Titre: Title:	Adaptive Piecewise-Smooth Modelling for Gaussian Noise Reduction
Auteur: Author:	Noel Plascencia Diaz
Date:	2026
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Plascencia Diaz, N. (2026). Adaptive Piecewise-Smooth Modelling for Gaussian Noise Reduction [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/76859/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/76859/
Directeurs de recherche: Advisors:	Antoine Saucier
Programme: Program:	Doctorat en Mathématique

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Adaptive Piecewise-Smooth Modelling for Gaussian Noise Reduction

NOEL PLASCENCIA DIAZ
Département de mathématiques et de génie industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Mathématiques

Mai 2026

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Adaptive Piecewise-Smooth Modelling for Gaussian Noise Reduction

présentée par **Noel PLASCENCIA DIAZ**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Luc ADJENGUE, président
Antoine SAUCIER, membre et directeur de recherche
Jonathan JALBERT, membre
Marco Arieli VALDEZ membre externe

DEDICATION

To everyone that contributes to
harmonic existence in a world
often defined by conflict and confusion.
I am infinitely grateful to my wife.
Mitsui: your love, tenderness and
patience keep me motivated every day.

"Perseverance yields favorable results
to those who are firm and unyielding."
-King Wen of Zhou

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who have supported me throughout this journey. First and foremost, to my wife Mitsui, for her endless love, patience, and encouragement. Your smile always keeps me going and doing my best. I thank my parents, my brother Erick and my sister Clau, for their continuous support, love and guidance. I am thankful to Mitsui Saito-Quezada and Isaias Salgado-Hugarte for providing a constant source of motivation and support.

I am profoundly grateful to my research advisor, Professor Antoine Saucier, whose creative mindset and unique approach to science were both challenging and inspiring. His guidance not only helped shaping this work but also pushed me to grow as a professional researcher and as an individual.

I also extend my sincere thanks to Professors Luc Adjengue, Richard Labib and Marco Arieli Herrera-Valdez for their critical observations and insightful comments, which were invaluable to the development and improvement of this work.

My appreciation goes to the friends I met in Montréal, as well as those who have been by my side for many years. I am also thankful to Dr. Abhinav Sharma, and everyone who, in one way or another, helped me stay patient and focused, keeping me determined to complete this degree.

This work was made possible through the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC), for which I am deeply grateful. I also acknowledge École Polytechnique de Montréal for providing a stimulating and enriching environment that fostered my academic growth.

RÉSUMÉ

Cette thèse de doctorat aborde le problème de la réduction du bruit blanc gaussien additif (BBGA) dans les signaux 1D et les images 2D. Bien que la littérature englobe des méthodes de réduction du bruit (RB) allant des filtres de base aux modèles avancés d'apprentissage automatique, les approches actuelles se heurtent souvent à des limitations dans l'équilibre entre l'adaptabilité, l'efficacité computationnelle et la RB sans artefacts.

Nous avons introduit de nouvelles méthodes de réduction du bruit gaussien conçues pour intégrer une RB efficace et performante, tout en assurant une large applicabilité à des types de signaux variés grâce à une modélisation adaptative aux données. Le développement de cette thèse est guidé par les objectifs suivants :

O1 : Développer une méthode de réduction du BBGA sans apprentissage pour les signaux 1D par morceaux lisses.

O2 : Étendre le cadre développé dans **O1** aux images 2D.

Dans les deux cas, la méthodologie doit : 1) réduire le bruit tout en préservant les détails structurels, 2) reposer principalement sur un réglage des paramètres adaptatif aux données, et 3) être computationnellement efficace, permettant une application pratique.

Nous avons développé le Modèle par Morceaux Lisses (MML), qui formule la réduction du bruit comme un problème de sélection de modèle. Le signal sans bruit est estimé en ajustant des polynômes de degré ≤ 2 sur tous les intervalles possibles et en sélectionnant une séquence optimale de morceaux non chevauchants par minimisation de critères d'information. Nous avons démontré empiriquement qu'en agrégeant plusieurs modèles MML locaux obtenus dans des fenêtres glissantes chevauchantes, il est possible de surpasser des versions optimales de plusieurs méthodes de RB de la littérature. Nous avons fourni des algorithmes pour estimer les paramètres du modèle directement à partir des données bruitées et introduit une formulation par Plus Court Chemin pour résoudre efficacement le problème de minimisation du critère d'information.

Pour réduire le bruit des images, nous avons introduit S3F (Fusion de Lissage et d'Auto-Similarité), un cadre sans apprentissage qui oriente adaptativement les pixels vers des stratégies de réduction du bruit complémentaires. S3F se compose de trois composantes couplées : un estimateur du niveau de bruit, un débruiteur de régions lisses (MML appliqué sur des fenêtres glissantes sur l'image linéarisée), et un débruiteur de régions texturées basé sur l'auto-similarité.

Le MML appliqué sur des fenêtres glissantes (MML-FG) génère une carte de variance par pixel qui sert de détecteur de texture pour partitionner l'image. Les pixels dépassant un seuil de variance (indiquant une texture à haute fréquence) sont traités par Régression à Rang Réduit (RRR), qui regroupe des patchs similaires et utilise une Décomposition en Valeurs Singulières tronquée avec sélection du rang basée sur les données pour la reconstruction. L'image finale est assemblée en utilisant l'estimée MML-FG pour les pixels lisses et en faisant la moyenne des estimées MML-FG et RRR pour les pixels texturés.

Une motivation fondamentale de ce travail est la reconnaissance que la plupart des techniques de RB existantes sont insuffisantes sur au moins l'un des aspects suivants :

- Réduction du bruit de haute qualité : une RB efficace sans introduction d'artefacts
- Adaptabilité : produire une RB de haute qualité pour une grande variété de signaux
- Efficacité computationnelle : atteindre une RB de haute qualité tout en évitant le recours à des algorithmes gourmands en ressources ou à de larges jeux de données d'entraînement.

Applications et Perspectives

Les applications du cadre MML comprennent la réduction du bruit thermique dans les capteurs, les traces sismographiques et les électrocardiogrammes. Étant donné que le MML a la capacité d'identifier des segments non lisses, son utilisation pour la détection de pics pourrait constituer une autre application prometteuse, notamment dans les signaux biomédicaux et l'analyse de séries temporelles. Les applications de S3F incluent la réduction du bruit dans les images biomédicales et de microscopie.

Les travaux futurs pourraient se concentrer sur l'extension des méthodes à des modèles de bruit non gaussiens ou dépendants du signal, des extensions multi-échelles du modèle RRR, et des schémas de pondération optimaux pour l'inférence multi-modèles sous des structures de corrélation inconnues.

ABSTRACT

This doctoral thesis addresses the problem of reducing additive white Gaussian noise (AWGN) in 1D signals and 2D images. While the literature encompasses noise reduction (NR) methods ranging from basic filters to advanced machine learning models, current approaches often face limitations in balancing adaptability, computational efficiency, and artifact-free NR.

We introduced novel Gaussian noise reduction methods designed to integrate efficient and effective NR, while ensuring broad applicability to diverse signal types via data-adaptive modeling. The development of this thesis is guided by the following objectives:
O1: Develop a training-free AWGN reduction method for 1D piecewise-smooth signals.
O2: Extend the framework developed in **O1** to 2D images.
In both cases, the methodology should: 1) reduce noise while preserving structural details, 2) rely primarily on data-adaptive parameter tuning, and 3) be computationally efficient, allowing practical application.

We developed the Piecewise-Smooth Model (PSM), which frames noise reduction as a model selection problem. The noise-free signal is estimated by fitting polynomials of degree ≤ 2 across all possible intervals and selecting an optimal sequence of non-overlapping pieces via information criteria minimization. We empirically demonstrated that by aggregating multiple local PSM models obtained in overlapping sliding windows, it is possible to outperform optimal versions of multiple NR methods in the literature. We provided algorithms to estimate model parameters directly from the noisy data and introduced a Shortest Path optimizer formulation to efficiently solve the information criterion minimization problem.

To reduce noise from images, we introduced S3F (Smoothness and Self-Similarity Fusion), a training-free framework that adaptively routes pixels to complementary noise reduction strategies. S3F consists of three coupled components: a noise-level estimator, a smooth-region denoiser (PSM applied on sliding windows over the linearized image), and a textured-region denoiser based on self-similarity.

The PSM applied over sliding windows (SW-PSM) generates a per-pixel variance map that acts as a texture detector to partition the image. Pixels exceeding a variance threshold (indicating high-frequency texture) are processed via Reduced-Rank Regression (RR), which groups similar patches and uses truncated Singular Value Decomposition with data-driven rank selection for reconstruction. The final image is assembled by using the SW-PSM estimate for smooth pixels and averaging the SW-PSM and RR estimates for textured pixels.

A fundamental motivation for this work is the recognition that most existing NR techniques fall short in at least one of the following aspects:

- High-quality NR: Effective NR without introducing artifacts
- Adaptability: Producing high-quality NR for a wide variety of signals
- Efficient Computation : Achieving high-quality NR while avoiding the need for resource-intensive algorithms or large training datasets.

Applications and Future Directions

Applications of the PSM framework include the reduction of thermal noise in sensors, seismographic traces, and electrocardiograms. Since PSM has the capacity of identifying non-smooth segments, using it for spike detection could be another promising application, particularly in biomedical signals and time-series analysis. Applications of S3F include noise reduction from biomedical and microscopy images.

Future work may focus on extending the methods to handle non-Gaussian or signal-dependent noise models, multiscale extensions of the RR model, and optimal weighting schemes for multi-model inference under unknown correlation structures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

1D	One-Dimensional
2D	Two-Dimensional
AIC	Akaike Information Criterion
AWGN	Additive White Gaussian Noise
BM3D	Block-Matching and 3-D Filtering
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
IMF	Intrinsic Mode Functions
IVW	Inverse Variance Weighting
ML	Machine Learning
MLE	Maximum Likelihood Estimation
MSE	Mean Squared Error
NR	Noise Reduction
PSNR	Peak Signal-to-Noise Ratio
PSM	Piecewise-Smooth Model
RMSE	Root Mean Squared Error
SD	Spatial Domain
SNR	Signal-to-Noise Ratio
SNRI	Signal-to-Noise Ratio Improvement
SSL	Self-Supervised Learning
SW-PSM	Sliding-Window Piecewise-Smooth Model
TD	Transform Domain
UW	Uniform Weighting
VMD	Variational Mode Decomposition

TABLE OF SYMBOLS

Symbol	Description
$d_k \in \mathbb{N}$	Number of curve-fitting parameters for the k^{th} acceptable fit.
E_k	Sum of square errors for the k^{th} acceptable fit.
J_A	Set of all acceptable fits for all the possible intervals in the signal considered.
$\ell_k \in \mathbb{N}$	Support size for the k^{th} acceptable fit.
$m \in \mathbb{N}$	Total number of acceptable curve fittings.
$N \in \mathbb{N}$	Dimension of signal x, with $N >> 1$ typically.
S	Support of the discrete signal x, equal to $\{0,\ 1,\ 2, ...,\ N-1\}$.
$w \in \mathbb{R}^N$	Discrete noise, assumed to be a zero-mean constant-variance σ^2 AWG noise.
$x \in \mathbb{R}^N$	Discrete signal to be estimated.
$y \in \mathbb{R}^N$	Discrete noisy signal defined as $y = x + w$.
$\hat{\sigma} \in \mathbb{R}$	Noise-level estimate associated with y.
$M \in \mathbb{N}$	Total of smooth pieces used in a synthetic noise-free signal x.
$\Lambda \in \mathbb{N}$	Total number of signal model parameters.
$\theta_k \in \{0, 1\}$	Binary variable indicating if the k^{th} acceptable fit is used in the global covering.
$\Xi[i]$	Set of PSM point-estimates associated with the i^{th} noisy observation.
$X \in \mathbb{R}^{n \times n}$	Discrete image to be estimated ($n \in \mathbb{Z} >> 1$).
$W \in \mathbb{R}^{n \times n}$	Discrete noise signal, assumed to be a zero-mean constant-variance σ^2 AGW noise.
$Y \in \mathbb{R}^{n \times n}$	Discrete noisy image $Y := X + W$.
$\hat{\sigma} \in \mathbb{R}$	Noise-level estimate.
$\hat{x} \in \mathbb{R}^{n'}$	Estimate of noisy vector $y \in \mathbb{R}^{n'}$.
$\Delta\hat{x} \in \mathbb{R}^{n'}$	Vector containing the variance of each element in $\hat{x}$.
$V(Y)$	Vectorization operator applied to Y.
$V_p[Y]$	Subset of vectorized signals whose supports include the pixel p.
$\hat{X}[p]$	Final pixel-value estimate associated with te pixel at position p.
$\sigma[\hat{X}[p]]$	Standard deviation estimate of $\hat{X}[p]$.
$\hat{X} \in \mathbb{R}^{n \times n}$	SPSM-V image estimate.

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Noise reduction (NR), particularly for Additive White Gaussian noise (AWGN), is a critical task in signal processing across multiple domains. While numerous techniques have been proposed to suppress AWGN in acquired signals, both traditional and modern approaches often struggle to balance noise suppression, detail preservation, computational efficiency, and cross-domain performance. This thesis addresses these challenges by introducing the Piecewise-Smooth Model (PSM) method, and its extension, the Sliding-Window Piecewise-Smooth Model (SW-PSM).

SW-PSM is a principle-based framework that integrates fast curve-fitting, a novel model selection approach, and adaptive signal representation via local model aggregation. The SW-PSM framework is designed to reduce noise while preserving the fine-scale details of the signal via data-adaptive modeling, without requiring computing-heavy routines or external datasets. Furthermore, the first-principled design of PSM enables generalization to higher-dimensional signals, such as images. This combination of effectiveness, generality, and efficiency represents a significant advancement in signal NR.

1.1 Basic Concepts

Noise reduction (NR) represents a fundamental challenge in signal and image processing, focusing on the enhancement of signal quality through the suppression of unwanted noise components while preserving the structural features of the signal. These features include edge discontinuities, fine-scale textures, and localized signal variations that may carry essential information. Effective NR becomes particularly significant in specialized applications where signal quality directly impacts analytical outcomes and decision-making processes. This includes communication systems [1–3], speech processing applications [4, 5], and biomedical signal analysis [6, 7],

The AWGN model represents one of the most widely investigated noise distributions in signal processing, owing to both its mathematical tractability and its frequent occurrence in practical applications [8].

This work addresses the problem of recovering a piecewise continuous signal $x \in \mathbb{R}^n$ from its noise-corrupted observation $y \in \mathbb{R}^n$, where the degradation process follows the AWGN model. Formally, we model the noisy signal y as

$$y = x + w, \tag{1.1}$$

where the entries in the noise vector w are independently and identically distributed (*i.i.d.*) Gaussian random variables with zero mean and unknown non-zero standard deviation $\sigma \in \mathbb{R}$. NR methods produce an estimate $\hat{x} \in \mathbb{R}^n$ of the unknown noise-free signal x. The NR performance associated with a signal estimate $\hat{x}$ will be assessed quantitatively with the Signal-to-Noise Ratio Improvement (SNRI), defined as

$$\text{SNRI} = \frac{||y - x||^2}{||\hat{x} - x||^2},$$

A signal estimate $\hat{x}$ demonstrates improved quality over the noisy observation y when SNRI $>$ 1, with higher SNRI values corresponding to more effective noise reduction performance. For additional quality assessment, we consider the Peak Signal-to-Noise Ratio (PSNR), defined as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\max(x)^2}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2} \right).$$

The PSNR measures the ratio between the maximum possible signal power, given by $\max(x)^2$ and the mean squared error of the reconstruction. It is expressed in decibels (dB), where higher values indicate better reconstruction quality. A PSNR of 0 dB corresponds to a reconstruction whose mean squared error equals the peak signal power, while each increase of 10 dB corresponds to a tenfold reduction in mean squared error.

1.1.1 Definition of signals

Definition 1 *A function $f(v) : I \subseteq \mathbb{R} \to \mathbb{R}$ is said to be piece-wise smooth in an interval $I \subseteq \mathbb{R}$ if it is continuous in I except for a finite number of jump discontinuities, and if in addition its first derivative $f'(v)$ is piecewise continuous.*

The discrete noise-free signal is denoted by $x \in \mathbb{R}^n$, where $n \gg 1$ is an integer. The signal x is assumed to be obtained by sampling an unknown piecewise smooth function $f : \mathbb{R} \to \mathbb{R}$, representing a time-varying (or spatially varying) quantity. In the discrete signal context, the i^{th} component of x is $x[i] := f(i\,\Delta t)$ for $i \in S$, where the set $S = \{0, 1, \ldots, n-1\}$ will be called the signal support, and $\Delta t > 0$ is a constant denoting the spacing between consecutive samples of f.

Since $f(v)$ is piecewise smooth, it can be well approximated by a Taylor series on each smooth piece. Consequently it should be possible to find many intervals within S where the signal x can be approximated by a polynomial. In particular, there should exist a collection of adjacent and non-overlapping intervals that covers the signal support S in such a way that x can be well approximated by a polynomial on each interval. Note that considering disjoint sets facilitates the modeling of discontinuities that may occur between adjacent smooth pieces.

1.1.2 Akaike Information Criterion and Bayesian Information Criterion

A key challenge in statistical modeling is selecting an appropriate model from a set of candidates, as models with too many parameters relative to available data frequently lead to overfitting. Given observed data $y \in \mathbb{R}^n$ and a parametric model $p(y \mid \theta)$ with parameter vector $\theta \in \mathbb{R}^d$, maximum likelihood estimation (MLE) seeks

$$\hat{\theta} = \arg\max_{\theta} \, \mathcal{L}(\theta \mid y),$$

where $\mathcal{L}(\theta \mid y) = \log p(y \mid \theta)$ is the log-likelihood. While MLE is asymptotically efficient, it tends to favor overly complex models that fit the noise rather than the underlying signal when d is large relative to n [9].

This motivates the use of penalized likelihood methods, which complement the log-likelihood with a term that penalizes overparameterization. A framework for deriving such penalties is the Kullback-Leibler (KL) divergence, which measures the information lost when a model $p(y \mid \theta)$ is used to approximate the true data-generating distribution $p^*(y)$:

$$D_{\mathrm{KL}}(p^* \| p) = \int p^*(y) \log \frac{p^*(y)}{p(y \mid \theta)} \, dy.$$

Minimizing the KL divergence is equivalent to maximizing the expected log-likelihood $\mathbb{E}_{p^*}[\log p(y \mid \theta)]$, which forms the theoretical basis for the Akaike Information Criterion (AIC) [10]. AIC provides a bias-corrected estimate of the expected KL divergence, and is defined as [9]

$$\mathrm{AIC} = -2\mathcal{L}(\hat{\theta} \mid y) + 2d,$$

where the first term measures the goodness of fit via the maximized log-likelihood, and the second term penalizes model complexity through the number of free parameters d. A closely related criterion is the Bayesian Information Criterion (BIC) [11], derived from a Laplace approximation to the marginal likelihood $p(y) = \int p(y \mid \theta)p(\theta)\, d\theta$, and defined as

$$\mathrm{BIC} = -2\mathcal{L}(\hat{\theta} \mid y) + d \log n,$$

where n is the sample size. Compared to AIC, BIC imposes a stronger penalty on model complexity when $n > 8$, resulting in sparser models and offering greater consistency in model selection as $n \to \infty$. By balancing goodness of fit against model complexity, both AIC and BIC mitigate overparameterization in regression tasks, reducing overfitting risk and promoting more generalizable models.

