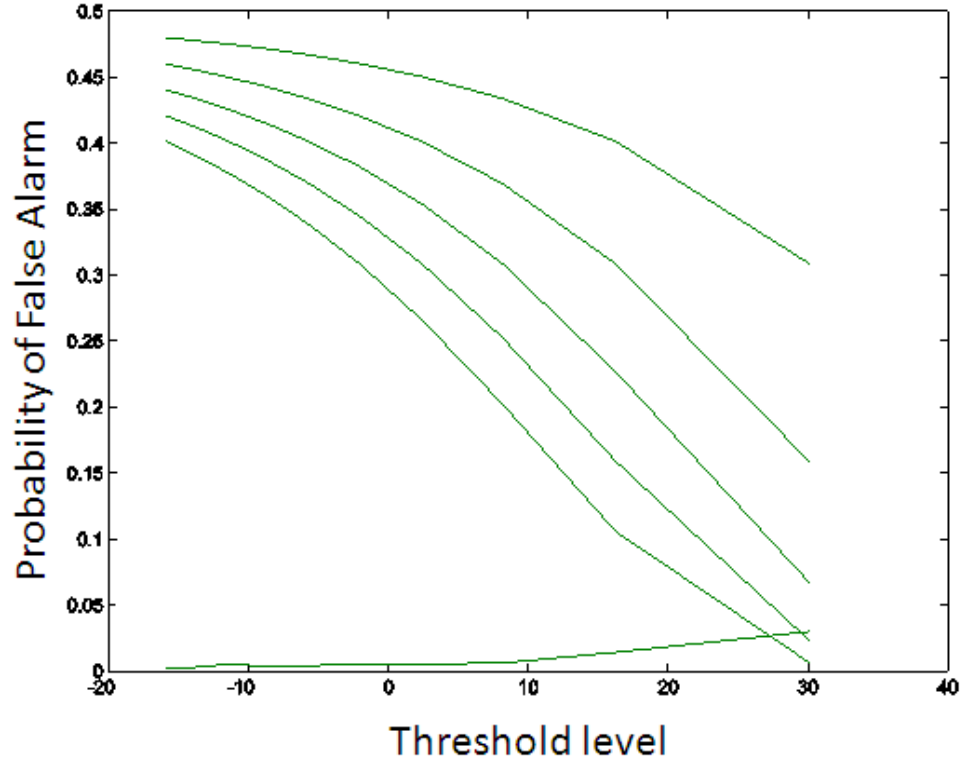


Figure 2.60: The probability of false alarm for theoretical thresholds and for the actual stopping criterion about the case of piecewise constant signal.

2.10 Texture Recovery by Adaptive Patch Averaging

Total variation based texture recovery techniques have problem about robustness, because they generally rely on the fine-tuning of a penalty function. This thesis defines a more robust technique by synthesizing total variation minimization and a new method so-called iterative pasting. As a first step, the edges and flat steps in a one dimensional signal is detected by total variation with a large coefficient. Then, existence of periodic structure is investigated in the residual. This is done by estimating the period in an interval iteratively. In this interval, the period other than which only noise remains is selected as the period. If the residual of none of the periods have lower energy then a threshold, no periodicity is concluded. Periodicity search is done in an adaptive window whose length

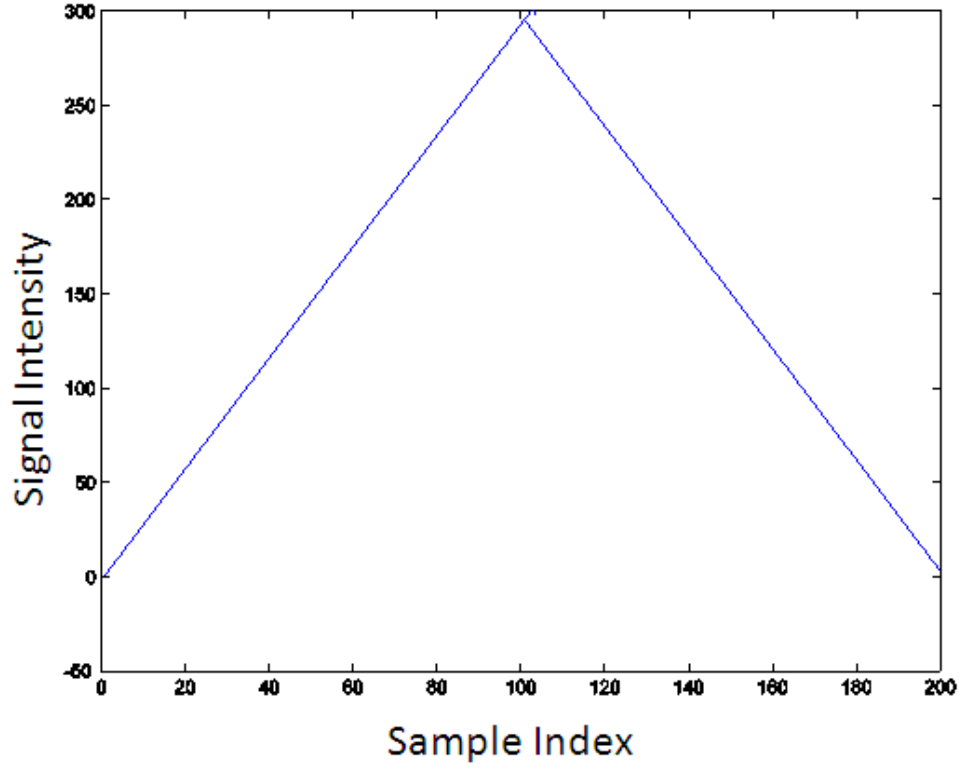


Figure 2.61: Original synthetic triangular signal.

enlarges by the detection of the same periodic structure nearby. After detection of a period which has a monotonic texture, all periods are averaged so that noise is removed. Then, this prototype period is pasted iteratively in the window. This technique is depicted in Figure 2.72. This technique was tested on one dimensional signals, but it can be applied on $2 - D$ images as well. For this purpose, Radon Transform can be used to carry the image to one dimension.

2.10.1 Formulation and Algorithm

The observed signal $x'[n]$ is the noisy version of the textured signal $x[n]$ such that:

$$x'[n] = x[n] + u[n], \quad (2.70)$$

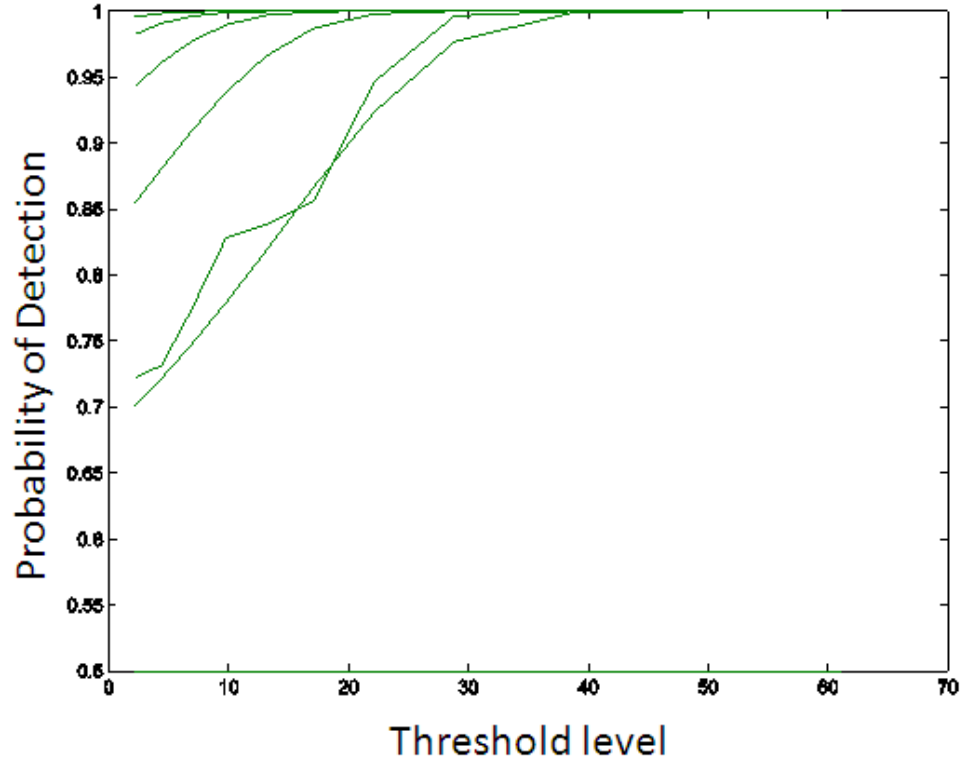


Figure 2.62: The probability of detection for theoretical thresholds and for the actual stopping criterion about the triangular signal.

where $u[n]$ is Gaussian white noise with a standard deviation σ and a mean of 0.

As a first step, for edge and step detection, $x'[n]$ is passed through a total variation denoising system with a σ' higher than the estimated standard deviation σ .

$$\hat{x}[n] = TV_{\sigma'}\{x'[n]\}, \quad (2.71)$$

Then, to obtain the residual, the difference from the noisy signal is found.

$$d[n] = x'[n] - \hat{x}[n]. \quad (2.72)$$

For example, let us assume that the original vector consists of three different segments, one of which is periodic, the other is again periodic with a different

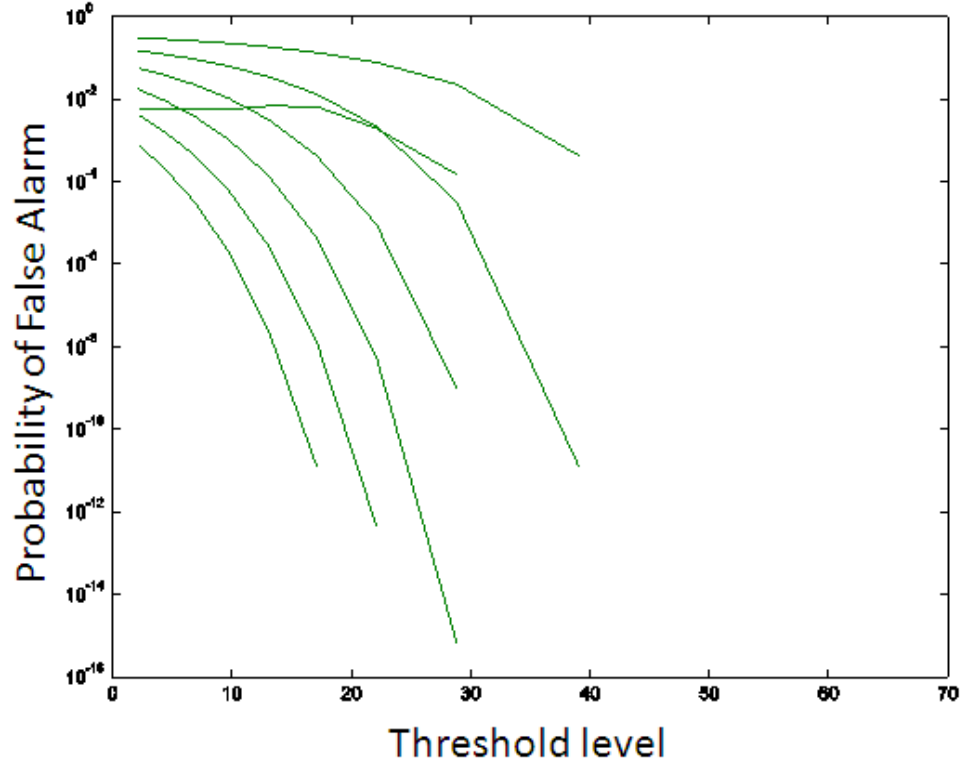


Figure 2.63: The probability of false alarm for theoretical thresholds and for the actual stopping criterion about the triangular signal.

period, and the last is non-periodic but may have some non-periodic variations. Note that these segments of the vector may or may not be in the same level. That is,

$$x[n] = \left\{ \begin{array}{lll} p[n] & \text{if } n < n_1, & ; \\ r[n] & \text{if } n_1 \leq n < n_2, & ; \\ s[n] & \text{if } n_2 \leq n < n_3, & . \end{array} \right\}, \quad (2.73)$$

where $p[n]$ is periodic with period T_1 with a certain structure, $r[n]$ is periodic with period T_2 with some other structure and $s[n]$ is non-periodic. Note that the algorithm is robust against the changes in the arrangement of these sub-vectors and their step levels. We have the noisy version of this signal. That is,

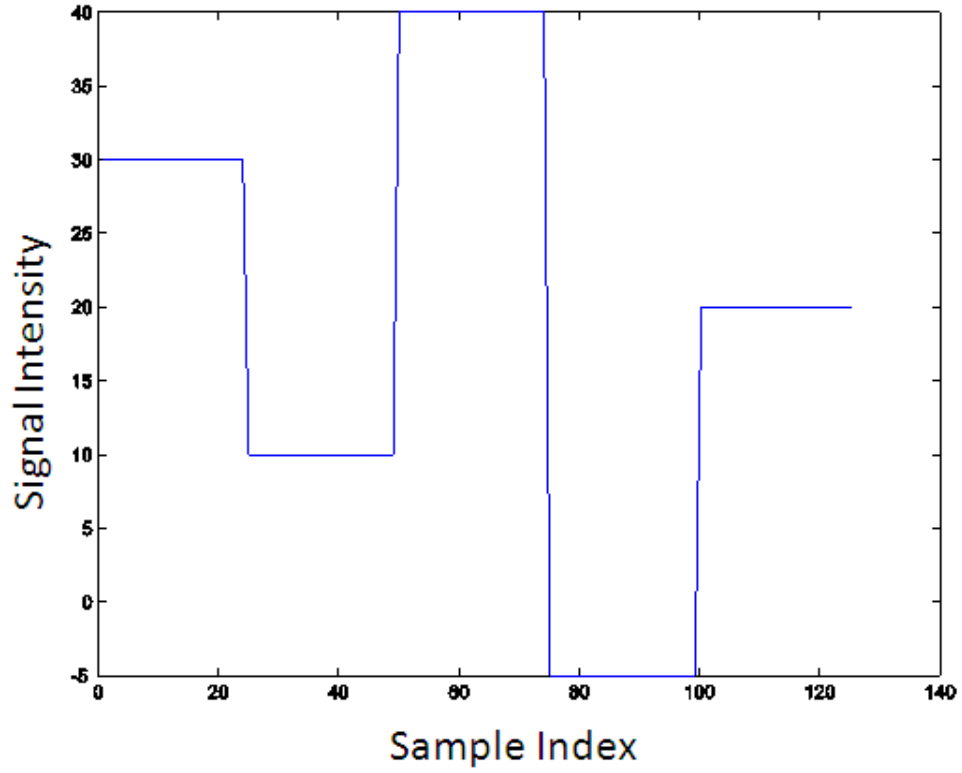


Figure 2.64: Original synthetic signal with piecewise constant parts.

$$x'[n] = \left\{ \begin{array}{ll} p'[n] \text{ if } n < n_1, & ; \\ r'[n] \text{ if } n_1 \leq n < n_2, & ; \\ s'[n] \text{ if } n_2 \leq n < n_3, & . \end{array} \right\}, \quad (2.74)$$

where $x'[n]$ indicates the noisy signal like stated before. If an edge is detected between these sub-arrays, each step is processed distinctly. As put in the above lines, this crude edge detection is accomplished by total variation reduction. If the sub-arrays have the same level, the process is started from the right-most index with a window size of L , and the window is enlarged adaptively until a new edge or a new periodic or non-periodic structure is reached. This enlargement is to increase success, because larger window means more periods to be averaged to remove noise. This will be explained in more detail below.

Once the edges are detected by total variation, the process starts from the first index $n = 1$ by windowing the residual so that:

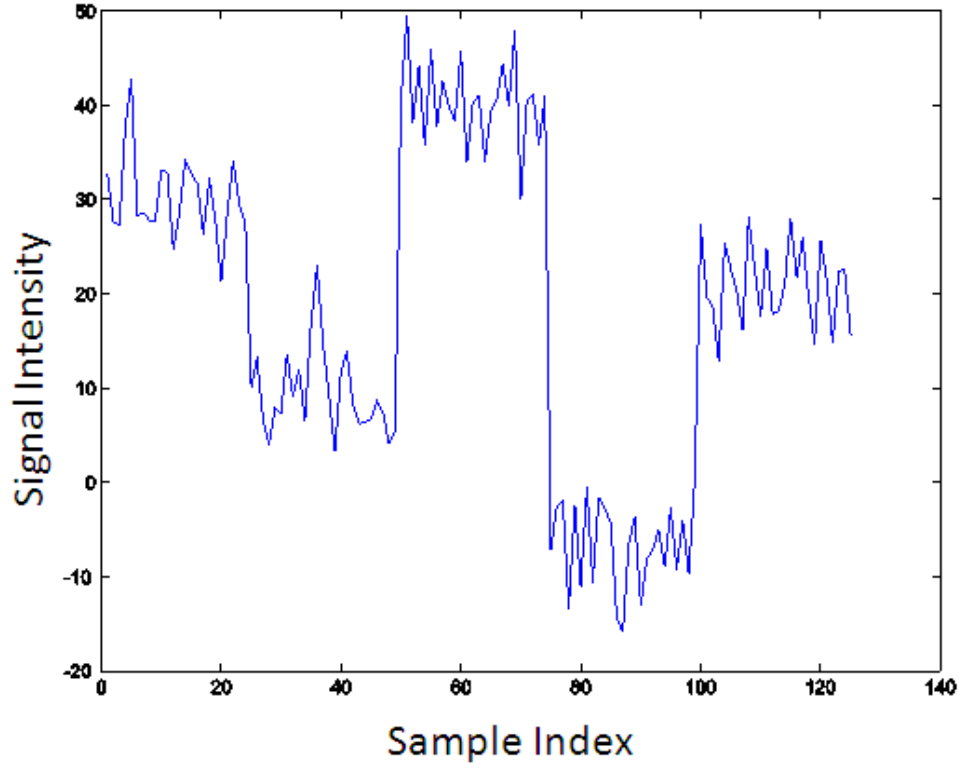


Figure 2.65: Noisy piecewise constant signal with an additive Gaussian noise with a standard deviation of 5.

$$f[n] = w[n].d[n], \quad (2.75)$$

$$w[n] = \begin{cases} 1 & \text{if } n \leq L, \\ 0 & \text{if } n > L, \end{cases} \quad (2.76)$$

The next task is to investigate periodicity in this segment $f[n]$. This is achieved by taking the autocorrelation of $f[n]$. It is a well known fact that this autocorrelation signal makes impulses if the signal has a periodicity. This is also true for a noisy periodic signal as well. (Noise should not be too high.) On the other hand, if the signal is not periodic, the autocorrelation signal decays. This implies a difference of total variation between autocorrelation of periodic and non-periodic noisy signals. Thus, it is reasonable to compare the total variation rate of the autocorrelation signals to decide periodicity. That is,

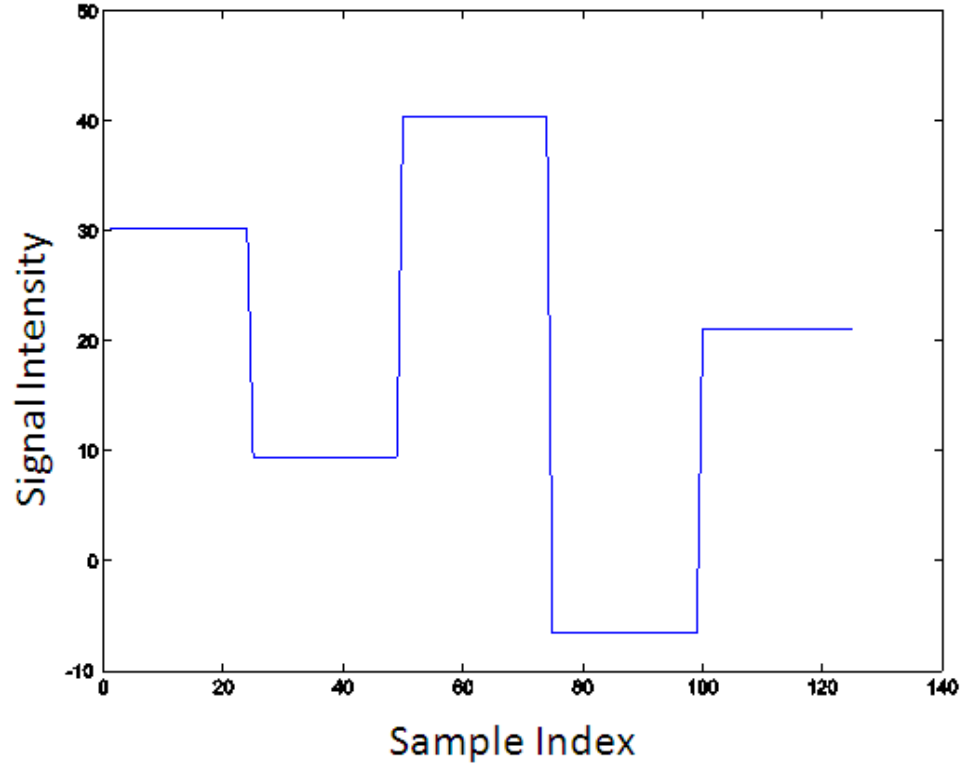


Figure 2.66: Piecewise constant signal denoised by the proposed technique.

$$r_f[k] = \frac{1}{N+1} \cdot \sum_{n=1}^N f[n-k] \cdot f[n] \quad (2.77)$$

$$tv\{r_f[k]\} = \sum_{k=2}^N |r_f[k] - r_f[k-1]| \quad (2.78)$$

$$v = \frac{tv\{r_f[k]\}}{K}. \quad (2.79)$$

If v is greater than a threshold, the vector is decided to contain periodic components. If not, it is assumed to be non-periodic, therefore, total variation process with an optimum estimated standard deviation is followed in order to reconstruct the details. That is,

$$\tilde{x}[n] = TV_{\sigma}(x'[n]). \quad (2.80)$$

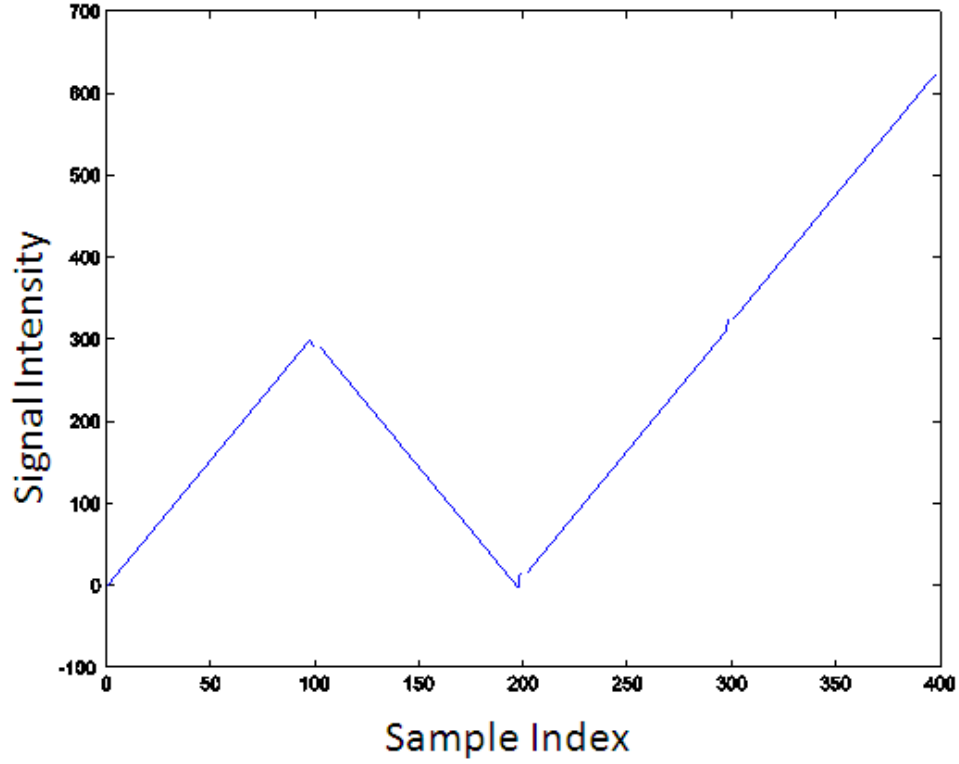


Figure 2.67: Original synthetic signal with piecewise linear parts.

If the periodicity is confirmed for $p'[n]$, a range of periods $T_1, T_2 \dots T_m$ are checked one by one to find the true period. Since doing this by exhaustive brute force search brings an enormous computation burden, PSO intelligence is used. For this purpose, for each T , the subsequent sub-arrays of length of T are averaged. Candidate of one period is obtained as shown below:

$$t_T[n] = \frac{\sum_{j=1}^K f[(j-1).T + n]}{K} \quad \text{if } 1 \leq n \leq T. \quad (2.81)$$

Since noise has typically zero mean, this averaging process removes the noise with great success. For each period segment estimated by using a period length T , the periods patches are pasted side by side to obtain the periodic estimated signal.

$$\tilde{p}_T[n] = \sum_{j=1}^K t_T[n - (j-1).T]. \quad (2.82)$$

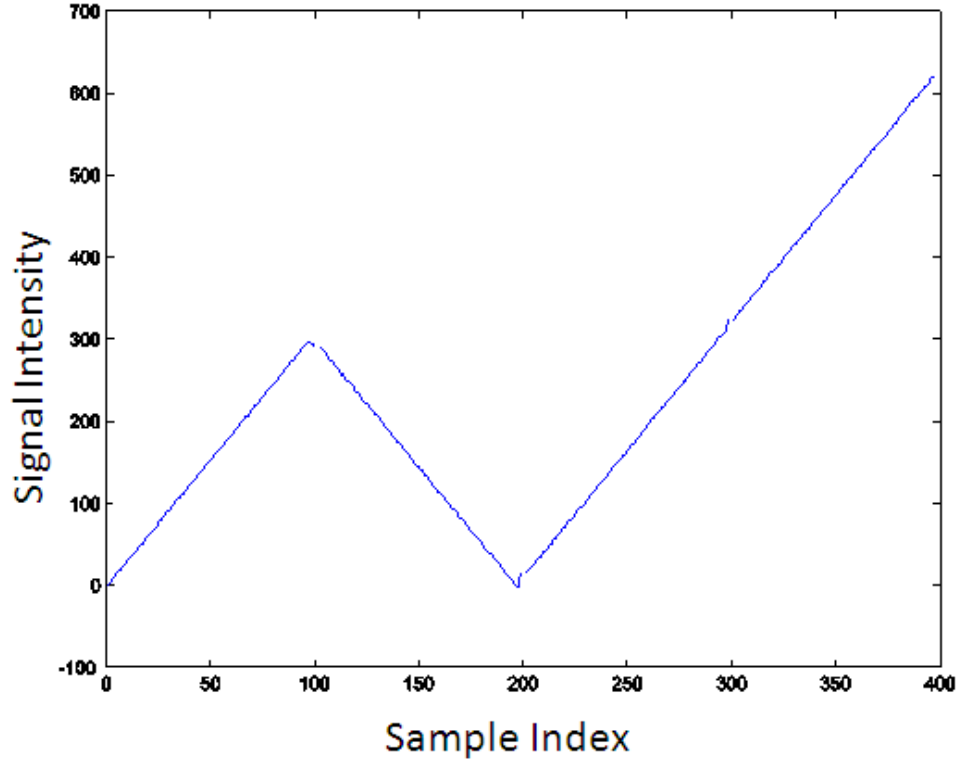


Figure 2.68: Noisy piecewise linear signal with an additive Gaussian noise with a standard deviation of 1.

The cost incurred for this periodic signal obtained is:

$$T_f = \underset{T}{\operatorname{argmin}} \{ \|p'[n] - \tilde{p}_T[n]\| \}. \quad (2.83)$$

By this way, the period with the least error energy is assigned as the true period T_f .

If an edge starts from $n > L$, it is processed distinctly, because it can not be the continuation of the previous periodic structure. If the same average level of the vector goes on, the same process is followed as described above. If a periodic structure is detected in the neighboring window, it is compared with the periodic structure of the previous window. This is achieved by cross-correlation technique. If the periods are the same, the cross-correlation of the corresponding periods have high variation attaining local minima and maxima successively. This means

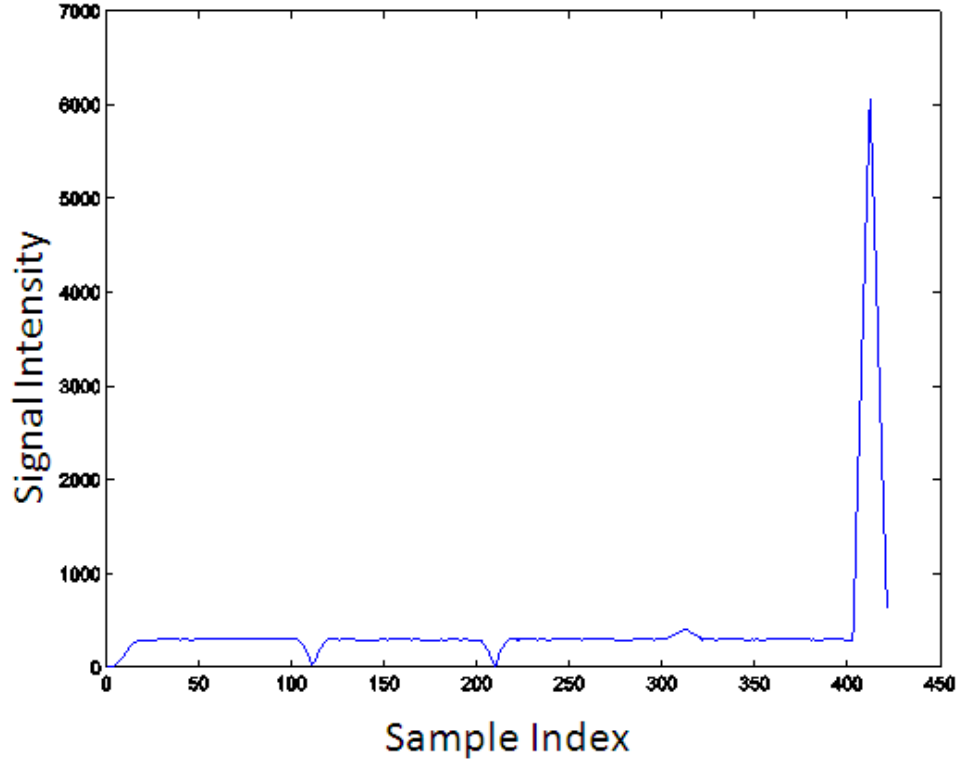


Figure 2.69: Output of convolution at the first stage.

that total variation of the cross-correlation rate is large if the periods are the same. In the j^{th} window this can be expressed as:

$$\frac{TV\{C\{t_f^{j-1}[n], t_f^j[n]\}\}}{N} < \varepsilon, \quad (2.84)$$

where C is a cross-correlation system and N is the length of cross-correlation sequence. For two signals $f[n]$ and $g[n]$, the cross correlation estimation is performed as:

$$(f \star g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m] g[n + m]. \quad (2.85)$$

If the periods are found to be the same, two windows, w_j and w_{j-1} are combined and the averaging process is repeated to increase success. Success increases, because the more periods are averaged, the better noise is suppressed.