1.1.3 Sliding-Window Processing and Estimate Aggregation

Sliding window processing (SWP) is a technique used to process sequential data by dividing it into overlapping segments that progressively move or *slide* across the signal. This methodology enables processing data through smaller, more manageable blocks for more detailed analysis, while maintaining contextual continuity. By operating on localized data subsets (windows), SWP facilitates optimized memory utilization, and reduced computational demands. In this work, the implementation of SWP induces a many-to-one relationship where individual data points in the noisy signal y may be associated with multiple estimates. To combine these multiple estimates, we employ an inverse variance weighting (IVW) scheme giving higher weights to more accurate estimates, thereby producing a robust composite result.

1.2 Problem Statement

High-quality NR methods are characterized not only by achieving a PSNR higher than that of the noisy observation y, but also by preserving signal details while avoiding the introduction of artifacts and oversmoothing. Achieving high-quality NR requires selecting an adequate model for the noise-free signal, which usually requires model parameter tuning. Developing methodologies that balance high-quality NR with effectiveness across multiple datasets, while remaining computationally efficient, remains an open problem in signal processing.

The main elements of the AWGN reduction problem are addressed in detail in the following subsections.

1.2.1 Artifacts and Oversmoothing

Oversmoothing is a phenomenon characterized by the elimination of signal information by NR methods, erroneously replacing subtle variations, or sharp transitions in x by smooth curves. One of the main challenges for NR methods revolves around the preservation of sharp transitions or discontinuities in the signal [12]. Improper modeling of such abrupt variations may result in the emergence of artifacts in the form of spikes or spurious oscillations near sharp edges or, conversely, oversmoothing across discontinuities, leading to the loss of crucial information [12]. It follows that an accurate localization of discontinuities is essential for mitigating the adverse effects of oversmoothing and ringing artifacts in signal NR.

1.2.2 Selection of Model Parameters

Noise Level Estimation

To adequately guide NR methods in discriminating noise from signal, most AWGN reduction methods require information about the standard deviation σ of the Gaussian noise distribution that corrupts x. Most NR methods focus primarily on signal estimator design and assume that σ is known. However, this assumption is often not met in practical scenarios where the noise source and characteristics are unknown beforehand, introducing the additional challenge of computing an accurate noise level estimate $\hat{\sigma}$.

Choosing a Basis

Many NR methods operate by approximating the noisy signal y using a signal model $\hat{x}$. If this model closely matches the true underlying signal, high-quality noise reduction is achievable. In this setting, y is represented in an alternative domain via a transform $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$, which facilitates the separation between noise and signal components. Specifically, $\mathcal{F}(y)$ produces a set of components $\{c_k\}_{k=1}^n$ that capture different features of y in the transformed domain. The NR procedure is then performed on $\mathcal{F}(y)$ by zeroing out or attenuating the components c_k associated with noise. The signal estimate $\hat{x}$ is then constructed using the retained components.

However, a limitation of this approach lies in the selection of an appropriate transformation. Transforms relying on fixed basis functions, for example, Fourier or Wavelet transform, can be too rigid, failing to accurately represent the underlying signal x, thus limiting NR quality. Alternative approaches relying on data-driven signal representations also require careful hyper-parameter tuning (for example, the number of elements in the basis and threshold value), with direct impact on the final result. Therefore, adequate signal representation remains as an open challenge in AWGN reduction tasks.

1.2.3 Balancing Generalization - Parameter Estimation and Computational Efficiency

The increasing adoption of machine learning (ML) and artificial intelligence (AI) has significantly raised in developing robust methods capable of handling diverse data distributions while maintaining high performance. Models which exhibit minimal performance degradation across varying conditions, are said to *generalize well* [13]. In the context of Additive White Gaussian Noise (AWGN) reduction, generalization means delivering high-quality NR across a wide array of signal types and noise levels. Achieving this goal requires a combination of adaptive parameter selection, flexible signal representation, and computational efficiency. However, existing methods tend to prioritize one of these aspects over the others. For example, data-driven approaches may improve adaptability, but they come with higher computational costs. To date, no AWGN reduction technique successfully integrates all these features. Bridging this gap remains an open challenge, highlighting the need for deeper exploration of data-driven NR models that balance adaptability, automation, and efficiency.

1.3 Research Objectives

- **Objective 1**: Develop a fully-adaptive AWGN reduction method for 1-dimensional signals that produces high quality NR.

  Sub-objective 1.1: Achieve effective NR (SNRI>1) for a wide family of signals and different noise levels.

  Sub-objective 1.2: Obtain SNRI values higher or comparable to the ones produced by classical AWGN reduction methods.

- **Objective 2**: Extend the method in **Objective 1** to images. maintain efficiency, and reduce noise without introducing artifacts (especially blurring and oscillations) .

  Sub-objective 2.1: Achieve effective NR (SNRI>1) for a wide family of signals and different noise levels.

1.4 Thesis Outline

This thesis is organized as follows: Chapter 2 offers a literature review for AWGN reduction methods for 1D signals and images. Covering both, classical and modern approaches along with their respective strengths and limitations. Chapter 3 contains early methodological approaches that led to the development of the final methods proposed in this thesis. Chapter 4 presents the first article produced in my studies, entitled: *Adaptive Noise Reduction via Optimal Multiscale PiecewiseSmooth Modelling*. This article contains a detailed methodological formulation of PSM, and SW-PSM, along with extensive experiments. Chapter 5 presents the second research article: *Gaussian Noise Reduction in Images via Smoothness and Self-Similairty Fusion*, in which we extended the application of the SW-PSM method to 2D images and introduce a hybrid methodology that combines SW-PSM with a Reduced Rank Regression method to reduce noise from smooth and non-smooth noisy regions in an image. A general discussion of the work is provided in Chapter 6. Conclusions and final remarks regarding this thesis are offered in Chapter 7.

CHAPTER 2 LITERATURE REVIEW

This section provides a comprehensive critical review of fundamental developments in AWGN for both 1D signals and images. This review covers the validity and implications of the AWGN assumption in practical scenarios, main NR methodologies with special attention to data-driven representation techniques, and current challenges posed by open research questions in AWGN reduction.

The relevance of NR methods emerges from two fundamental factors: the need for high-fidelity data in sensitive applications, and the vast diversity of signal characteristics and noise profiles encountered in practice. These demands continue to motivate active research in NR methodologies, driving the development of increasingly sophisticated approaches with a particular emphasis in widely applicable methods. Particular fields of application include communication systems, [1,3] speech processing [4,5] biomedical signal analysis [6,7], astronomy [14, 15], and remote sensing [16, 17].

2.1 Additive White Gaussian Noise

Among various noise models, additive Gaussian White Noise (AWGN) is one of the most extensively studied due to its mathematical tractability and prevalence in practical applications [8]. Many sensor readings naturally exhibit thermal (allso called Johnson–Nyquist) noise, which originates from thermal agitation of electrons in resistive components, and can be well-approximated by a Gaussian (normal) distribution. Furthermore, this kind of noise is uncorrelated, and under suitable acquisition conditions the constant noise-variance assumption holds [18]. Shot noise, which arises from the discrete nature of charge or energy carriers, such as electrons in electronic circuits or photons in optical systems, is usually modeled as a Poisson distribution, but when the number of electrons/photons is large, the Poisson distribution can be suitably approximated by a Gaussian distribution. [19, 20]. Since shot and Nyquist noise are independent of the signal, they can be suitably modeled as additive noise. For particular cases, multiplicative noise owed to the effects of signal-dependent noise is also addressed by NR models [21, 22]. Active imaging systems, i.e., imaging systems that illuminate the target with its own controlled source of radiation, such as laser, microscope, synthetic aperture radar (SAR) imaging, represent typical scenarios of contamination by multiplicative noise [23].

Particular cases of non-constant noise-variance owed to the effects of temperature-dependent noise, or varying non-controllable experimental conditions, are also addressed in NR models [24–26]. The constant-variance assumption may serve as an initial approximation for the development of mathematically tractable NR frameworks that can be extended to tackle the non-constant variance AWGN scenario [27–29].

In this dissertation, I focus on the problem of estimating a piecewise continuous signal $x \in \mathbb{R}^N$ from a noisy observation $y \in \mathbb{R}^N$, where noise contamination follows an additive Gaussian white noise model, i.e.

$$y = x + w, \tag{2.1}$$

where the noise components in w are independently and identically distributed (i.i.d.) normal variables with zero mean and unknown standard deviation σ. The goal of NR is to produce a signal estimate $\hat{x}$ that closely matches the unknown noise-free signal.

Similarly, I consider the additive white Gaussian noise model for the contamination of two-dimensional (2D) images. In this scenario, I assume that the noisy image $Y \in \mathbb{R}^{n\times n}$ is generated through the degradation of a clean image $X \in \mathbb{R}^{n\times n}$ by additive white Gaussian noise (AWGN) with zero-mean and noise level $\sigma \in \mathbb{R}$, i.e.,

$$Y = X + W, \tag{2.2}$$

where $W \in \mathbb{R}^{n\times n}$ represents the noise matrix whose entries are independently and identically distributed sampled from a zero-mean Gaussian distribution with standard deviation σ.

Similarly to the 1D case, the image noise reduction (INR) problem is formulated as an inverse problem that seeks to recover an accurate estimate $\hat{X} \in \mathbb{R}^{n\times n}$ of the noise-free image X.

With the intention of reducing notation, the noise model presented in equation 2.2 explicitly considers square images. This is done without loss of generality, since the proposed methodology can be directly applied to non-square images.

2.2 Spatial Domain Noise Reduction Methods

Spatial Domain methods represent some of the earliest and most intuitive noise-removal (NR) techniques. These methods directly operate on pixel intensities using local neighborhoods to estimate the clean signal. Specifically, they operate by defining the estimate $\hat{x}[i]$ associated with the i^{th} element of $x \in \mathbb{R}^n$ as the output of a mapping $f : \mathbb{R}^L \to \mathbb{R}$, applied to a local neighborhood $v_i = (y[i-r], \dots, y[i+r]) \in \mathbb{R}^L$ of size $L = 2r + 1$ centered at index i in the noisy observation y, i.e.,

$$\hat{x}[i] = f(v_i).$$

The analogous formulation extends naturally to 2D images. The estimate at pixel (i, j) is defined as the output of a mapping $g : \mathbb{R}^{L \times L} \to \mathbb{R}$, applied to a square patch of size $L \times L$ centered at (i, j) yielding the estimate $\hat{X}[i, j] = g(v_{ij})$. Typical choices for f include simple statistical estimators such as the mean, median, and mode. A classical example of a more sophisticated function is the Wiener filter, which incorporates both local variance estimation and prior assumptions about the signal and noise distributions. Spatial Domain methods tend to perform reasonably well when the underlying signal is smooth and lacks abrupt intensity changes. However, their effectiveness diminishes in the presence of sharp transitions or fine-scale details, where such methods often lead to blurring or edge degradation [12].

Non-linear spatial domain methods such as median-based filters and non-linear total variation minimization [30], have been developed to mitigate the oversmoothing artifacts commonly associated [12, 31]. Nonetheless, these methods still fail to accurately adapt to rapid signal changes, resulting in signal estimation estimation errors, and the introduction of artifacts [32]. Examples of spatial domain methods for 2D data are mean, median, Gaussian, and Weiner filters. The 2D extensions of spatial domain NR methods reduce noise contamination to some extent, however, they tend to produce blurry image estimates, especially in fast-varying image regions [33].

2.3 Transform Domain Methods

Transform domain (TD) methods employ a signal transform $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$ to represent the noisy signal y in a different basis, yielding the transformed representation $\mathcal{F}(y) \in \mathbb{R}^n$. Typically, the change of basis is chosen so that the noiseless signal x and the noise w correspond to different elements of the new basis, thus facilitating their separation.

Generally, TD methods follow a three-stage process:
1) Transform: The input signal y is projected onto a new basis by applying $\mathcal{F}$ to y. The transformed signal is denoted by $y' = \mathcal{F}(y)$.
2) *Thresholding:* In this stage, noise reduction is performed directly on the transformed signal y'. The main idea is to diminish the contributions of components in y' that are likely to correspond to noise. For instance, when $\mathcal{F}(y)$ is the discrete Fourier transform of y, it is common to assume that high-frequency components correspond to rapid fluctuations caused by noise. Accordingly, frequencies higher than a threshold $\tau \in \mathbb{R}$ are attenuated or eliminated. Thresholding can be implemented in two principal ways: *Hard thresholding* where suspected noise-related coefficients are set to zero, and *Soft thresholding* (shrinkage): where the magnitude of these coefficients is reduced without nullifying them entirely [34, 35]. The thresholded transformed signal is denoted by by y'_τ

3)*Reconstruction:* Finally, the noise-free signal estimate $\hat{x}$ is reconstructed by applying the inverse transform to y'_τ, i.e., $\hat{x} = \mathcal{F}^{-1}(y'_\tau)$, where $\mathcal{F}^{-1}(y) : \mathbb{R}^n \to \mathbb{R}^n$, denotes the inverse function of $\mathcal{F}$.

2.3.1 Noise Reduction with Fourier Transform and Wavelet Transform

The discrete Fourier transform (DFT), and the discrete wavelet transform (DWT) are classical choices of $\mathcal{F}$ for one-dimensional signals [36–38] and images [39–42].

Discrete Fourier Transform

In the context of noise reduction, the Discrete Fourier Transform decomposes a signal into its frequency components, representing it as a sum of sinusoidal basis functions with specific frequencies, amplitudes, and phases. This frequency-domain representation enables noise reduction through the elimination of components with small coefficients, based on the principle that noise typically manifests as low-amplitude, high-frequency components across the entire frequency spectrum [43]. By applying a threshold to the DFT coefficients and setting those below this threshold to zero, noise components can be effectively removed while preserving the dominant signal features represented by large-amplitude coefficients. This thresholding approach assumes that significant signal information is concentrated in a relatively small number of frequency components with substantial magnitudes, while noise is distributed across many coefficients with smaller amplitudes. However, the DFT's global frequency analysis provides poor time localization, making it less suitable for signals with transient or non-stationary characteristics where the temporal context of coefficient magnitudes becomes crucial for distinguishing between noise and signal components.

An illustrative example of DFT-based noise reduction is presented in Figure 2.1. The left panel displays the noise-free *heavisine* [44] test signal x (black curve) with sample size $N = 512$, the noisy signal y (grey curve) with additive Gaussian noise ($\sigma = 0.1$), and the DFT reconstruction (blue curve) obtained by retaining only the frequency components shown in the right panel. In this example, frequency components with Fourier coefficients below the threshold $\tau = 8.3$ were eliminated by setting them to zero, where τ represents the optimal threshold value that minimizes the root-mean-squared-error (RMSE) between the reconstructed signal $\hat{x}$ and the noise-free signal x, defined as

$$\mathrm{RMSE}(x, \hat{x}) = \frac{1}{\sqrt{N}} ||x - \hat{x}||_2,$$

where $||a||_2$ denotes the Euclidean norm of a.

The sinusoidal waveforms in the right panel correspond to the frequency components preserved after thresholding, and its sum produces the blue curve in the left panel. A limitation of the Fourier-based approach is evident in the reconstruction: the sharp discontinuities near points 150 and 370 in x are inadequately represented by the global sinusoidal basis functions, resulting in oversmoothing near these transitions.

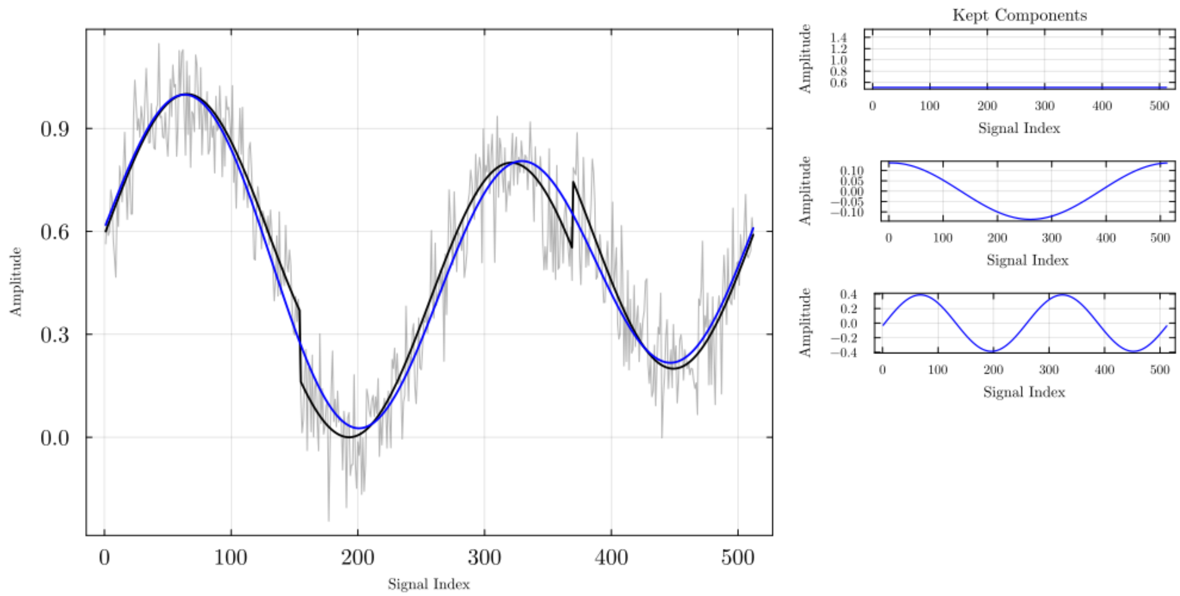

Figure 2.1 **Left:** Signals x (black curve), y (grey curve, $\sigma = 0.1$), and thresholded Discrete Fourier reconstruction (blue curve, RMSE=0.04274). **Right:** Sinusoidal components employed in the reconstruction.

Discrete Wavelet Transform

Empirical and theoretical studies consistently demonstrate that wavelet-based NR outperforms Fourier-based counterparts, [36, 43, 45, 46]. This superiority arises from the wavelet transform's use of localized waveforms, which improves the signal representation in both time and frequency. The specific waveform that is employed to represent the signal in the wavelet basis is called the *mother wavelet.* Figure 2.2 displays six common mother wavelets used in wavelet analysis: Daubechies (**db2**, **db10**), Symlet (**sym4**), Haar (**haar**), Coiflet (**coif4**), and Battle-Lemarie (**batt4**) wavelets. Each wavelet is plotted over the interval [0, 1] showing their distinct temporal characteristics and varying degrees of regularity. For example, the **haar** wavelet exhibits simple step-like discontinuities with constant plateaus, making it suitable for modeling signals containing piecewise constant regions and sharp transitions. The **batt4** wavelet exhibits a sharp variation suggesting its appropriateness for signals characterized by the presence of spike-like features. The choice of mother wavelet significantly impacts the signal reconstruction [43], as different wavelets are suited for different signal characteristics.

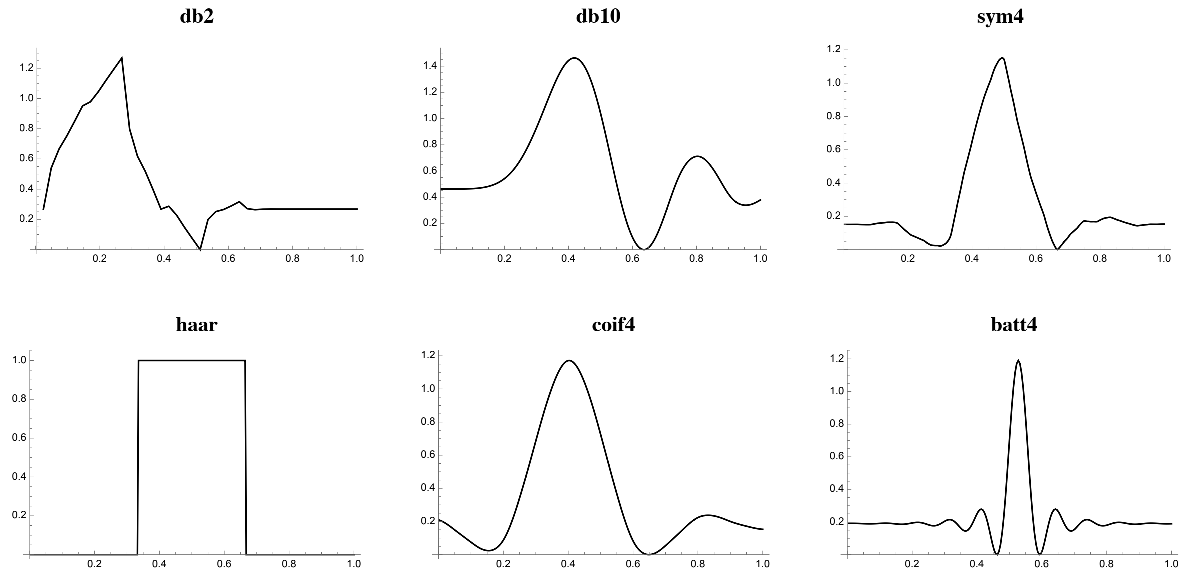

Figure 2.2 Examples of typical mother wavelets used in wavelet analysis.

As mentioned in the previous sub-section, DWT-based methods require a thresholding step. The main drawback of the hard-threshold approach is the introduction of spurious discontinuities in the signal estimate caused by the abrupt elimination of coefficients [47]. To address this limitation, multiple soft-thresholding techniques [48–50] have been proposed, demonstrating improved performance in edge preservation. One of the most popular thresholding approaches is *VisuShrink*, where the wavelet coefficients smaller than $T = \sigma\sqrt{2\log(N)}$, are set to zero [44], where σ is the noise standard deviation and N is the number of samples in the signal. Another parameter that needs to be specified in the context of wavelet analysis is the decomposition level, which determines the hierarchical scales at which the signal is analyzed through successive filtering and downsampling operations. At each level, the signal is separated into approximation coefficients (low-frequency content) and detail coefficients (high-frequency content), creating a multi-resolution representation.

Wavelet-based frameworks exhibit high sensitivity to the number of decomposition levels [45], a low number level may fail to isolate low-frequency signal trends from noise, whereas an excessive level may introduce boundary artifacts, cause over-smoothing, and unnecessarily increase execution time. This sensitivity requires careful selection based on the signal characteristics and the specific application requirements.

An illustrative example of Discrete Wavelet Transform in the noise reduction context is displayed in Figure 2.3, displaying the noise-free signal x (black curve), the noisy signal y (grey curve $\sigma = 0.1$), and the thresholded Discrete Wavelet Transform reconstructions corresponding to the mother wavelets in Fig. 2.2. In this example, the number of decomposition levels for each signal estimate was set to minimize the reconstruction error between the estimated signal and x, and the *VisuShrink* [45] threshold was used to eliminate small coefficients.

Figure 2.3 illustrates that the **haar** wavelet (top-left), due to its step-function definition, accurately preserves the first vertical jump near index 150, but produces multiple discontinuity artifacts (e.g. near indices 25 and 175) and a "staircase" effect on the smooth curves, resulting in the highest reconstruction error. The **db2** wavelet (top-right) recovers the smooth component more faithfully than the **haar** wavelet, but introduces triangular artifacts in some regions of the estimate, e.g., around index 200 and 480. The **db10** (center-left) wavelet provides a much smoother reconstruction but introduces a ringing artifact near index 480. The **sym4**, **coif4** and **batt4** (center-right, bottom-left, and bottom-right panels, respectively) wavelets avoid the introduction of ringing artifacts, but still fail to capture the signal feature in regions near discontinuities.

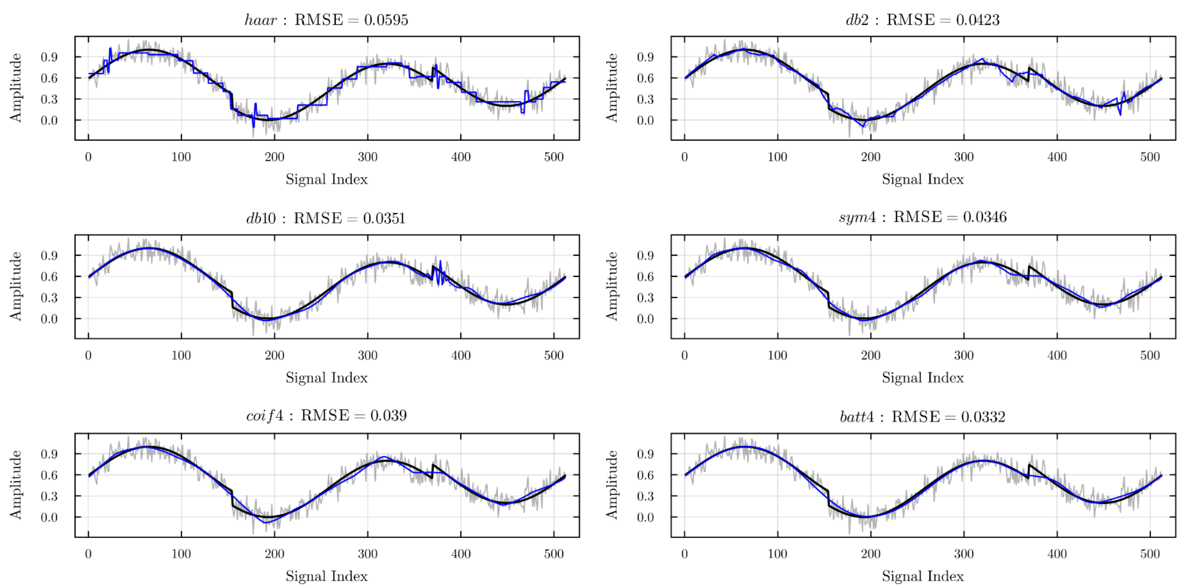

Figure 2.3 Signals x (black curve), y, (grey curve, ($\sigma = 0.1$)), and wavelet reconstructions (blue curves) with the mother wavelets in Fig. 2.2.

Figure 2.3 illustrates how an inappropriate choice of wavelet (e.g., **haar**) can degrade reconstruction quality, and highlights a fundamental limitation of wavelet methods: their difficulty in adapting to signal discontinuities, which leads to ringing artifacts or oversmoothing. In practical NR applications, the noise-free signal is generally unknown, making the selection of an optimal mother wavelet a non-trivial challenge, regardless of the thresholding strategy employed.

To circumvent the limitations that come with the selection of a unique mother wavelet, some authors have proposed multi-wavelet approaches employing filter banks, in which multiple wavelets are combined to achieve more robust NR performance [51, 52]. Nonetheless, the efficacy of most multi-wavelet schemes remains highly dependent on several user-defined parameters, including the number of decomposition levels, thresholding strategies per level, and the mother wavelets that conform the filter bank [53, 54].

2.4 Data-Driven Representations for Noise Reduction

The reliance on fixed, pre-defined basis functions imposes a rigid modeling framework that can limit the generalization capacity of traditional noise-reduction methods. To overcome these limitations, a variety of data-driven representation strategies have been developed. These methods use information from the noisy signal to guide the construction of representations that are more effective for noise reduction. Adopting terminology from the machine learning literature, these methods can be broadly categorized into supervised and non-supervised representation frameworks.

2.4.1 Supervised Data-Driven Representations:

These approaches aim to learn a function that produces an accurate signal estimate $\hat{x}$ using paired noise-free and noise-contaminated signals.

In this context, domain-specific signal datasets have been used to train neural networks to learn smooth signal representations for 1D signals [55, 56], and images [57–60]. Supervised data-driven representation methods for noise reduction offer several advantages, including fast execution after training, and adaptive parameter selection. However, they also face practical limitations:

L1) *Generalization to unseen data*: supervised methods tend to operate by assuming identical statistical distributions between training and testing data [61–63]. However, this assumption often breaks down in practical applications, where noise characteristics and image content demonstrate significant variability.

L2) *Reliance on noise-free data*: Supervised methods require access to clean ground-truth data [55–60]. In many practical settings clean reference signals are unavailable, making supervised training infeasible.

L3) *Computational Resources*: Conducting effective supervised learning for image noise reduction requires large training datasets [57, 58], this comes with a substantial demand of computational resources. As a result, deploying these methods in resource-constrained environments remains a significant challenge.

The limitations L1) and L2) motivated the development of self-supervised learning techniques for noise reduction. Self-supervised methods operate without requiring clean training data, and empirical studies have demonstrated that self-supervised image noise reduction methods are able to achieve performance comparable to their supervised counterparts [64, 65], while improving the generalization to unseen data [66, 67].

Nonetheless, limitation L3) has not been completely addressed by self-supervised noise reduction methods. These methodologies tend to be computationally intensive [63], as they require large datasets to match or improve the performance of supervised methods [68], which limits their use in real-time applications [69–71]. These constraints reduce the practicality of learning-based models in many real-world settings, where data sources and/or computational resources are limited. In these scenarios, training-free image NR techniques maintain their relevance by providing a balance between implementation ease, performance, interpretability, and computational efficiency.

2.4.2 Non-training Data-Driven Representation Methods for Noise Reduction

Non-training Data-Driven Representation methods for NR provide a noise reduction approach that operates without training phases or external datasets, directly exploiting the inherent structure of noisy input data.

These methods typically adapt their representations based exclusively on the input signal. While their application comes with a decreased performance compared to learning-based methods [57, 58], the unsupervised design of training-free methodologies circumvents the limitations of external data dependence and high computational demands (L2 and L3). At the same time, training-free methods offer data-adaptability, allowing their application to different domains (addressing L1), and making them particularly attractive for resource-constrained applications.

Non-Training Data-Driven Methods for 1D Signal Noise Reduction

In the context of 1D signals, popular data-driven representation approaches that do not rely on train-test stages include Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD). EMD decomposes the noisy signal into a set of additive components referred to as intrinsic mode functions (IMFs). In the reconstruction phase the signal estimate $\hat{x}$ is obtained by recombining IMFs that capture signal characteristics, discarding those dominated by noise [72, 73].

In it's simpler versions, EMD demonstrates a lack of robustness in the presence of noise non-stationary noise, and in cases where the noise and signal spectrums overlap [74]. In such cases, noise can disrupt the decomposition process, resulting in inaccurate IMF estimation, and mode mixing, where frequency components leak into multiple IMFs, leading to non-unique IMFs [75, 76]. Furthermore, EMD lacks a formal mathematical foundation, and the stopping criteria often requires empirical tuning or heuristic methods that may limit its generalization [77, 78].

Variational Mode Decomposition (VMD) was proposed as an improvement of the EMD method to deal with the mode-mixing problem [79]. Unlike EMD, VMD formulates the decomposition as an optimization problem, yielding IMFs with greater robustness to both noise and mode mixing [53, 79].

The primary limitations of EMD and VMD include the reliance on band-limited components [72] and sensitivity to user-defined parameters. These parameters include both the number of IMFs for decomposition [80] and, particularly for VMD, the bandwidth constraint for IMFs [72]. Such manual parameter selection may restrict both the applicability and generality of mode decomposition approaches.

Figure 2.4 illustrates an application of the VMD method. The number of decomposition levels and the bandwidth constraint were selected to maximize the RMSE between the reconstructed and the noise-free signal. Notably, VMD achieved a lower RMSE than the wavelet and Fourier noise reduction methods presented in Figures 2.1 and 2.3, despite relying on a representation basis of only two elements. The first element captures finer-scale details, such as the discontinuity near index 150 (top-right panel of Figure 2.4), while the second corresponds to a mostly smooth curve that reflects the general shape of x without accounting for discontinuities.

Data-driven sparse representation methods with statistical foundations have also been developed to address these challenges. For instance, Saucier et al. proposed an adaptive framework for constructing multiscale orthogonal basis functions [81]. While this approach demonstrated superior compression performance compared to conventional wavelet transforms, it incurred significantly higher computational costs.

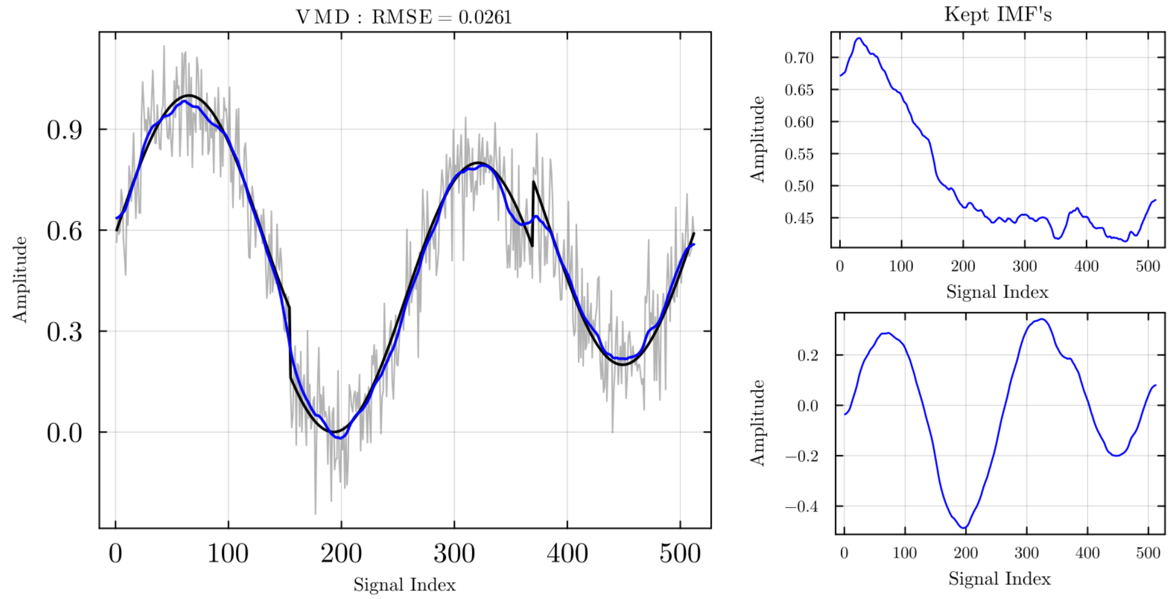

Figure 2.4 **Left:** Signals x (black curve), y (grey curve), and VMD signal estimate (blue curve). **Right:** Intrinsic mode functions composing the VMD signal estimate.

Non-Training Data-driven Representation Methods for Image Noise Reduction

For 2D signals, particularly images, widely used data-driven representation methods that do not rely on a training stage include patch-matching approaches such as Block-Matching and 3D Filtering (BM3D) [82], its extension, shape-adaptive BM3D [83], and dictionary learning methods like K-means Singular Value Decomposition (K-SVD) [84]. Both K-SVD and BM3D gained significant attention in the field by showing substantial NR quality improvements over traditional spatial and transform domain methods [33].

BM3D groups similar patches across the image and jointly reduces noise on them via collaborative filtering in a transformed domain constructed from the data itself [83, 85]. Among non-learning techniques, BM3D achieves state-of-the-art denoising quality for natural images. Nonetheless, it relies on multiple parameters such as patch size, search window, and thresholds for grouping and filtering. These are typically selected using default configurations derived from empirical studies [85, 86]. BM3D is relatively easy to use in practice, and depending on the chosen parameters, relatively fast. However, its cross-domain effectiveness is limited by the need for either manual selection, or domain-specific optimization of multiple parameters [86], which depend on both the image and noise level. Furthermore, BM3D assumes the presence of repeated structures in the image, limiting its effectiveness in images with highly textured irregular patterns.

K-SVD is a dictionary learning algorithm that iteratively updates a set of basis atoms to represent small overlapping patches of the image [84]. Similar to BM3D, these K-SVD methods tend to rely on the manual selection of noise-level-dependent parameters such as dictionary size, limiting its applicability across different domains [70].

While learning-based methods dominate recent literature, classical non-learning approaches such as BM3D remain as competitive alternatives for Gaussian noise reduction and continue to serve as benchmarks.

2.5 Current trends and future directions in AWGN reduction

Regardless of methodological diversity in signal noise reduction, model tuning remains a persistent challenge, limiting both domain generalization and practical applicability. In response, hybrid data-driven strategies that integrate robustness, simplicity, interpretability, and high NR performance emerge as a promising solution

Recent advances in signal noise reduction reflect a growing interest in methods that are adaptive, efficient, and generalizable. Notably, there is a convergence between traditional or unsupervised signal processing and deep learning, with modern approaches incorporating signal priors such as sparsity, smoothness, and multi-scale structures into trainable neural networks. This fusion improves both interpretability and generalization capabilities [87–89]. In the context of 1D signals, adaptive frameworks like EMD and VMD have been integrated with neural architectures to enhance NR performance [90, 91].

Another emerging focus in signal noise reduction is self-supervised learning. In particular, blind noise reduction has become a central challenge, driven by the need to handle real-world noise without prior knowledge of its distribution. These methods relax traditional assumptions about noise characteristics, aiming for a more flexible and robust performance. Recent examples of blind NR models incorporate model architectures like generative adversarial networks, convolutional neural networks, and transformers [61, 92–95].

Emerging directions that address both efficiency and generalization include few-shot and zero-shot NR [96–99]. A central, unresolved challenge is the design of lightweight, general-purpose denoisers that retain strong performance under minimal supervision. In this context, a relevant challenge for modern approaches lies in providing an adequate balance between accuracy, adaptability, and computational efficiency.

CHAPTER 3 EARLY AND ALTERNATIVE METHODOLOGICAL APPROACHES

3.1 Introduction

Prior to the development of the methods presented in this thesis, several alternative approaches were explored. The final methods emerged from an iterative process of design, implementation, and refinement of multiple modeling attempts. This chapter reviews earlier methodological approaches that, while not ultimately adopted, provided valuable insights and informed the development of the final NR models.

One early approach was based on a signal envelope framework derived from the Lipschitz constant of the signal, which provided guaranteed bounds on the reconstructed signal. While this approach offered theoretical guarantees, it was ultimately not adopted its inability to accommodate signal discontinuities.

The need to explicitly model signal discontinuities led to the development of early *signal covering* approaches, which are cruder versions of the final Piecewise Smooth Modeling (PSM) method introduced in this thesis. Initial versions modeled y via piecewise linear regression, extensions introduced the use of constant and polynomial models, and, finally, the use of information criteria minimization to determine an optimal piecewise regression.

The next section describes the initial formulation of the PSM method, whose optimal solution relied on a mixed-integer linear programming optimizer. This formulation served as an initial approach to finding an optimal piecewise regression for y; however, it was not included in the final version of the method, as the derivation was specific to the Akaike Information Criterion. The final formulation presented in Article 1 includes a unified derivation covering both the Akaike Information Criterion and the Bayesian Information Criterion for piecewise regressions with clearer notation. Furthermore, it is designed to be solved directly via a shortest-path optimizer, which requires the use of information criteria that admit an additive decomposition with respect to the global model.

3.2 Estimation via Mixed Integer Linear Programming Optimization

The discrete signal of interest is denoted by $x \in \mathbb{R}^N$, where $N >> 1$ is an integer. The signal x is assumed to be obtained by sampling an unknown piecewise smooth [1] function $f : \mathbb{R} \to \mathbb{R}$,

[1] i.e. differentiable with respect to time t everywhere except at a finite number of localized points.

i.e. the n^{th} component of x is $x[n] := f(n\,\Delta t)$, $n \in \{0,\, 1,\, ...,\, N-1\} =: S$. The set S will be called the signal support, $t = n\,\Delta t$ denotes the continuous time variable and $\Delta t > 0$ is the uniform sampling interval. The signal x is assumed to be degraded by additive zero-mean discrete white Gaussian noise $w \in \mathbb{R}^N$ with standard deviation σ, i.e., the measured signal $y \in \mathbb{R}^N$ is defined as $y = x + w$, where the equality holds component-wise for each $n \in S$. The components of w will be assumed to be mutually independent. It is also assumed that the function f may have a few localised irregularities, i.e. a few non-smooth regions having a relatively small extent, e.g. spikes. In this context, the problem is to estimate x using only y as input, assuming that both f and w are unknown.