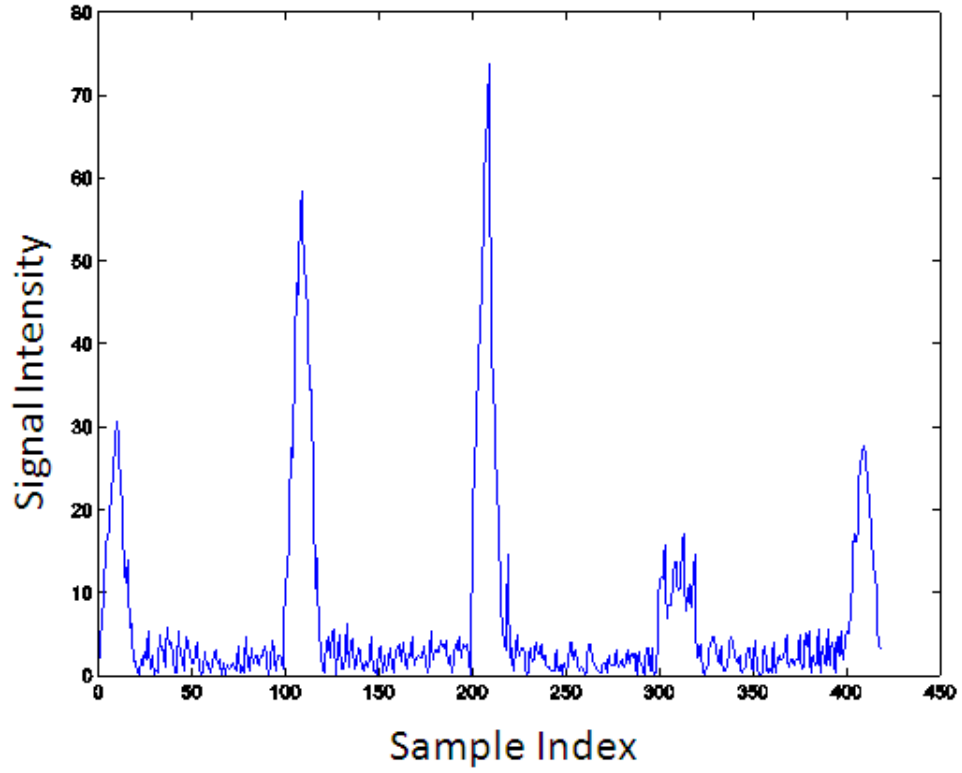


Figure 2.70: Output of convolution at the second stage.

If the $(j + 1)^{th}$ window is found to have the same period, then the window is enlarged again. The enlargement goes on until an edge, a window that contains a different periodic structure, a non-periodic structure, or the end of the one dimensional signal is reached.

For the illustrated example signal $x[n]$ given above, there will not be any enlargement of the window at the second iteration since the periods are not the same. Procedure of the second iteration for $r[n]$ will be similar to the first one. At the third iteration, the windowed signal $s[n]$ will not be periodic, hence total variation will turn back with a smaller optimal coefficient to keep the details.

It will be useful to summarize all the above steps in a flow diagram. The flow diagram is as in Figure 2.73.

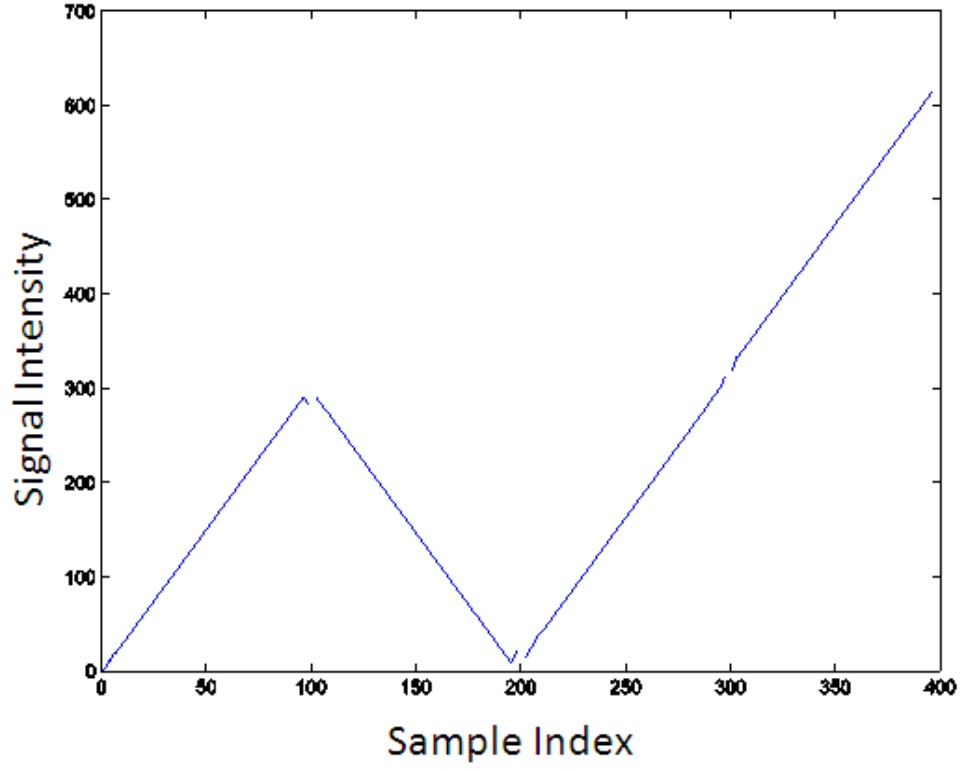


Figure 2.71: Piecewise linear signal denoised by the proposed technique.

2.10.2 Denoising Process

At first, a textured signal was generated as shown in Figure 2.74.

Gaussian noise with a standard deviation of 1 was added on the one dimensional signal. The signal is as in Figure 2.75.

Total variation process was followed and three different steps were observed. This smooth graph was used for the purpose of edge detection as well. Edge detection is critical to separate the vector into different parts to be treated differently. The output is as shown in Figure 2.76.

This last vector was subtracted from the first one in order to obtain the residual. This residual is as shown in Figure 2.76.

At this point, the period was assumed to be known, and one period of periodic structure was found by iteratively summing up all the periods. This yields

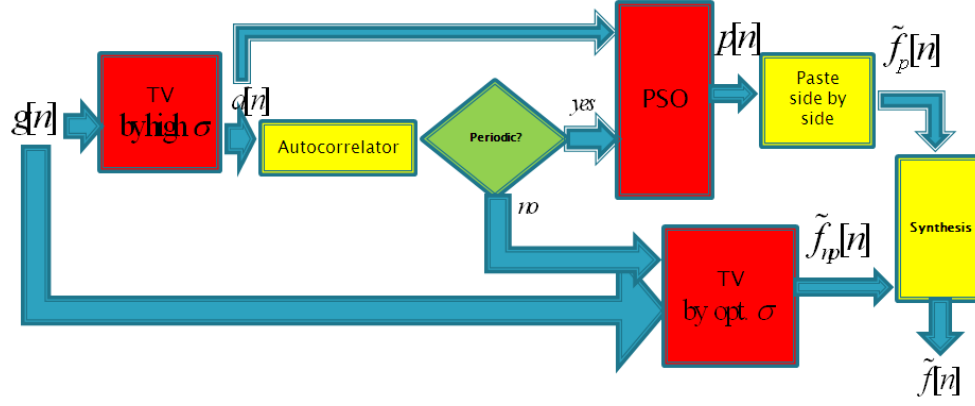


Figure 2.72: Flow diagram of the proposed Patch Averaging based Texture Recovery Technique.

acceptable result, because the noise is zero-mean which means that the average of all the periods converge to the original period. The enlarged graphs of the recovered three periods are as shown in Figure 2.78.

These three periods were iteratively pasted side-by-side to the smooth graph to recover these periodic structures. The result is as shown in Figure 2.81

Let us examine the situation that the signal has both a periodic part and a non-periodic step with a similar amplitude as shown in Figure 2.82.

All the methods above were based on the assumption that the period length was known. However, it is possible to find the period length in an automated manner. A period range was given and the averaging process was followed for each estimated period, then the residual for each process was taken and the energy of these residuals were calculated and the estimated period with the least residual energy was taken as the true period, since the residual is the noise. The period was found by this technique and the process above, averaging and pasting the periods was followed for the periodic part and routine total variation process was followed for the non-periodic part. In order to differentiate between the periodic and non-periodic parts, total variation with a large estimated noise variance coefficient was followed as a first step and two steps were obtained, so-called the non-periodic step and the periodic step as shown in Figure 2.84.

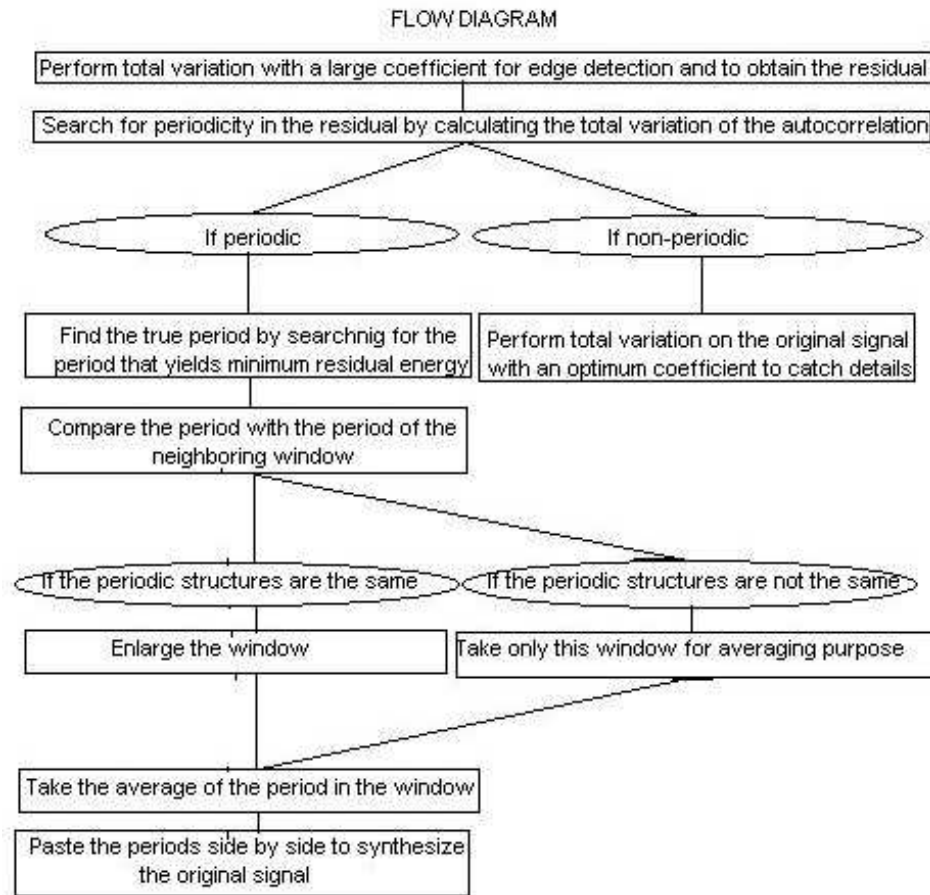


Figure 2.73: Flow diagram of texture recovery by adaptive windowing.

The residual of the two parts were taken and their autocorrelation was found for the purpose of detection of periodicity. The absolute value of the autocorrelation of the first part and second parts are as shown in Figure 2.85.

As one can observe from the plots in Figure 2.85 and 2.86 respectively, the periodicity results in impulses in the autocorrelation which means rapid changes. However, autocorrelation of the non-periodic residual is a monotonously decreasing function whose total variation is small. This difference of total variation of the autocorrelation function was used for periodicity checking. The part with a total variation of autocorrelation above a threshold was decided to be periodic and vice versa. In the periodic part, the process introduced above was followed

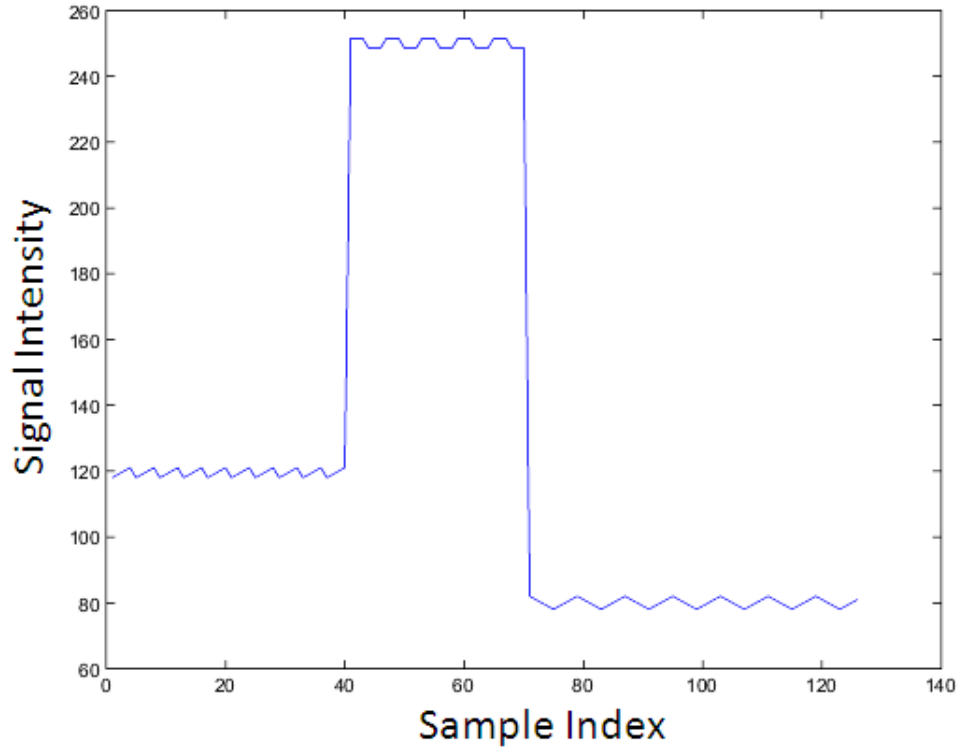


Figure 2.74: The synthetic signal with three different texture steps.

and in the non-periodic part, routine total variation with a smaller coefficient was followed and the step was successfully kept. The output of this process is as in Figure 2.87.

The efficiency of the texture recovery by the method of iterative pasting increases if the window that it is performed is enlarged adaptively. This method was tried and the process gave outstanding outputs. The synthetic signal in Figure 2.88 was used. Gaussian noise was added and the vector in Figure 2.89 was obtained.

To denoise this signal, frames of length 20 were used. In these windows, at first periodicity was checked by computing the total variation of the autocorrelation of the window. If no periods are observed, a total variation of a small coefficient was done and the step at the 170th pixel was saved. If periodicity is detected, an index is incremented and no else performance was done. In a neighboring window, if again periodicity is checked, it is investigated if these windows both

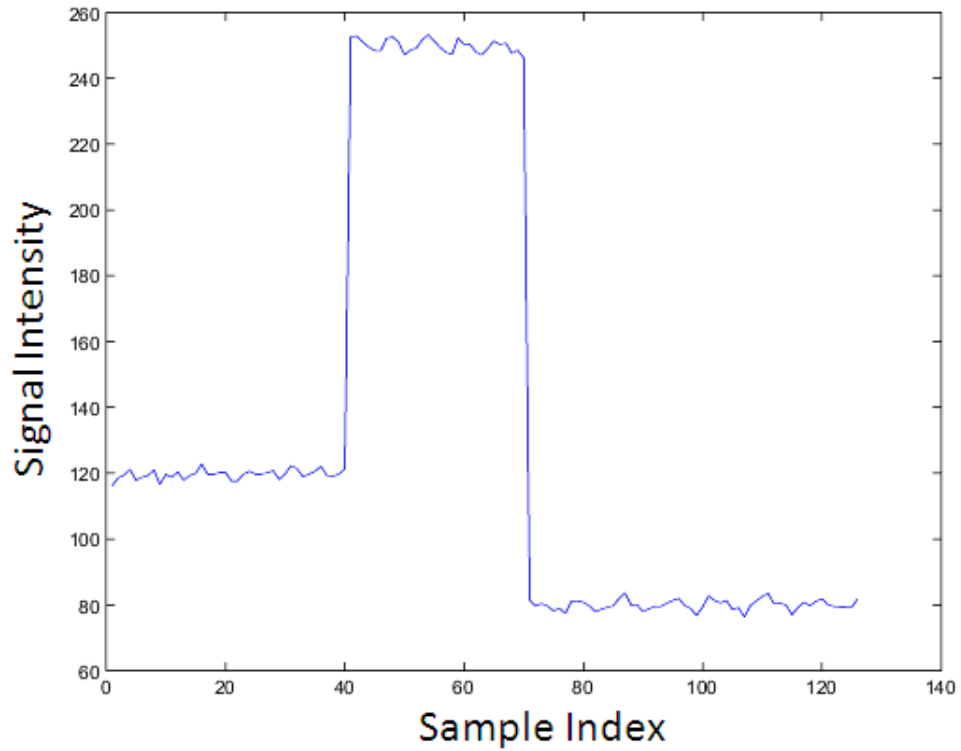


Figure 2.75: Noisy signal with three different texture steps.

have the same periods. This is achieved by the method of cross-correlation. If the total variation of the cross-correlation of the period of the previous window with the new period is above a threshold value, (since it should contain impulses). The two windows are assumed to have the same periods, so the window is enlarged, and the method of iterative pasting is performed in this large window. The window is enlarged until a window with a new period or without finding the periodicity. The output of this adaptive windowing process is as in Figure 2.90.

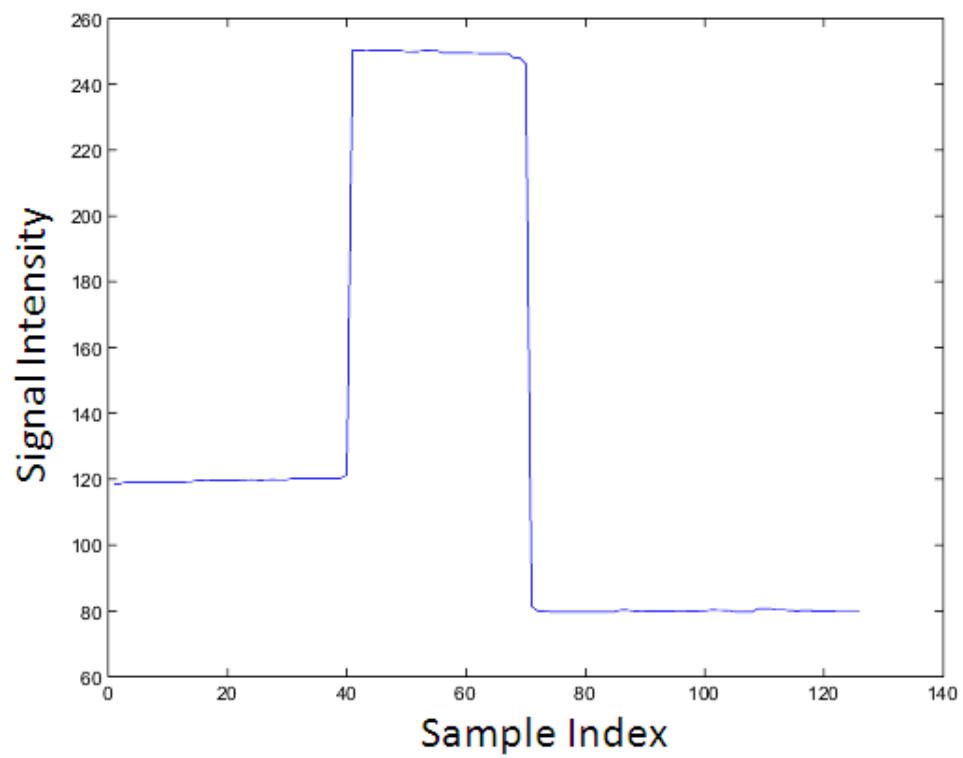


Figure 2.76: The signal with three smoothed steps.

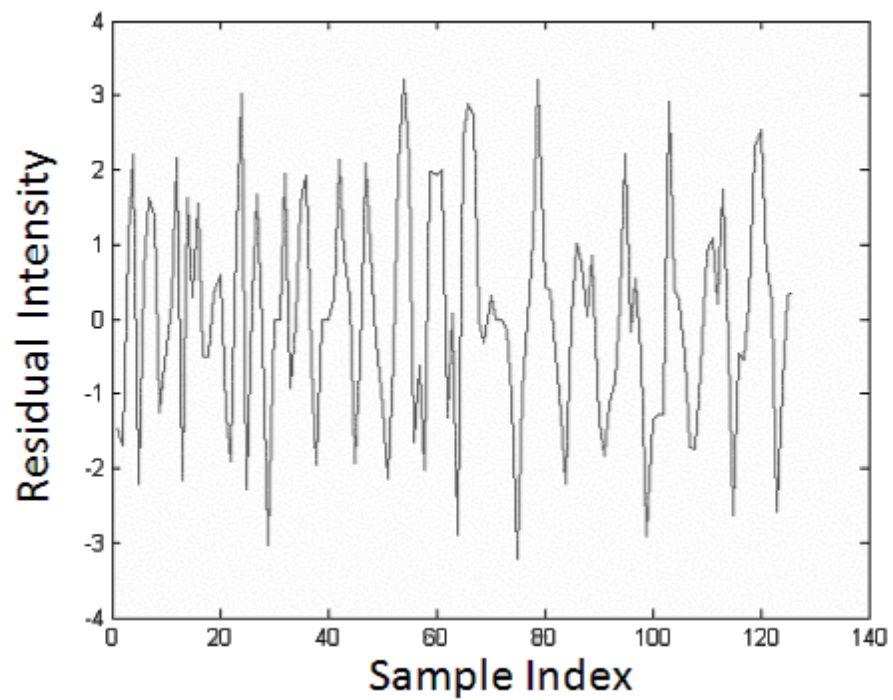


Figure 2.77: The residual of the signal with three different texture steps.

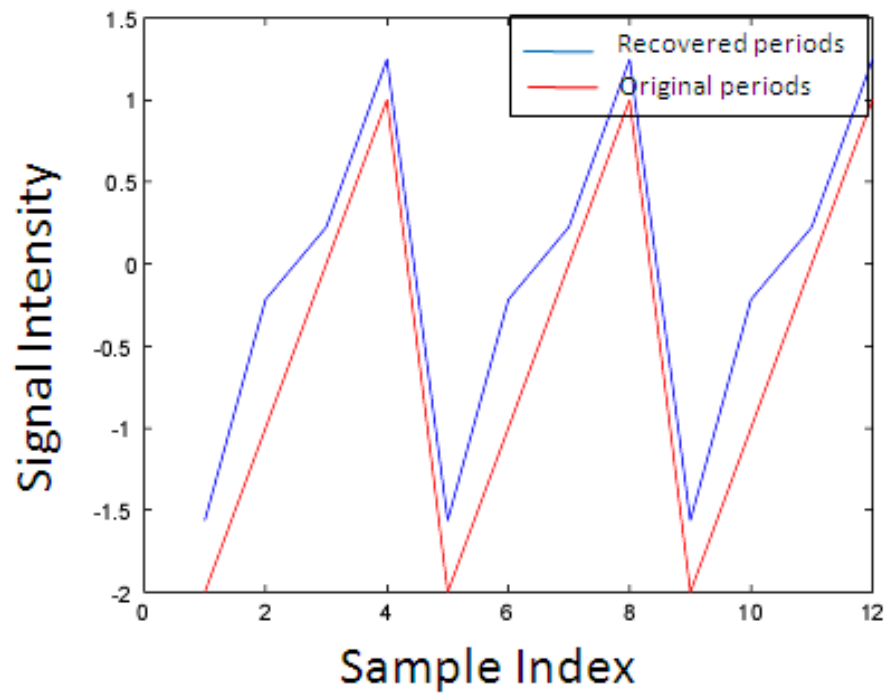


Figure 2.78: Enlarged graph of the first periodic texture. Three periods are shown for convenience.

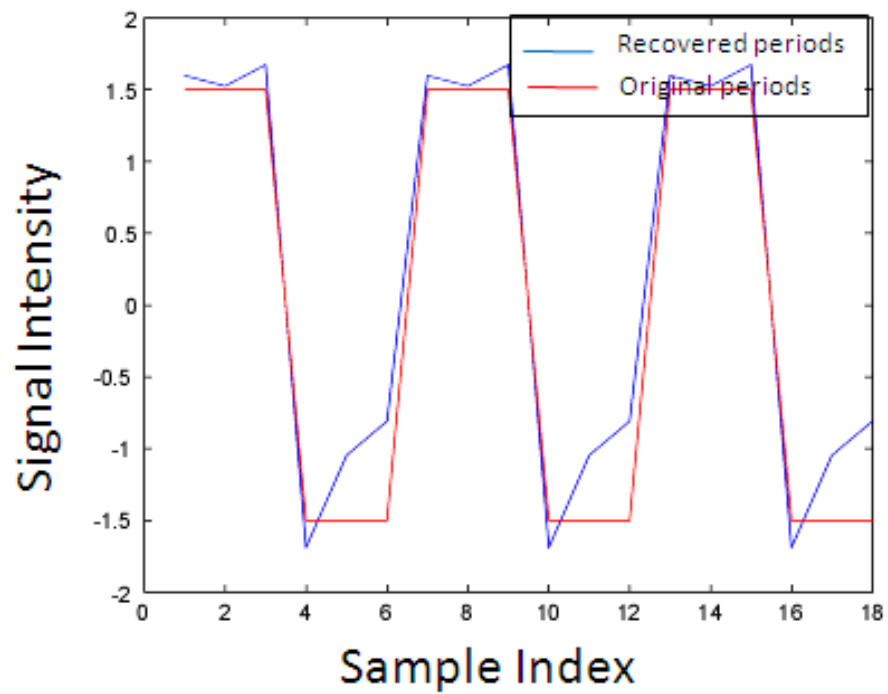


Figure 2.79: Enlarged graph of the second periodic texture. Three periods are shown for convenience.

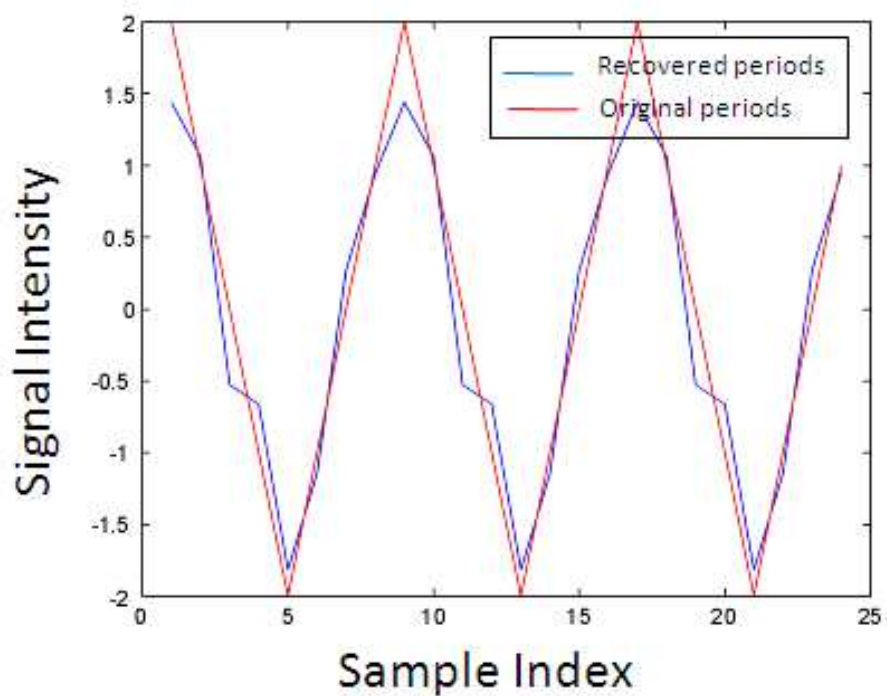


Figure 2.80: Enlarged graph of the third periodic texture. Three periods are shown for convenience.