3.2.1 Definition of coverings for the signal support

Since f is piecewise smooth, it has a Taylor series on each piece. It follows that it should be possible to find many intervals within S for which the discrete signal x can be approximated by a polynomial. More generally, there should exist a *covering*, which we define as a sequence adjacent but disjoint intervals that covers the signal support S in such a way that x is approximatly constant, linear or parabolic on each interval. Note that we choose disjoint sets to allow for possible discontinuities between adjacent smooth pieces.

The problem consists in finding $\mathcal{C}$, a covering of S that provides the best possible model for the noisy signal y. For this purpose, we propose to use the Akaike Information Criterion (AIC), which provides a systematic way to define an optimal compromise between goodness-of-fit and model simplicity.

The sets that we use to cover the support S will be called *covering-sets*. They are composed of *consecutive* elements of S. For example, if $N = 3$, then the sets used to build coverings of the support $S = \{0, 1, 2\}$ are $\{0\}$, $\{1\}$, $\{2\}$, $\{0, 1\}$, $\{1, 2\}$, and $\{0, 1, 2\}$. The number of all possible covering-sets for $S = \{0, 1, ..., N-1\}$ is $N + (N-1) + (N-2) + ... + 1 = N\,(N+1)/2$.

While the signal of interest x may be smooth almost everywhere, real one-dimensional signals may contain non-smooth regions. These uncommon regions, manifest as spikes composed of a few consecutive signal values. To accommodate such exceptions, we consider covering-sets consisting of only one point.

We define a covering $\mathcal{C}$ of S as a union of adjacent and disjoint covering-sets that covers S. For example, for the support $S := \{0, 1, 2\}$, all the possible coverings of S with these covering-sets are given by $\{\{0\}, \{1\}, \{2\}\}$, $\{\{0, 1\}, \{2\}\}$, $\{\{0\}, \{1, 2\}\}$, and $\{\{0, 1, 2\}\}$. More generally, if S_i denotes a covering-set for S, where $i \in \{1,\, 2,\, ...,\, N(N+1)/2\} =: J$ is an integer index, then a covering $\mathcal{C}$ of S is defined as a collection of covering sets $\mathcal{C} := \{S_i \subseteq S, i \in J' \subseteq J\}$

that respects both the non-intersection condition

$$S_i \cap S_j = \emptyset \text{ for all } (i, j) \in J' \times J' : i \neq j,$$

and the covering condition

$$\bigcup_{i \in J'} S_i = S.$$

3.2.2 Performing all possible curve-fittings

We perform least-squares curve-fittings (constant, linear, and parabolic) for several covering-sets of S. Curve-fitting on the k^{th} covering set $\{n_{k,1},\, n_{k,1}+1, ...,\, n_{k,2}\} \subseteq S$ consists in finding the parameter vector $a \in \mathbb{R}^d$, where $d \in \{1, 2, 3\}$, that minimizes the sum of square errors $e_k(a)$ defined as

$$e_k(a) := \sum_{i=n_{k,1}}^{n_{k,2}} \left(\frac{y[i] - f_k(i;a)}{\sigma} \right)^2. \tag{3.1}$$

In (3.1), the form of the function $f_k : [n_{k,1}, n_{k,2}] \times \mathbb{R}^d \to \mathbb{R}$ depends on the model type, i.e. constant, linear and parabolic. For the constant model, $d = 1$, $a = a_0$ and $f_k(i;a) := a_0$. For the linear model, $d = 2$, $a = (a_0\,,\, a_1)$ and $f_k(i;a) := a_0 + a_1\, i$. For the parabolic model, $d = 3$, $a = (a_0\,,\, a_1\,,\, a_2)$ and $f_k(i;a) := a_0 + a_1\, i + a_2\, i^2$. The parameter vector $a \in \mathbb{R}^d$ that minimizes $e_k(a)$ will be denoted by a^*. The minimal value of $e_k(a)$ is

$$E_k := e_k(a^*). \tag{3.2}$$

If the model and the fit are good, then each term $(y[i] - f_k(i;a))/\sigma$ in (3.1) is approximately a zero-mean unit-variance Gaussian random variable and E_k, which is the sum of their squares, is a random variable that follows a chi-square distribution with $n_{k,2} - n_{k,1} + 1 - d$ degrees of freedom. The quality of fit is measured via the chi-square test, i.e. the fit is regarded as acceptable if

$$E_k < p_{\nu_k,\,\alpha}, \tag{3.3}$$

where $\nu_k := n_{k,2} - n_{k,1} + 1 - d$. The critical value $p_{\nu,\,\alpha}$ is defined as $\mathrm{P}\left[\chi^2(\nu) > p_{\nu,\,\alpha}\right] = \alpha$, where $\chi^2(\nu)$ denotes a random variable following the chi-square distribution with ν degrees of freedom. The value $\alpha = 0.05$ is typically used. The fits that satisfy (3.3) will be called *acceptable fits*.

The acceptable fits are labeled by an integer $k \in \{1,\, 2,\, ...,\, m\}$, where m denotes the total number of acceptable fits. For k^{th} acceptable fit in the interval $[n_{k,1},\, n_{k,2}]$, this preprocessing provides a list of three parameters:

$$\begin{cases} \lambda_k \in \{1,2,3\} & : \quad \text{number of curve-fitting parameters} \\ \ell_k \in \{1,\dots,n\} & : \quad \ell_k := n_{k,2} - n_{k,1} + 1 \text{ is the support size;} \\ E_k \geq 0 & : \quad \text{minimal sum of square errors,} \end{cases} \tag{3.4}$$

where the number of curve-fitting parameters is 1 for constant, 2 for linear, 3 for parabolic.

There may be signal regions where the piecewise-smooth-signal hypothesis fails. For example, the signal y may contain high intensity localized *spikes* or highly irregular signal variations. Such points will be called *outliers*. In the context of our approach, such outliers manifest themselves as points that do not belong to any acceptable-fit interval. Since the piecewise-smooth hypothesis does not apply to these points, and also because they may carry relevant information, we choose to leave these signal values unchanged. Consequently, to each outlier at position i, we associate the signal estimate $\hat{x}[i] := y[i]$. Note that for an outlier of index k, we use $\lambda_k = 1$ and $E_k = 0$.

3.2.3 Defining the optimal signal model via the Akaike information criterion (AIC)

The k^{th} acceptable fit occurs in the covering set $S_{i(k)}$, where $i(k)$ is an integer index. For all acceptable fits, we define the indicator binary variables $\theta_k \in \{0,1\}$, $k = 1, 2, ..., m$, and the corresponding vector $\theta := (\theta_1, \theta_2, ..., \theta_m) \in \{0,1\}^m$. If $\theta_k = 1$, then the covering-set $S_{i(k)}$ associated with the k^{th} fit is part of the covering, otherwise it is not.

By definition, the covering-sets are disjoint and cover the signal support S. To take these constraints into account, we introduce for each acceptable fit of index k an indicator array $U_k \in \{0,1\}^n$ defined as

$$U_k(n) = \begin{cases} 1 & \text{if } n \in S_{i(k)}, \\ 0 & \text{otherwise,} \end{cases}$$

for $n \in \{0, 1, ..., n-1\} =: S$. Each point in the support S must be covered exactly one time. This leads us to define the following N linear constraints,

$$\sum_{k=1}^{m} \theta_k \; U_k(n) = 1, \;\; n = 0, 1, ..., N-1, \tag{3.5}$$

that ensure that S is covered with disjoint and adjacent covering-sets. The covering constraints (3.5) can be satisfied if the set of local curve-fittings in J_A is rich enough.

The total number of parameters Λ used in the global signal model is given by

$$\Lambda = \sum_{k=1}^{m} \theta_k \ \lambda_k + \sum_{k=1}^{m} \theta_k - 1, \tag{3.6}$$

where $\sum_{k=1}^{m} \theta_k \ \lambda_k$ is the total number of curve-fitting parameters, and $\sum_{k=1}^{m} \theta_k - 1$ is the total number covering sets required to define $\mathcal{C}$. Indeed, if the covering of S takes the form $\mathcal{C} = \{1,\ j_1\}, \{i_2,\ j_2\}, ..., \{i_m,\ j_m\}$, then it can be fully specified by the $m-1$ integers $i_2, i_3, ..., i_m$ since the first covering set is always given by $\{1,\ i_2 - 1\}$.

The Akaike Information Criterion (AIC), is defined as $\text{AIC} = -2\ln(\hat{L}) + 2k$, where $\hat{L}$ is the maximized likelihood of the model and k is the number of estimated parameters. In the context of linear regression with independent and identically distributed zero-mean Gaussian errors with constant variance, the AIC takes the form [10]

$$\text{AIC} = n \ln\Big(\frac{1}{n} \sum_i (y_i - \hat{x}_i)^2\Big) + 2k,$$

where n is the sample size. An AIC-optimal piecewise regression is obtained by minimizing the AIC.

If the noise standard deviation σ is known or estimated, then the AIC takes the form

$$\sum_{k=1}^{m} \theta_k \ E_k + 2\,\Lambda, \tag{3.7}$$

using the same E_k given by (3.2).

The complete optimization problem can be written explicitly as follows:

$$\begin{cases} \min\limits_{\theta \in \{0,1\}^m} \ \sum\limits_{k=1}^{m} \theta_k \ E_k + 2 \left(\sum\limits_{k=1}^{m} \theta_k \ \lambda_k + \sum\limits_{k=1}^{m} \theta_k - 1 \right), \\ \text{subject to} \\ \sum\limits_{k=1}^{m} \theta_k \ U_k(n) = 1, \ n = 0,\ 1,\ ...,\ N-1, \end{cases} \tag{3.8}$$

The objective function in (3.8) balances fittings errors, measured by the first term $\sum_{k=1}^{m} \theta_k \ E_k$s, and model complexity, measured by the second term $2\,(\sum_{k=1}^{m} \theta_k \ \lambda_k + \sum_{k=1}^{m} \theta_k - 1)$. This problem can be solved as a mixed integer linear program (MILP), optimizing over the m binary variables θ_k that select which acceptable fits compose the optimal covering of S.

Figure 3.1 illustrates the four stages of the PSM covering method. Starting from a noisy signal y (blue dots, first row) corrupted by AWGN around an unknown piecewise-smooth true signal (dashed line,first row), all $N(N+1)/2$ contiguous intervals of the support S (second row) are considered as candidate covering-sets with varying lengths. Each interval is fitted by a constant, linear, or parabolic polynomial and retained only if the residual sum of squares passes a chi-square goodness-of-fit test; isolated outliers such as spikes are assigned a dedicated single-point set. Finally, a MILP optimizer selects from these acceptable fits a subset that is disjoint, adjacent, and covers each element of S exactly once, minimizing the AIC objective, with discontinuities between adjacent pieces deliberately preserved rather than smoothed.

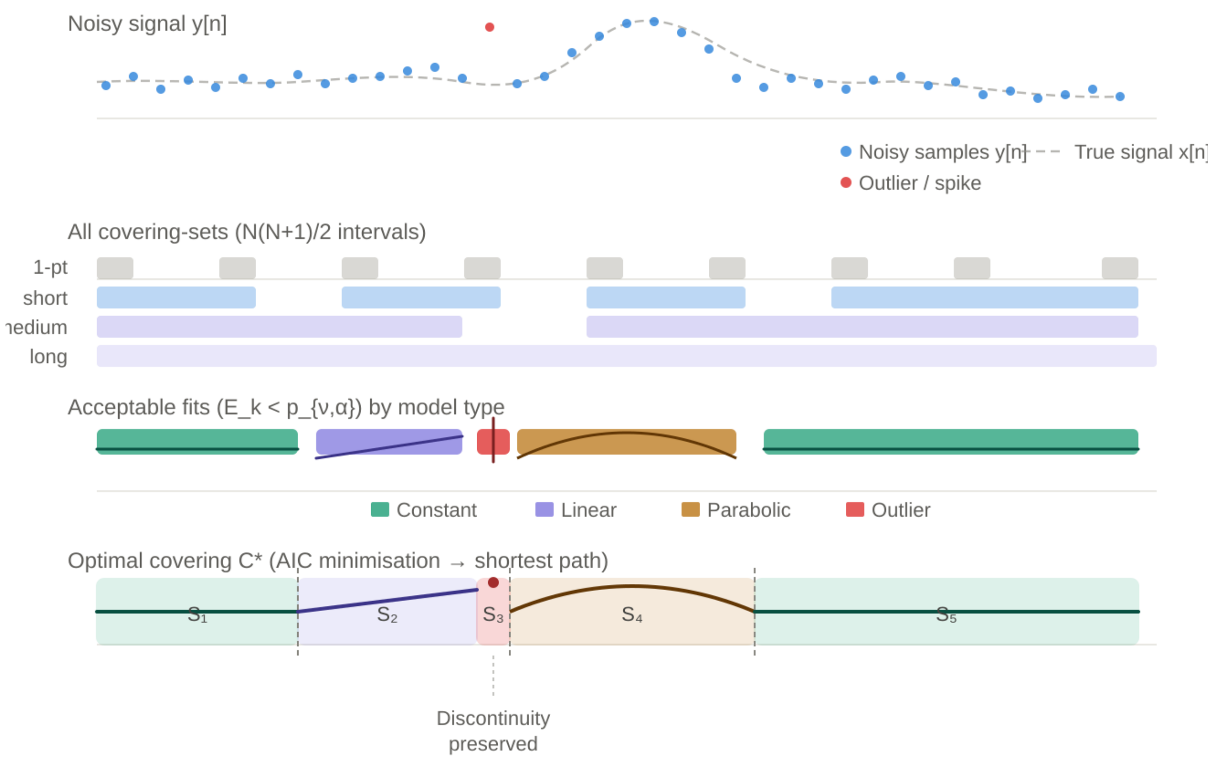

Figure 3.1 Diagram illustrating the Piecewise Smooth Model selection procedure using coverings.

3.3 SWPSM: Aggregation of Signal Estimates with different values of L

We explored the idea of combining multiple PSM estimates, each characterized by a single curve-fitting support size several strategies for combining these individual estimates. This approach was later replaced by the inverse variance weighting of PSM estimates obtained via sliding window processing, which resulted in a shorter processing time and higher reconstruction quality.

3.4 Adaptive Image Linearization

An early approach to image linearization was explored, based on reformulating the search for a minimal variation pixel sequence as a shortest path problem in the image graph. Dijkstra's algorithm was used to construct a Shortest Path Tree rooted at a given pixel, which provided an efficient and globally optimal solution within the restricted search space of tree paths. This approach was ultimately abandoned due to several limitations, the most critical being its sensitivity to noise. Since edge weights were computed directly from raw pixel intensities, local intensity fluctuations caused by noise tended to dominate path costs, causing the linearizations to follow noise rather than meaningful image structures. Additional limitations included sensitivity to the choice of root pixel, restriction of the search space to tree paths, and scalability concerns for high-resolution images.

3.5 PSM 2D

Another explored approach consisted of extending the PSM method to 2D data. Specifically, it involved extracting image patches, identifying homogeneous regions within each patch, and performing a 2D covering by fitting two-dimensional polynomial models over each region. While effective for patches with a smooth structure, this approach produced oversmoothed estimates in textured regions where no discernible smoothness was present. It was ultimately replaced by applying PSM to a vectorized representation of the 2D data, which proved more practical as it does not enforce smoothness in non-smooth regions.

CHAPTER 4 ARTICLE 1: ADAPTIVE NOISE REDUCTION VIA MULTISCALE PIECEWISE SMOOTH MODELLING

Submitted on July 3, 2025 to "Digital Signal Processing". Authors: Noel Plascencia-Diaz and Antoine Saucier.

Introduction to Article 1

The article presented in this chapter introduces the PSM method, which conforms the fundamental methodological element of my dissertation. This article aims to provide a comprehensive exposition of the PSM methodology including: the formulation of the model selection problem to select a piecewise smooth estimate of y, its solution via shortest-path algorithm, pointwise variance estimation for the curve-fittings, and a fast curve-fitting method using cumulants. The article also introduces SW-PSM as an extension to PSM, which produced substantial improvements over classical NR methods at multiple noise-levels. Methods for window-size and step-size selection are also provided. The authors of this article are Noel Plascencia-Diaz and Antoine Saucier. The manuscript was submitted to the *Expert Systems with Applications* journal and is currently under assessment by reviewers.

4.1 Abstract

We introduce Piecewise Smooth Model (PSM) selection as a novel solution to the piecewise regression problem, alongside its extension, the Sliding-Window Piecewise Smooth Model (SW-PSM), for signal noise reduction. PSM is formulated as an information-theoretic optimization problem that partitions a signal into adjacent smooth segments, simultaneously identifying breakpoints and performing model selection for each segment. SW-PSM improves upon PSM by aggregating multiple local PSM estimates. We empirically demonstrate that SW-PSM effectively recovers signals with non-polynomial or high-degree features using only low-degree polynomial fits (degree ≤ 2). Experiments show that data-driven SW-PSM outperforms oracle-optimized versions of established denoising methods across low to moderately-large noise levels and multiple signal instances. Sliding-window processing, fast curve-fitting, and a shortest-path optimizer ensure high computational efficiency.

4.2 Keywords

piecewise regression, multi-model inference, information criterion, noise reduction

4.3 Introduction

4.3.1 Background and context

Noise reduction (NR) is a fundamental problem in signal processing that aims to enhance signal quality by suppressing noise while preserving structural components. Effective NR requires a compromise between noise attenuation and the preservation of relevant signal features, since excessive filtering may remove fine-scale structures, whereas insufficient NR may leave residual noise. This challenge is critical in multiple applications where noise contamination impacts the reliability of acquired data [1, 3, 4] Among noise models, additive white Gaussian noise (AWGN) is widely studied due to its mathematical tractability and prevalence in acquisition systems [8].

4.3.2 Problem statement

In this work, the NR problem is formalized as the estimation of an underlying noise-free signal while preserving fine-scale structures, including discontinuities, sharp transitions, and localized variations, and without introducing artifacts such as over-smoothing or ringing.

One of the main challenges in NR lies in the preservation of fast-varying features and discontinuities. Methods that fail to adapt to abrupt variations in the noise-free signal often produce ringing artifacts or excessive smoothing [12]. Furthermore, since the noise-free signal is inaccessible in practice, parameter selection constitutes an additional difficulty. Inappropriate parameters may lead to undersmoothing or oversmoothing, making their robust determination a key issue in NR. These challenges have motivated the development of a wide range of NR methods, which differ in their modeling assumptions, signal representations, and parametrization strategies. In the following subsections, we review the main classes of existing NR approaches.

4.3.3 Spatial domain methods

Spatial filters estimate each sample of the underlying signal by applying a function to neighboring noisy samples within a local window. These methods typically perform well when the underlying signal varies smoothly. and lacks sharp transitions, but tend to incur in oversmoothing otherwise [12]. Classic examples include mean, median, and Wiener filters [100]. To alleviate over-smoothing, non-linear spatial domain methods such as median-based filters and total variation minimization [12, 30, 31] have been proposed. While they preserve details better than linear filters, they still fail to accurately preserve sharp transitions [32].

4.3.4 Transform-domain noise reduction methods

Transform-domain NR methods reduce noise by substituting noisy data with fitted models [101]. In particular, these methods represent y using a set of basis functions where signal and noise contribute to distinct coefficients, facilitating their separation. NR is then performed by reducing noise-related coefficients through thresholding [101]. The effectiveness of this approach depends critically on the choice of the transformation. Since the shape of noise-free signal is generally unknown and varies across applications, selecting an appropriate transformation is nontrivial.

In the context of wavelet-based NR methods, various thresholding strategies have been proposed [47, 48, 50], with varying success in preserving edges. However, independent of the thresholding strategy, the selection of an appropriate mother wavelet remains a challenge. An unsuitable choice limits the representation of signal features and degrades NR performance [52]. To mitigate this limitation, some approaches employ wavelet banks [51, 52]. The performance of these methods depends on user-defined parameters, including the number of decomposition levels, threshold values, and the set of selected wavelets [53, 54].

To circumvent the limitations associated with fixed bases in transform-domain NR methods, alternative data-driven signal representations have been proposed. A prominent example of methods for constructing data-driven bases is Empirical Mode Decomposition (EMD), which decomposes y into intrinsic mode functions (IMFs). Its apploication to NR involves discarding the noise-dominated components, and reconstructing the signal from the remaining IMFs [72, 73]. However, EMD lacks robustness in noisy settings, leading to non-unique IMFs, which degrades signal reconstruction quality [74–76]. Variational Mode Decomposition (VMD) mitigates these issues by formulating the decomposition as an optimization problem [53, 79], although its performance remains sensitive to user-defined parameters such as the number of IMFs and bandwidth constraints [72].

While data-driven representations may allow more flexible and accurate signal representations, they still lack mechanisms for adaptive parameter selection. Moreover, reliance on user-defined parameters can be a time-consuming task. We propose a novel NR method in which we combine optimal piecewise curve fitting to preserve discontinuities in x with a model aggregation strategy to expand the representation capabilities of the final signal estimator.

Furthermore, all the parameters needed by our signal estimator are systematically computed from y, removing the need of manual parameter tuning.

4.4 Definition of signals

The discrete signal of interest is denoted by $x \in \mathbb{R}^N$, where $N >> 1$ is an integer. The signal x is assumed to be obtained by sampling an unknown piecewise smooth [1] function $f : \mathbb{R} \to \mathbb{R}$, i.e. the n^{th} component of x is $x[n] := f(n\,\Delta t)$, $n \in \{0, 1, ..., N-1\} =: S$. The set S will be called the signal support, $t = n\,\Delta t$ denotes the continuous time variable and $\Delta t > 0$ is the uniform sampling interval. The signal x is assumed to be degraded by additive zero-mean discrete white Gaussian noise $w \in \mathbb{R}^N$ with standard deviation σ, i.e., the measured signal $y \in \mathbb{R}^N$ is defined as $y = x + w$, where the equality holds component-wise for each $n \in S$. The components of w will be assumed to be independent and identically distributed (*iid*). We adhere to the *iid* noise assumption which has been widely adopted in the NR literature [44, 45, 79, 82, 102]. In practice, this assumption holds because the aggregate effect of numerous independent stochastic sources, such as thermal fluctuations and electronic jitter in sensors converges toward a Gaussian distribution according to the Central Limit Theorem [8,19,20]. Examples of applications in which the uncorrelated Gaussian noise assumption serves as a suitable approximation include the instrumentation noise in digital photography [103], magnetic resonance imaging signals at high signal-to-noise ratios [104] and electrocardiography (ECG),where high-frequency stochastic interferences can be effectively modeled as Gaussian fluctuations [105].

It is also assumed that the function f may have a few localised irregularities, i.e. a few non-smooth regions having a relatively small extent, e.g. spikes. In this context, the problem is to estimate x using only y as input, assuming that both f and w are unknown.

We propose a signal estimator capable of selecting an optimal piecewise smooth signal model in which both the support and the underlying curve associated with each smooth segment of x are estimated. Specifically, we employ a piecewise curve-fitting strategy based on non-overlapping polynomial fits of degree ≤ 2. The optimal combination of piecewise fits is determined through information criterion minimization. This problem can be solved using mixed integer linear programming [106, 107]; however, we propose a novel formulation of the problem which allows using a shortest path optimizer to find the global solution more efficiently, significantly reducing the computing time compared to mixed integer linear programming.

[1] i.e. differentiable with respect to time t everywhere except at a finite number of localized points.

4.5 The Piecewise Smooth Model

In the regions where the function f is piecewise smooth, f has a Taylor series on each piece. It follows that it should be possible to find many intervals within S for which x can be approximated by a fit to a polynomial model. Generally speaking, our method consists in defining our signal estimate as a suitable sequence of local polynomial fits on y. More precisely, for an interval $S_{i,j} = \{i, i+1, \ldots, j\}$, where $(i,j) \in S^2$, a fit is obtained by performing a linear regression to the model $y[n] = \sum_{k=0}^{d} c_k \; x^k + w[n],\; n \in S_{i,j}$, where $d \in \{0, 1, 2\}$ is the degree of the polynomial and the c_ks are real parameters. The fit, or estimate, of the signal x within $S_{i,j}$ is then defined as

$$\hat{x}[n] := \sum_{k=0}^{d} \hat{c}_k \; n^k,\; n \in S_{i,j}, \tag{4.1}$$

where the $\hat{c}_k$s are the parameter estimates obtained via least squares minimisation. We will say that an interval is *smooth* if at least one of the three regressions (i.e. for $d = 0, 1, 2$) provides an acceptable fit, as determined by a χ^2-test at the 5% significance level. Otherwise, the interval is *non-smooth*. The set of all such smooth intervals will be denoted by $\mathcal{I}$.

Since x may contain a few localized irregularities, there may exist some signal points that do not belong to any smooth interval. Such points, which may be associated with spikes or local irregular textured patterns, will be called *outliers*. Since the piecewise-smooth hypothesis does not apply to these points, and also because they may carry relevant information, we choose to leave these signal values unchanged in the final signal estimate. More precisely, if $O \subset S$ denotes the set of all the time indices associated with outliers, then we define the signal estimates at these locations by

$$\hat{x}[n] := y[n],\; n \in O. \tag{4.2}$$

Combining both (4.1) and (4.2) allows to produce signal estimates for all points of S.

To construct an estimate of x over the whole support S, we propose to cover S with adjacent and disjoint intervals. These intervals are typically smooth, but can also be single-point if they are associated with an outlier. For each smooth interval, the signal estimate is obtained via (4.1). For each single-point interval, the signal estimate is obtained via (4.2).

The set $\mathcal{F}_{\alpha,\beta}$ of smooth intervals, which are associated with acceptable polynomial fits, is defined as

$$\mathcal{F}_{\alpha,\beta} := \{S_{i,j} \in \mathcal{I},\; (i,j) \in S^2 : \alpha \leq |S_{i,j}| \leq \beta \leq N-1\}, \tag{4.3}$$

where $|S_{i,j}|$ denotes the cardinality of $S_{i,j}$. Note that we typically use $\alpha \geq 4$ to ensure that the quadratic fit, which uses three parameters, is statistically meaningful. The set of single-point intervals, which are associated with outliers, is defined as

$$\mathcal{F}_o := \{\{i,i\},\ i \in O\}. \tag{4.4}$$

Note that these single points intervals are denoted as pairs $\{i,i\}$ in (4.4) because they will represent edges in a graph (section 4.7). We consider all the sequences of disjoint and adjacent intervals taken in

$$\mathcal{F} := \mathcal{F}_{\alpha,\beta} \bigcup \mathcal{F}_o$$

that completely cover S. Such a sequence will be called a *covering*. A collection of local polynomial models associated with a covering constitutes what will be called a *Piecewise Smooth Model* (PSM) of the unknown signal x. For each interval of a covering, there may be as much as three possible models, i.e. polynomials of degree 0, 1 and 2, hence there are generally several PSMs for each covering.

4.6 Information Criteria for a PSM

An information criterion (IC) is a model selection metric that quantifies the trade-off between goodness of fit and model complexity through a penalized likelihood function. In model selection scenarios involving multiple candidate models, a widely used approach involves selecting the model that minimizes an information criterion value. Different criteria weight the trade-off between goodness of fit and model complexity in distinct ways, with each criterion applying its own penalty structure for model complexity. This approach is appropriate for noise reduction, as the trade-off between fidelity and parsimony favors smoother models that avoid overfitting to noise components in the data.

Consider a PSM fitted on $y \in \mathbb{R}^N$, and composed by $m \in \mathbb{N}$ covering intervals that we label by the index $k \in \{1, 2, \dots, m\}$. Each covering interval is identified by the indices (i_k, j_k, d_k), where i_k and j_k denote the initial and final indices of the interval over which the k^{th} polynomial regression is performed, and d_k is the degree of the polynomial regression.

Assuming that σ is known, i.e. previously estimated with good accuracy, the curve-fitting parameters associated with the k^{th} interval are the polynomial coefficients, which are stored in the array $\theta_k := (c_{k,0}, c_{k,1}, ..., c_{k,d_k}) \in \mathbb{R}^{d_k+1}$, we provide a fast curve-fitting procedure in Supplementary Material S1. For a given PSM, all the curve-fitting parameters are stored in an array $\theta := (\theta_1, \theta_2, ..., \theta_m) \in \mathbb{R}^{\Delta}$, where $\Delta = \sum_{k=1}^{m}(d_k + 1)$. The PSM model also includes as parameters the $m-1$ position indices $j_1, j_2, ..., j_{m-1}$, which can be viewed as break-points that fully define the covering of a PSM. It follows that the total number Λ of parameters for a PSM is

$$\Lambda = (m-1) + \sum_{k=1}^{m} (d_k + 1). \tag{4.5}$$

The likelihood function associated with the k^{th} covering interval will be denoted by $\mathcal{L}(y_{i_k,j_k}|\theta_k)$, where $y_{i_k,j_k} := (y_{i_k}, y_{i_k+1}, ..., y_{j_k})$. Since noise values are mutually independent and the covering intervals are disjoint, the likelihood functions $\mathcal{L}(y_{i_k,j_k}|\theta_k)$ associated with a given PSM covering are independent. It follows that the likelihood function $\mathcal{L}(y|\theta)$ of a PSM takes the form

$$\mathcal{L}(y|\theta) = \prod_{k=1}^{m} \mathcal{L}(y_{i_k,j_k}|\theta_k). \tag{4.6}$$

For a given model, the log-likelihood function is maximized with respect to the parameter array θ. Using the notations $\ell(y|\theta) := \ln \mathcal{L}(y|\theta)$ and $\ell(y_{i_k,j_k}|\theta_k) := \ln \mathcal{L}(y_{i_k,j_k}|\theta_k)$, the maximized log-likelihood function takes the form

$$\ell(y|\hat{\theta}) = \sum_{k=1}^{m} \ell(y_{i_k,j_k}|\hat{\theta}_k), \tag{4.7}$$

where $\hat{\theta}$ and $\hat{\theta}_k$ denote the estimated values of the parameters θ and θ_k, respectively.

4.6.1 Akaike Information Criterion

Since the AIC is defined as $\text{AIC} = -2\,\ell(y|\hat{\theta}) + 2\,N_p$, where N_p is the total number of parameters [108], using (4.5) and (4.7) yields

$$\text{AIC} = -2 \sum_{k=1}^{m} \ell(y_{i_k,j_k} \mid \hat{\theta}_k) + 2\Big((m-1) + \sum_{k=1}^{m} (d_k + 1)\Big). \tag{4.8}$$

Since additive constants do not alter the optimal solution in minimization problems, we omit the term -2 in the penalty term to obtain an effective AIC, denoted by AIC_0, which can be written in the equivalent form

$$\text{AIC}_0 = \sum_{k=1}^{m} \Big[-2\ \ell(y_{i_k,j_k} \mid \hat{\boldsymbol{\theta}}_k) + 2(d_k + 2)\Big]. \tag{4.9}$$

The noise components are mutually independent and Gaussian with zero mean, hence the maximum likelihood for the k^{th} interval is

$$\begin{aligned}\mathcal{L}(y_{i_k,j_k}|\hat{\theta}_k) &= \Pi_{n=i_k}^{j_k} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\left(\frac{y[n]-\hat{x}[n]}{\sigma}\right)^2\right), \\ &= \frac{1}{(2\pi\sigma^2)^{n_k/2}} \exp\left(-\frac{1}{2}\frac{\| \, y_{i_k,j_k}-\hat{x}_{i_k,j_k} \, \|^2}{\sigma^2}\right).\end{aligned} \tag{4.10}$$

where $n_k := j_k - i_k + 1$. Taking the logarithm of this expression yields the maximized log-likelihood function, which takes the form

$$\ell(y_{i_k,j_k}|\hat{\theta}_k) = -\frac{n_k}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\frac{\| \, y_{i_k,j_k}-\hat{x}_{i_k,j_k} \, \|^2}{\sigma^2}. \tag{4.11}$$

Substituting (4.11) into (4.9) yields

$$\mathrm{AIC}_0 = \sum_{k=1}^{m}\left(n_k \ln(2\pi\sigma^2) + \frac{\| \, y_{i_k,j_k}-\hat{x}_{i_k,j_k} \, \|^2}{\sigma^2} + 2(d_k+2)\right). \tag{4.12}$$

Since $\sum_{k=1}^{m} n_k = N$ for all models, (4.12) becomes

$$\mathrm{AIC}_0 = N \ln(2\pi\sigma^2) + \sum_{k=1}^{m}\left(\frac{\| \, y_{i_k,j_k}-\hat{x}_{i_k,j_k} \, \|^2}{\sigma^2} + 2(d_k+2)\right). \tag{4.13}$$

In (4.13), the first term $N \ln(2\pi\sigma^2)$ is an additive constant that does not depend on the model, hence it can be ignored to compare different PSMs. Dropping the term $N \ln(2\pi\sigma^2)$ in (4.13) yields the following effective expression of the AIC, denoted by AIC_1:

$$\mathrm{AIC}_1 = \sum_{k=1}^{m}\left(\frac{\| \, y_{i_k,j_k}-\hat{x}_{i_k,j_k} \, \|^2}{\sigma^2} + 2(d_k+2)\right). \tag{4.14}$$

Note that for one-point segments associated with outliers, we use $d_k = 0$ in (4.14).

For a given fixed covering, AIC_1 is minimal if the weight $\frac{\|y_{i_k,j_k}-\hat{x}_{i_k,j_k}\|^2}{\sigma^2} + 2(d_k+2)$ is minimal for each smooth segment of the covering. The estimated signal $\hat{x}_{i_k,j_k}$ is actually a function $\hat{x}_{i_k,j_k}(d)$ of the degree d of the polynomial fit. This leads us to define d_k^* by

$$d_k^* \in \mathrm{argmin}_{d\in\{0,1,2\}} \frac{\| \, y_{i_k,j_k}-\hat{x}_{i_k,j_k}(d) \, \|^2}{\sigma^2} + 2\,(d+2). \tag{4.15}$$

Using the notation $\hat{x}^*_{i_k,j_k} := \hat{x}_{i_k,j_k}(d_k^*)$, this leads us to the final effective expression of the

AIC, denoted by $\overline{\mathrm{AIC}}$:

$$\boxed{\overline{\mathrm{AIC}} = \sum_{k=1}^{m} \left(\frac{\| y_{i_k,j_k} - \hat{x}^*_{i_k,j_k} \|^2}{\sigma^2} + 2\,(d_k^* + 2) \right).} \tag{4.16}$$

4.6.2 Bayesian Information Criterion (BIC)

Since the BIC is defined by $\mathrm{BIC} = -2\,\ell(y|\hat{\theta}) + \ln(N)\,N_p$, where N_p is the number of parameters [11], using (4.5) and (4.7) yields

$$\mathrm{BIC} = -2 \sum_{k=1}^{m} \ell(y_{i_k,j_k} \mid \hat{\theta}_k) + \ln(N) \left[(m-1) + \sum_{k=1}^{m} (d_k + 1) \right], \tag{4.17}$$

In (4.17), the term $-\ln(N)$ is an additive constant that does not depend on the model, hence it can be ignored to compare different PSMs. Dropping this term yields an effective BIC, denoted by BIC_0, which can be written in the equivalent form

$$\mathrm{BIC}_0 = \sum_{k=1}^{m} \left(-2\;\ell(y_{i_k,j_k} \mid \hat{\theta}_k) + \ln(N)\,(d_k + 2) \right). \tag{4.18}$$

Using (4.11), (4.18) becomes

$$\mathrm{BIC}_0 = \sum_{k=1}^{m} \left(n_k\;\ln(2\pi\sigma^2) + \frac{\| y_{i_k,j_k} - \hat{x}_{i_k,j_k} \|^2}{\sigma^2} + \ln(N)\,(d_k + 2) \right). \tag{4.19}$$

As previously, $\sum_{k=1}^{m} n_k = N$ for all coverings. Dropping this term from the summation yields another effective BIC, denoted by BIC_1, and given by

$$\mathrm{BIC}_1 = \sum_{k=1}^{m} \left(\frac{\| y_{i_k,j_k} - \hat{x}_{i_k,j_k} \|^2}{\sigma^2} + \ln(N)\,(d_k + 2) \right). \tag{4.20}$$

Note that for one-point segments associated with outliers, we use $d_k = 0$ in (4.20).

For a given fixed covering, BIC_1 is minimal if the weight $\frac{\|y_{i_k,j_k} - \hat{x}_{i_k,j_k}\|^2}{\sigma^2} + \ln(N)\,(d_k + 2)$ is minimal for each smooth segment of the covering. As for AIC, the estimated signal $\hat{x}_{i_k,j_k}$ is a function $\hat{x}_{i_k,j_k}(d)$ of the degree d of the polynomial fit. This leads us to define d_k^* by

$$d_k^* \in \mathrm{argmin}_{d \in \{0,1,2\}}\; \frac{\| y_{i_k,j_k} - \hat{x}_{i_k,j_k}(d) \|^2}{\sigma^2} + \ln(N)\,(d + 2). \tag{4.21}$$

Using the notation $\hat{x}^*_{i_k,j_k} := \hat{x}_{i_k,j_k}(d_k^*)$, this leads us to the final effective expression of the

BIC, denoted by $\overline{\mathrm{BIC}}$:

$$\boxed{\overline{\mathrm{BIC}} = \sum_{k=1}^{m} \left(\frac{\| y_{i_k,j_k} - \hat{x}^*_{i_k,j_k} \|^2}{\sigma^2} + \ln{(N)}\, (d^*_k + 2) \right).} \tag{4.22}$$

4.7 Optimal PSM obtained via a Shortest Path Algorithm

To each PSM corresponds a covering of S and the corresponding local polynomial fits. Each PSM composed of m segments can be characterized by a set of parameters of the form

$$\{\{i_k, j_k, d^*_k\},\ k = 1, 2, ..., m\}, \tag{4.23}$$

where i_k and j_k, $i_k \leq j_k$, are the boundaries of the k^{th} interval, and d^*_k is the degree of the polynomial fit that minimises the local weight, i.e. equations (4.15) and (4.21) for AIC and BIC respectively. The covering intervals are adjacent and disjoint, hence $i_k = j_{k-1} + 1$ for $2 \leq k \leq m$, with $i_1 = 0$ and $j_m = N - 1$. If we label all the PSM models by an integer index $p \in P$, then the expressions $\overline{\mathrm{AIC}}$ and $\overline{\mathrm{BIC}}$ given by (4.16) and (4.22) become functions $\overline{\mathrm{AIC}}(p)$ and $\overline{\mathrm{BIC}}(p)$ of the model index p. It follows that the best PSM models are solutions of the problems

$$\min\nolimits_{p \in P}\ \overline{\mathrm{AIC}}(p) \tag{4.24}$$

if we use the AIC, and

$$\min\nolimits_{p \in P}\ \overline{\mathrm{BIC}}(p) \tag{4.25}$$

if we use the BIC.