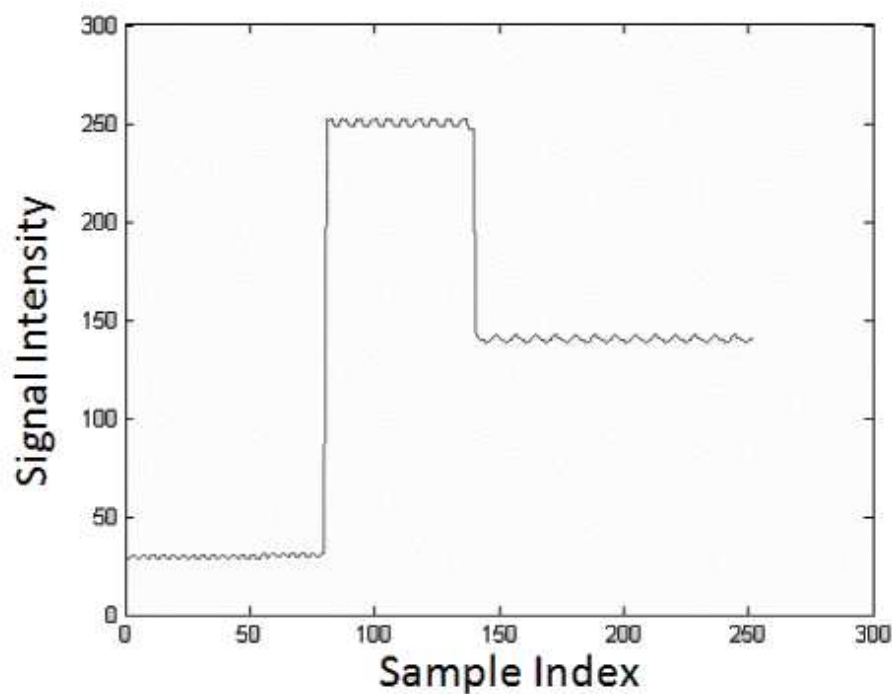


Figure 2.81: Signal recovered by the method of patch averaging.

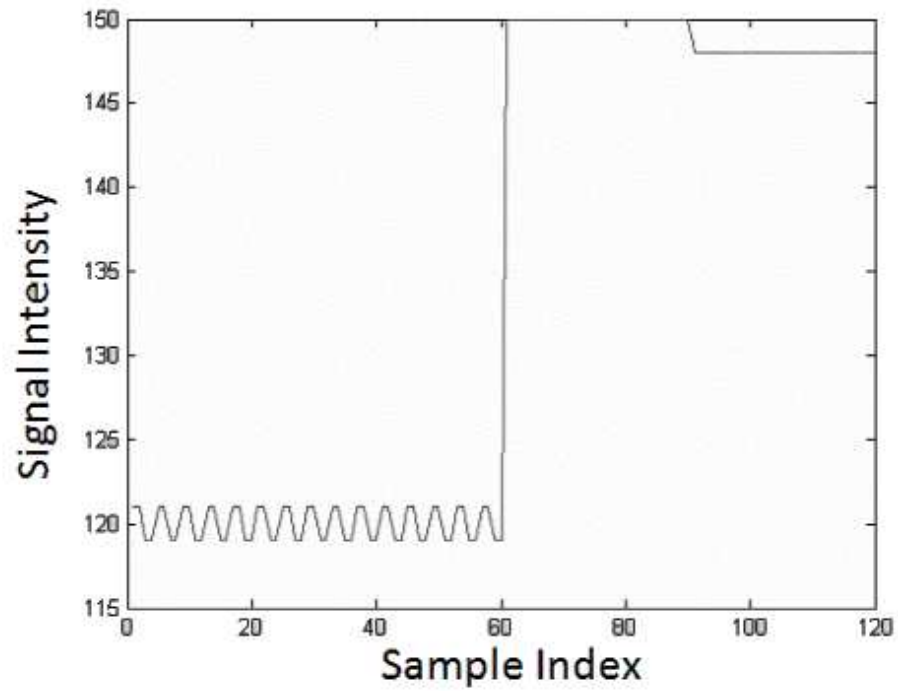


Figure 2.82: A syntetic signal with a periodic and a non-periodic step.

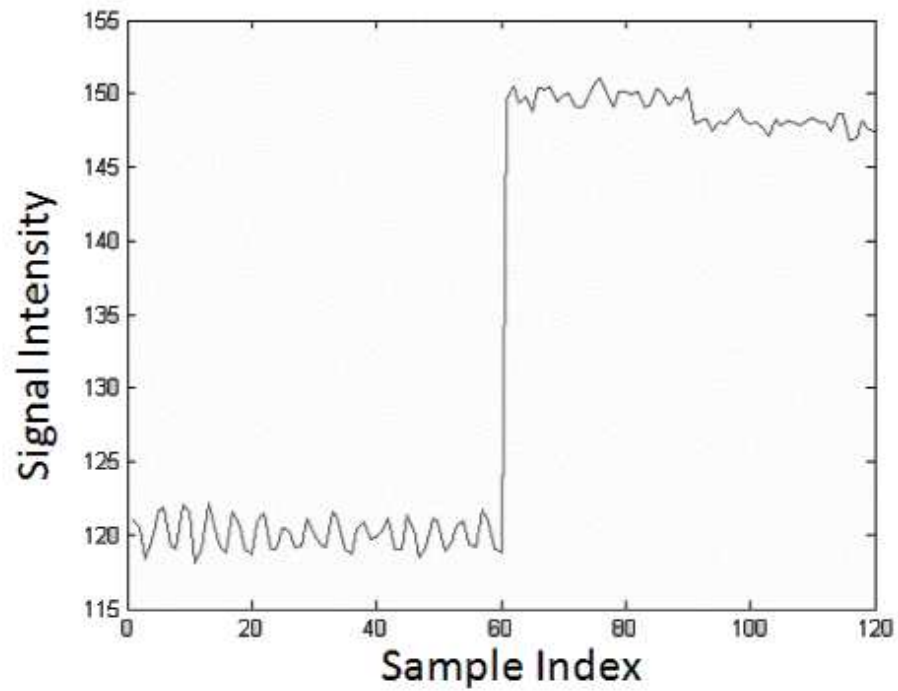


Figure 2.83: The noisy version of the signal with a periodic and a non-periodic step.

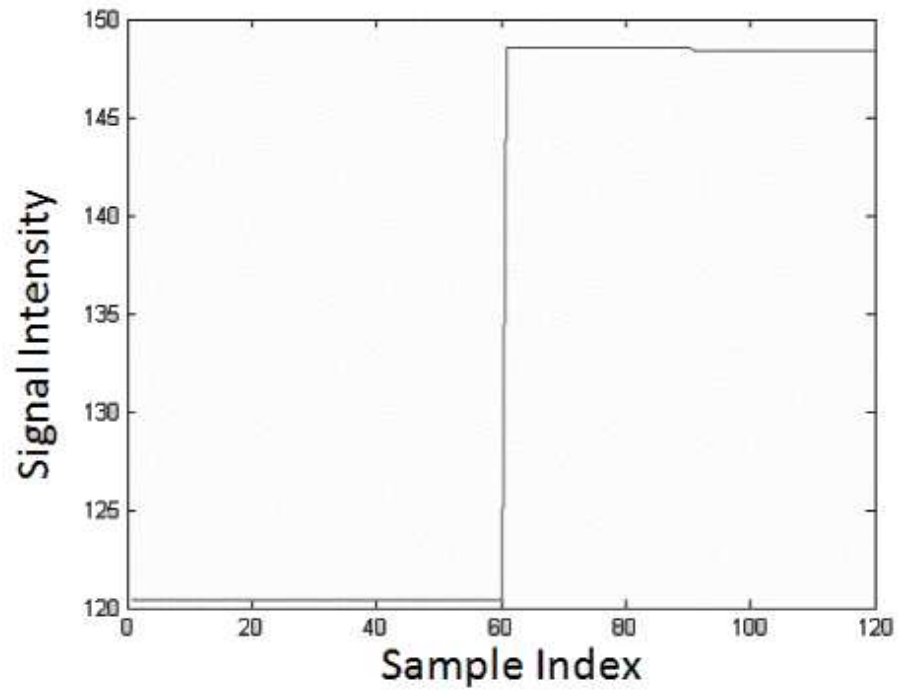


Figure 2.84: The smoothed version of the signal with a periodic and a non-periodic step.

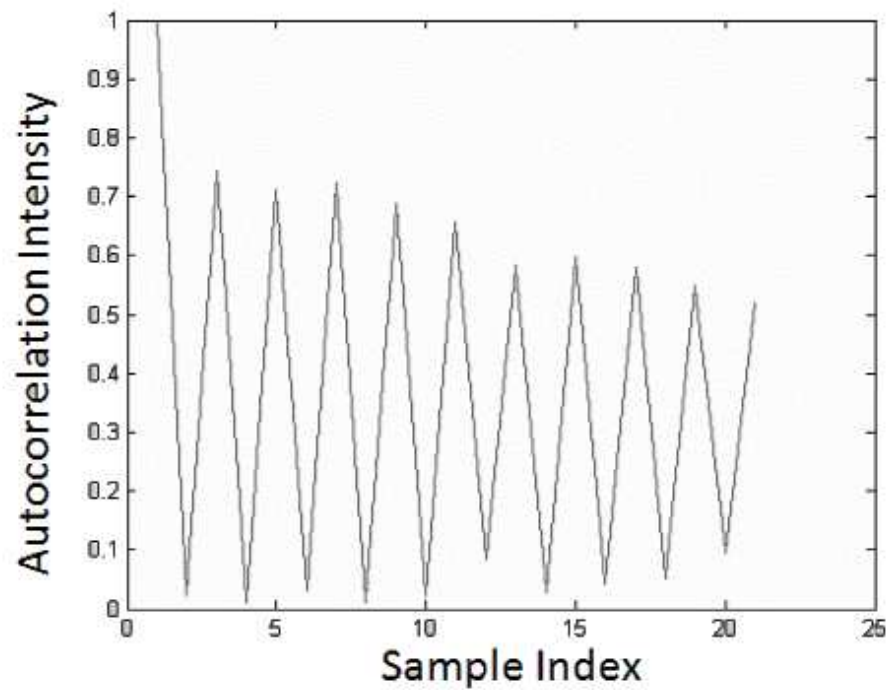


Figure 2.85: Autocorrelation of the first (periodic) part.

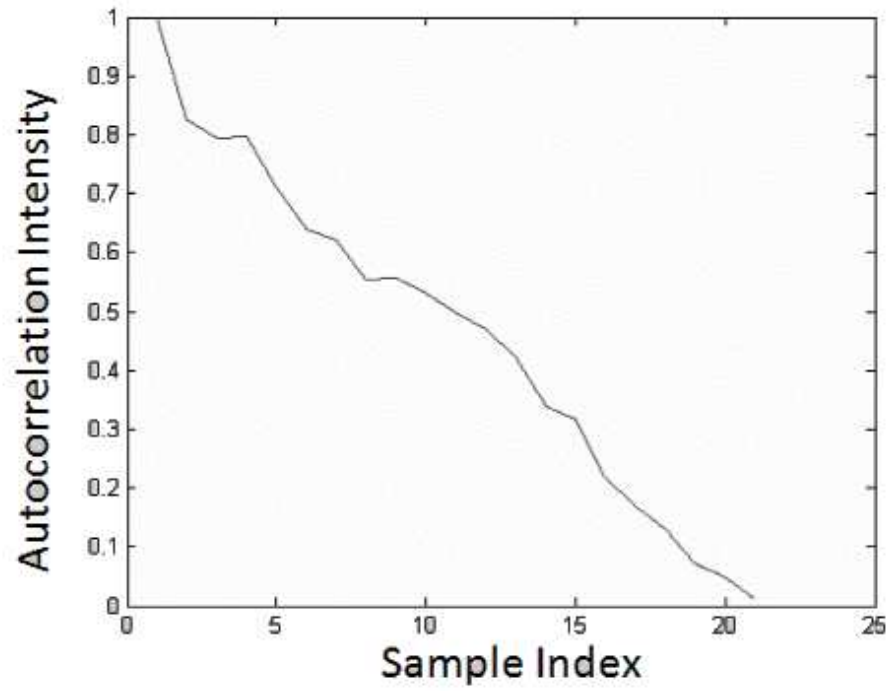


Figure 2.86: Autocorrelation of the second (non-periodic) part.

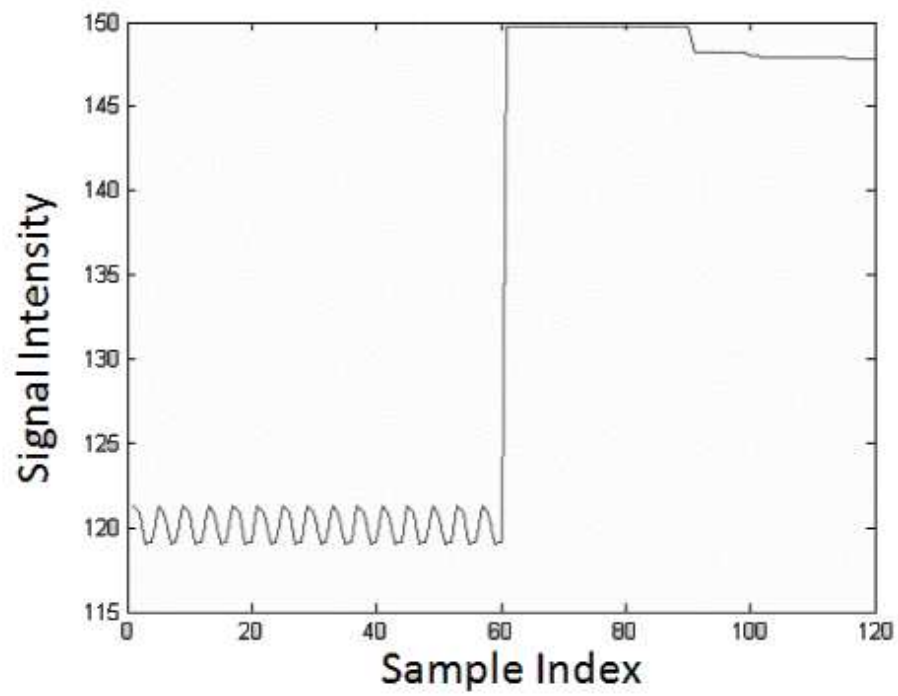


Figure 2.87: The signal recovered by patch averaging.

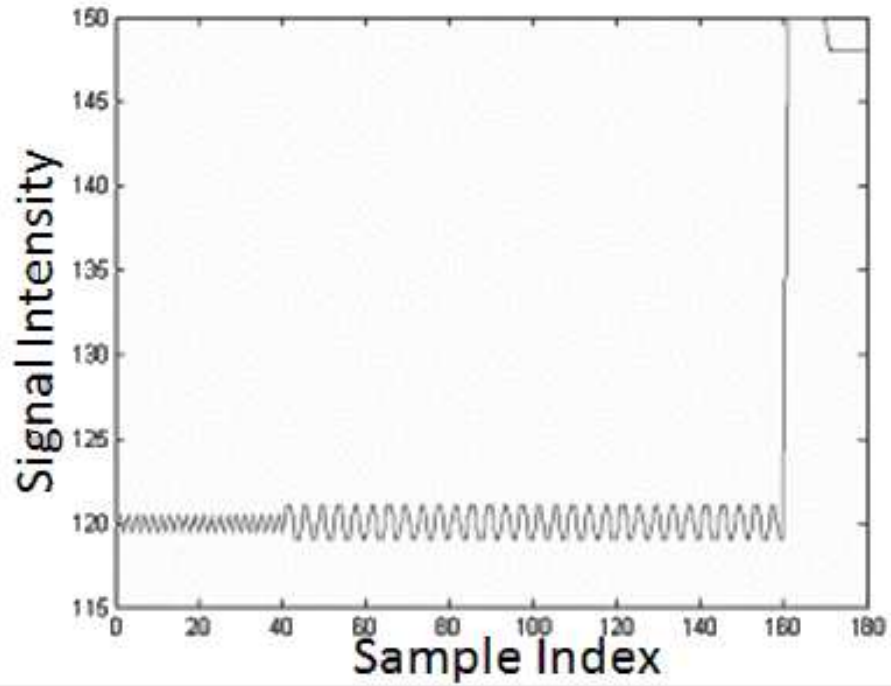


Figure 2.88: A synthetic signal consisting of four various parts.

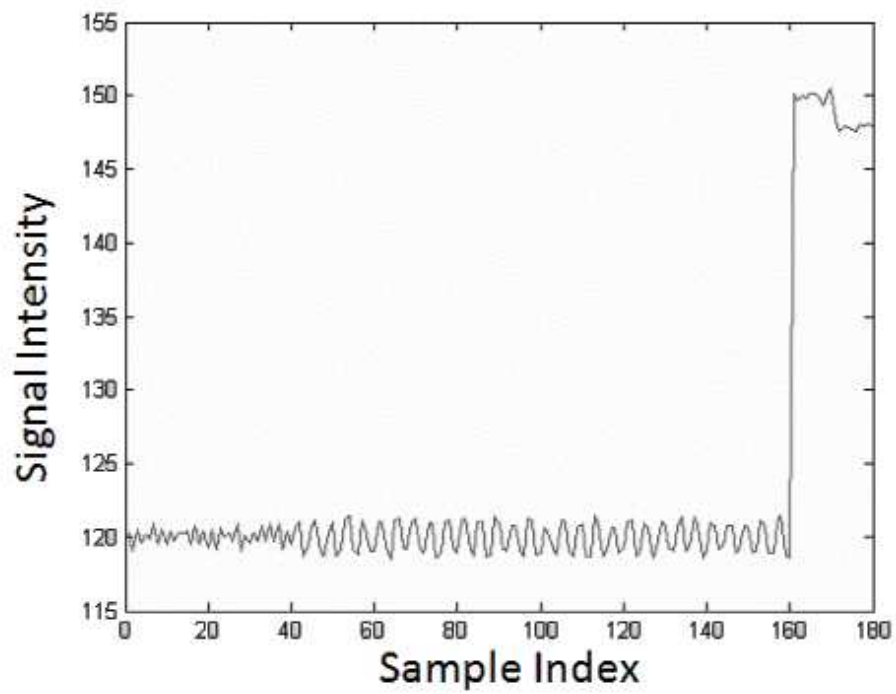


Figure 2.89: Noisy version of the signal consisting of four parts.

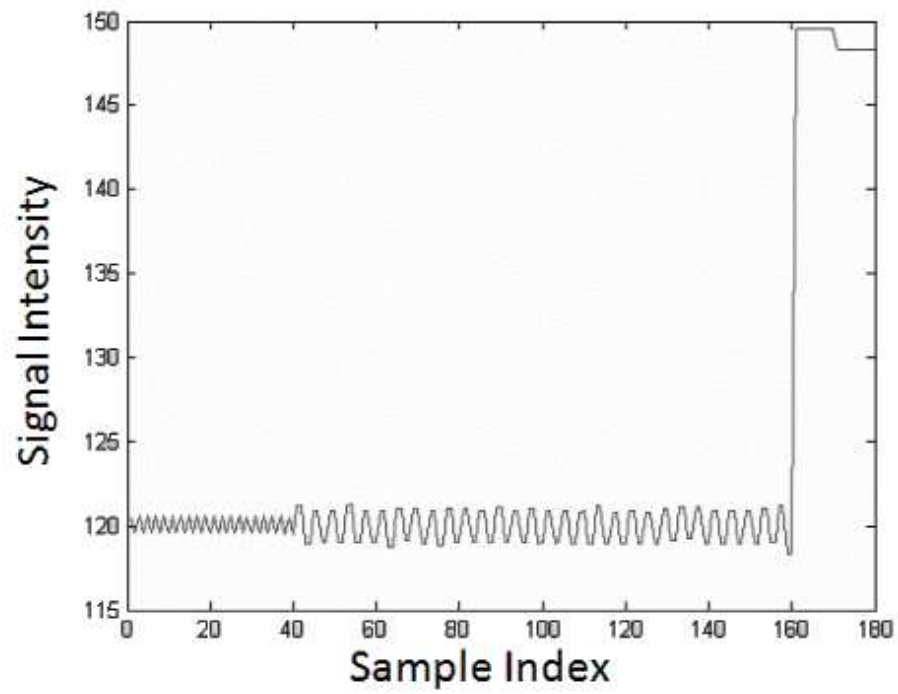


Figure 2.90: The signal recovered by patch averaging.

Chapter 3

IMAGE DENOISING

Image denoising has various applications ranging from remote sensing imaging, medical imaging to graphics design. It serves for both visual quality and a precondition for image analysis. This crucial area of signal processing is a challenging active research area. It is challenging because the processing that can be done in one dimension relatively easier such as edge detection, polynomial fitting can not be accomplished in two dimensions in a straight forward manner. Moreover, the optimality of matched filtering and total variation can not be carried to two dimensions. Applying globally optimal techniques such as dynamical programming to two dimensions is computation-demanding. In this chapter, carrying total variation which is optimal in one dimension to two dimensions by intelligent methods will be discussed. Straight forward but promising Slice TV and Separable TV techniques and rotation independent R-TV technique are going to be introduced in this chapter.

3.1 2 – D Noise Standard Deviation Estimation

As denoted in the one dimensional estimation, the estimated σ is used as a stopping criterion for smoothing processes of images. Images actually have more correlated structures than a 1 – D signal. Moreover, a sufficiently large image is extremely likely to contain piecewise constant blocks. In [55] this fact is exploited. Another approach uses the noise model to estimate the noise statistic [56]. If the noise is Gaussian, the amplitude of the noise histogram is a Gaussian curve. The standard deviation of this Gaussian curve is the final estimate.

In the technique proposed in this thesis, the first difference of each column is median-filtered with a window of length 7. Then the standard deviation of output of each row divided by $\sqrt{2}$ is taken and these values for all rows are averaged. The mathematical formulas are like this:

$$f_m[n] = F[m, n], \quad (3.1)$$

$$g_m[n] = f_m[n + 1] - f_m[n], \quad (3.2)$$

$$h_m[n] = \mu\{g_m[n]\}, \quad (3.3)$$

$$\tilde{\sigma}_m = \frac{\sigma_{h_m[n]}}{\sqrt{2}}, \quad (3.4)$$

$$\hat{\sigma} = \frac{1}{M} \sum_{m=1}^M \tilde{\sigma}_m. \quad (3.5)$$

The algorithm proposed was applied on a synthetic noisy image shown in Figure 3.2. The noise σ is 18 here. It is estimated as 18.4273 by the proposed technique.

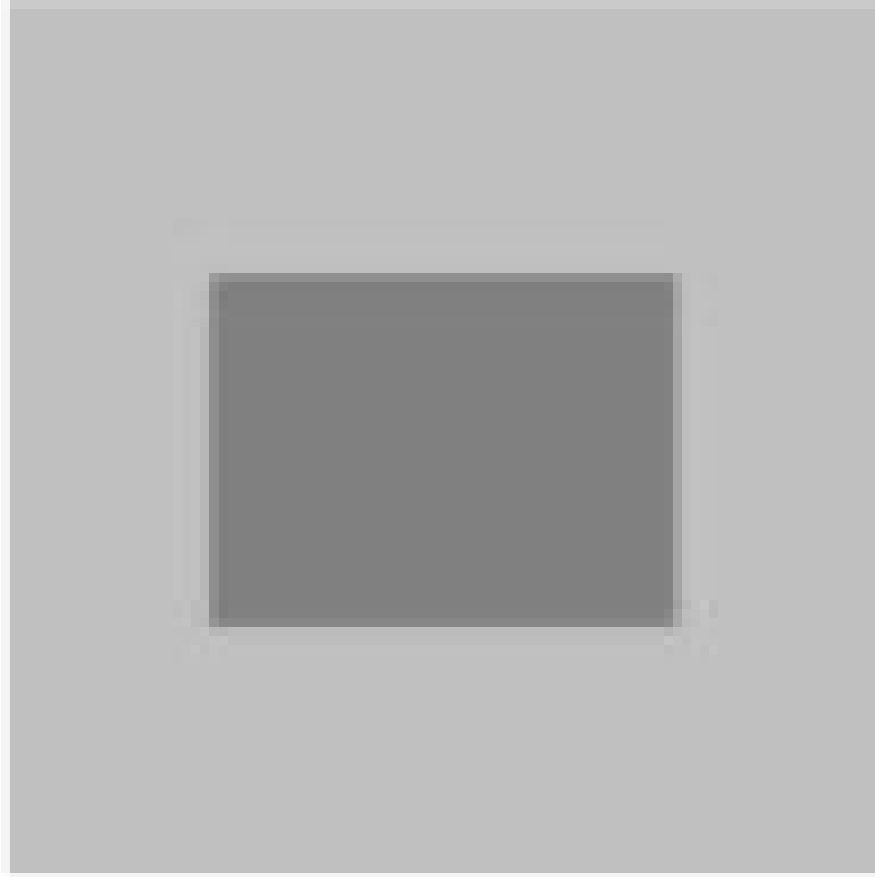


Figure 3.1: Original synthetic image.

3.2 Nonlinear Total Variation Based Noise Removal Algorithms

Total variation based algorithms are widely used in signal denoising. The aspects that make them advantageous are their edge-preserving characteristic and optimal solutions they provide due to different problem definitions. The total variation method proposed by Rudin, *et. al.* [35] was revolutionary for signal

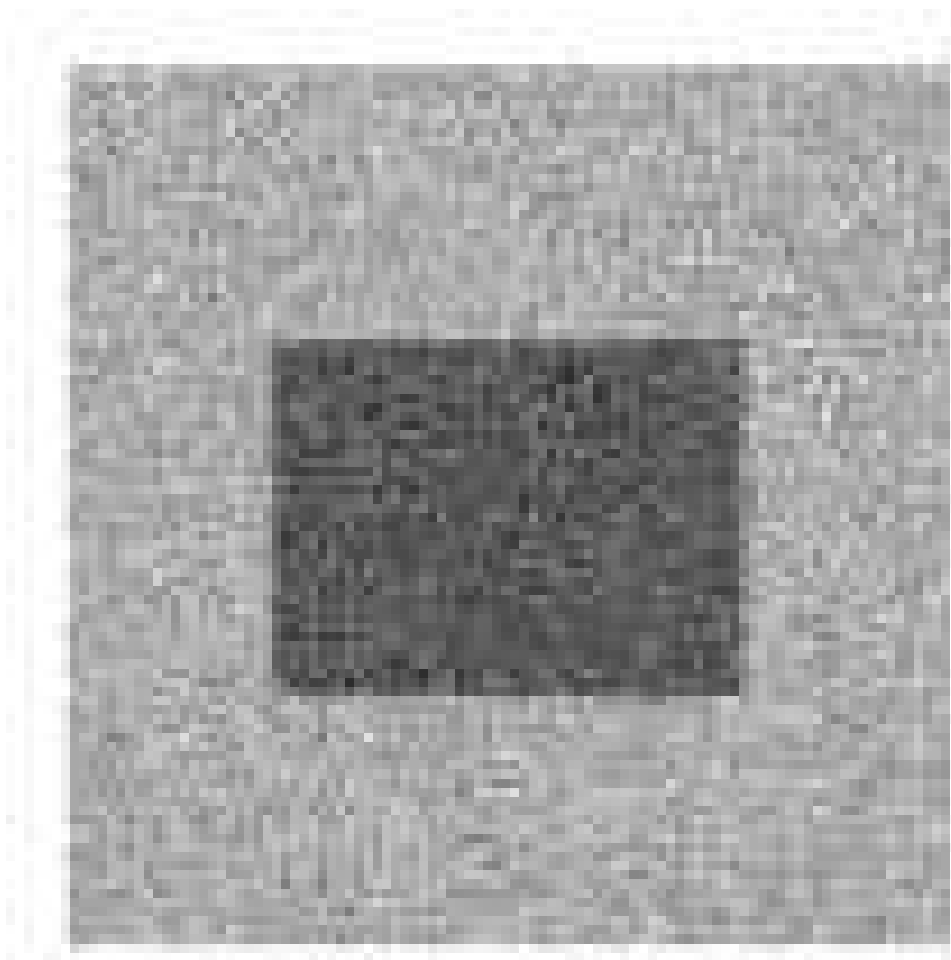


Figure 3.2: Noisy synthetic image whose noise σ is estimated.

denoising community. Its natural constraints due to noise statistics and meaningful cost of total variation of the signal produced impressive results. Reducing the total variation of the signal makes the signal smoother step by step and knowing the noise standard deviation provides a means to stop the smoothing process. Actually this optimization problem was solved analytically by Euler-Lagrange equations which solves the minimization using Lagrange multipliers. This method is well-known as gradient-projection method. A numeric algorithm based on this analytic solution was presented.

3.2.1 Formulation

The constraints of the optimization problem is the result of zero mean noise and a constant noise standard deviation for all samples. These are imposed as:

$$\int \int F_T(x, y) dx dy = \int \int F_0(x, y) dx dy, \quad (3.6)$$

$$\int \int \frac{1}{2} (F_T(x, y) - F_0(x, y))^2 dx dy = \sigma^2. \quad (3.7)$$

Here, F_T refers to the final image, F_0 corresponds to the initial image, and also the time during the process will be denoted as t .

To solve this problem, the constraints are imposed by adding the minimization cost with a Lagrange multiplier. This functional is solved by gradient-descent method.

$$F_t(x, y) = \frac{\partial}{\partial x} \left(\frac{F_{tx}(x, y)}{|\nabla F_t(x, y)|} \right) + \frac{\partial}{\partial y} \left(\frac{F_{ty}(x, y)}{|\nabla F_t(x, y)|} \right) - \lambda_t (F_t(x, y) - F_0(x, y)). \quad (3.8)$$

The λ_t above evolves in time due to gradient projection method [57]:

$$\lambda_t = \frac{-1}{2\sigma^2} \int \int \left[|\nabla F_t(x, y)| - \left(\frac{F_{0x}(x, y)F_{tx}(x, y)}{|\nabla F_t(x, y)|} + \frac{F_{0y}(x, y)F_{ty}(x, y)}{|\nabla F_t(x, y)|} \right) \right] dx dy. \quad (3.9)$$

The numeric algorithm relies on the discretisation of the evolution equations and is omitted here.

3.3 Curvelet Transform Based Denoising

Curvelet transform is extensively studied in recent years. The application of curvelet transform on denoising [3] was a novel approach. Curvelet based denoising outperformed FIR based smoothing techniques. Because, crude averaging provided by FIR filtering does not keep the edges sharp. Also, block-wise averaging or running average degrades the resolution as well. Instead of coarse averaging, wavelet transform keeps the features better.

The curvelet smoothing procedure consists of two steps. These steps are rectangular smoothing and polar smoothing. This combination yields more natural results free of artifacts. For rectangular smoothing, $2 - D$ wavelet transform is exploited. After this step, the polar smoothing is achieved in Radon projection domain. Radon transform takes the projection of the image to a vector at each polar angle. The projections are smoothed by one dimensional wavelet transform. The initial dyadic squares of smoothing blocks are enlarged step by step and over-smoothing is avoided by stopping when the error norm is equal to the noise standard deviation.

3.4 Comparison of Viable Techniques

3.4.1 Examples Concerning the Comparison of the Output Images of Five Different Algorithms

The five algorithms are curvelet by USFFT, curvelet by wrapping, enhanced curvelet, [3] total variation, total variation enhanced [35]. Images are of dimension 128×128 pixels.

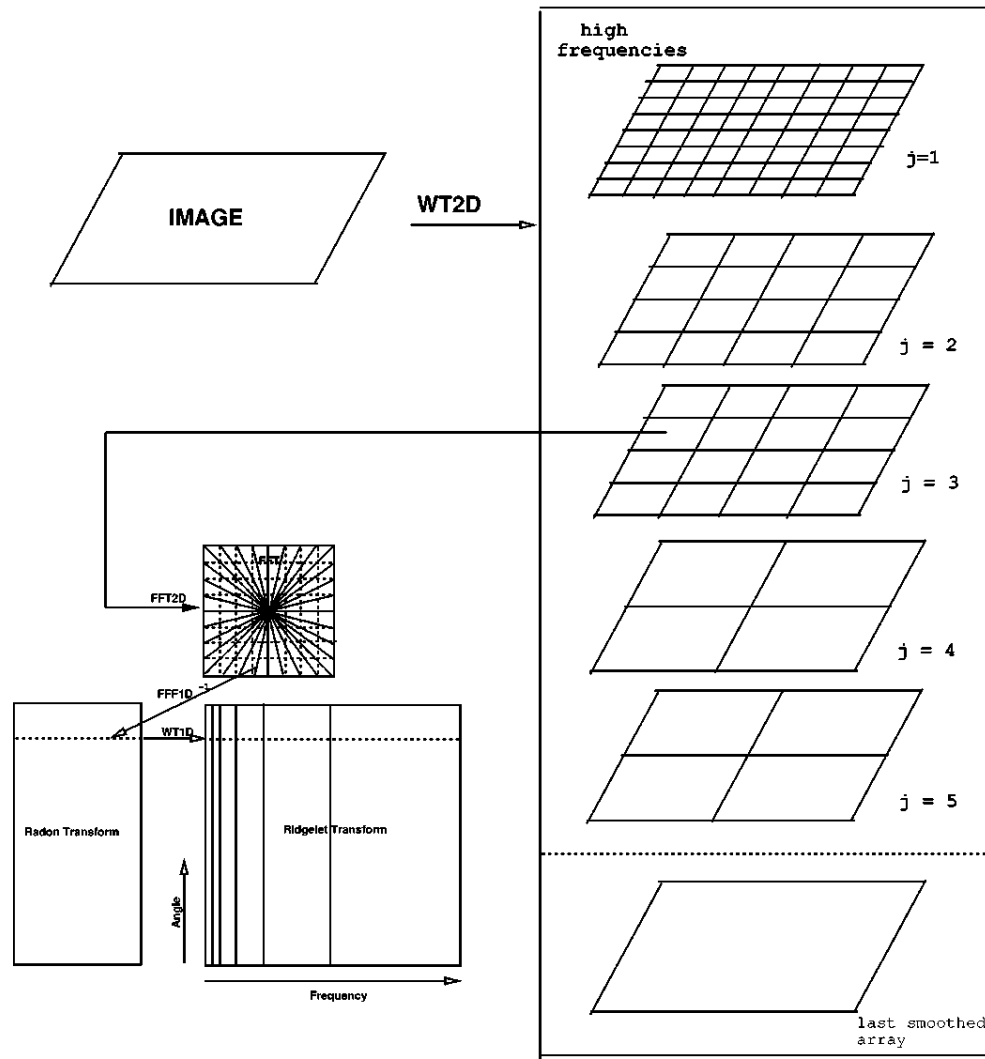


Figure 3.3: Curvelet transform building blocks for image denoising [3].

3.4.2 Comparison of Curvelet and other benchmark denoising techniques

The benchmark techniques to be compared are various implementations of Total Variation by Rudin *et al.* and curvelets. To sum up, curvelet output has some artifacts such as blackening white spaces and forming speckle noise. The advantage of curvelet decomposition based denoising is its rotation-independence. On the other hand, although total variation is edge preserving and free of artifacts, its enhanced implementation which regularizes the output is slower than the

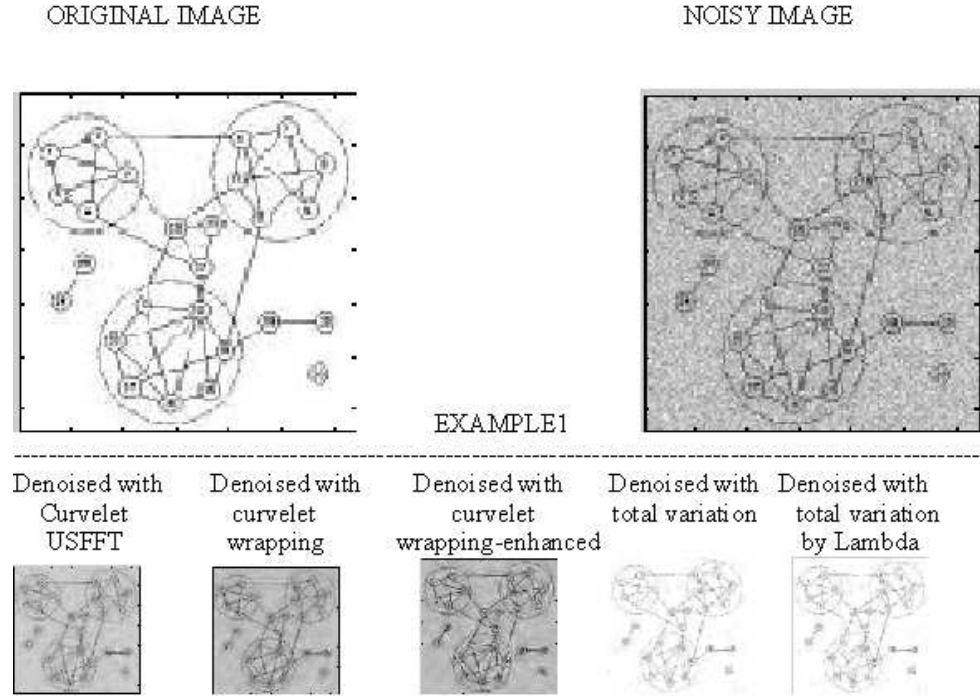


Figure 3.4: The denoised outputs of five different algorithms for a synthetic image.

curvelet based counterparts. The conclusions summarized in 3.1 can be drawn out of those simulations.

3.5 $2-D$ Denoising by Averaging Row-Wise and Column-Wise $1-D$ Total Variation Outputs

There has been many exciting developments about total variation based denoising techniques recently. However, there are still problems of optimization and efficiency on performing total variation reduction in two dimensions. This need has arisen, because total variation minimization can be applied optimally in one dimension only. This thesis defines a method that carries nonlinear total variation reduction to two dimensions. This is accomplished by averaging row-wise

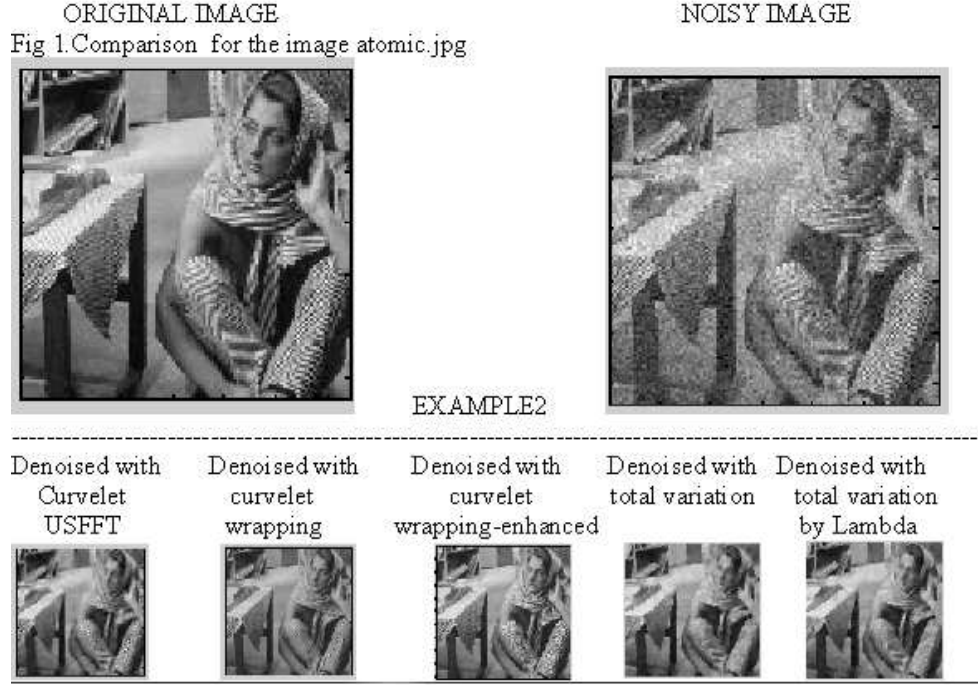


Figure 3.5: The denoised outputs of five different algorithms for the classical Barbara image.

and column-wise $1 - D$ total variation outputs. This approach has given promising results. Methods such as curvelets also gives outstanding results. However, noise estimation result can be used more efficiently in the technique that we propose, since it is relatively more challenging to make use of estimated noise in radon transform domain. Moreover, total variation based denoising provides outputs free of speckle noise. Texture is kept better, whereas noise is better removed. All the advantages of one dimensional total variation minimization are benefited in this method. The problem of discontinuities between row-wise and column-wise outputs are resolved by crude averaging these outputs. This averaging drops the need to use gradients in two dimensional implementation. Theoretically, the output is rotation dependent, however, this is not visible in the outputs. Our method estimates the noise variance on the image by sliding-mean technique and uses this data in the denoising process. Success increases if the noise variance is known *a priori*. The technique is illustrated in Figure 3.6 as a flow diagram. Simulations have given promising results.

CURVELET USFFT ALGORITHM	CURVELET WRAPPING ALGORITHM	ENHANCED WRAPPING ALGORITHM	TOTAL VARIATION ALGORITHM	ENHANCED TOTAL VARIATION
Average speed	Very fast	Fast	Fast	Too slow
The least high pass effect, smoothed edges	Average high pass effect, sharper edges than usfft algorithm	The most high pass effect, sharp edges	No observable high-pass effect, good at smooth places, smoother edges than enhanced total variation	No observable high pass effect, but a bit low pass effect, good at smooth places, sharper edges than total variation
Poor performance at photos	Average performance at photos	Good at photos	average performance at photos	Average performance at photos

Table 3.1: The comparison and comments for five renowned denoising techniques.

3.5.1 Formulation and Algorithm

The denoising technique introduced here is based on the claim that if the $1 - D$ slices of an image are denoised, the noise of the image itself is also removed. It is possible to implement the denoising procedure on row-wise and column-wise slices. Since all the slices are denoised independently, there will be discontinuities between neighboring slices. This can be removed with by averaging these two outputs.

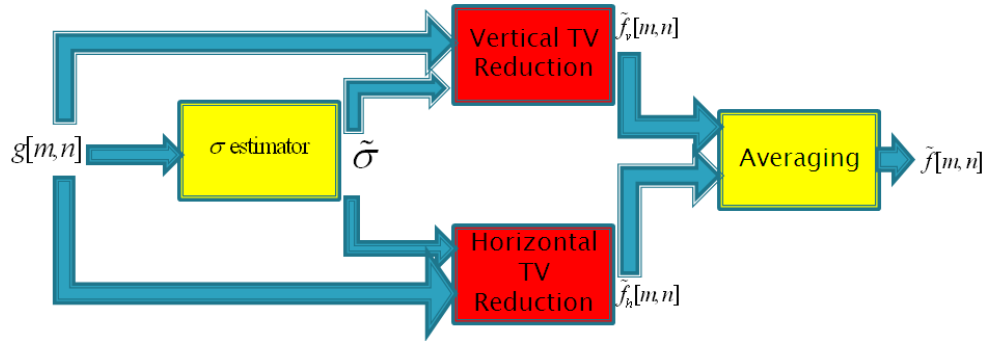


Figure 3.6: Flow diagram of the proposed Slice TV Technique.

In this denoising algorithm, all the slices are processed independently. The method used here to minimize the total variation is the algorithm proposed by Willsky, *et al.* which is explained in detail in the Chapter 2.4. The mathematical problem that this algorithm solves can be posed like this:

$$\min(TV(\mathbf{u})), \quad (3.10)$$

$$s.t. \|\mathbf{u} - \mathbf{u}_f\| = \sigma \sqrt{N}. \quad (3.11)$$

Here, σ is the standard deviation of the noise that is assumed to be known *a priori*, N is the number of samples and $\mathbf{u}_f$ is the output vector of the total variation reduction process. If the standard deviation of the noise is not known, you can apply the noise statistic estimation procedure explained in the Chapter 3.1 to a flat part of the image so that the source of variation is only noise.

Here, note that i is the slice index. In heterogeneous parts, this will yield too high estimation since it will detect the texture and details as noise. Hence, heterogeneousness detection is essential. For scale of heterogeneity, the C_{mat} matrix proposed in [58] is used. In this thesis, this matrix was used for the detection of heterogeneous parts. Denoising is achieved by averaging in sliding windows. Averaging effect is reduced for the parts that C_{mat} matrix takes a large value, and it is increased for the homogeneous parts that the C_{mat} matrix takes a small value. The average of this matrix is computed in windows of size M . Also the noise is estimated in these windows, as well. It is assumed that the regions at which C_{mat} matrix attains a value smaller than a threshold are homogeneous and the noise estimation in these regions are more reliable. These windows are detected and the average of the estimated noise is taken in standard deviation in these windows.

The algorithm proposed here has two steps to take. First one is the vertical total variation reduction.

$$F_v[m, \cdot] = TV_\sigma\{F[m, \cdot]\} \quad (3.12)$$

Here, σ is the estimated noise standard deviation which provides a criterion for total variation process to stop. Moreover, $F[m, n]$ is the noisy image. Next step is the horizontal total variation reduction on the initial noisy image which can be explained like this.

$$F_h[\cdot, n] = TV_\sigma\{F[\cdot, n]\} \quad (3.13)$$

The 1-D slices of these two parallel outputs are optimally denoised by total variation reduction. However, since these slices are processed independently, they may have discontinuities. One straight forward solution is the crude averaging of the two outputs which may lessen the discontinuities. This averaging process can be formulated as:

$$F_d[m, n] = \frac{F_v[m, n] + F_h[m, n]}{2} \quad (3.14)$$

Simulations have shown that discontinuities are removed significantly.

3.5.2 Comparison of Experiments

The techniques that are going to be compared in this section are total variation by Rudin *et al*, its enhanced version and our technique, that will be referred to as the Slice-TV.

It will be proper to start the comparison by a rather simple image that consists of only a rectangle.

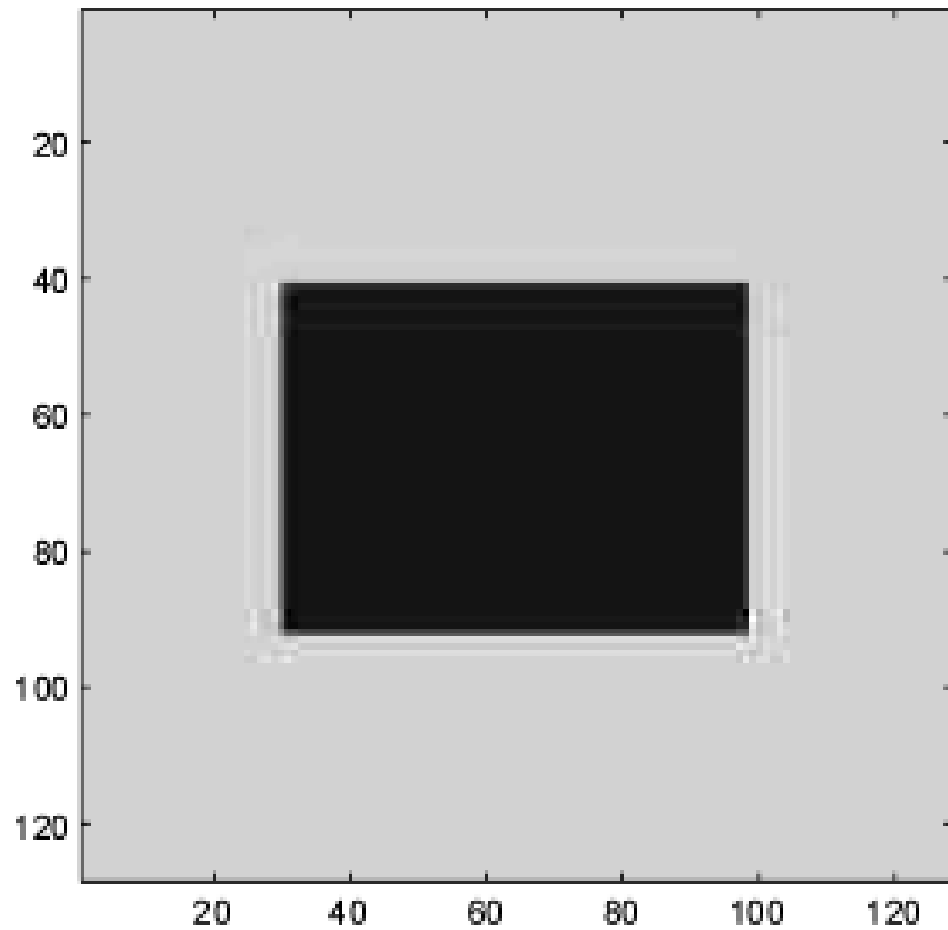


Figure 3.7: Original synthetic image.

Total variation output has some artifacts in it. Edges are preserved to some extent, there are artifacts on the edges. Artifacts are more remarkable on the black part. No observable high pass effect which is desired quality. The process takes a short period of time.

It takes a long period of time to denoise the image by enhanced total variation. As seen in Figure 3.10, the corners are smoothed out which implies low-pass action. Flat places are almost perfectly recovered. It is reasonable to state that the overall quality is the best.

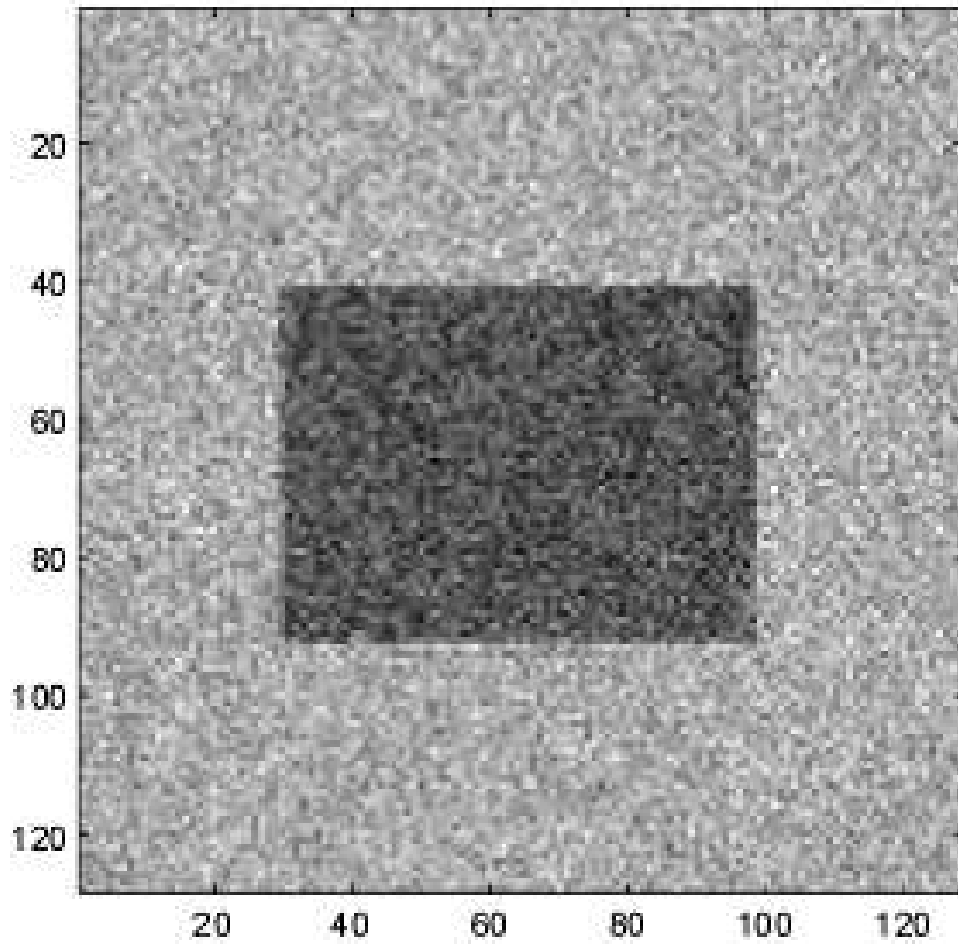


Figure 3.8: Noisy synthetic image with a Gaussian noise of $\sigma=15$.

Slice-TV process provides a reliable reconstruction of the image. There are no artifacts or speckle noise in the output. There is no high pass effect at smooth and white places. There is no blurring as is the case for enhanced total variation. Edges are perfectly preserved.

Besides these advantages, there are also drawbacks of Slice-TV. Since slices are denoised independently, noise is not perfectly cleaned. This is especially observed at flat parts. The process has a moderate computation time.

Now let us examine what the difference between original image and recovered images through these three processes is; i.e. it is useful to compare the error image which can be expressed as:

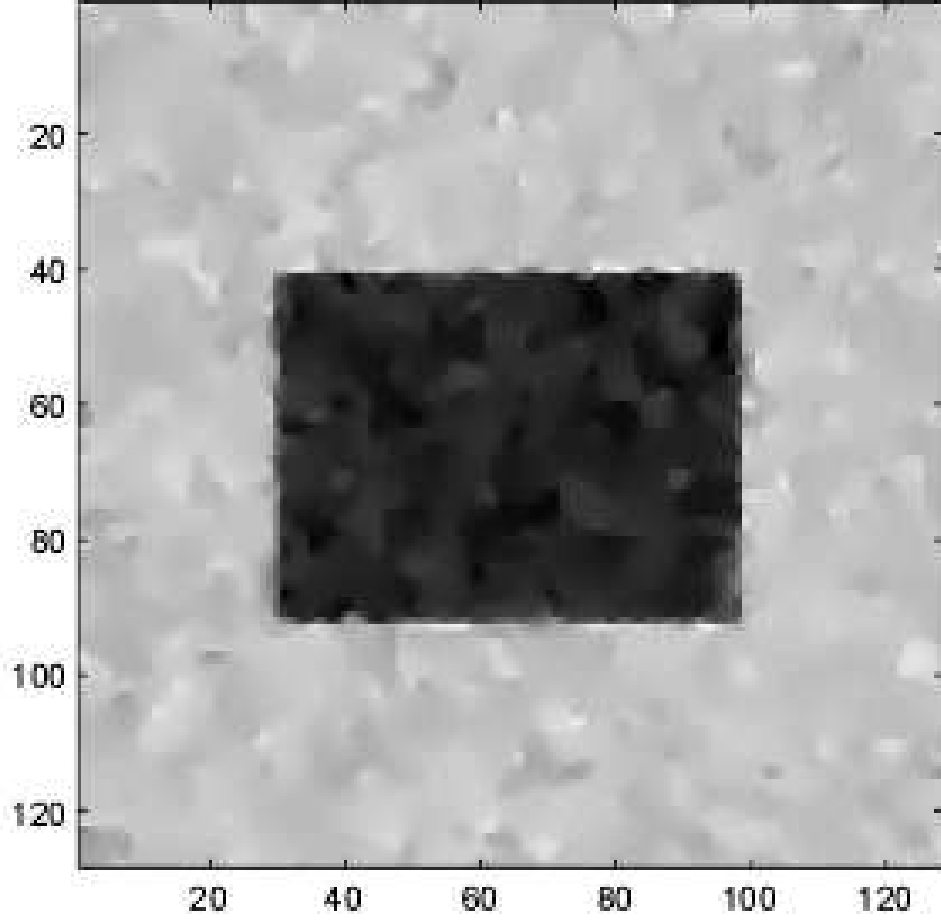


Figure 3.9: Synthetic image denoised with total variation by Rudin *et al.*

$$\mathbf{X}_e = \mathbf{X}_{org} - \mathbf{X}_r, \quad (3.15)$$

where $\mathbf{X}_{org}$ is the original image, $\mathbf{X}_r$ is the recovered image and $\mathbf{X}_e$ is the error image.

For total variation by Rudin *et al.*, at small iteration parameter, $\mathbf{X}_e$ behaves like a random noise which is expected. If we increase the number of iterations, the error becomes bivariate Gaussian and this makes artifacts on the denoised image.

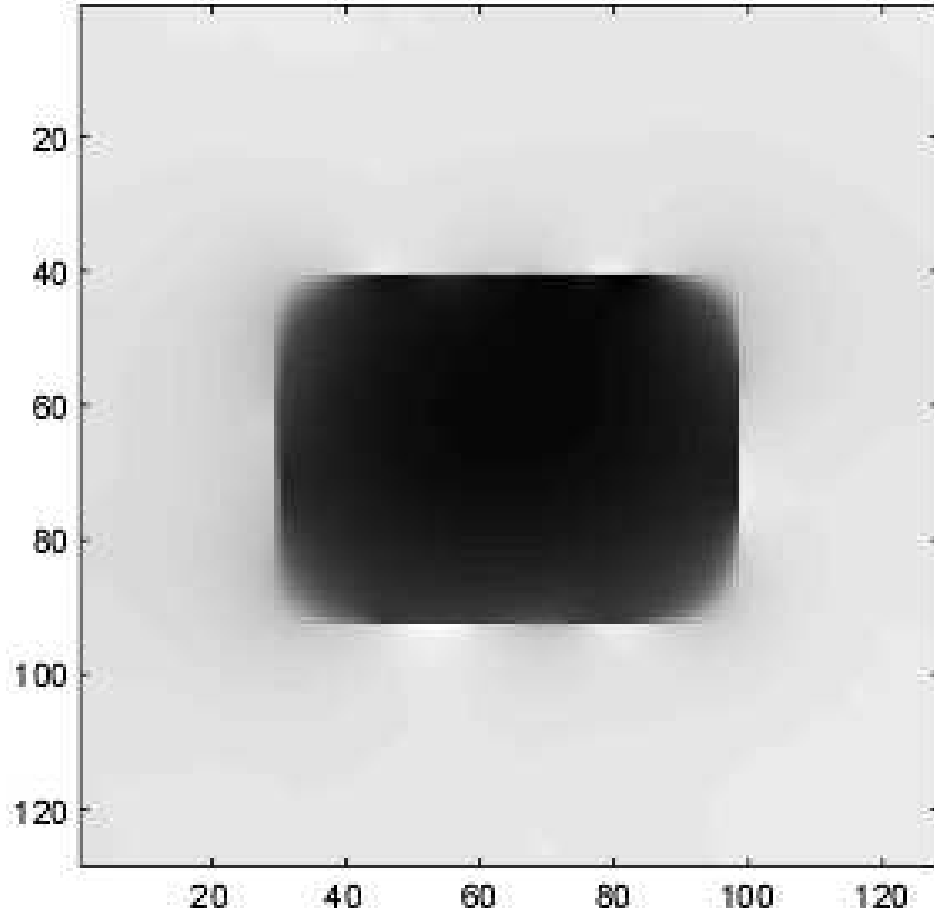


Figure 3.10: Synthetic image denoised with enhanced total variation.

It is seen that the error behavior is similar for small parameter. If estimated σ is increased, the error starts to contain the edges which means a low pass effect on the original image. (Error becomes high-pass)

Similar conclusions as enhanced total variation can be drawn for Slice-TV as well. Error becomes high pass if the parameters are increased and hence, edges are lost through the process.

The error images themselves can be misleading, since they do not give any information about the norm of the error. Because, MATLAB scales the pixel values to gray scale and shows the scaled image. Therefore, it could be useful to give the graphs of the Frobenius norms of the error images versus estimation parameters.

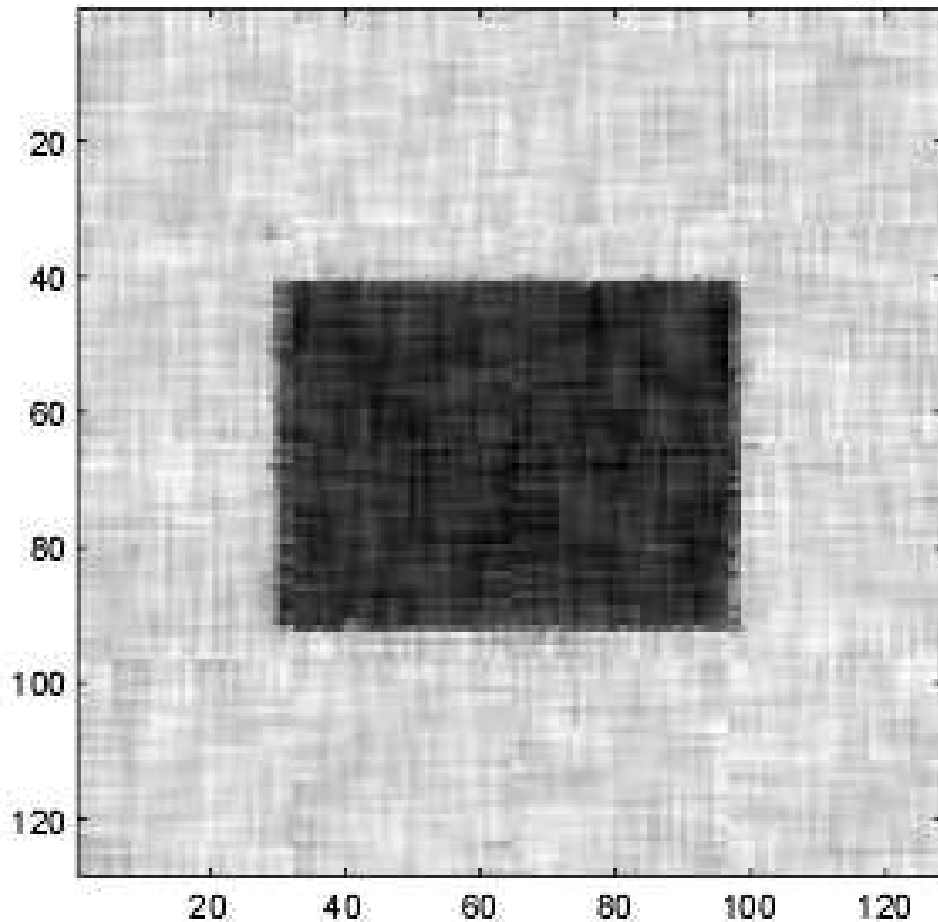


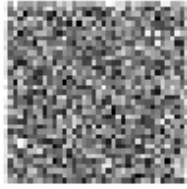
Figure 3.11: Synthetic image denoised with Slice-TV.