Both $\overline{\mathrm{AIC}}$ and $\overline{\mathrm{BIC}}$ are expressed as a sum of local contributions that depend only on the properties of each individual segment of a given model. For $\overline{\mathrm{AIC}}$, the contribution Ω_k of the k^{th} segment is

$$\Omega_k := \frac{\| y_{i_k,j_k} - \hat{x}^*_{i_k,j_k} \|^2}{\sigma^2} + 2(d^*_k + 2). \tag{4.26}$$

For $\overline{\mathrm{BIC}}$, the contribution of the k^{th} segment is

$$\Omega_k := \frac{\| y_{i_k,j_k} - \hat{x}^*_{i_k,j_k} \|^2}{\sigma^2} + \ln{(N)}\, (d^*_k + 2). \tag{4.27}$$

Each covering can be regarded as a path composed of edges that link the starting points of consecutive segments having weights given by (4.26) for AIC or by (4.27) for BIC. From this standpoint, the optimization problems (4.24) and (4.25) can be regarded as shortest-path problems in a weighted graph $G(V, E, W)$, where $V \subset S$, $E \subset S^2$ and W denote the vertices, edges and weights sets, respectively. These problems can be solved efficiently using Dijkstra's algorithm. In the following, we define E, V and W.

In Section 4.5, we denoted by $\mathcal{F}$ the set of all intervals $\{i, j\} \in S^2$ for which the signal is either smooth (i.e., a polynomial fit is adequate) or an outlier (i.e., no acceptable polynomial fit exists and $i = j$). Each $\{i, j\} \in \mathcal{F}$ defines an edge between the vertices i and $j+1$, i.e. each edge joins the starting points of consecutive adjacent segments in $\mathcal{F}$. Formally, the set of edges is defined as

$$E := \{\{i, j+1\} : \{i, j\} \in \mathcal{F}\}.$$

The corresponding set V of vertices is composed of the starting points of all the segments in $\mathcal{F}$, i.e.

$$V := \{i \in S : \{i, j\} \in \mathcal{F}\} \cup \{N\},$$

where one additional vertex at coordinate N must be added to include the last segment. Note that one may have $\{i, j\} = \{i, i\}$ if $\{i, j\}$ corresponds to an outlier.

For AIC, the set of weights is

$$W := \left\{ \frac{\| y_{i,j} - \hat{x}^*_{i,j} \|^2}{\sigma^2} + 2\,(d^*_{i,j} + 2) : \{i, j\} \in \mathcal{F} \right\}.$$

For BIC, the set of weights is

$$W := \left\{ \frac{\| y_{i,j} - \hat{x}^*_{i,j} \|^2}{\sigma^2} + \ln(N)\,(d^*_{i,j} + 2) : \{i, j\} \in \mathcal{F} \right\}.$$

Figure 4.1 illustrates the construction of the model graph for a noisy piecewise-constant signal. The left panel shows the noise-free signal x (black curve), which consists of three constant segments with values 1, 2, and 1; the noisy observations y (gray "$\times$" markers) with noise level $\sigma = 0.1$; and the resulting PSM estimate $\hat{x}$ (dashed curve). The right panel shows the model graph $G(V, E, W)$, in which each gray directed edge represents a finite-weight segment model. The optimal path consists of the three edges $\{1, 9\}$, $\{9, 17\}$ and $\{17, 25\}$ (black arrows), which correspond to the AIC-optimal PSM. This model comprises three constant fits over the consecutive intervals $\{1, 8\}$, $\{9, 16\}$, and $\{17, 24\}$, respectively.

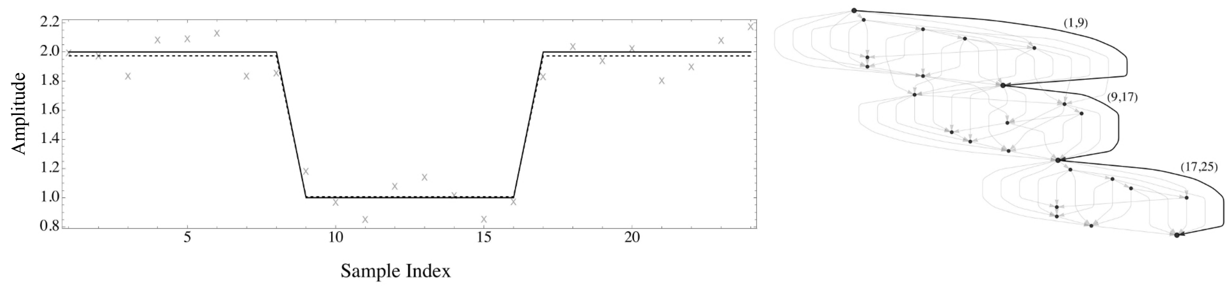

Figure 4.1 **Left**: Signals: x (solid black curve), y ($\times$), and AIC-optimal PSM $\hat{x}$ (dashed curve). **Right**: Smooth model graph $G(V, E, W)$ for y; vertices (black dots), edges (grey arrows), and optimal path (black arrows).

4.8 Results for the PSM noise reduction method

4.8.1 Numerical Experiments: description and motivations

In this section, we evaluate the performance of the method when y is a periodic signal contaminated with additive white Gaussian noise. We consider square, triangular, and sinusoidal waves with various periods and noise levels, and we compare the noise-reduction performance obtained when PSM selection uses either AIC or BIC as the objective function.

Our motivation for selecting these signals is to assess PSM's ability to address three types of challenges in noise-reduction estimation: discontinuities (square waves), angular points (triangular waves), and smooth variations that are neither constant, linear, nor parabolic (sinusoidal waves). Periodicity allows the estimation challenge —such as accurately detecting the location of a discontinuity— to recur multiple times within a single waveform, thereby allowing for a more reliable and comprehensive evaluation of performance.

To quantify the relative strength of x with respect to the noise w, we use the signal-to-noise ratio (SNR). In the context where the noise standard deviation σ is known, the SNR can be defined as

$$\mathrm{SNR} = \left(\frac{\| x \|^2}{\mathbb{E}(\| w \|^2)}\right)^{1/2} = \frac{\| x \|}{\sqrt{N}\,\sigma},$$

where $\| \ldots \|$ denotes the Euclidian norm. After noise reduction, the noise is effectively reduced to the estimation error $\hat{x} - x$ so that the effective SNR increases to a value SNR* given by

$$\mathrm{SNR}^* = \left(\frac{\| x \|^2}{\| \hat{x} - x \|^2}\right)^{1/2} = \frac{\| x \|}{\| \hat{x} - x \|}.$$

The corresponding SNR improvement, denoted by SNRI, is defined as

$$\mathrm{SNRI} := \frac{\mathrm{SNR}^*}{\mathrm{SNR}} = \frac{\sqrt{N}\,\sigma}{\|\,\hat{x} - x\,\|},$$

where SNRI > 1 indicates that $\hat{x}$ is a more accurate representation of x than y, and larger SNRI values correspond to stronger noise suppression.

To ensure fair comparisons across different signal shapes, all noise-free signals are normalized to satisfy $\frac{\|x\|}{\sqrt{N}} = 1$, yielding SNR $= 1/\sigma$. This normalization decouples the noise level from the particular choice of x, thus allowing noise-level regimes to be controlled only by σ.

The PSM estimation method was applied to periodic signals of length $N = 512$, and period $T \in \{8, 12, ..., 64\}$. For the noisy signals, we used $\sigma \in \{0.01, 0.025, 0.05, 0.1, 0.2\}$. The corresponding set of SNR values is $\{100, 40, 20, 10, 5\}$, spanning low to moderately large noise scenarios.

4.8.2 Results for SNRI

Figure 4.2 presents the average SNRI, computed over 200 independent noise realizations, as a function of T. Results are shown for both AIC-based (black curves) and BIC-based (grey curves) estimation, applied to square (left), triangular (center), and sinusoidal (right) waveforms at SNR $= 20$. It can be observed that SNRI increases with T. Across signal types, the highest and lowest SNRI values were achieved for square, and sinusoidal waves, respectively. For other SNR values, we similarly observe that SNRI increases with T, and that BIC-based PSM outperforms AIC-based PSM for smoother signals.

For very small periods ($T < 12$), the average SNRI tends to fall below 1 for both AIC- and BIC-based selection. This behavior is not unexpected, as such small periods lead to a substantial increase in the number of PSM parameters relative to the sample size. In particular, when $T = 12$ and $N = 512$, the number of smooth segments in x is $512/6 \approx 85$. Thus, x consists of 84 breakpoints and at least one parameter defining the curve within each segment, resulting in a minimum of $k = 85 + 84 = 169$ parameters. In this setting, $N/k \approx 3 < 40$, which corresponds to a small-sample regime [109], where both AIC and BIC are known to be prone to modeling errors.

Figure 4.3 displays SNRI as a function of SNR for the three signal types, with $T = 16$ For square waves (left panels), BIC-based selection (grey curves) consistently outperforms AIC-based selection (black curves) across all SNR levels. The performance gap widens at $T = 64$, where BIC attains SNRI values above 5, compared to approximately 2.5 for AIC. For

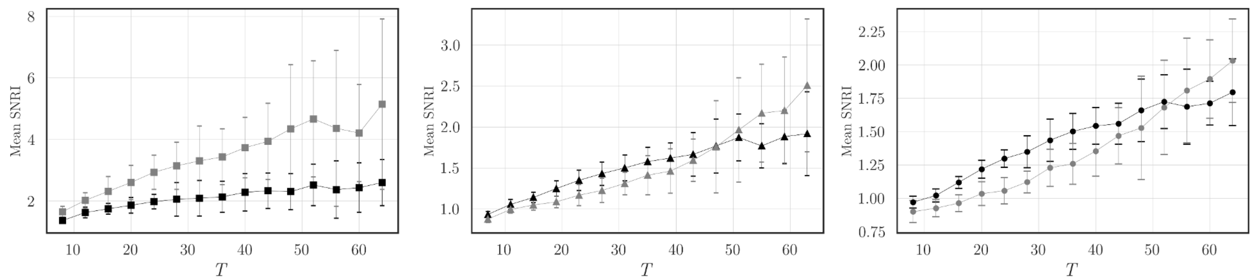

Figure 4.2 Mean SNRI ($n = 200$) versus period T for square (left), triangular (center), and sinusoidal (right) waves at SNR = 20. Black and grey curves correspond to AIC and BIC minimization, respectively. Error bars represent ±2 standard deviations.

triangular waves at $T = 16$ (top-center panel), AIC slightly outperforms BIC, particularly at intermediate SNR values. However, for the larger period $T = 64$ (bottom-center panel), BIC again achieves higher SNRI.

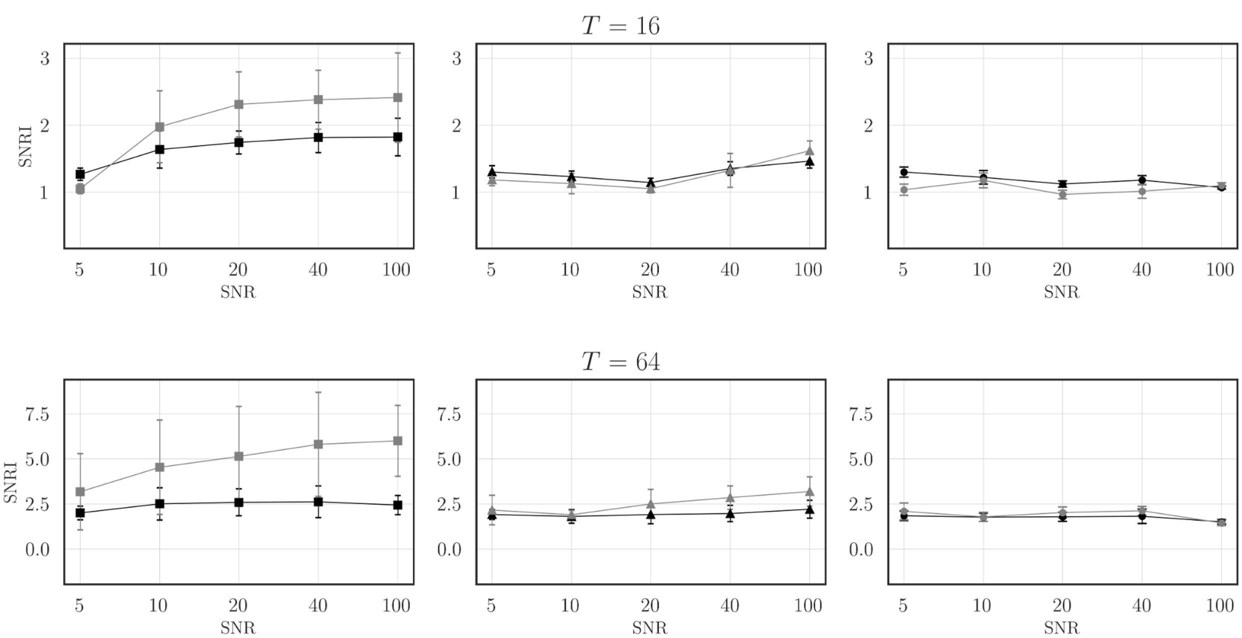

Figure 4.3 Average SNRI is shown as a function of SNR for $T = 16$ (top) and $T = 64$ (bottom). The curves represent AIC-based (black) and BIC-based (grey) PSM selection for square (left), triangular (center), and sinusoidal (right) waves. Error bars indicate ±2 standard deviations.

Figures 4.2 and 4.3 suggest that for simpler signals (those with fewer parameters) and larger periods T, BIC yields a higher SNRI than AIC. This result aligns with the inherent penalties of each criterion: for a model with K parameters and sample size N, AIC adds a $2K$ term

(Eq. (4.16)), whereas BIC adds $\ln(N)\,K$ (Eq. (4.22)). Consequently, for $N \geq 8$, BIC penalizes complexity more strictly than AIC since $\ln(8) > 2$, leading it to favor simpler models.

4.8.3 Interpretation

To better understand the NR performance of PSM across the three periodic signals, Figures 4.4 and 4.5 provide representative estimation examples for $T = 16$ and $T = 64$, respectively. In both figures, the AIC-minimal estimates are displayed in the left column, while BIC-minimal estimates appear in the right.

The PSM estimation errors can be categorized into three types: oversmoothing, discontinuity artifacts, and errors arising from incorrect local-model selection.

Oversmoothing errors may occur when BIC is applied to rapidly varying signals (e.g., small periods T). For instance, the right column of Figure 4.4 shows that BIC visibly rounds the peaks of the triangular and sinusoidal waves (middle and bottom panels, respectively). This suppression of fine-scale variations explains why the AIC-minimal PSM yields a slightly superior SNRI for these signals, as shown in the top-center panel of Figure 4.3.

Discontinuity artifacts may arise because the local models used in PSM do not overlap, which can lead to discontinuities between adjacent models. For example, such artifacts can be observed in the top-left panel of Figure 4.4, as well as in Figure 4.5 (left column and bottom-right panel). These artifacts are less likely to occur when BIC is used.

Errors in local model selection stem from three primary causes. First, the limited repertoire of local models —comprising constant, linear, or parabolic segments— restricts adaptability. Second, ambiguity occurs when different global models exhibit statistically comparable AIC values. Finally, discontinuity artifacts can also precipitate the selection of suboptimal local models. For instance, the top-left panel of Figure 4.5 shows several constant plateaus modeled by linearly varying segments. Similarly, the bottom-left panel illustrates the use of linear segments to approximate the sinusoid near samples 325, 350, and 380.

We conclude that the AIC-minimal PSM is generally more effective than the BIC-minimal version when the signal x contains a large number of segments relative to the sample size. In contrast, BIC yields superior results as the number of pieces in x decreases (e.g., for larger periods T of a periodic signal).

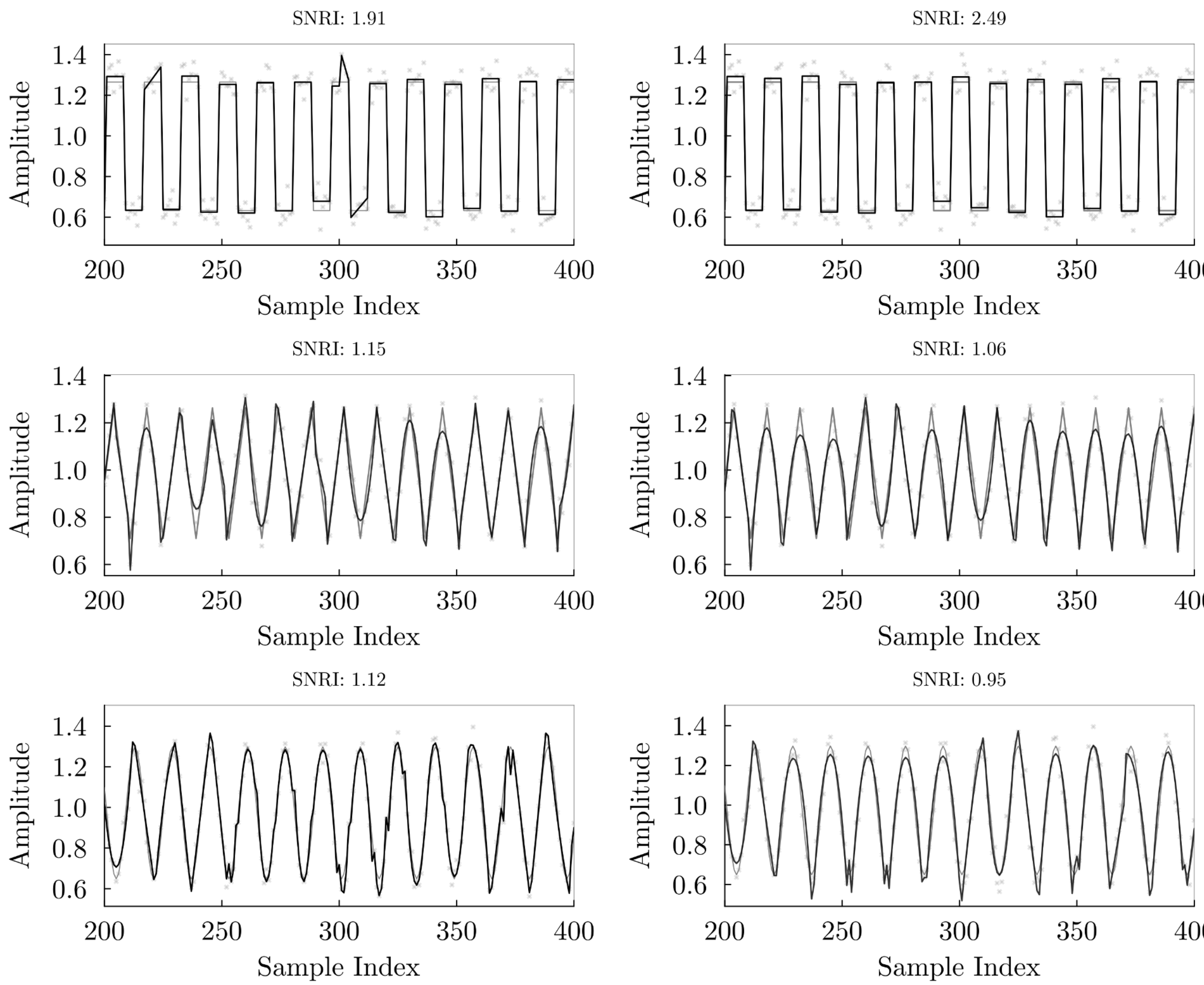

Figure 4.4 Zoomed-in signal comparison ($T = 16$): original signal x (grey curve), noisy observation y with SNR = 20 (grey $\times$ markers), and PSM estimate $\hat{x}$ (black curve) obtained by minimizing AIC (left column) and BIC (right column). Rows corespond to square (top), triangular (middle), and sinusoidal (bottom) signals.

4.9 The sliding-window PSM noise reduction method

4.9.1 PSM selection optimisation via sliding-window processing

In the previous section, we showed that PSM can produce discontinuity artifacts and local model-selection errors. Such errors may arise because several slightly different models can exhibit comparable statistical performance when evaluated using an information criterion.