The error norm for total variation by Rudin *et al* does not have a large correlation with the number of iterations. It is extremely robust since the error norm changes by 300 for iterations in the interval from 1 to 90. The least value is 1950 which is the worst among three methods.

Enhanced total variation is not robust against the variation of estimated σ . There is a difference of 1700 between the highest and smallest error norm. The least error norm achieved is 600 which seems moderate.

Slice-TV is more robust than enhanced total variation against the deviations of the estimation parameter. The norm value changes by 1400 for a range of

Iteration=10

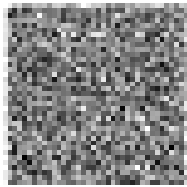


Iteration=80



Figure 3.12: Images of difference between the original image and the denoised image at extreme parameters for total variation by Rudin *et al.*

Iteration=3



Iteration=18

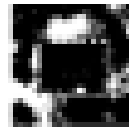


Figure 3.13: Images of difference between the original image and the denoised image at extreme parameters for enhanced total variation.

estimated variance. The error is more stable against deviations for large estimated variance. The norm of the error is approximately equal to the norm of the enhanced total variation.

It might be useful to repeat the comparison with a much more complicated image, the classical Barbara image, since this image has periodic texture and a big variety of pixel values.

Conclusions about the Barbara image are similar to those of synthetic image. The artifacts of the total variation by Rudin *et al* are more obviously disturbing.



Figure 3.14: Images of difference between the original image and the denoised image at various parameters for Slice-TV.

Besides its high time complexity, enhanced total variation gives the best results. Slice-TV has also given promising result with no artifacts around.

3.6 Image Denoising by Total Variation Minimization in Radon Transform Domain

Total variation minimization has been found as an effective tool to denoise $1 - D$ signals. Total variation based denoising of images is also extensively studied. Alternatively, radon transform domain denoising techniques such as curvelets also provide promising results. We propose a new radon transform domain denoising technique where the $1 - D$ projection signals are denoised by optimal $1 - D$ total variation minimization technique. This way, we capture the benefits of denoising in the radon transform domain as well as the effective $1 - D$ denoising achieved by using total variation minimization. The proposed method adaptively estimates the noise variance on the image and utilizes this information in the denoising. Another constraint is the zero mean constraint which is parallel with the zero mean noise. The technique is illustrated in Figure 3.23. Simulation based comparison of the proposed technique with alternatives revealed that the proposed technique is a highly efficient and effective denoising technique.

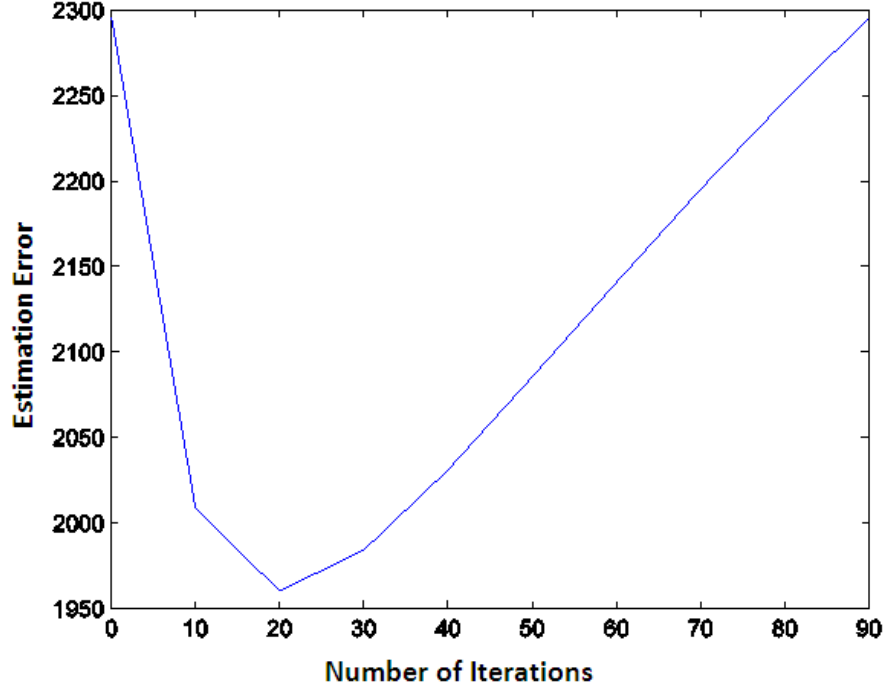


Figure 3.15: Frobenius norm of estimation error for total variation by Rudin *et al.*

In Section 3.6.1, the proposed Radon transform domain total variation minimization technique, R-TV detailed. In Section 3.6.2, comparison of the R-TV with both the curvelet and $2-D$ total variation denoising techniques is provided.

3.6.1 Formulation and Algorithm

As shown in Figure 3.24, the proposed R-TV denoising technique consists of three stages: Radon transformation, total variation denoising of the obtained projections and the inverse Radon transformation. Thus, the first step is to find the one dimensional projection vectors of the $2-D$ image via Radon transformation.

For an image matrix $F[m, n]$ of size M, N ; the discrete Radon transform is found by the summation given in the (3.16). The geometry of this discrete Radon Transform is depicted in Figure 3.25.

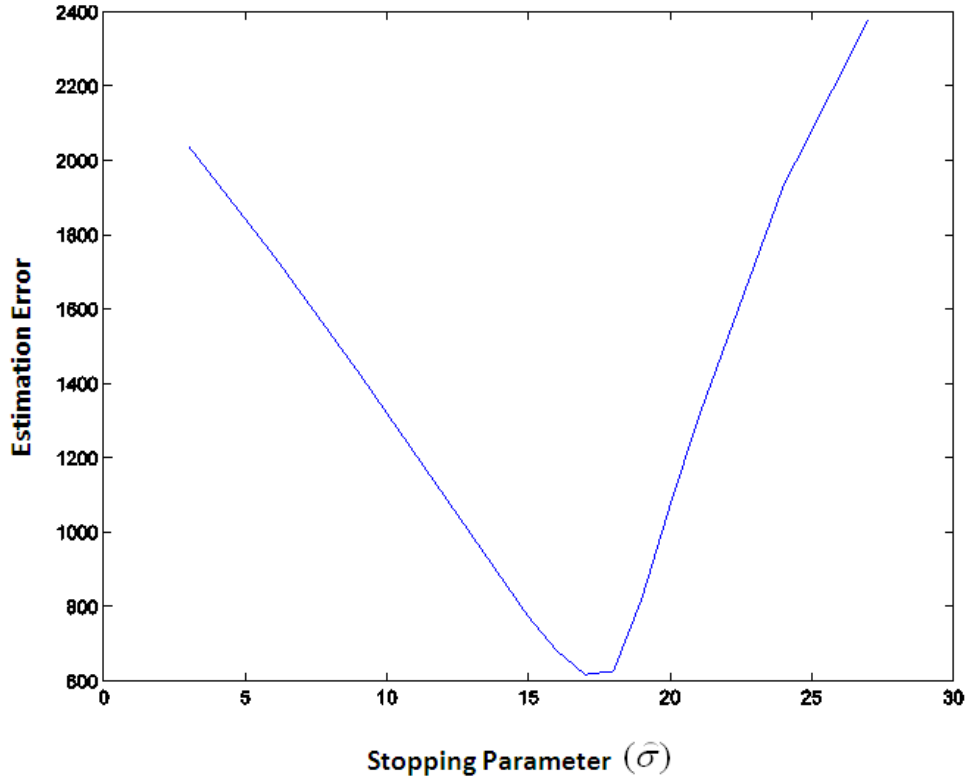


Figure 3.16: Frobenius norm of estimation error for enhanced total variation.

$$g[\phi, l] = \sum_{m=1}^M \sum_{n=1}^N F[m, n] \delta(m \sin \phi - n \cos \phi - l), \quad (3.16)$$

where l is the distance of the projection line from the origin and ϕ is the projection angle at which projection is taken.

Now, we pose that, if we reduce the total variation of each projection $u[l]$ at each angle and recover the image from these projections, the total variation of the $2-D$ image is also reduced. The cost and the constraint for one dimensional denoising problem is the same as told in the (3.5)

The difference norm constraint in that problem assumes that the noise standard deviation is known. If the standard deviation of the noise is not known, the formula explained in the (3.5) can be applied.

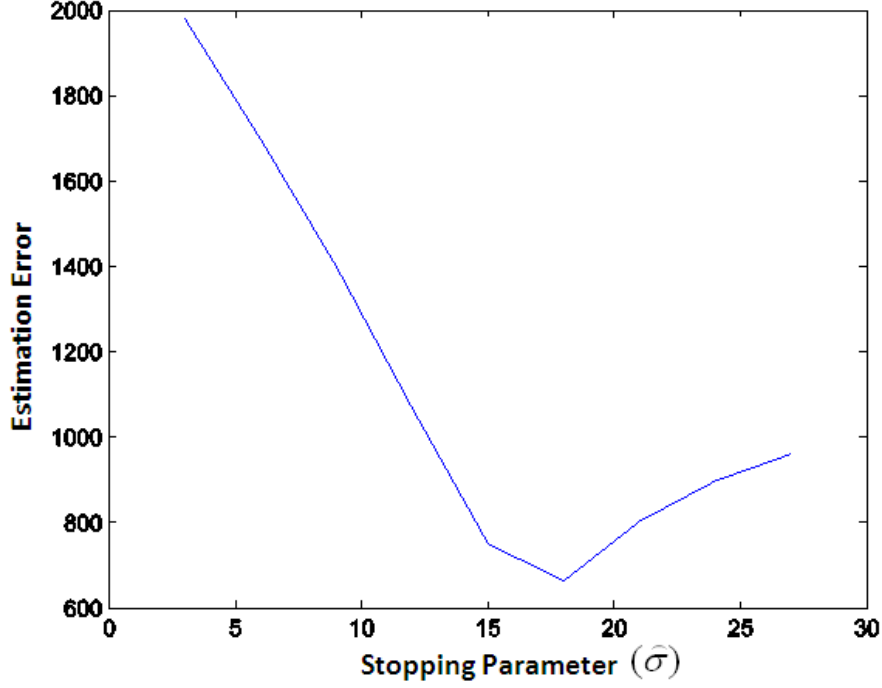


Figure 3.17: Frobenius norm of estimation error for total variation by Slice-TV.

For total variation reduction purpose, the optimal algorithm introduced in [4] was used. The algorithm is omitted here. It is explained in Sect. 2.4 in a detailed manner.

Now, the task is to realize the inverse Radon transform to get the image matrix. We define a new function, $G[\phi, \omega]$ such that:

$$G[\phi, \omega] = \sum_{l=-L}^L e^{-j\omega l} g[\phi, l]. \quad (3.17)$$

Now, the inverse Radon transform can be taken by using this new function and Fourier Slice Theorem.

$$F[m, n] = \frac{1}{4\Pi^2} \sum_{\phi=0}^{\Pi} \sum_{w=-W}^W G(\phi, \omega) e^{j\omega(x \sin \phi - y \cos \phi)} |\omega| d\omega d\phi \quad (3.18)$$

The steps can be summarized as the following.

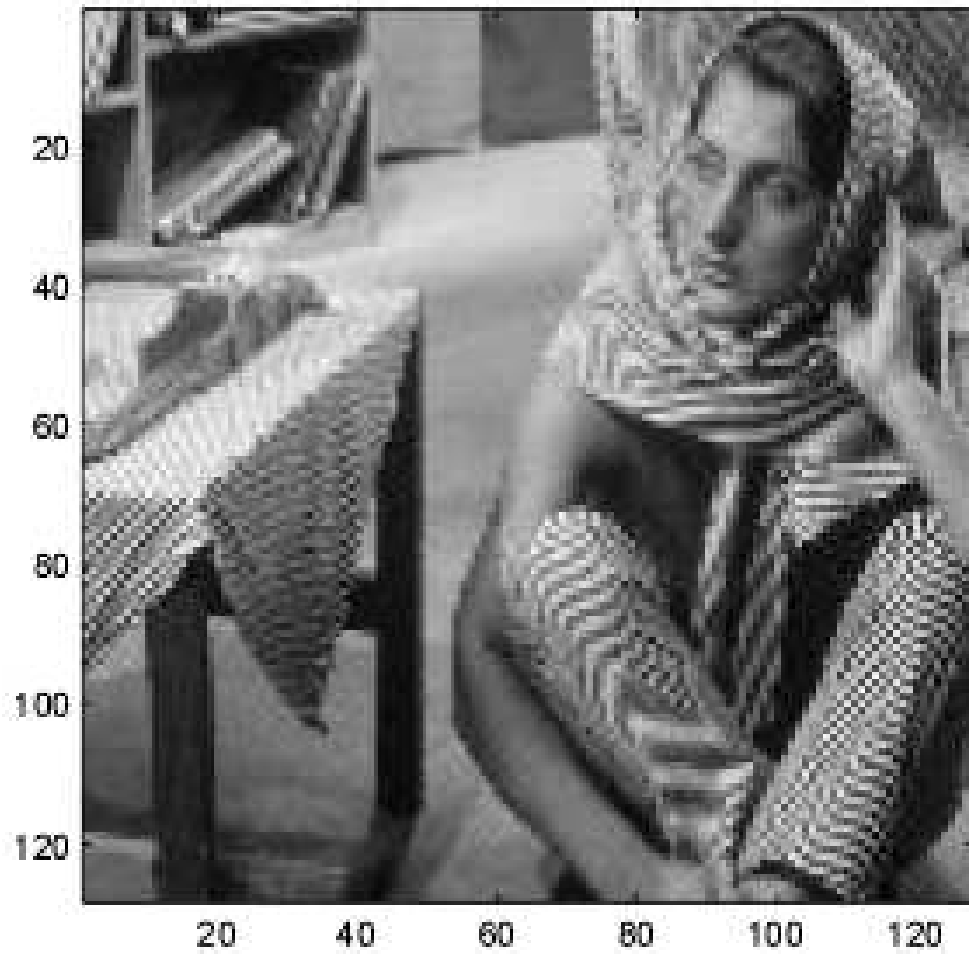


Figure 3.18: Original Barbara image.

1. **Projection.** Take Radon Transform to find $1 - D$ projections.
2. **Total variation reduction.** Denoise the projections by using $1 - D$ technique. Stop the iterations when $\|u[N] - u[0]\|_2^2 = \sigma \cdot \sqrt{N}$
3. **Back Projection.** Compute inverse-Radon transform of the denoised slices.

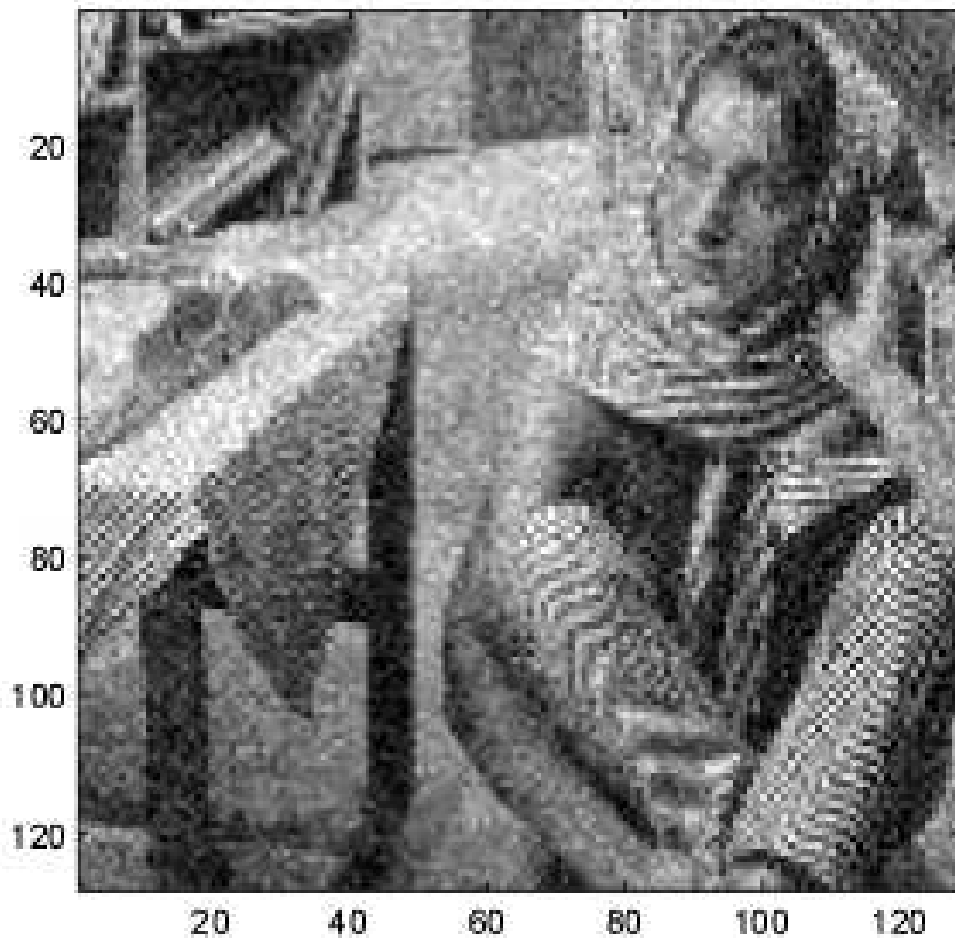


Figure 3.19: Noisy Barbara image with Gaussian noise of $\sigma = 15$.

3.6.2 Comparison of the Techniques

In this section, we provide denoising results obtained by the proposed R-TV technique, and compare them with the results obtained by using the curvelets in [36] and the 2-D total variation techniques in [35].

It will be proper to start the comparison by a rather simple image that consists of only a rectangle.

Total variation by Rudin *et al* is implemented by a constraint of finite iterations. At first, noise was estimated and the corresponding iteration number was found by trials based on the comparison of curvelet output with the same estimation standard deviation parameter. As shown in Figure 3.28, total variation of

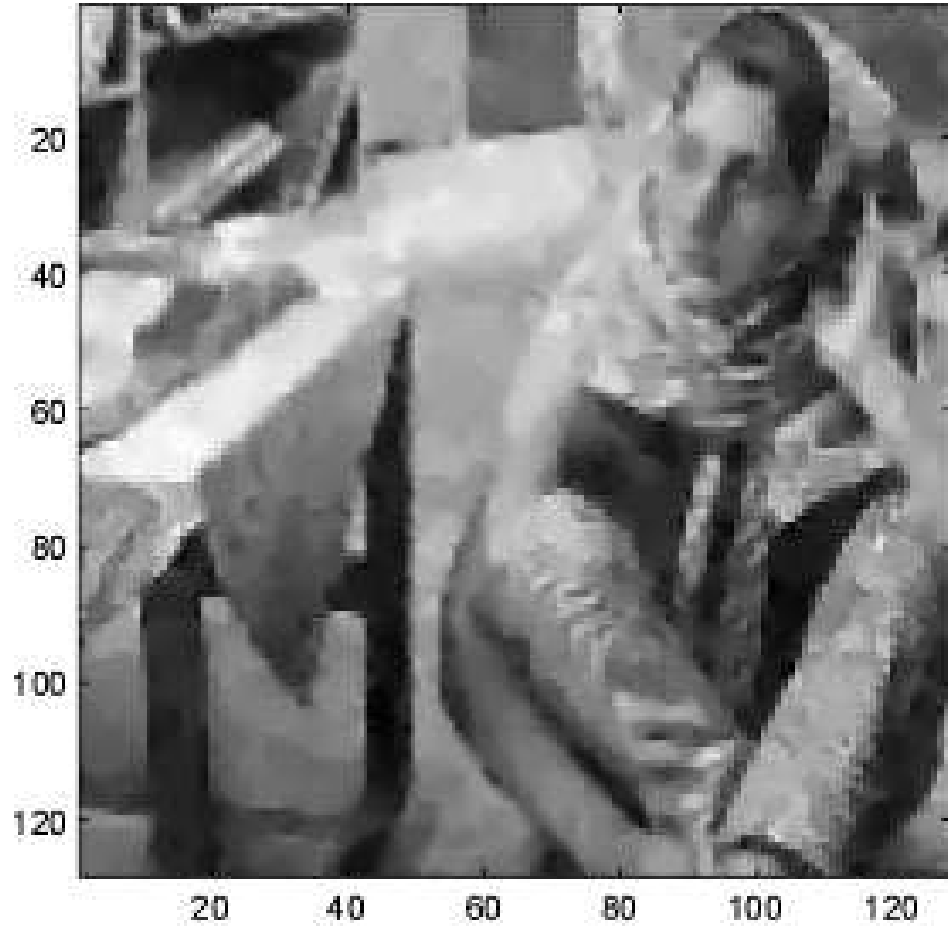


Figure 3.20: Image denoised with total variation by Rudin *et al.*

the $2 - D$ total variation output has some artifacts in it. Edges are not preserved through the process, because there is a bit low pass effect at the edges. Flat places are sharp which means that noise is cleaned successfully. No observable high pass effect in the flat places which is a desired quality. The process takes about 6.03 sec which is a moderate value.

Curvelet technique was applied based on the estimated standard deviation parameter. This was estimated by the method proposed in the previous parts. As shown in Figure 3.29, provides a highly denoised image where the edges are preserved and it is free from speckle noise. Smooth places are a bit darkened which means high pass action in the smooth places, since high pass action yields low pixel values in the smooth parts of the image.

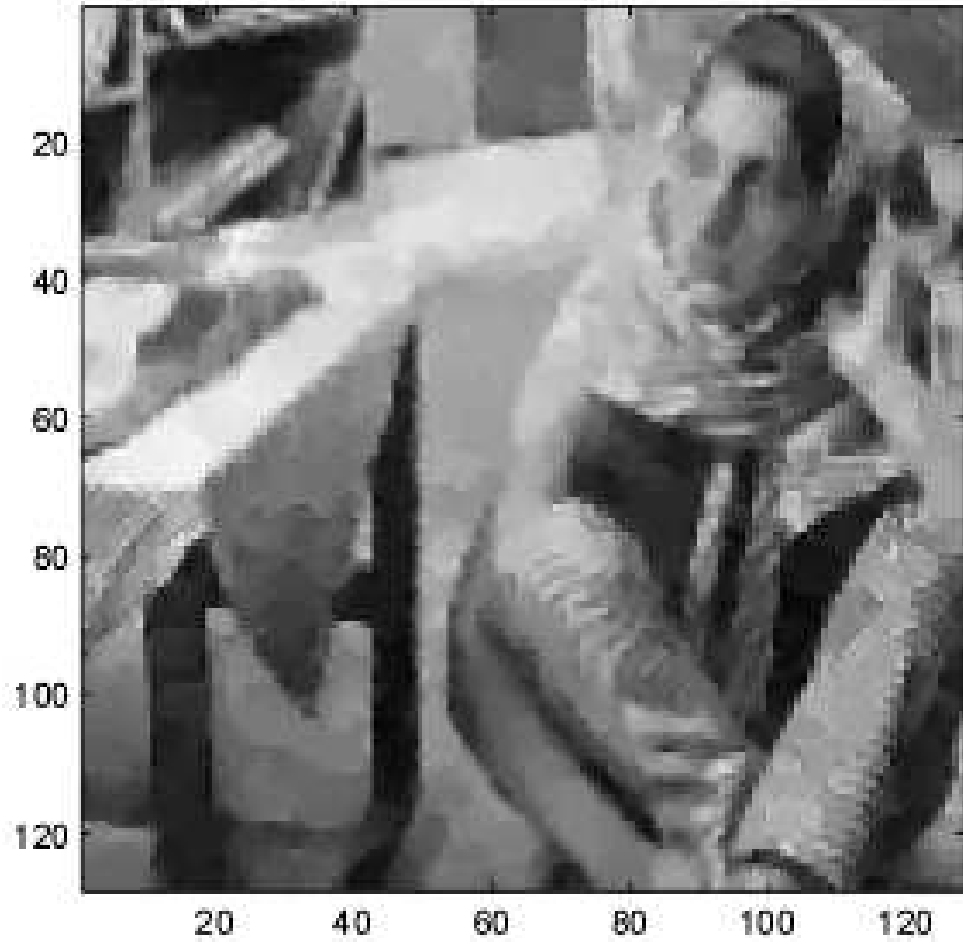


Figure 3.21: Image denoised with enhanced total variation.

As shown in Figure 3.30 provides a highly denoised R-TV process does not spoil the image. There are no artifacts or speckle noise in the output. There is not data loss with respect to the original image. For example, there is no high pass effect at smooth and white places. The output is rotation independent which could be an important quality for SAR image denoising.

Besides these advantages, there are also drawbacks of R-TV. Since projections are denoised independently, noise is not perfectly cleaned. This is especially observed at flat parts. The image is also smoothed out. Edges are a bit lost. Moreover, the process takes about 13.5 sec which is the slowest among three.

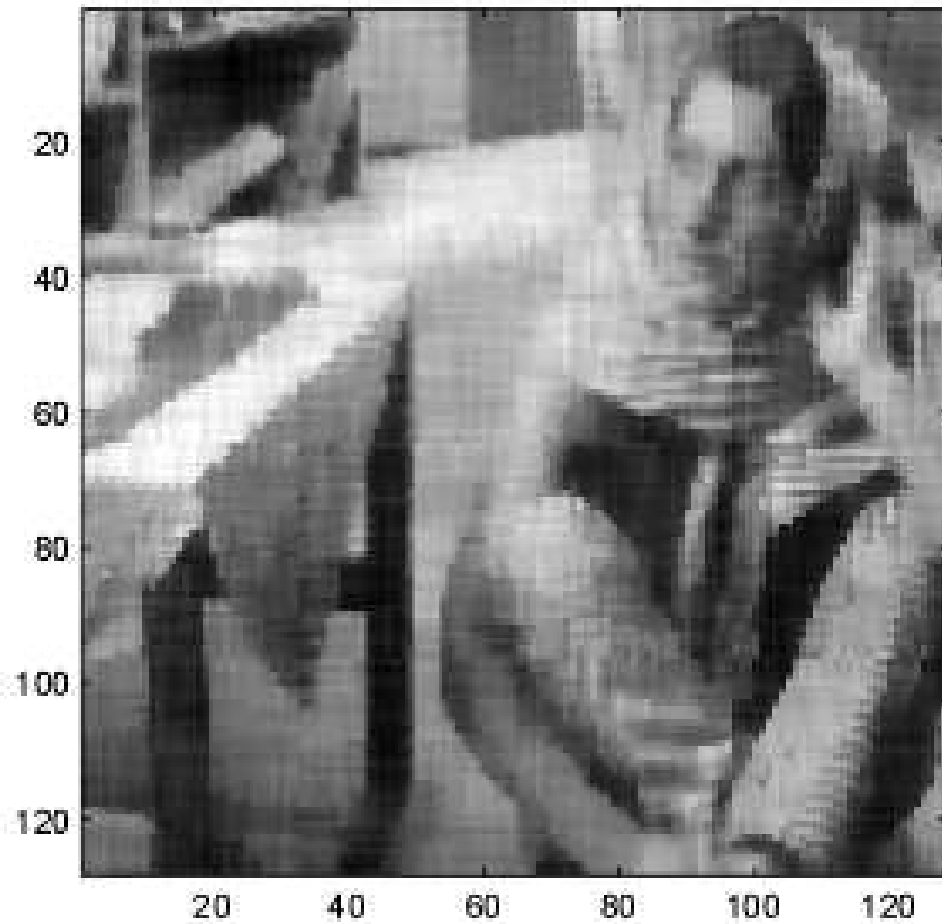


Figure 3.22: Image denoised with with Slice-TV.

It might be useful to repeat the comparison with a much more complicated image, the classical Barbara image, since this image has periodic texture, flat parts and a big variety of pixel values.

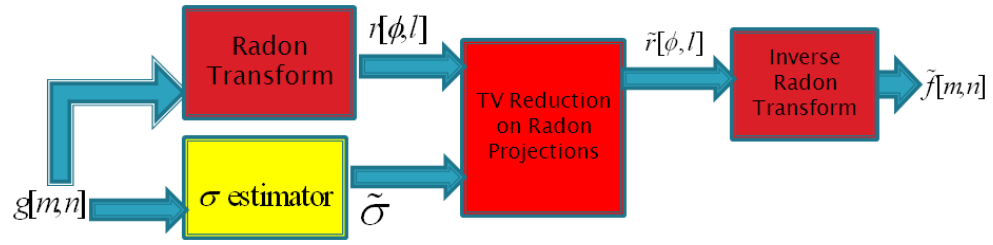


Figure 3.23: Flow diagram of the proposed Radon TV Technique.

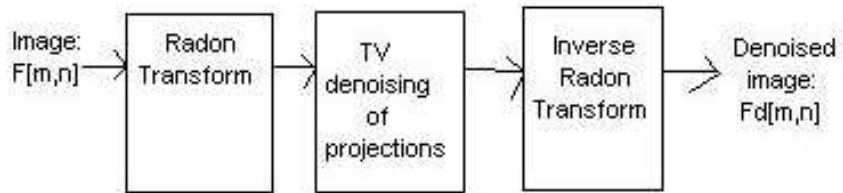


Figure 3.24: Block diagram of the denoising process.