To improve the signal estimation accuracy of the PSM method while reducing the computational cost of solving (4.24) and (4.25), we propose combining PSM-based model selection with a sliding-window (SW) approach. It will be shown that PSM can produce significantly

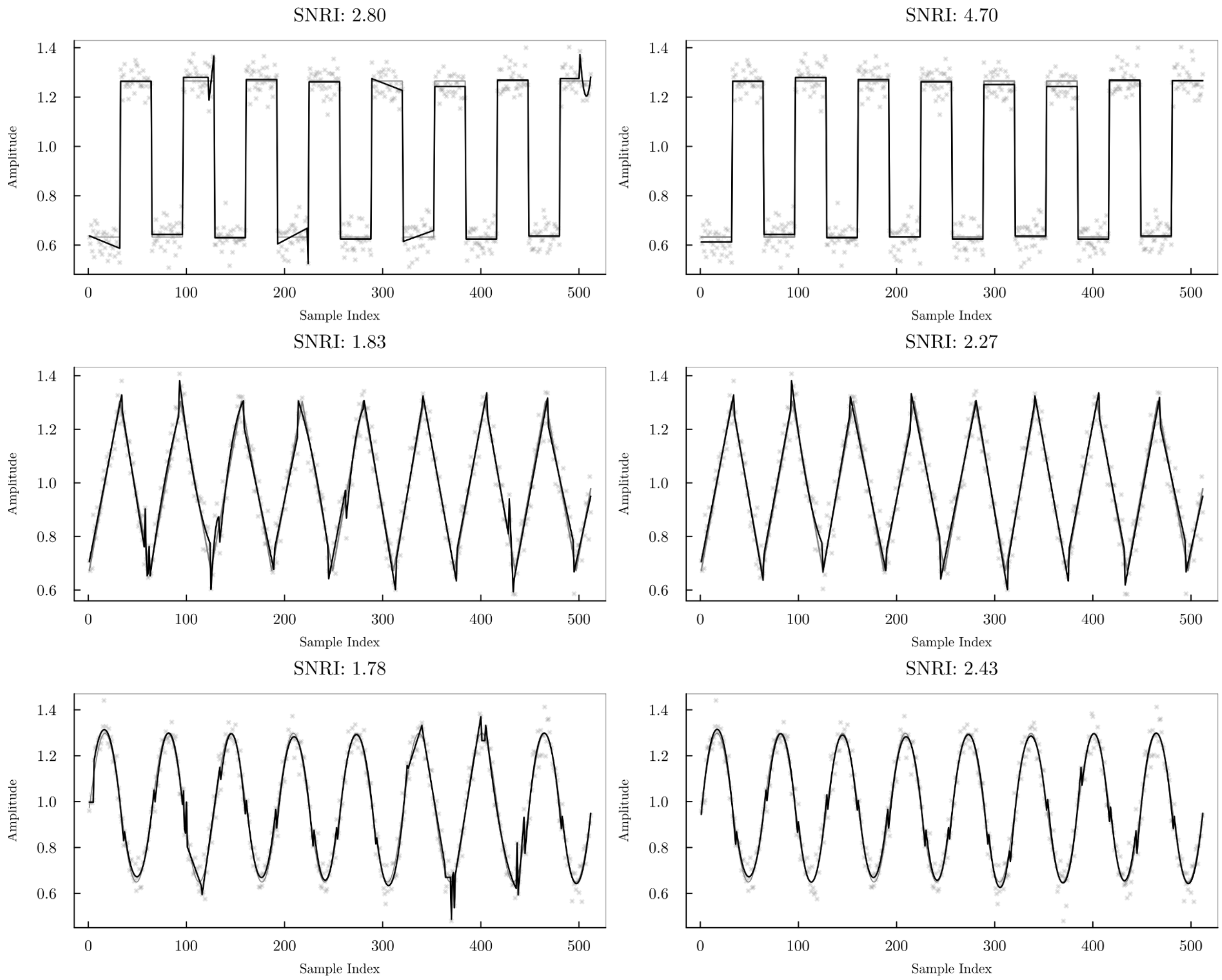

Figure 4.5 Signal comparison for $T = 64$: original signal x (grey curve), noisy observation y with SNR $= 20$ (grey $\times$ markers), and PSM estimate $\hat{x}$ (black curve) obtained by minimizing AIC (left column) and BIC (right column). Rows corespond to square (top), triangular (middle), and sinusoidal (bottom) signals.

different signal estimates for overlapping windows. This occurs because multiple candidate models often have information criterion values that are sufficiently close to the minimum to be considered statistically indistinguishable. By combining estimates obtained from different windows, this model-selection ambiguity can be reduced, which is particularly beneficial for complex signals.

Rather than computing a single PSM estimate, we define the estimator $\hat{x}$ by suitably combining multiple local PSM estimates derived from overlapping windows. We refer to the resulting framework as the Sliding Window Piecewise-Smooth Model (SW-PSM). This strategy allows, for each pixel, the selection of the combination of local PSM models that provides higher estimation accuracy.

The first step of SW-PSM consists in defining evenly-spaced windows of size L. Generally, L should be at least as large as the smallest available curve-fit, i.e., such that $L \geq \alpha$. The n^{th} window is the interval $[n\,\Delta,\, n\,\Delta + L - 1]$, where $n \in \{0, 1, ..., \overline{n}\}$, $\Delta \in \mathbb{N}$, with $\Delta \geq 1$ is the step-size (the overlap between consecutive windows is $L - \Delta$), and $\overline{n} := \lfloor (N - L)/\Delta \rfloor$. If $\overline{n}\,\Delta + L - 1 < N - 1$, then an additional window $[N - 1 - (L - 1),\, N - 1]$ is added to ensure coverage of the final signal samples. In the second step, PSM estimation is applied to each windowed signal segment, yielding signal estimates for each window. In practice, the local signal estimates are computed with the minimum and maximum curve-fitting scales set to 4 and L respectively.

Since a given pixel location $i \in S$ may belong to several overlapping windows, the SW-PSM method generates multiplex estimates of $x[i]$ for each $i \in S$. We denote the k^{th} of these estimates by $\hat{x}_k[i]$. For each $i \in S$, we define the set $X_i := \{\hat{x}_1[i], \hat{x}_2[i], ..., \hat{x}_{n_i}[i]\}$, where $n_i \in \mathbb{N}$ is the number of windows containing the signal value $y[i]$. To give greater influence to more precise estimates in the final reconstruction, we define the final estimate $\hat{x}[i]$ as $\hat{x}[i] := \sum_{k=1}^{n_i} \omega_{i,k}\ x_k[i]$, where the weights $(\omega_{i,1}, \omega_{i,2}, , ..., \omega_{i,n_i})$ are determined using inverse-variance weighting. Specifically, each $\hat{x}_k[i]$ is assigned the weight $\omega_{i,k} := \sigma_{i,k}^{-2} (\sum_{k=1}^{n_i} \sigma_{i,k}^{-2})^{-1}$, where $\sigma_{i,k}$ denotes the estimated standard deviation of $\hat{x}_k[i]$, as provided by the curve-fitting procedure (Eq. (S8) in the Supplementary Material). While IVW is optimal when estimates are independent, the standard variance formula tends to underestimate true variance in the presence of correlation. Furthermore, using an excessively small delta should be avoided, as it can artificially inflate the correlation between estimates.

To prevent the smoothing of spikes that may contain relevant features of x, each X_i is used to compute the number γ_i of instances in which $\hat{x}_k[i]$ is classified as an outlier that cannot be satisfactorily represented by a smooth curve. If all estimates in X_i are outliers ($\gamma_i = |X_i|$), the original signal value $y[i]$ is retained, i.e., $\hat{x}[i] := y[i]$. Otherwise, when $\gamma_i < |X_i|$, $\hat{x}[i]$ is calculated as a weighted average based solely on the fitted polynomials, excluding outliers. This approach reduces the influence of unreliable estimates and promotes signal estimation grounded in models with stronger statistical support.

To evaluate the performance of our method, we consider a family of 500 piecewise polynomial signals of size $N = 512$. The k^{th} polynomial piece is defined by a random degree d_k uniformly sampled from 3 to 6, and a random support size uniformly sampled from 20 to 80. To generate the kth segment, we select $d_k + 1$ equally spaced points within its support and assign them random values uniformly sampled from the interval $[1, 2] \subset \mathbb{R}$. The degree-$d_k$ polynomial corresponding to the kth segment is obtained using Van der Monde polynomial

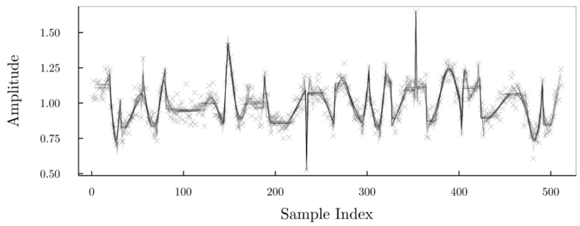

Figure 4.6 **Left**: Noisy signal y (grey $\times$, SNR $= 10$), and AIC-optimal local PSM estimates (grey, overlapping curves) ($L = 20$, $\Delta = 1$) . **Right**: y (grey $\times$, SNR $= 10$), x (grey), and SW-PSM estimate (black), i.e., aggregation of local estimates.

interpolation [110] through the randomly generated points. The piecewise polynomial signal x is obtained by juxtaposing M polynomial pieces generated using the method above. To ensure that the concatenated signal attains the desired length, we concatenate polynomial segments until at least 512 points are obtained, and then retain the first 512 samples. To simulate the presence of outliers, we generated five random and independent spike locations uniformly distributed on the support $\{0, 1, ..., N-1\}$. At each spike location, a random increment with a value sampled uniformly from $[-1, 1]$ was added to x. To generate each noisy signal y, we added synthetic white Gaussian noise to each piecewise polynomial signal.

Figure 4.6 provides an illustration of the SW-PSM method using AIC as model selection criterion, window size $L = 20$, and step-size $\Delta = 1$. The noise-free signal x (grey curve in right panel) is one of the piecewise-polynomial signals and y (grey $\times$'s in both panels) is the corresponding noisy signal at SNR $= 10$. Figure 4.6 suggests that a suitable aggregation of the local signal estimates (grey curves, left panel) produces a global signal estimate (black curve, right panel) that can describe smooth functions that are not necessarily included in the set of local models (constant, linear, and parabolic).

4.9.2 Effect of window size L on noise reduction quality and computing time

SW-PSM requires the selection of a window size L and a step size Δ, which influence both computational cost and NR performance. A larger L allows the PSM selection procedure to incorporate a broader signal context into each local estimate, but increases computation time by increasing the number of edges in $G(V, E, W)$. A smaller step size Δ increases window overlap and the number of estimates per $x[i]$, potentially improving precision at the expense of higher computational cost. However, small values of Δ may also produce more strongly correlated estimates, which can render inverse-variance weighting suboptimal [111].

We examine the effect of L on SNRI and computational time by applying the SW-PSM method to the family of piecewise-polynomial signals described in the previous section, at SNR $= 10$. The value SNR$= 10$ is chosen as a representative intermediate noise level, avoiding both noise-dominated regimes at low SNR and trivial estimation scenarios at high SNR. The procedure is repeated for $L \in \{10, 15, ..., 50\}$, with the step size fixed at $\Delta = 1$ for all values of L. Figure 4.7 shows the mean SNRI (left panel) and the mean execution time in seconds

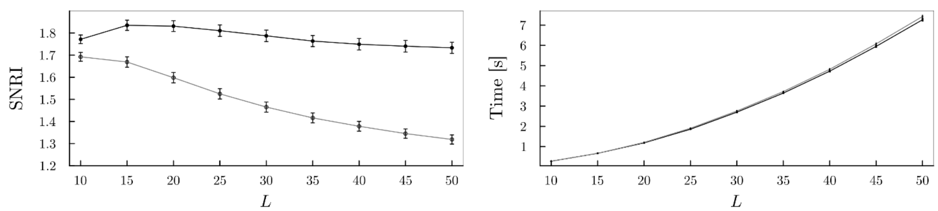

Figure 4.7 Mean SNRI of SW-PSM ($\Delta = 1$) over 500 piecewise polynomial signals (SNR $=$ 10) versus L (left), and computing time versus L (right). Black and grey curves denote AIC- and BIC-minimal estimation.

(right panel), both computed over the 500 signals, as functions of L. The error bars represent two standard deviations of the corresponding mean values. On the left panel, we observe that for AIC-optimal SW-PSM estimation (black curve), the mean SNRI increases with L over the range $L \in [10,\ 20]$, and is maximal for $L \approx 20$. For BIC-minimal estimation (grey curve), the mean SNRI decreases as L increases. Furthermore, AIC-optimal estimation outperformed its BIC counterpart for all values of L. The computation time exhibits a nearly identical dependence on L for both AIC- and BIC-minimal SW-PSM. For these signals, $L > 25$ leads to a significant increase in computation time, e.g., $L = 40$ results in an approximately fivefold time increase compared with $L = 20$. Experiments conducted at other SNR values exhibit similar qualitative trends in both SNRI and computational time.

In Figure 4.7, the lower SNRI at small L arises from reduced local sample sizes, which cause the information criterion to select overly complex models, leading to overfitting. Consequently, the estimate follows the noise rather than the true signal. As shown in Figure 4.8 ($L = 10$, top panel), this results in small, spurious oscillations, e.g., near indices 100, 200, and 300. Conversely, large L values decrease SNRI because the information criterion favors low-degree (≤ 2) polynomial fits over wide intervals, which can result in oversmoothing, as seen in Figure 4.8 ($L = 50$, bottom panel), e.g., near indices 50, 100, and 180.

The window size $L = 20$ (Figure 4.8, center) achieves the highest SNRI because the signals considered are typically smooth over intervals of this length, making our simple local models (polynomials of degree 0–2) appropriate.

Representative SW-PSM estimates ($L = 20$, $\Delta = 1$, SNR $= 10$) for a piecewise polynomial

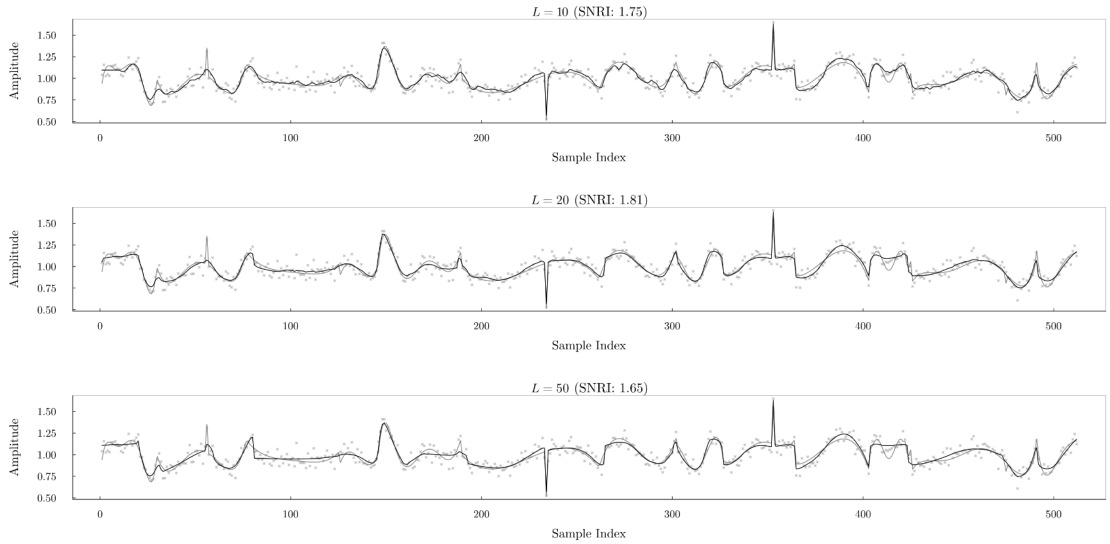

Figure 4.8 x (grey curve), y (grey $\times$'s, SNR $= 10$), and AIC-minimal SW-PSM estimate (black curve) for $L \in \{10, 20, 50\}$ (top to bottom) and $\Delta = 1$. The corresponding SNRI values are displayed above each panel.

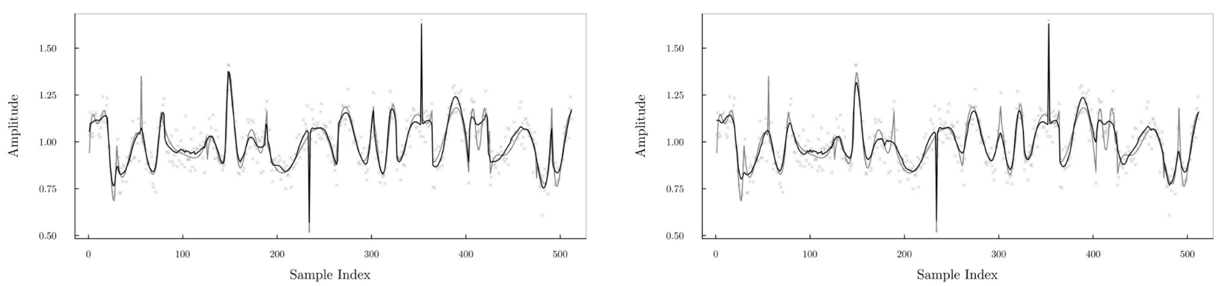

Figure 4.9 **Left**: AIC-minimal SW-PSM estimate (black; SNRI = 1.80), pure signal x (grey curve) and noisy signal y with $\sigma = 0.1$ (grey $\times$'s). **Right**: BIC-minimal SW-PSM estimate (black; SNRI = 1.54) for the same x (grey) and y (grey $\times$'s) at SNR=10.

signal are shown in Figure 4.9 for AIC- and BIC-minimal SW-PSM estimates. While AIC-minimal estimation captures all local spikes (left panel), the BIC-minimal variant (right panel) can oversmooth these features, e.g. failing to detect the rapid transitions near indices 400 and 490. Consequently, AIC-minimal estimation typically achieves a higher SNRI (1.80) than BIC (1.54) for this signal family. These findings are consistent with the BIC-induced oversmoothing of fast-varying signals noted in Section 4.8.

4.9.3 Choosing L and Δ

The window size L should chosen so that the local models allow for a sufficiently accurate estimate of x in each window. One approach is to select L such that the signal is typically smooth (approximately linear) over intervals of length L. This increases the likelihood that low-degree polynomials (degrees 0,1,2) yield a good fit to y in each window.

Since the window size L bounds the support of the local models used to fit y within each window, L must be large enough to ensure accurate local fits and effective noise reduction. As shown in Figure 4.7 for our reference signals, i.e. the family of piecewise polynomials described previously, the SNRI peaks between $15 \leq L \leq 25$ and declines only marginally for $L \geq 25$. Accordingly, we adopt $L \geq 16$ as a suitable lower bound for our reference signals.

These observations motivate using SW-PSM with a window size equal to a characteristic smoothness scale L^*, defined as the typical support length over which x can be considered locally smooth. More precisely, we consider a collection $\mathcal{L}_A$ of *acceptable linear fits* on y over all the intervals in $\mathcal{F}_{\alpha=16,\beta=32}$, where acceptability is determined using a χ^2-test at significance level 0.05. We then define L^* as the median of the support sizes of all fits in $\mathcal{L}_A$. The parameter $\alpha = 16$ serves as an empirical lower bound based on the observations in Figure 4.7, indicating that $L = 15$ is close to optimal for our reference signals. The upper bound $\beta = 32$ is used promote data-adaptability and limit computing cost.