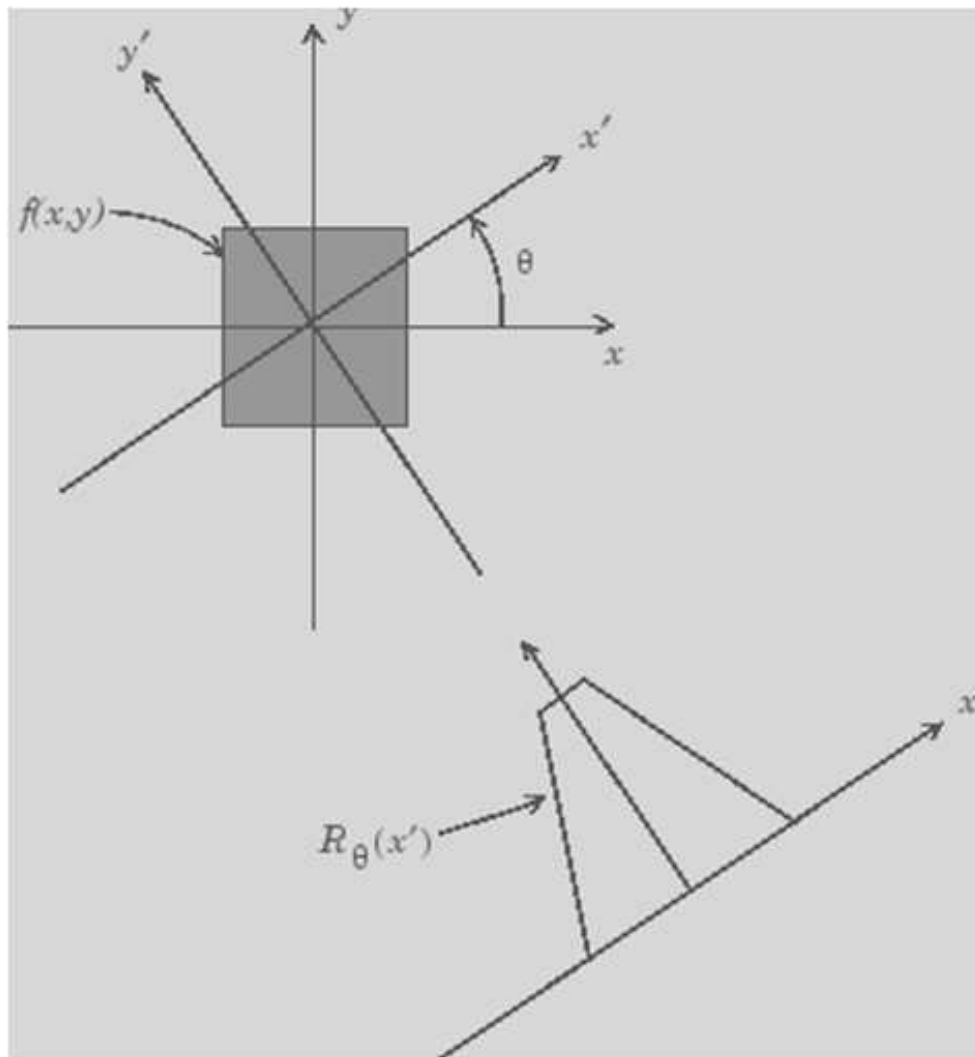


Figure 3.25: The geometry of Radon Transform (taken from MATLAB Radon Transform help).

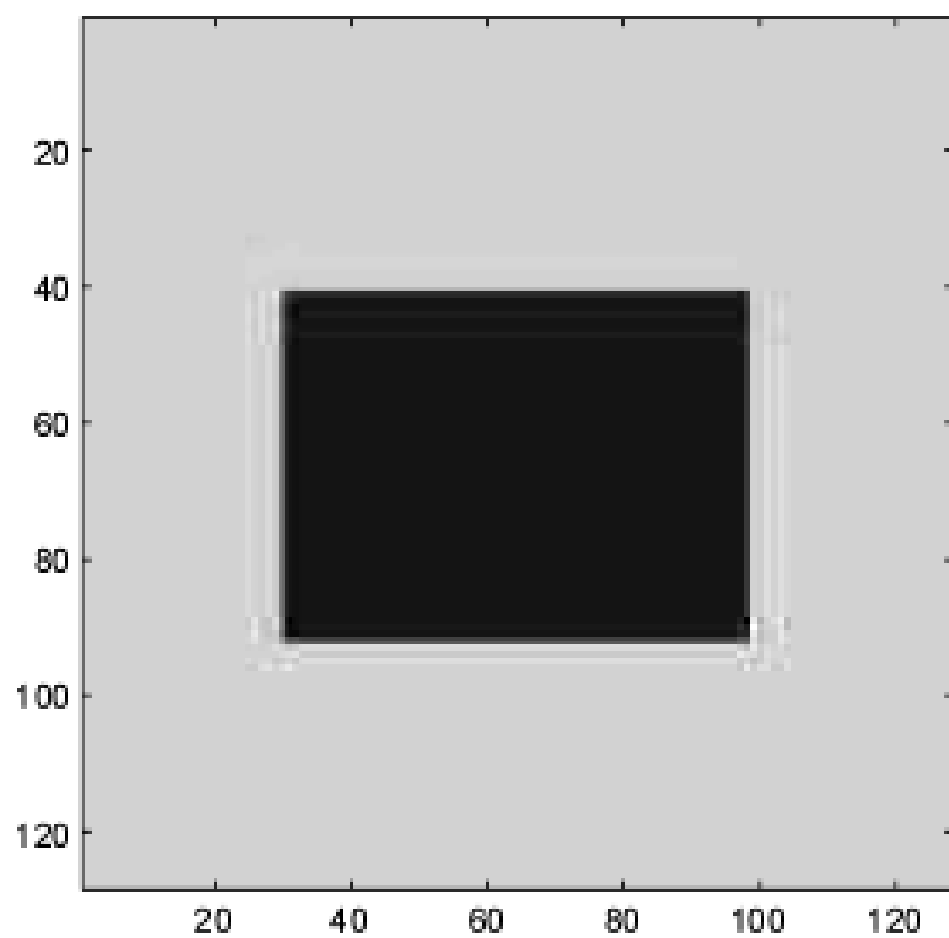


Figure 3.26: Original synthetic image.

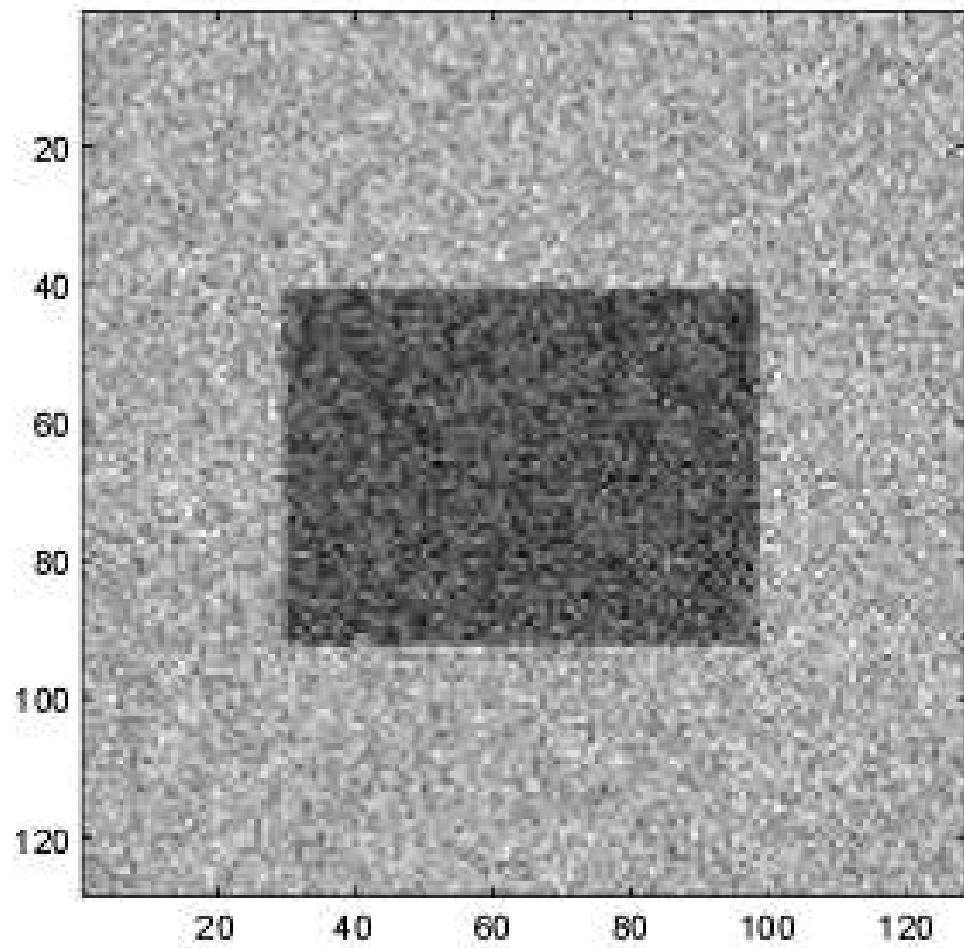


Figure 3.27: Noisy synthetic image with a Gaussian noise of $\sigma=15$.

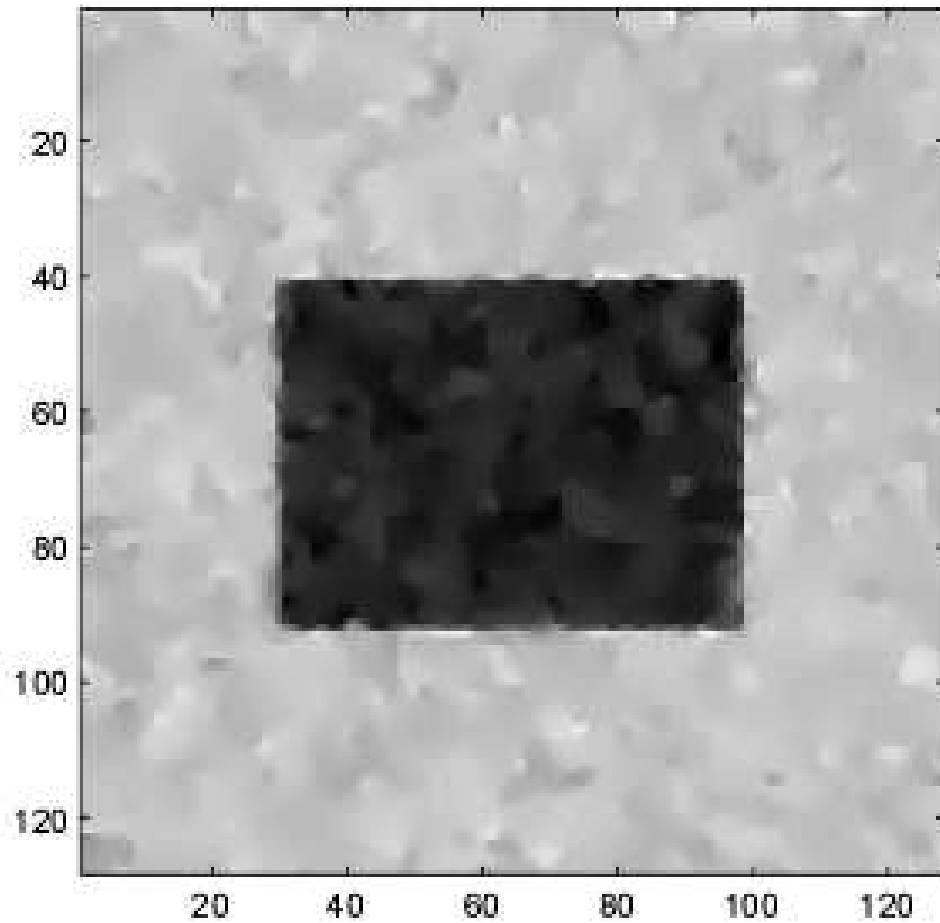


Figure 3.28: Synthetic image denoised with total variation by Rudin *et al.*

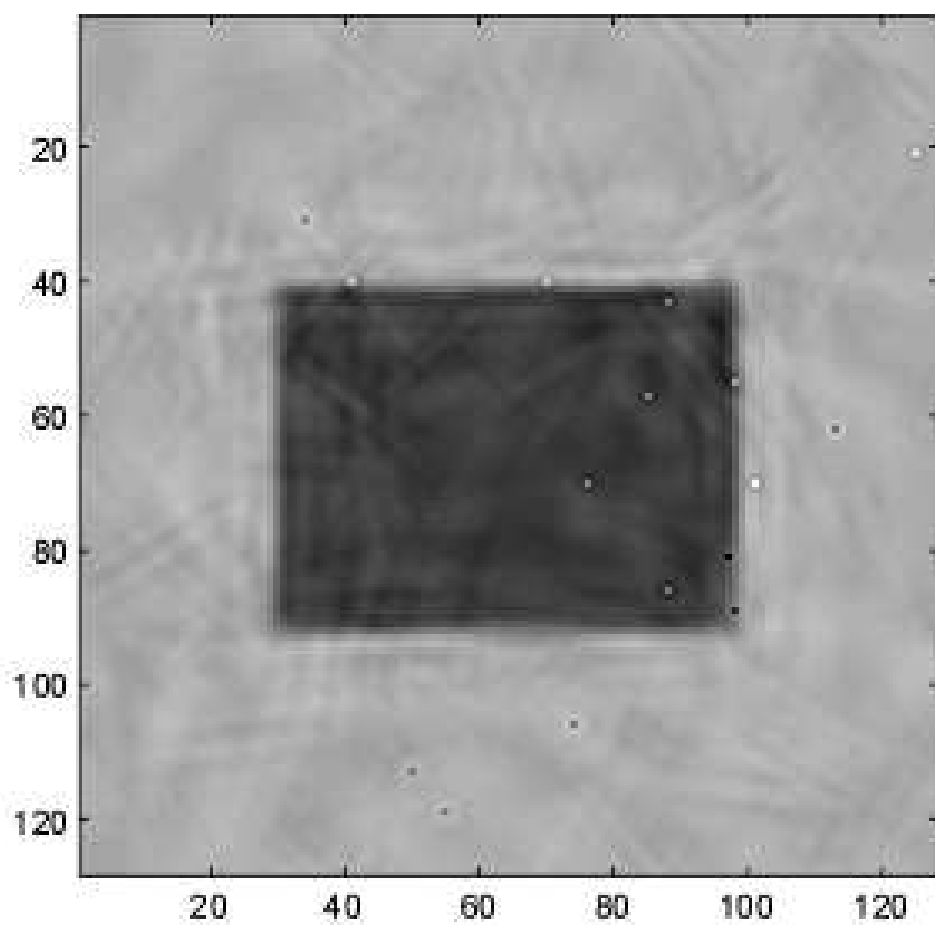


Figure 3.29: Synthetic image denoised with curvelet.

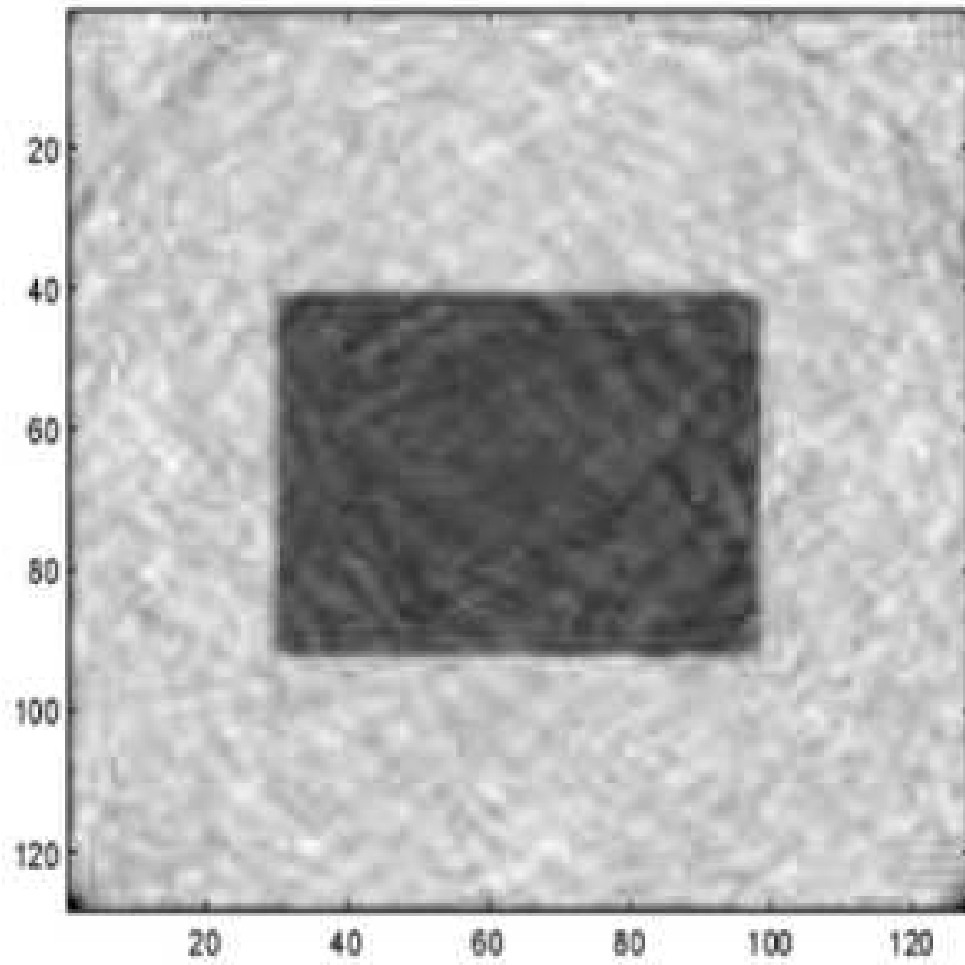


Figure 3.30: Synthetic image denoised with proposed R-TV.

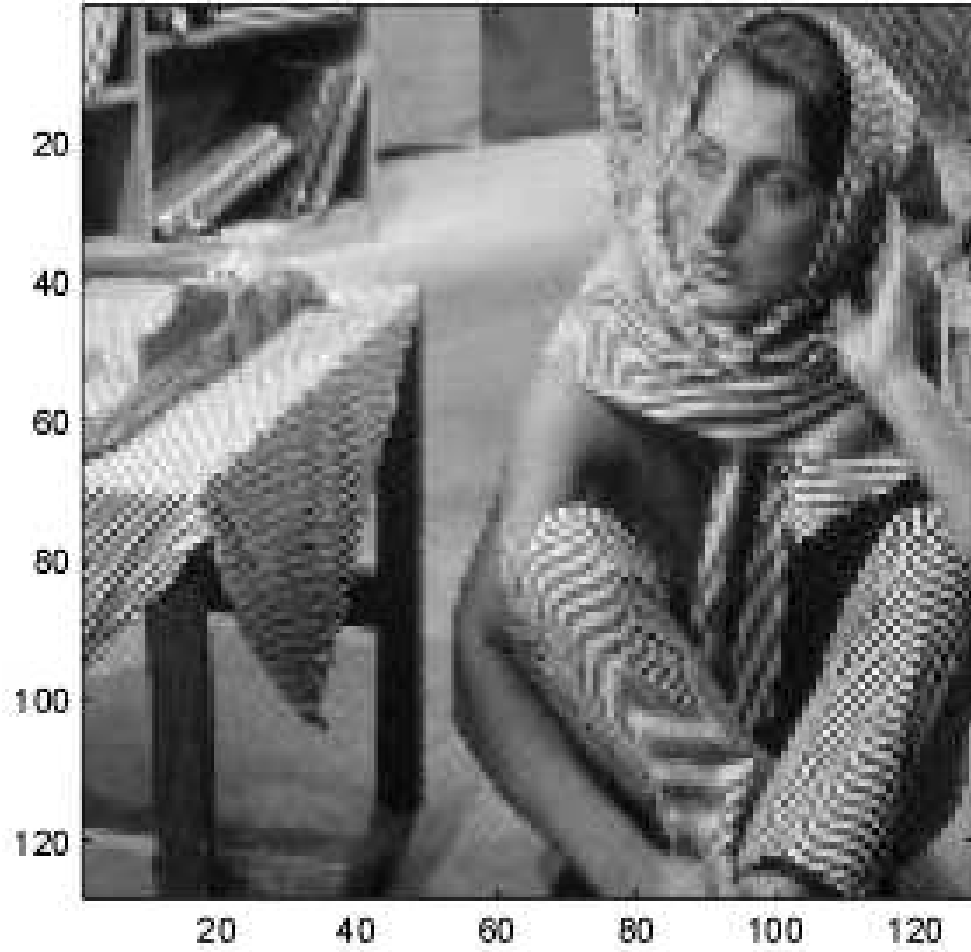


Figure 3.31: Original Barbara image.

Conclusions about the Barbara image are similar to those of synthetic image. The artifacts of the total variation by Rudin *et al* are more obviously disturbing. Curvelet has speckle noise and observable wavelet curves in it, but other than these drawbacks, it gives the best result. R-TV did not perfectly remove the noise, however, the shape is preserved with no artifacts around.

Now let us examine what the difference between original image and recovered images through these three processes is. That is, it is useful to compare the error image which can be expressed as:

$$\mathbf{X}_e = \mathbf{X}_{org} - \mathbf{X}_r, \quad (3.19)$$

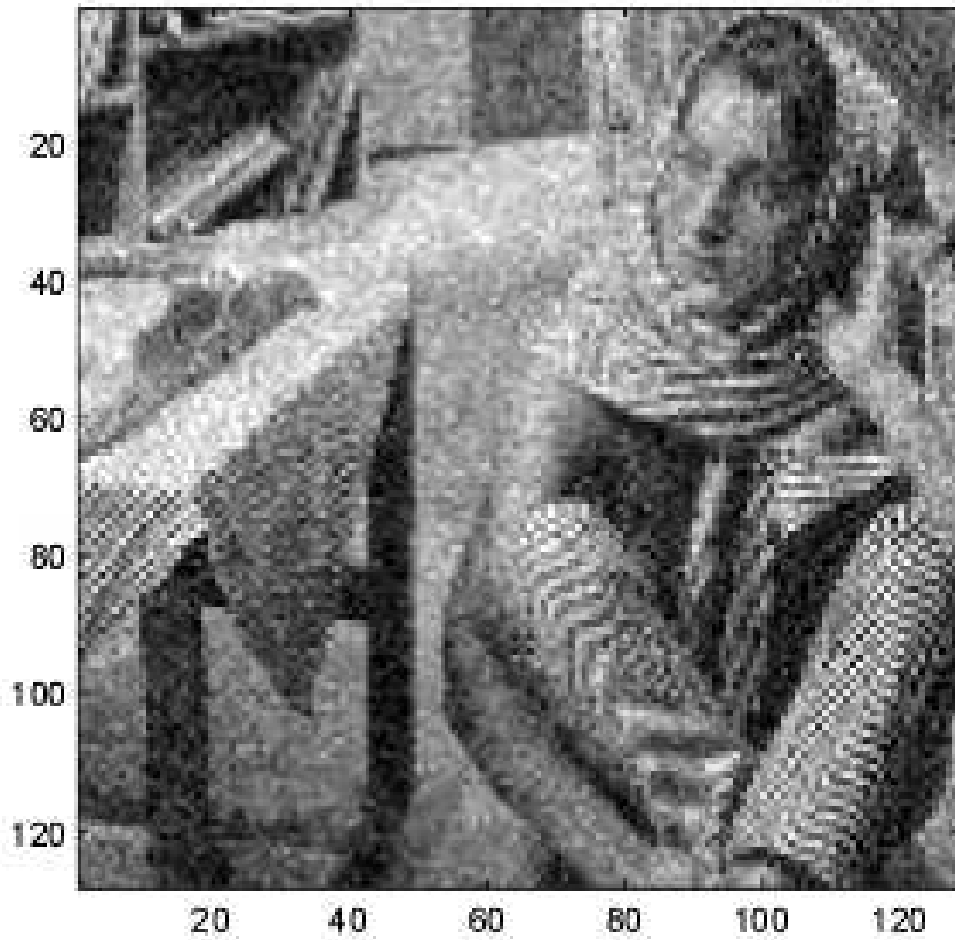


Figure 3.32: Noisy Barbara image with Gaussian noise of $\sigma = 15$.

where $\mathbf{X}_{org}$ is the original image, $\mathbf{X}_r$ is the recovered image and $\mathbf{X}_e$ is the error image.

For total variation by Rudin *et al*, at small iteration parameter, $\mathbf{X}_e$ behaves like a random noise which is expected. If we increase the number of iterations, the error becomes high pass which takes the edges out.

Similar observations can be drawn out of the Figure 3.37 for curvelet transform. If we increase estimated σ , the error starts to contain the edges which means a low pass effect on the original image.

As seen above in Figure 3.38, R-TV is the most robust of all three against deviations of the estimated parameter. Error remains the same whether the

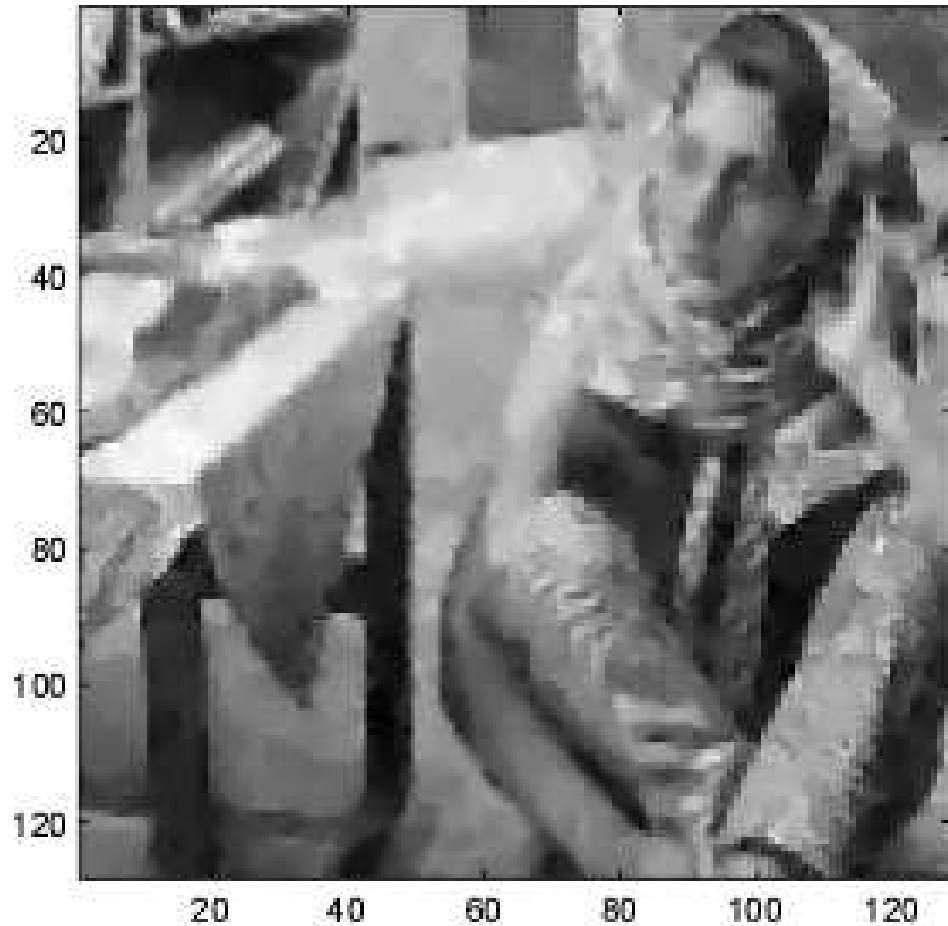


Figure 3.33: Barbara image denoised with total variation by Rudin *et al.*

estimated σ is changed. Error is always high pass and hence, edges are lost a little through the process.

The error images themselves can be misleading, since they do not give any information about the norm of the error. Because, MATLAB scales the pixel values to gray scale and shows the scaled image. Therefore, we thought that it could be useful to give the graphs of the frobenius norm of the error images versus estimation parameters.

The error norm for total variation by Rudin *et al* have a large correlation with the number of iterations. It is not robust since the error norm changes by 300 for iterations in the interval from 0 to 90. The least value is 1950 which is the least among three methods.

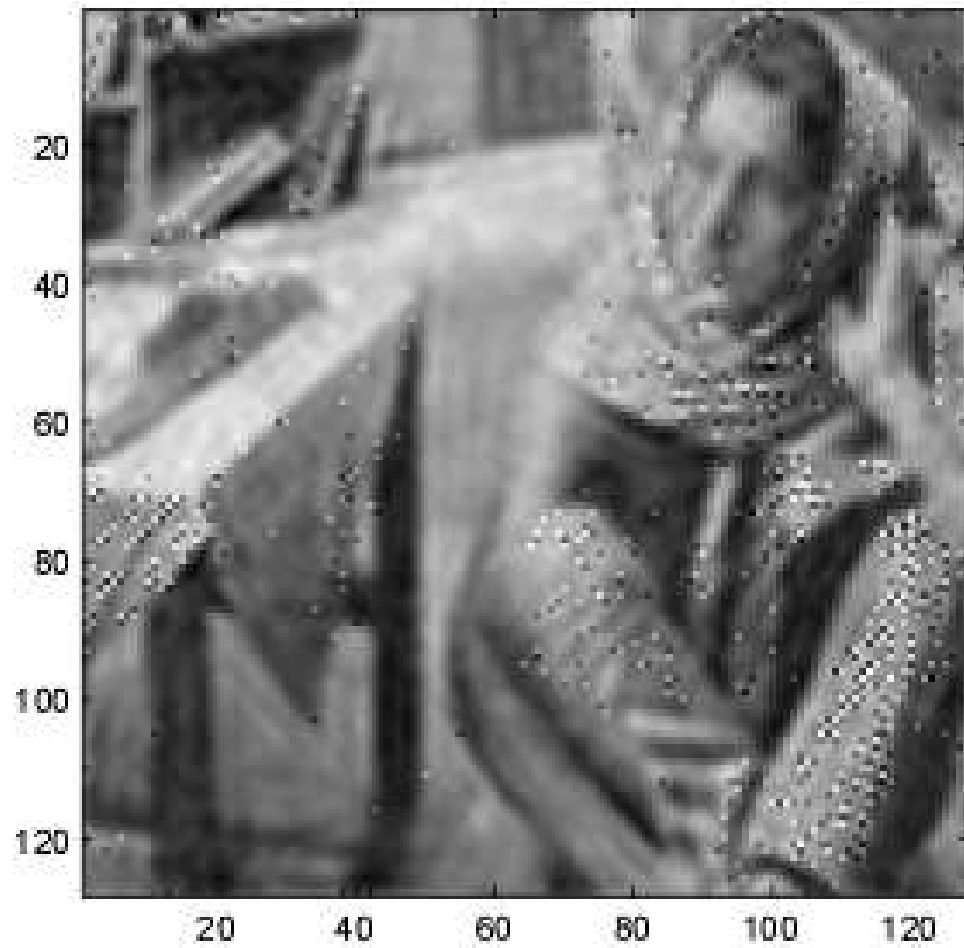


Figure 3.34: Barbara image denoised with curvelet transform.

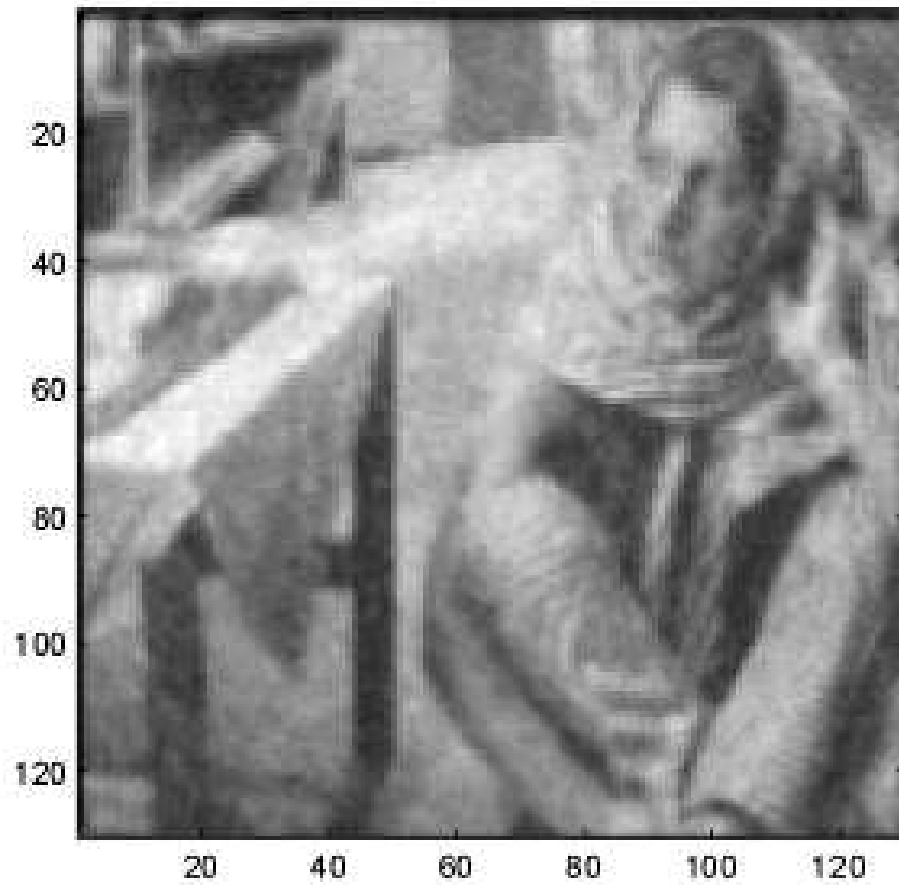
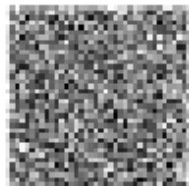


Figure 3.35: Barbara image denoised by R-TV.

Iteration=10



Iteration=80



Figure 3.36: Images of difference between the original image and the denoised image at various parameters for total variation by Rudin *et al.*



Figure 3.37: Images of difference between the original image and the denoised image at various parameters for curvelet transform.



Figure 3.38: Images of difference between the original image and the denoised image at various parameters for R-TV.

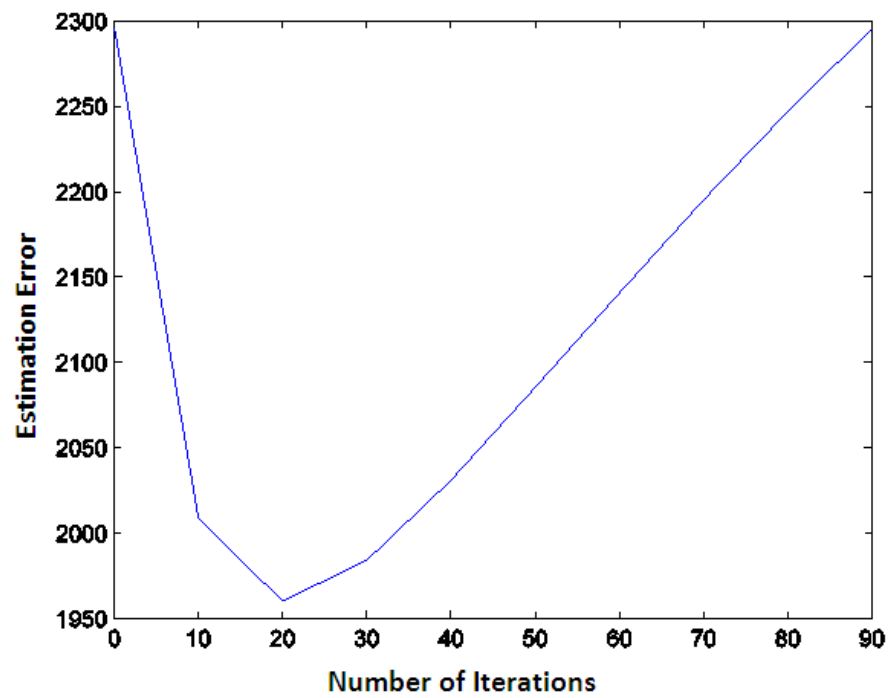


Figure 3.39: Frobenius norm of estimation error for total variation by Rudin *et al.*

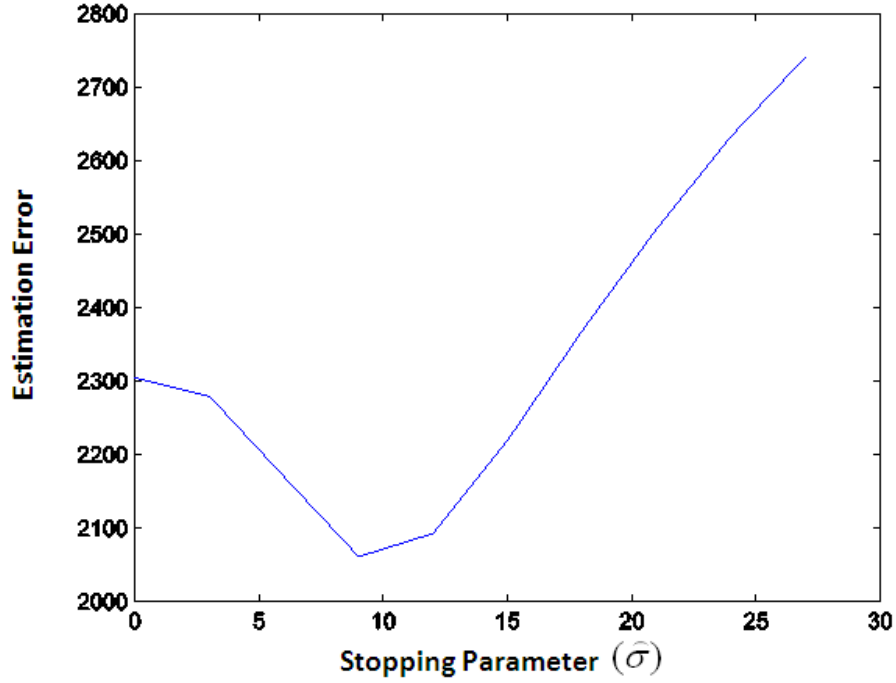


Figure 3.40: Frobenius norm of estimation error for curvelet transform.

Curvelet is not robust against the variation of estimated σ either. There is a difference of 700 between the highest and smallest error norm. The least error norm achieved is 2100 which seems moderate.

The denoising performance of the R-TV technique is highly robust to the errors in the estimated σ . The norm value changes only by 200 for a range of estimated variance. The average norm of the error is high which means that the loss at the edges is high. This might imply that this technique may also be used for robust edge detector.

3.7 Image Texture Recovery by Separable Implementation of 2 – D Total Variation

Direct implementation of 2 – D LTI digital filters do not have a memory requirement. It can be implemented online. However, the direct implementation

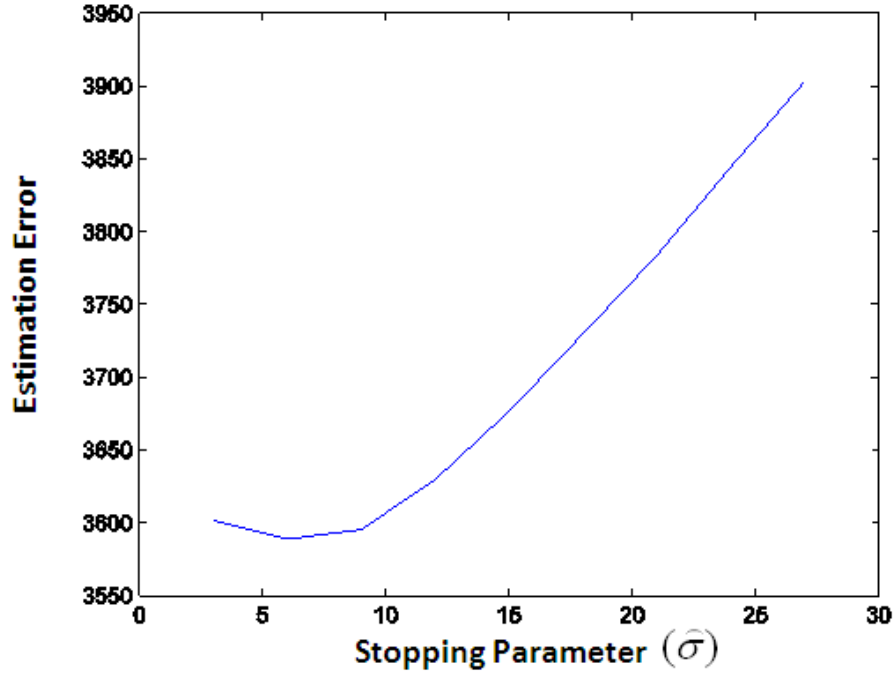


Figure 3.41: Frobenius norm of estimation error for total variation by R-TV.

is relatively slow. In order to speed up the filtering process, separable implementation can be used. A $2-D$ filter is separable if its impulse response can be expressed as the outer product of two $1-D$ impulse responses such that:

$$\mathbf{A} = \mathbf{u}\mathbf{v}^T. \quad (3.20)$$

It is straight-forward to show that the $2-D$ kernel matrix $\mathbf{A}$ should be rank-1 in order to be separable. Separable implementation is performed in two steps. First, each row is convolved by the filter $\mathbf{u}$ and the matrix $\mathbf{B}$ is obtained. This $\mathbf{B}$ matrix should be kept in a memory. Next, the columns of the matrix $\mathbf{B}$ is convolved by the filter $\mathbf{v}^T$. The mathematical formulation is like this:

$$G[m, n] = F[m, n] * A[m, n], \quad (3.21)$$

for $n = 1$ to N

$$B[m, n] = F[m, n] * u[m],$$

for $m = 1$ to M

$$G[m, n] = B[m, n] * v[n].$$

If the size of the convolution kernel $A[m, n]$ is $a \times b$, the cost of the direct implementation is $MNab$. On the other hand, the cost of the separable implementation is $MN(a + b)$. If a or b are greater than 2, then separable implementation is more efficient. This gain increases if M and N increase.



Figure 3.42: The original Barbara image.

Since there is an optimal implementation of $1-D$ total variation reduction [4], carrying it to $2-D$ by separable implementation was tried in the research of this thesis. However, the technique did not yield decent results since the separability conditions are not satisfied. First condition is that the filter should be an LTI



Figure 3.43: The noisy Barbara image with a noise σ of 8.

filter. Since total variation reduction is a nonlinear process, this requirement is not met. The second constraint is that the $2 - D$ filter should have a rank equal to 1. Since there is no $2 - D$ counterpart of the total variation reduction [4], the second is not met either. Although these are the mathematical drawbacks, the output of this separable implementation of total variation produces visually pleasing outputs despite the unsatisfactory error norm. Since total variation is an edge preserving algorithm, the $2 - D$ separable implementation has high preservation of texture. This merit makes this proposed approach a $2 - D$ texture recovery technique. The original image, noisy image and the denoised image are as shown in the Figure 3.42, 3.43 and 3.44. The highly textured anonymous Barbara image is a proper test for texture recovery.



Figure 3.44: Barbara image denoised with the separable total variation technique.

Chapter 4

CONCLUSIONS AND FUTURE WORK

4.1 1 – D Signal Denoising

The running average filter turned out to be an efficient way of estimating noise standard deviation of one dimensional signals. It is a fast and robust tool for noise estimation. It can be used as a stopping criterion for iterative smoothing algorithms.

After comparison of simulations, it is rather viable to state that the PWL technique we found is a rather robust, efficient and a reliable denoising technique. It is highly successful for signals that are formed of piecewise continuous lines. It's time complexity is $O(N\log N)$ which is the same as Willsky's but worse than Rudin's.

As a future work, we could separate the signal into parabolas instead of lines. For that purpose, we'd better work at the second difference domain instead of first difference domain. Moreover, in the future, we can separate an image into

optimal planes. In that work, finding the boundaries of the planes will be crucial. Hopefully, we will be able to carry our algorithm to video denoising as well.

The Piecewise Parabolic and Cubic signal denoising performed well for signals that are formed of piecewise cubic or parabolic signals, however it works rather well for arbitrary signals as well. It has the same order of time complexity as the proposed PWL technique.

After the simulations, it is reasonable to state that the denoising technique by matched filtering that we found is a superb denoising technique. It has low false alarm rates. Its probability of detection is at reasonable values as well. It can denoise difficult structures such as two adjacent lines with the same slope. It can perfectly denoise signal structures with piecewise constant parts achieving a very small error norm. Moreover, no prior information on the degree of the polynomials is needed.

There are some points in our study which requires further research. For example, the probability of false alarm for the signal with a jump is observed to increase with increasing SNR which is unexpected. This should be investigated and the reason should be found. Moreover, two kinds of thresholds are used in our technique. One is to understand if the output of convolution makes impulses. If we can formulate a value for this threshold instead of manually selecting it, it would automatize our technique. Second, a threshold is used to find the maximums of the impulses. We should formulate the starting point for this as well. After this, it is reduced automatically.

For the classical problem of denoising signals with piecewise polynomial models, a new technique is proposed. In the proposed technique, the boundaries of the polynomial segments are obtained by using fast converging PSO technique. As demonstrated by a set of simulations, the proposed denoising technique has a

robust performance that is better than that of the total variation denoising technique. For the determination of model parameters, a criterion based on average fit error is proposed. Statistical significance of this criterion will be investigated in a future study.

In this thesis, $1 - D$ texture recovery is accomplished by adaptive windowing. Total variation is used to detect steps and edges. In the residual, periodicity is investigated and the period is estimated. The noise is removed by averaging the noisy periods and this prototype is pasted in place of the noisy ones. To increase success, adaptive window is used so as to enlarge if the same periodic structure is found in the neighboring window. Because of this relatively complex algorithm, this can not be performed in a two dimensional image. Therefore, the first future work should be carrying this algorithm to two dimensions. Radon transform might be a tool for this purpose. Increasing the efficiency is an important task, too. Also, this algorithm assumes that the periods are perfectly same. However, in real life, there is a concept of perspective so that periods do not look perfectly the same. So, probabilistic approach can be given in the case of perspectivity. Moreover, this algorithm could be synthesized with two dimensional total variation reduction. If the image is heterogeneous and has texture in it, texture recovery algorithm may be used. If the image is relatively smooth and homogeneous total variation reduction techniques such as R-TV could be used. The C matrix in the Frost denoising algorithm could be used to detect heterogeneity.

4.2 Image Denoising

There are various image denoising techniques that needs to know the noise variance *a priori* for proper operation. The median-filter based approach that we found is a robust non-parametric technique that gives nearly optimal results. As

a future work, this noise estimation method will be used to stop the processes of the parametric denoising techniques.

The Slice-TV technique turn out to be a reliable and effective tool on image denoising due to the simulation results. Because, minimization of total variation had already been proven to be the optimal method in one dimension, and it also showed itself to be applicable in a fast and robust manner. Two constraints; zero mean and noise variance and a cost of total variation has given promising results. With the technique proposed, noise can be estimated accurately. Using this estimate gives outstanding results, whereas the algorithm is robust against the choice of the parameter.

The discontinuity problem of row-wise and column-wise outputs have been resolved by averaging these two outputs. Also, mathematically, the overall output is still rotation dependent. This dependence might cause some problems about remote sensing images such as SAR images. Therefore, besides these two orthogonal angles, i.e rows and columns, total variation minimization can be performed in some angles in between. This will remove discontinuities better, and it will not change the order of the time complexity.

Radon projection turned out to be a reliable and an efficient method to provide the processing of an image in one dimension. It must be kept in mind that in digital medium it does not provide perfect reconstruction because of discretization. Minimization of total variation had proven to be the optimal method in one dimension, and it also showed itself to be applicable in one dimension with a fast and robust manner. Combining these two methods, that is, Radon transform and total variation yields a promising new technique that we can call R-TV. Estimation of the signal is not perfectly accurate, however, the algorithm is rather robust against the choice of this parameter. Only two constraints, total variation and the estimated variance do not allow perfect reconstruction of textures and details. Therefore, a third condition has to be taken into account. Since noise

can be differentiated in the frequency domain, Fourier transform and its digital signal processing filtering can be applied. As a second option, we can consider taking correlation as a cost. In nature, remote sensing images have uncorrelated noise. On the other hand, texture is usually correlated so that meaningful shapes show up. Taking this condition, the correlation as a constraint or cost will let us preserve the texture, therefore this is the main point that should be worked on.

Optimization of Radon transform is a research area. Radon transform is one of the basics of tomography imaging. Dynamic programming is a widely used tool in order to increase the efficiency of Radon transform. So, studying this research area can provide us to develop more efficient Radon based algorithms. See the papers of Basu and Bresler on dynamic programming and tomography imaging for details.

For $2 - D$ texture recovery total variation by Willsky is exploited. The separable implementation of this optimal $1 - D$ algorithm carries the technique to images. Moreover, this method turns out to preserve the edges with a surprisingly high success. This is because of the edge-preserving nature of total variation.

APPENDIX A

PROCESS AND OPTIMALITY OF PIECEWISE LINEAR POLYNOMIAL FITTING

Notation:

By $u[n]$ we mean the original spatial domain signal that function of n . Here, it is assumed that $u[n]$ is a noisy signal that is close to a piecewise linear signal $\check{u}[n]$. Our purpose is to obtain $\check{u}[n]$. By $d[n, t)$ it is meant the signal which is at the beginning equal to first difference of signal $u[n]$ with respect to spatial coordinate n and which evolves in time by total variation process given in [4]. Mathematically:

$$d[n, 0) = u[n] - u[n - 1] \tag{A.1}$$

and $d[n, t)$ should satisfy the following differential equation.

$$\dot{d}[n, t] = \frac{1}{m_i(d[n, t])} \{ \text{sgn}[\mu_{i+1}(d[n, t]) - \mu_i(d[n, t])] - \text{sgn}[\mu_i(d[n, t]) - \mu_{i-1}(d[n, t])] \}, \quad (\text{A.2})$$

for

$$n = k_i(d[n, t]), \dots, k_{i+1}(d[n, t]) - 1, \quad (\text{A.3})$$

$$i = 1, \dots, N_r(d[n, t]). \quad (\text{A.4})$$

Here, “ $\dot{d}$ ” denotes the derivative with respect to time. $m_i(d[n, t])$ is the length of region i . A region is the flat part of a signal at which the signal takes constant value $\mu_i(d[n, t])$. It can have a length of 1 as well. $k_i(d[n, t])$ is the index of left endpoint of region i . $N_r(d[n, t])$ is the number of regions in $d[n, t]$. We will call $\|\mathbf{s} - \mathbf{r}\|^2$ as $\Lambda(\mathbf{s}, \mathbf{r})$. By $u[m : n]$ it is meant the part of u between the spatial coordinates m and n . By $l_{p,p+1}$ we mean the signal constructed by concatenating l_p and l_{p+1} where l_p is a line signal. The region boundaries of $d[n, t]$ can be used to segment the original signal $u[n]$ in successive non-overlapping intervals that cover the whole spatial interval $1 \leq n \leq N$. In each of these intervals, the optimal least squares fit to the original signal by a line segment can be performed as given in equations (2.38) to (2.28). Then we can concatenate these line segments to obtain a piecewise linear approximation to $u[n]$ given by:

$$F_t(\mathbf{u}[n]) = \sum_{k=1}^K p \left(\frac{x - m[k, t]}{m[k+1, t] - m[k, t]} \right) (\alpha[k].n + \beta[k]). \quad (\text{A.5})$$

where, $m[n, t]$ is the signal of right endpoints of regions of $d[n, t]$. The length of $m[n, t]$ is K . $p(x)$ is the rectangular pulse of unit amplitude with support in $(0, 1)$ interval. Here, $\alpha[k]$ and $\beta[k]$ are the minimizers of $\|\mathbf{y}[m[k] : m[k+1]] - (\beta[k].\mathbf{1} + \alpha[k].\mathbf{x}[m[k] : m[k+1]])\|_2$. which can be obtained similar to (2.38) to (2.28).

In this example as shown in Figure A.1, $u[n]$ is chosen as a noisy version of piecewise linear signal $\check{u}[n]$.

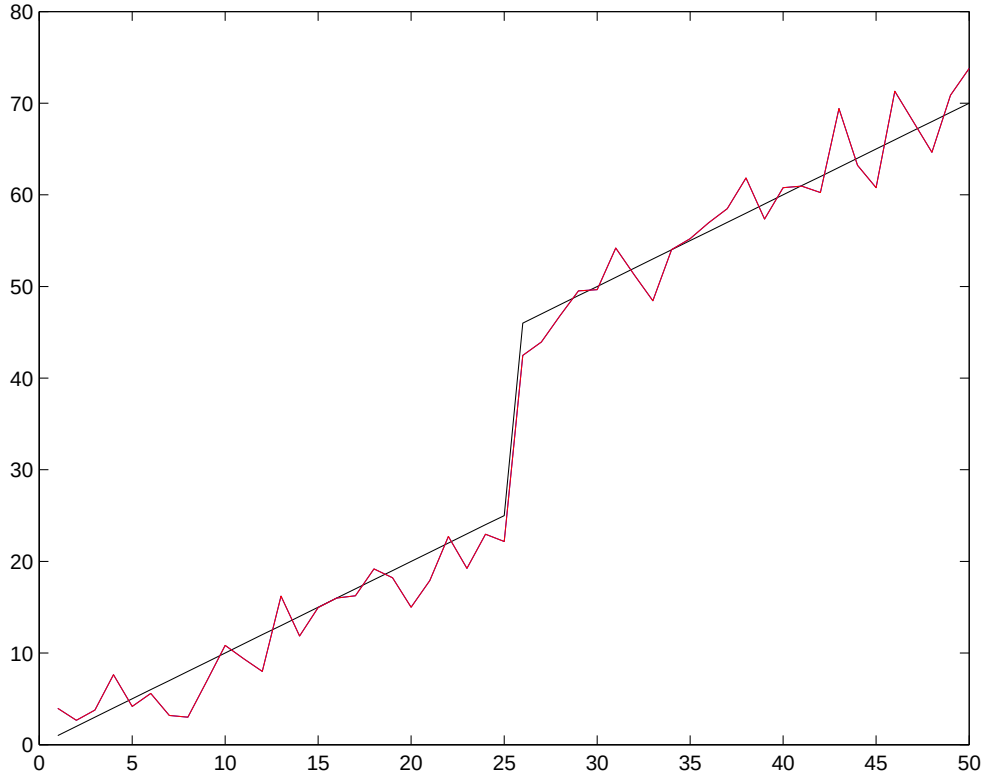


Figure A.1: The original signal $\check{u}[n]$ and the $\mathbf{u}[n]$ signal which is noisy.

Let us show the process with an example. The original signal $\check{u}[n]$ and its noisy version $\mathbf{u}[n]$ are shown in Figure A.1. $d[n, 0]$ which is equal to the first difference of $\mathbf{u}$ is shown in (A.2).

$F_t(u[n])$ and $d[n, t]$ after 34 iterations are shown in Figure A.3 and A.4, respectively. It can be seen that, $F_t(u[n])$ is more similar to the original signal and the total variation of $d[n, t]$ is less than $d[n, 0]$.

$F_T(u[n])$ which is the final solution is shown in Figure A.5 and $d[n, T]$ is shown in Figure A.6. After the total variation process, $d[n, t]$ has four regions. $F_T(u[n])$ is found by finding the line fitting of $u[n]$ in those regions of $d[n, t]$. To determine T , the stopping time of the total variation iterations, we use the noise standard deviation based stopping criterion introduced in [4]. We assumed that the noise standard deviation is known. Actually, by using robust estimation techniques, the noise standard deviation can also be estimated along with the iterations as well.

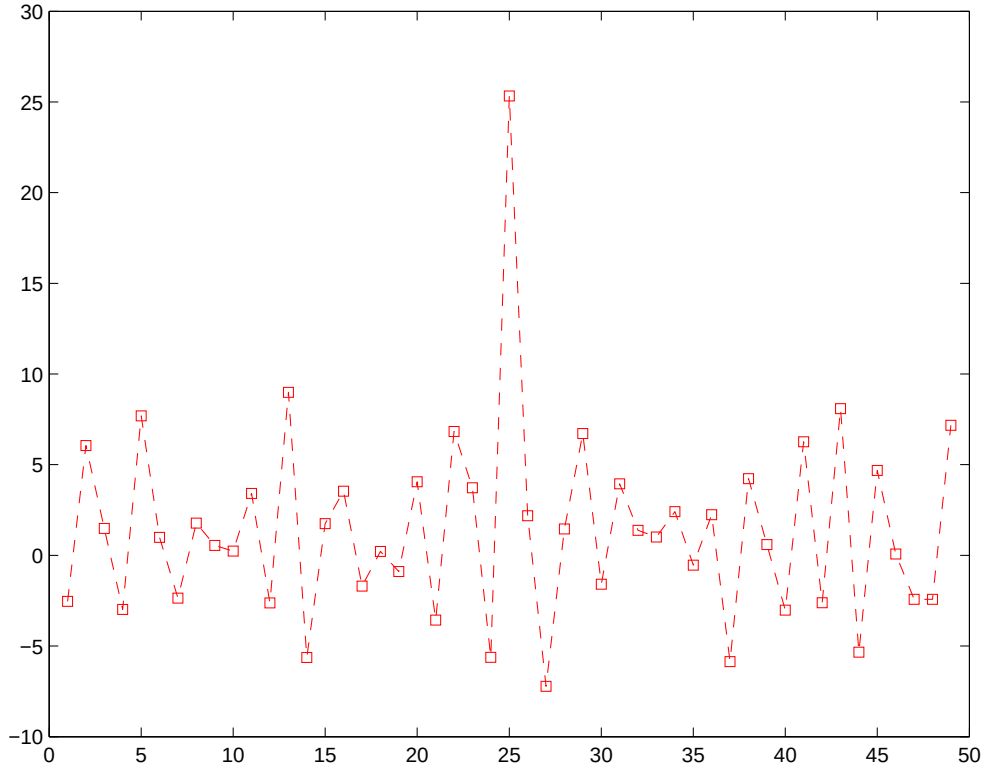


Figure A.2: $d[n, 0]$ which is equal to the first difference of $\mathbf{u}$.

The optimal solution to the problem

The total variation based piecewise linear signal approximation problem can be formally defined as:

$$\text{Find } v_{pl}(\sigma^2) = \underset{\frac{1}{N} \|u - v_{pl}\|^2 < \sigma^2}{\text{argmin}} \{TV(dn, t)\}$$

such that v_{pl} is a linear fit of $u[n]$ in the regions of $d[n, t]$.

In the following we will prove that $v(\sigma^2) = F_T(\mathbf{u}[n])$

Proposition 1: $\|u[n] - F_t(u[n])\|^2$ is a monotonically increasing function of time.

Proof of proposition 1:

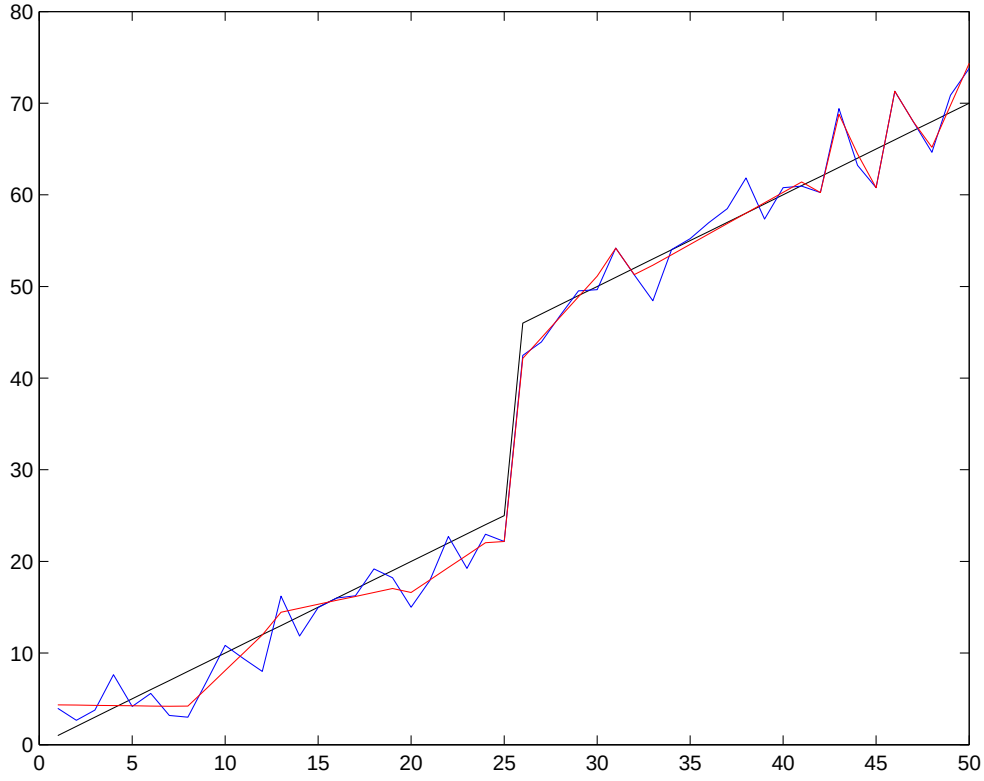


Figure A.3: The original signal and $F_t(\mathbf{u}[n])$.

As given in eqns. (2.33), (2.34), (2.35), (2.37) and (2.36). $F_t(u[n])$ evolves in time by merging two adjacent lines, therefore, as it will be shown below, $\|u[n] - F_t(u[n])\|^2$ increases during the evolution process.

If $\Lambda(u[m : p], l_1) < \Lambda(u[m : p], l_{2,3})$ then we can separate l_1 into two parts $l_1[m : n]$ and $l_1[n : p]$ and one of these lines has to be more similar to $u[n]$. That is, either $\Lambda(u[m : n], l_1[m : n]) < \Lambda(u[m : n], l_2)$ or $\Lambda(u[n : p], l_1[n : p]) < \Lambda(u[n : p], l_3)$

But these are not true. Therefore $\Lambda(u[m : p], l_1) > \Lambda(u[m : p], l_{2,3})$

The optimal technique for total variation minimization is the technique given in [4], so, that is used for total variation minimization of $d[n, t]$. The problem searches for piecewise linear fits of $u[n]$ in the regions of $d[n, t]$, therefore the fit with minimum difference norm is $F_T(u[n])$. For a fixed $TV(d[n, t])$ the minimum difference norm is with $F_T(u[n])$. Let us assume that for a signal v_{pl} with the

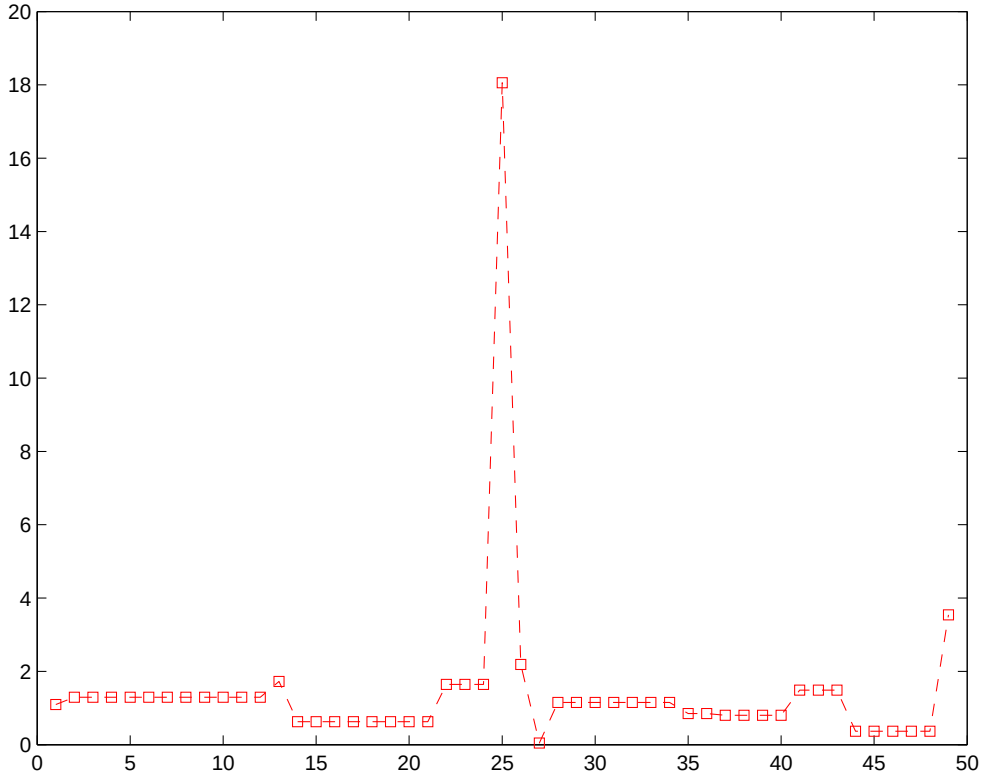


Figure A.4: $d[n, t]$ which is the result of total variation process on $d[n, 0]$.

same difference norm σ^2 as $F_T(u[n])$ the total variation of $d[n, t]$ is less. Since $TV(d[n, t])$ is a monotonically decreasing function of time, it means that time T' has passed T . At that time T' , the best fit is $F_{T'}(u[n])$ whose difference norm $\|u[n] - F_{T'}(u[n])\|^2$ is more than σ^2 , because $\|u[n] - F_t(u[n])\|^2$ is a monotonically increasing function of time. At the beginning, the difference norm was chosen to be the same but it turned out to be different and more, it means that, for a fixed difference norm, $TV(d[n, t])$ is the global minimum.

Proposition 2: $F_t(u[n])$ is not equal to the integral of $d[n, t]$.

Proof of Proposition 2:

If the signal $u[n]$ is taken as $\{0, 1, 6, 13, 15\}$ its first difference is equal to $d[n, 0] = \{1, 5, 7, 2\}$. Now, the issue is to perform total variation once in the first difference domain. The signal which stands for how many adjacent entries an entry is greater than is $\{-1, 0, 2, -1\}$. The signal that represents the absolute value

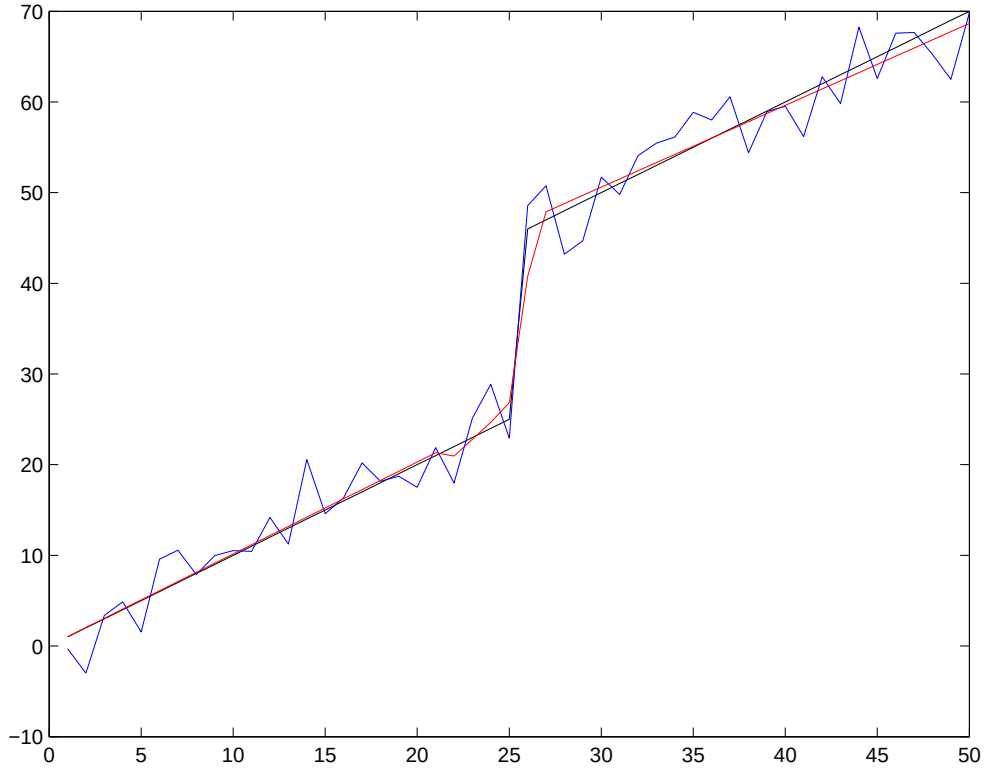


Figure A.5: The original signal and final solution $F_T(\mathbf{u}[n])$.

of first difference of $d[n, 0]$ is $\{4, 2, 5\}$. The maximum perturbation that can be made to the signal without changing the order of the entries is $\min(\{4, 1, 5/3\}) = 1$. Perturbing the signal directly proportional with 1, we find $\{2, 5, 5, 3\}$. This means that the best line of the 2nd, 3rd and 4th entries should be found because flat region in the first difference domain corresponds to line in the actual domain. If the line fitting for $\{1, 6, 13\}$ is found, the slope is found to be 6. This is as follows:

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \frac{12}{N_1(N_1-1)(N_1+1)} & \frac{-6}{N_1(N_1-1)} \\ \frac{-6}{N_1(N_1-1)} & \frac{2(2N_1+1)}{N_1(N_1-1)} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{N_1} i y_1(i) \\ \sum_{i=1}^{N_1} y_1(i) \end{pmatrix} \quad (\text{A.6})$$

for our problem:

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \frac{12}{3 \cdot 2 \cdot 4} & \frac{-6}{3 \cdot 2} \\ \frac{-6}{3 \cdot 2} & \frac{2 \cdot 7}{3 \cdot 2} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^3 i y_1(i) \\ \sum_{i=1}^3 y_1(i) \end{pmatrix}$$

Here, α is found to be 6. However, in the first difference domain the slope is given as 5. The proof is completed by this counterexample.

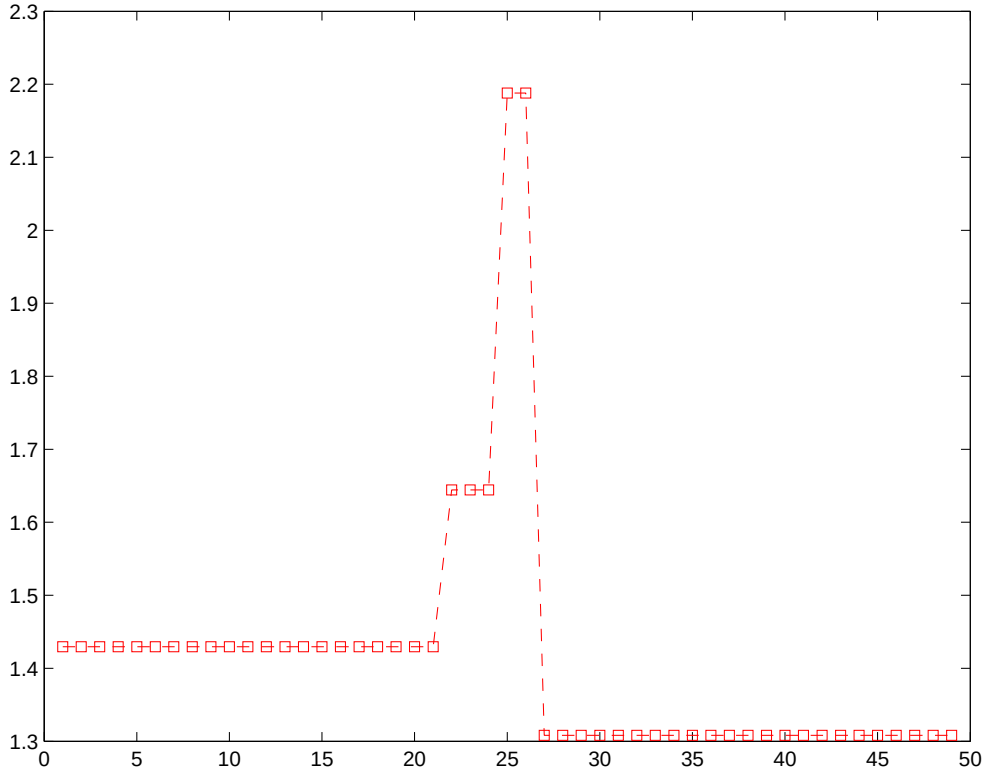


Figure A.6: $d[n, T]$ which is the result of total variation process on $d[n, 0]$.

Now, let us check optimality for a different problem.

$$\|\mathbf{u} - \mathbf{v}_{pl}(i_1, i_2, \dots, i_r)\|^2 \geq \|\mathbf{u} - \mathbf{v}_{pl}(i_{d_1}, i_{d_2}, \dots, i_{d_r})\|^2$$

The problem checks if difference norm for line fits in arbitrary regions is greater than the difference norm for line fits in the regions of $d[n, t]$.

Actually, it was checked for a signal whether the minimum difference norm is attained for two regions of $d[n, t]$. This turned out to be false, since the minimum difference norm is attained for finding linear fits in the regions separated by a boundary at the index 15 for a signal of length 20. However, two regions of $d[n, t]$ were separated at the index of 9. Our solution is heuristic from this perspective.

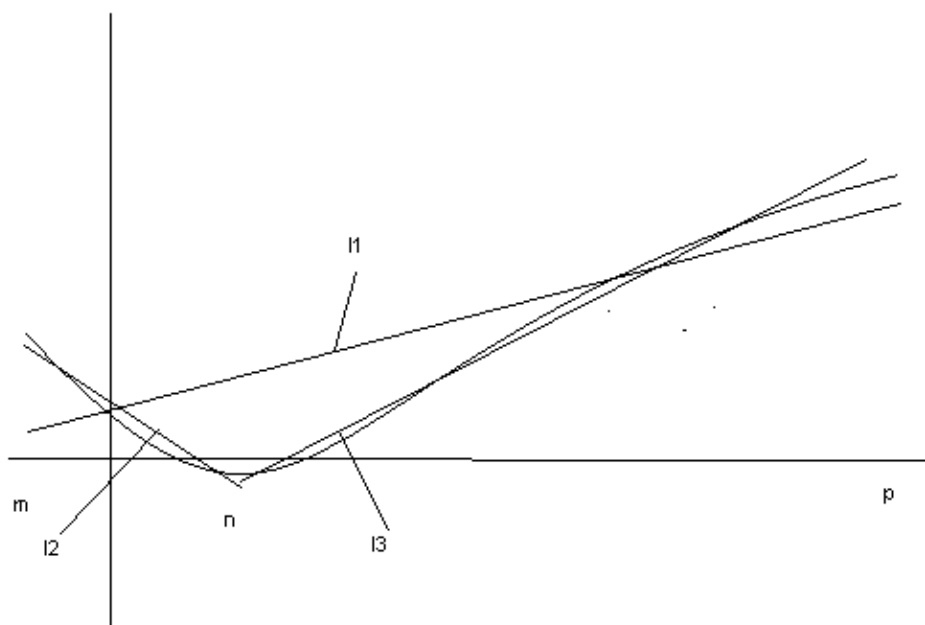


Figure A.7: Original signal, two lines to be merged and the merged line.

Bibliography

- [1] M. T. Heath, “Scientific computing lecture notes,” <http://www.cse.illinois.edu/heath/scicomp/notes/chap3.pdf>, 2002.
- [2] A. Yildiz and O. Arikan, “Signal denoising by piecewise continuous polynomial fitting,” *SIU*, April 2010.
- [3] J. L. Starck, E. J. Candes, and D. L. Donoho, “The curvelet transform for image denoising,” *IEEE Transactions for Image Processing*, vol. 11, pp. 670–684, 2002.
- [4] I. Pollak and A. S. Willsky, “Nonlinear evolution equations as fast and exact solvers of estimation problems,” *IEEE Transactions on Image Processing*, vol. 53, pp. 484–498, 2005.
- [5] U. Ozertem, D. Erdogmus, and O. Arikan, “Piecewise smooth signal denoising via principal curve projections,” *IEEE Workshop on Machine Learning for Signal Processing*, pp. 426–431, October 2008.
- [6] J. Linatti, “Performance comparison of ds signal matched filter acquisition using median filter and transversal filter as an aiding device in awgn channel and cw jamming,” *IEEE*, pp. 1181–1185, 1995.
- [7] J. R. C. Williamson and V. Dolat, “Reflective-array matched filter for a 16-pulse radar burst,” *IEEE*, 1975.

- [8] P. Tortoli, F.Guidi, and C.Atzeni, “Digital vs saw matched filter implementation for radar pulse compression,” *IEEE*, pp. 199–202, 1994.
- [9] A. Gunawardena and D. Longstaff, “A matched filter based synthetic aperture radar(sar) algorithm for stepped frequency ground penetrating radar,” *IEEE*, pp. 239–243, 1995.
- [10] K. Arora and R. K.Arora, “Matched filter response to a linear fm signal transmitted through a phased array,” *IEEE transactions on aerospace and electronic systems*, vol. AES-12, 1976.
- [11] R. K. M.R. Patel, “Array dispersion effects on matched filter output for a linear fm signal,” *IEEE transactions on aerospace and electronic systems*, 1978.
- [12] D.S.Zrnic and R.J.Doviak, “Matched filter criteria and range weighting for weather radar,” *IEEE transactions on aerospace and electronic systems*, 1978.
- [13] J. D. Rhodes, “Matched filter theory for dopple-invariant pulse compression,” *IEEE transactions on circuit theory*, vol. CT-19, pp. 53–59, 1972.
- [14] R. E. Millett, “A matched-filter pulse-compression system using a nonlinear fm waveform,” *IEEE transactions on aerospace and electronic systems*, pp. 73–78, 1970.
- [15] X.A.Shen, Y. Bai, and R.Kachru, “Texture preserving variational denoising using an adaptive fidelity term,” *OPTICS LETTERS*.
- [16] P. Courmontagne, “A review on stochastic matched filter based denoising methods for sas images despeckling,” *IEEE*, 2007.
- [17] P. Courmontagne, “Sas images denoising:the jointly use of an autoadaptive mean filter and the stochastic matched filter,” *IEEE*.

- [18] P. Courmontagne and F. Chaillan, "The adaptive stochastic matched filter for sas images denoising," *IEEE*, 2006.
- [19] T. Shibano, K. Lizuka, M. Miyamoto, M. Osaka, R. Miyama, and A. Kito, "Matched filter for ds-cdma of up to 50mchip/s based on sampled analog signal processing," *IEEE International Solid State Circuits Conference*, 1997.
- [20] S. Shiga and T. Ikegami, "Digital matched filter detection of gmsk/ds/ds signal," *IEEE*, pp. 1281–1284, 2000.
- [21] M.-L. Liou and T.-D. Chiueh, "A low power digital matched filter for direct sequence spread spectrum signal acquisition," *IEEE journal of solid state circuits*, vol. 36, pp. 933–943, 2001.
- [22] S.-H. Lin, Y.-M. Yeh, and B. Chen, "Exploiting polynomial fit histogram equalization and temporal average for robust speech recognition," *Inter-speech*, pp. 1–18, 2006.
- [23] P. J. Burt, "Moment images polynomial fit filters and the problem of surface interpolation," *IEEE*, pp. 144–152, 1988.
- [24] V.G.Spokoiny, "Estimation of a function with discontinuities via polynomial fit with an adaptive window choice," *The annals of statistics*, 1998.
- [25] M. T. Heath, *Scientific Computing*. New York: McGraw-Hill, 1997.
- [26] F. van den Bergh and A. P. Engelbrecht, "Cooperative learning in neural networks using particle swarm optimizers," *South African Comput. J.*, vol. 26, pp. 84–90, November 2000.
- [27] J. Mao, Z. Wu, and X. Wu, "A tdma scheduling scheme for many toone communications in wireless sensor networks," *Comput. Commun.*, vol. 30, pp. 863–872, 2007.

- [28] Z. Yang, D. Qi, and L. Yang, "Signal period analysis based on hilbert-huan transform anmd its application to texture analysis," *Proceedings of the Third International Conference on Image and Graphics*, 2004.
- [29] P. P. Kanjilal and S. Palit, "Extraction of multiple periodic waveforms from noisy data," *IEEE*, vol. II, pp. 361–364, 1994.
- [30] ISM2004, "Extraction of signals from noise," *PX4508(7) SNR*, 2004.
- [31] J. M. Orwell, J. F. Boyce, J. F. Haddon, and G. Watson, "Detecting periodic structure," *Crown copyright*, 1998.
- [32] O. G. Guleryuz, "Iterated denoising for image recovery," *Proceedings of the data compression conference, Epson Palo Alto Laboratory*, 2002.
- [33] G. Gilboa, N. Sochen, and Y. Y.Zeevi, "Texture preserving variational denoising using an adaptive fidelity term,"
- [34] D. Bratton and J. Kennedy, "Defining a standard for particle swarm optimization," *Proceedings of the 2007 IEEE Swarm Intelligence Symposium*, pp. 291–294, January 2007.
- [35] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
- [36] D. L. Donoho and M. R. Duncan, "Digital curvelet transform: Strategy, implementation and experiments," *Department of Statistics, Stanford University*, 1999.
- [37] J. L. Starck, F. Murtagh, E. J. Candes, and D. L. Donoho, "Gray and color image contrast enhancement by the curvelet transform," *IEEE Transactions on Image Processing*, vol. 12, pp. 706–717, 2003.
- [38] T.F.Chan, G.H.Golub, and P. Mulet, "A nonlinear primal-dual method for tv-based image restoration," in *Proc. ICAOS:Images, Wavelets, PDEs, Paris, France*, pp. 241–252, 1996.

- [39] S. Didas, *Higher order variational methods for noise removal in signals and images*. PhD thesis, Saarland Univ., Saarbrücken, Germany, 2004.
- [40] Y. Li and F. Santosa, “A computational algorithm for minimizing total variation in image restoration,” *IEEE Transactions on Image Processing*, vol. 5, pp. 987–995, 1996.
- [41] J. Darbon and M. Sigelle, “A fast and exact algorithm for total variation minimization,” *Telecom Paris*, 2005.
- [42] D. C. Dobson and C. R. Vogel, “Convergence of an iterative method for total variation denoising,” *SIAM J. Numer. Anal.*, vol. 34, pp. 1779–1791, 1997.
- [43] V. Agarwal, “Total variation regularization and l-curve method for the selection of regularization parameter,” *ECE*, 2003.
- [44] C. R. Vogel and M. E. Oman, “Fast, robust total variation-based reconstruction of noisy, blurred images,” *IEEE Transactions on Image Processing*, vol. 7, pp. 813–824, 1998.
- [45] P. Blomgren, T. F. Chan, and P. M. C. K. Wong, “Total variation image restoration: Numerical methods and extension,” *IEEE*, pp. 384–387, 1997.
- [46] P. L. Combettes and J.-C. Pesquet, “Image restoration subject to a total variation constraint,” *IEEE Transactions on Image Processing*, vol. 13, pp. 1213–1222, 2004.
- [47] P. Combettes and J. Pesquet, “Incorporating total variation information in image recovery,” *IEEE*, vol. 3, pp. 373–376, 2003.
- [48] F. Malgouyres, “Combining total variation and wavelet packet approaches for image deblurring,” *IEEE*, pp. 57–64, 2001.
- [49] T. F. Chan and H. Zhou, “Total variation improved wavelet thresholding in image compression,” *IEEE*, 2000.

- [50] R. Storn and K. Price, "Differential evolution, a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, 1997.
- [51] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm," *IEEE Trans. Evol. Comput.*, vol. 6, pp. 182–197, April 2002.
- [52] M.Clerc and J.Kennedy, "The particle swarm- explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 58–73, 2002.
- [53] U. Ozertem and D. Erdogmus, "Mean shift, mode seeking, and clustering," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790–799, January 1995.
- [54] U. Ozertem and D. Erdogmus, "Local conditions for critical and principal manifolds," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, vol. 62, pp. 291–294, January 2008.
- [55] D. H. Shin, R. H. Park, and S. Yang, "Block-based noise estimation using adaptive gaussian filtering," *IEEE Transaction on Consumer Electronics*, vol. 51, pp. 218–226, February 2005.
- [56] K. Rank, M. Lendl, and R. Unbehauen, "Estimation of image noise variance," *Proc. IEE Vis. Image Signal Process*, vol. 146, pp. 80–84, April 1999.
- [57] J. Rosen, "The gradient projection method for nonlinear programming," *J. Soc. Ind.Appl.Math.*, vol. 9, pp. 514–532, 1961.
- [58] A. Lopesm, R. Touzi, and E. Rezry, "Adaptive speckle filters and scene heterogeneity," *IEEE Trans. Geoscience and Remote Sensing*, vol. 28, pp. 992–1000, 1990.

- [59] J. Starck and F. Murtagh, “Multiscale entropy filtering,” *Signal Processing*, vol. 76, pp. 147–165, 1999.
- [60] P. Blomgren and T. Chan, “Modular solvers for constrained image restoration problems,” *Numer. Linear Algebra Applicat.*, vol. 9, pp. 347–358, 2002.
- [61] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, “Wavelet-based signal processing using hidden markov models,” *IEEE Transactions on Signal Processing (Special Issue on Wavelets and Filterbanks)*, 1998.
- [62] G. Steidl, J. Weickert, P. T. Brox, and M. Welk, “On the equalvalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides,” *SIAM J. Numer. Anal.*, vol. 42, pp. 686–713, 2004.
- [63] C. Bouman and K. Sauer, “An edge-preserving method for image reconstruction from integral projections,” in *Proc. Conf. Info. Sci. Syst. Baltimore, MD, Mar*, pp. 383–387, 1991.
- [64] C. Bouman and K. Sauer, “An edge-preserving method for image reconstruction from integral projections,” *Numer. Math.*, vol. 76, 1997.
- [65] L. Alvarez, P. Lions, and J. Morel, “Image selective smoothing and edge detection by nonlinear diffusion,” *SIAM J. Num.*, vol. 29, p. 845, 1992.
- [66] D. Geman and G. Reynolds, “Constraint restoration and the recovery of discontinuities,” *IEEE*, 1990.
- [67] E. Fatemi, S. Osher, and L. Rudin, “Removing noise without excessive blurring,” *Cognitech report 5*, vol. 60, pp. 259–268, 1992.
- [68] L. Rudin and S. Osher, “Reconstruction and enhancement of signals using non-linear non-oscillatory variational methods,” *Cognitech report 7*, 1993.
- [69] S. Kim, “Numerical modeling for the recovery of fine structures in pde-based image denoising,”

- [70] P. P. Kanjilal and S. Palit, "On multiple pattern extraction using singular value decomposition," *IEEE transactions on signal processing*, vol. 43, pp. 1536–1540, 1995.
- [71] J. Vanderschoot, D. Callaerts, W. Sansen, J. Vandwalle, G. Vantrappen, and J. Janssens, "Two methods for optimal mecg elimination and f,cg detection from skin delectrode signals," *IEEE Transactions on Biomedical Engineering*, vol. BME-34, 1987.
- [72] B. Widrow and S. D. Stearns, "Adaptive signal processing," *Prentice Hall Inc., Englewood Cliffs, N.J. 07632*, 1986.
- [73] A. R. Rao and G. L. Lohse, "Towards a texture naming system: Identifying relevent dimensions of texture.," *In Proc. IEEE Conf. Visualization,*, vol. 60, pp. 220–227, 1993.
- [74] P. Blomgren and T. F. Chan, *Total variation methods for restoration of vector valued images(Ph. D. thesis)*. PhD thesis, UCLA Dept. of Math., 1998.
- [75] A. Chambolle and P. L. Lions, "Image recovery via total variational minimization and related problems," *Numer. Math.*, vol. 76, pp. 167–188, 1997.
- [76] F. Catte, P. Lions, M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM J. Numer. Anal.*, vol. 29, pp. 182–193, 1992.
- [77] P. N. Blomgren and T. Chan, "Color tv: Total variation methods for restoration of vector valued images," *IEEE Trans. Image Processing*, vol. 7, pp. 304–309, 1998.
- [78] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "Using geometry and iterated refinement for inverse problems(1):total variation based image restoration," *CAM report Department of Mathematics UCLA*, pp. 4–13, 2004.

- [79] S. Kim, “Edge preserving noise removal: Motion by mean curvature,” *Submitted to SIAM J. Sci. Comput.*, vol. 60, pp. 259–268, 1992.
- [80] L. A. Vese and S. Osher, “Modeling textures with total variation minimization and oscillation patterns in image processing,” *UCLA CAM report*, pp. 02–19, 2002.
- [81] D. M. Strong and T. F. Chan, “Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing,” *UCLA CAM rep*, pp. 94–46, 1996.
- [82] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *PAMI 12*, vol. 7, pp. 629–639, 1990.
- [83] Y. Meyer, “Oscillatory patterns in image processing and nonlinear evolution equations,” *University Lecture Series*, AMS, vol. 22, 2001.
- [84] L. Blanc-Feraud, P. Charbonnier, G. Aubert, and M. Barlaud, “Nonlinear image processing: modelling and fast algorithm for regularization with edge detection,” *In Proc. IEEE ICIP-95*, vol. 1, pp. 474–477, 1995.
- [85] J. Vesterstrom and R. Thomsen, “A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems,” *Proceedings of the 2004 IEEE Congress on Evolutionary Computation*, vol. 2, pp. 1980–1987, 2004.
- [86] R. Eberhart and Y. Shi, “Comparison between genetic algorithms and particle swarm optimization,” *Proceedings of the 7th International Conference on Evolutionary Programming*, vol. VII, pp. 611–616, 1998.
- [87] U. Paquet and A. Engelbrecht, “Training support vector machines with particle swarms,” *Proceedings of the International Joint Conference on Neural Networks*, 2003.

- [88] S. DY and H. C.Y., “A hybrid particle swarm optimization for job shop scheduling problem. computers, industrial engineering 2006,” vol. 51, pp. 791–808.