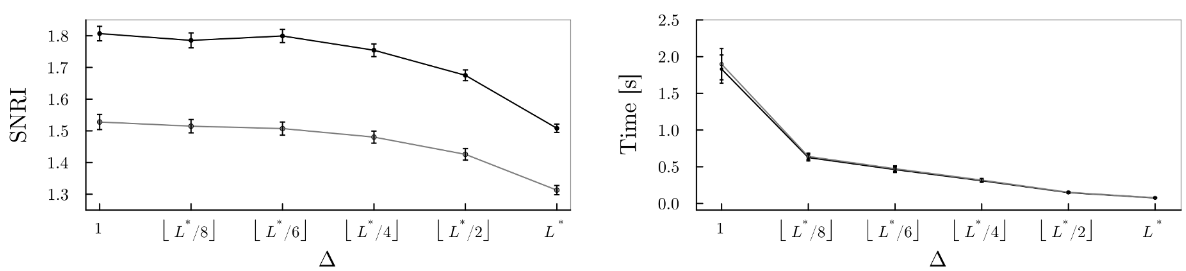

Figure 4.10 Mean SNRI of SW-PSM ($L = L^*$) over 500 piecewise polynomial signals (SNR = 10) as a function of Δ (left), and corresponding computional time versus Δ (right). Black and grey curves denote AIC- and BIC-minimal estimation.

To assess the effect of the window step-size Δ on NR performance and computation time, we measured the average SNRI as a function of Δ using $L = L^*$ for the same 500 piecewise polynomial of signals at SNR=10. The results are presented in Figure 4.10 for $\Delta \in \{1, \lfloor L^*/8 \rfloor, \lfloor L^*/6 \rfloor, \lfloor L^*/4 \rfloor, \lfloor L^*/2 \rfloor, L^*\}$. In the left panel, we observe that for both AIC-optimal and BIC-optimal SW-PSM estimation (black and grey curves, respectively), SNRI decreases as Δ increases. For $\Delta > \lfloor L^*/6 \rfloor$, the degradation in NR performance becomes more pronounced. In the right panel, computation time decreases as Δ increases, following nearly identical trends for AIC- and BIC-minimal SW-PSM. Moreover, setting $\Delta = \lfloor L^*/6 \rfloor$ significantly reduces computation time compared with $\Delta = 1$. For our reference signals, these experiments suggest that $\Delta = \lfloor L^*/6 \rfloor$, provides a suitable balance between estimation accuracy and computational cost. In the following experiments, we therefore use $L = L^*$ and $\Delta = \lfloor L^*/6 \rfloor$.

4.9.4 Comparison of PSM vs SW-PSM

In Section 4.8, we identified three main sources of reconstruction errors associated with signal estimation via PSM selection: local model limitations, where the local polynomial fits cannot accurately represent non-polynomial or fast-varying regions in x; model ambiguity, caused when information criteria yield similar values across candidates; and discontinuity artifacts resulting from the non-overlapping constraint in the PSM formulation. The SW-PSM framework is designed to mitigate all three effects by aggregating multiple locally optimal models estimated within overlapping windows, reducing sensitivity to model ambiguity, improving signal representation, and attenuating discontinuity artifacts.

We compare the performance of PSM and SW-PSM on the previously introduced periodic signals at SNR $\in \{10, 20\}$. In Table 4.1, we report the mean SNRI (over $n = 200$ noise realizations) for PSM and SW-PSM with AIC-optimal and BIC-optimal model selection, for periodic square, triangular, and sinusoidal signals at SNR $\in \{10, 20\}$ and periods $T \in \{16, 32, 64\}$, as well as for our reference piecewise polynomial signals. Bold type indicates the method yielding the highest mean SNRI for each signal. For square signals, BIC-optimal PSM achieves the highest SNRI values across all tested periods and noise levels, with SW-PSM yielding slightly lower performance. For triangular and sinusoidal signals, SW-PSM outperforms PSM for both AIC and BIC selection. For piecewise polynomial signals, AIC-optimal SW-PSM attains the highest SNRI values, outperforming both BIC-optimal SW-PSM and standard PSM.

The origin of the performance differences observed in Table 4.1 are clarified by the corresponding reconstruction figures. Figure 4.11 displays representative examples of the PSM (left column) and SW-PSM (right column) estimates for the periodic signals at $T = 64$ and SNR=20, and a piecewise polynomial signal. In line with the results displayed in Table 4.1, we used BIC-optimal model selection for periodic signals, and AIC-optimal selection for the piecewise polynomial example. The resulting SNRI value for each signal example is displayed at the top of each panel within the figure.

For square waves, Figure 4.11 shows that the BIC-minimal PSM is a global piecewise-constant model (black curve in the top-left panel), which matches x (grey curve) structure accurately.

Table 4.1 Mean SNRI ± (2 std) for PSM and SW-PSM with AIC/BIC model order selection at SNR = 20 and SNR = 10.

Signal	T	PSM (AIC)	PSM (BIC)	SW-PSM (AIC)	SW-PSM (BIC)
			SNR = 20		
Square	16	1.74 ± 0.17	$\mathbf{2.31 \pm 0.48}$	1.89 ± 0.02	2.15 ± 0.02
	32	2.13 ± 0.40	$\mathbf{3.32 \pm 0.94}$	2.44 ± 0.07	3.03 ± 0.09
	64	2.48 ± 0.65	$\mathbf{5.02 \pm 2.44}$	2.98 ± 0.05	3.93 ± 0.08
Triangular	16	1.14 ± 0.08	1.05 ± 0.06	$\mathbf{1.31 \pm 0.01}$	1.28 ± 0.01
	32	1.54 ± 0.21	1.31 ± 0.15	1.80 ± 0.02	$\mathbf{1.84 \pm 0.03}$
	64	1.91 ± 0.38	2.32 ± 0.64	2.37 ± 0.03	$\mathbf{2.72 \pm 0.03}$
Sine	16	1.12 ± 0.04	0.96 ± 0.06	1.43 ± 0.01	$\mathbf{1.44 \pm 0.01}$
	32	1.43 ± 0.12	1.23 ± 0.12	1.85 ± 0.02	$\mathbf{2.01 \pm 0.02}$
	64	1.80 ± 0.29	2.02 ± 0.43	2.49 ± 0.02	$\mathbf{3.34 \pm 0.05}$
Piecewise Polynomial	–	1.42 ± 0.22	0.99 ± 0.22	$\mathbf{1.79 \pm 0.01}$	1.57 ± 0.01
			SNR = 10		
Square	16	1.63 ± 0.28	$\mathbf{1.97 \pm 0.53}$	1.69 ± 0.02	1.79 ± 0.02
	32	2.02 ± 0.40	$\mathbf{3.26 \pm 1.09}$	2.38 ± 0.06	2.92 ± 0.09
	64	2.44 ± 0.68	$\mathbf{4.97 \pm 2.47}$	2.87 ± 0.05	3.79 ± 0.07
Triangular	16	1.23 ± 0.08	1.12 ± 0.15	1.48 ± 0.01	$\mathbf{1.55 \pm 0.03}$
	32	1.51 ± 0.14	1.63 ± 0.13	1.93 ± 0.02	$\mathbf{2.08 \pm 0.02}$
	64	1.84 ± 0.25	1.91 ± 0.20	2.53 ± 0.03	$\mathbf{2.95 \pm 0.03}$
Sine	16	1.22 ± 0.10	1.17 ± 0.11	1.55 ± 0.01	$\mathbf{1.65 \pm 0.01}$
	32	1.48 ± 0.12	1.42 ± 0.13	2.05 ± 0.02	$\mathbf{2.37 \pm 0.02}$
	64	1.81 ± 0.21	1.78 ± 0.20	2.71 ± 0.03	$\mathbf{3.24 \pm 0.04}$
Piecewise Polynomial	–	1.45 ± 0.25	0.98 ± 0.24	$\mathbf{1.79 \pm 0.01}$	1.50 ± 0.01

This is evidenced by the accurate localization of discontinuities between pieces and the flat plateau estimates. In contrast, SW-PSM relies on locally estimated models whose window boundaries can fall inside constant plateau intervals. When this occurs, one of the true plateau boundaries is excluded from the window, which can introduce discontinuity artifacts within an otherwise constant region, (e.g. near indices 230 and 390 in the SW-PSM estimate for the square wave) slightly degrading the aggregated reconstruction and leading to lower SNRI than PSM.

The PSM estimates for triangular and sinusoidal signals (left panels of the second and third rows in Figure 4.11) exhibit discontinuity artifacts between adjacent smooth curves. In contrast, the SW-PSM estimates for these signals (right panels of the same rows) are visibly smoother and exhibit fewer discontinuity artifacts. As a result, higher SNRI values are obtained by SW-PSM for these signals in comparison to PSM.

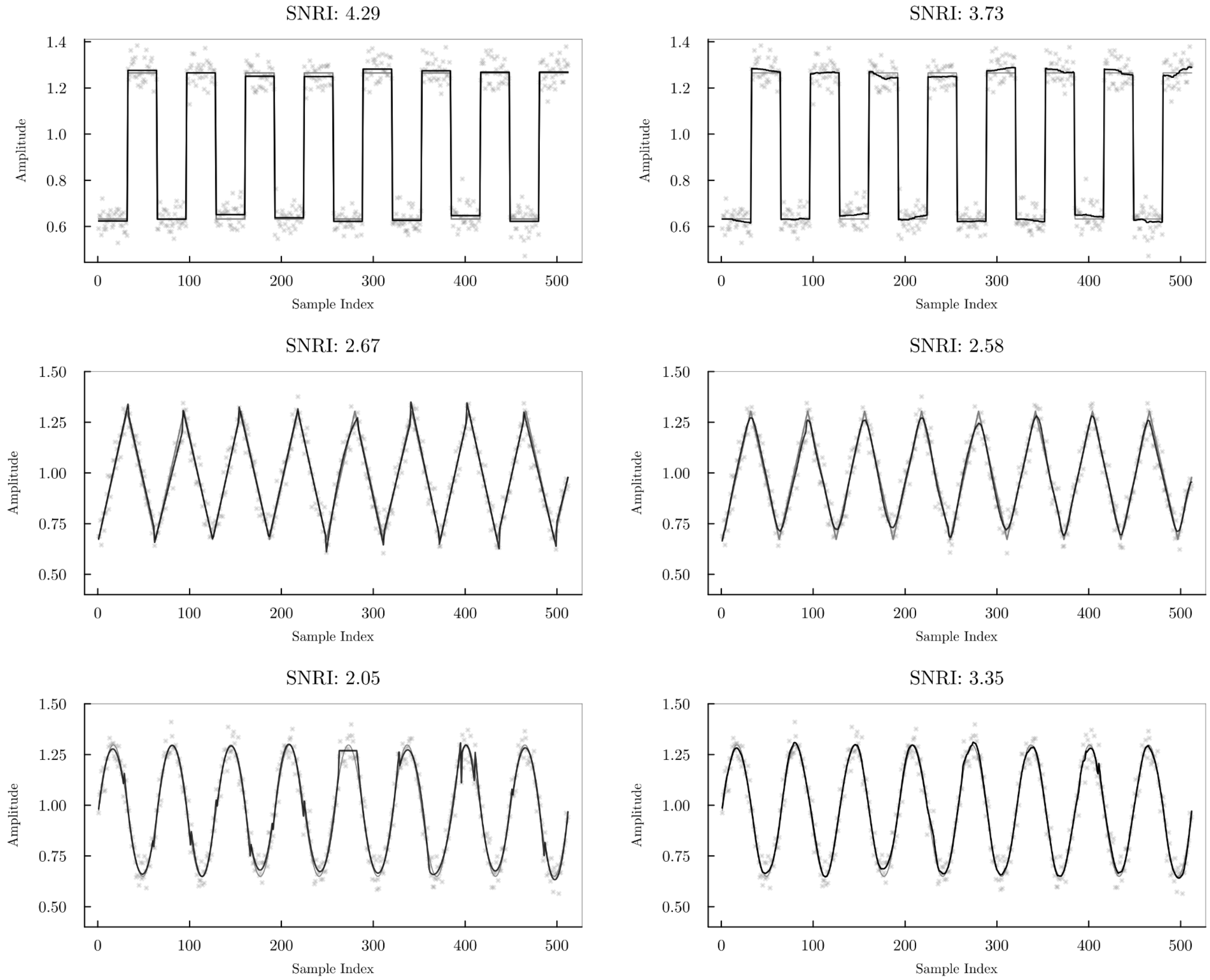

Figure 4.11 Signals: x (grey curve), noisy observations y (SNR=20, grey $\times$), and estimates (black curve). Columns: PSM estimates (left) and SW-PSM estimates ($L = L^*$ and $\Delta = \lfloor L^*/6 \rfloor$) (right). The SNRI of each estimated example is indicated ar the top of each panel.

Figure 4.12 shows the mean SNRI achieved by SW-PSM as a function of the period T for square, triangular, and sinusoidal signals at SNR = 20. Error bars represent ± 2 standard deviations computed over 200 independent noise realizations. For short periods ($T \leq 12$), both AIC-optimal and BIC-optimal SW-PSM achieve SNRI > 1, whereas PSM fails to reduce noise significantly, as shown in Figure 4.2. This behavior further highlights the benefit of SW-PSM in rapidly varying signals, where aggregation of locally optimal models reduces reconstruction error more effectively than PSM selection.

For piecewise polynomial signals, the PSM estimate (black curve in the left panel at the bottom of Figure 4.11) incurs in oversmoothing by representing some regions of x (grey curve) with overly simple local models. Examples of this are observed near indices 5, 375

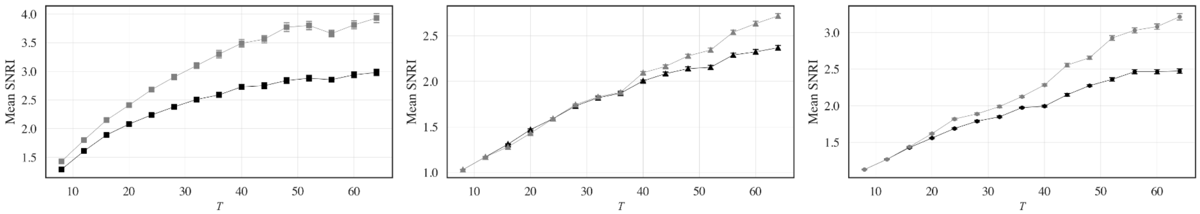

Figure 4.12 Mean SNRI (n = 200) via SW-PSM versus period T for square (left), triangular (center), and sinusoidal (right) waves at SNR = 20. Black and grey curves correspond to AIC and BIC minimization, respectively. Error bars represent ± 2 standard deviations.

and 450, where oscillations in x are represented by linear fits. Furthermore, PSM selection also incurs in discontinuity artifacts, for example, near indices 290 and 490. In contrast, AIC-optimal SW-PSM selects richer local models within individual windows, which are then aggregated into a more accurate global estimate. As observed in the right panel at the bottom of Figure 4.11, this aggregation reduces oversmoothing artifacts caused by the limitation of a single, local model while avoiding discontinuity artifacts, leading to the highest SNRI values.

Overall, for the class of synthetic signals considered here, the relative performance of AIC and BIC within the PSM and SW-PSM frameworks reflects the interplay between global regularity and local variability. These results suggest that BIC favors simpler models and is therefore advantageous for signals with large homogeneous regions. In contrast, AIC allows more complex local models to be selected, which is beneficial for signals exhibiting localized spikes or fast-varying regions. Furthermore, SW-PSM addresses key limitations of PSM, specifically discontinuity artifacts, local model constraints, and model selection ambiguity.

4.10 SW-PSM sensitivity to noise-level estimation

The SW-PSM method relies on the assumption that the noise level σ is either known or accurately estimated. To evaluate the sensitivity of the method to the noise-level estimate, we generated multiple SW-PSM estimates (AIC minimization) by defining the input noise parameter as $\hat{\sigma} = \sigma c$. We varied $c \in \mathbb{R}$ from 0.6 to 1.4 to simulate deviations in the noise-level estimation; these deviations were quantified by the relative percentage error $\epsilon = 100\frac{\hat{\sigma}-\sigma}{\sigma}$.

Figure 4.13 illustrates the impact of these deviations on the SNRI distribution, calculated across a sample of 200 piecewise-polynomial signals for each value of ϵ. For $\epsilon \in [-5\%, 5\%]$, the distributions remain nearly identical, suggesting that signal estimation quality is robust against small deviations from the exact noise level ($\epsilon = 0$).

For $-20\% < \epsilon < -5\%$, we observe a slight improvement in SNRI compared to $\epsilon = 0$. This is particularly evident when $\sigma = 0.1$ (Figure 4.13, bottom). Conversely, for $\epsilon > 5\%$, the median SNRI decreases with respect to as ϵ increases.

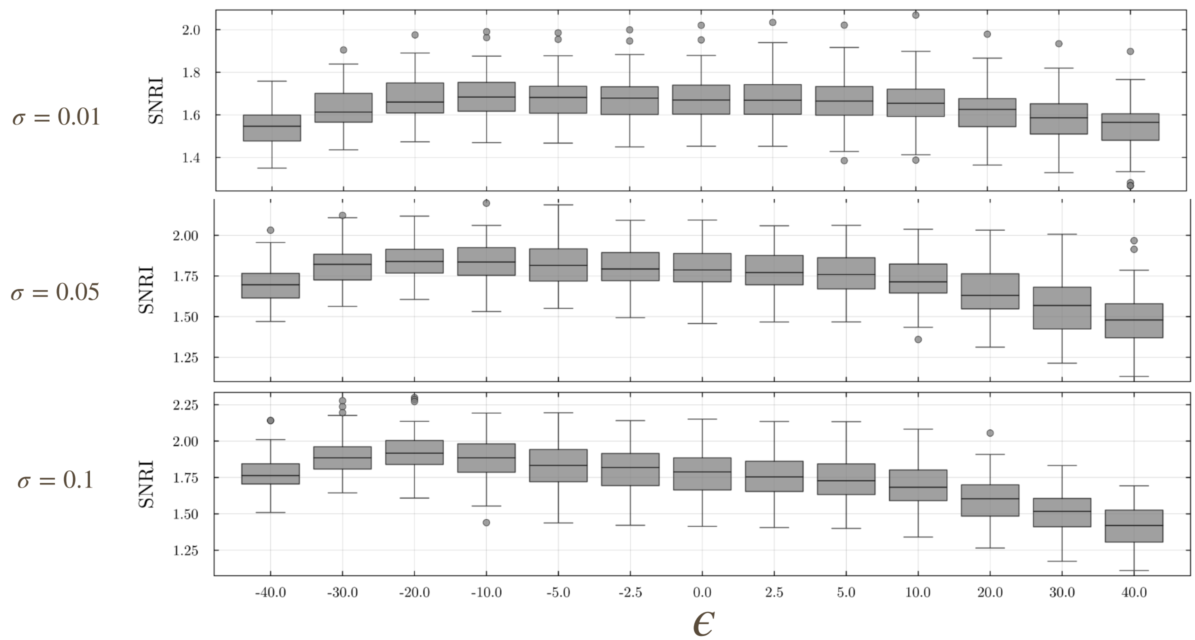

Figure 4.13 SNRI performance of SW-PSM estimates as a function of the noise-level estimation error ϵ for $\sigma = 0.01$ ($\epsilon = -20\%$, top), $\sigma = 0.05$ ($\epsilon = 0\%$, center), and $\sigma = 0.1$ ($\epsilon = +20\%$, bottom).

Figure 4.14 displays SW-PSM signal estimates for $\sigma = 0.1$ at ϵ values of -20%, 0%, and $+20\%$, corresponding to $\hat{\sigma} \in \{0.08, 0.1, 0.12\}$.

The highest SNRI was achieved with $\hat{\sigma} = 0.08$ ($\epsilon = -20\%$). As shown in the top row, this estimate (black curve) recovers the spikes near indices 100 and 320, as well as the high-frequency components around index 400, more accurately than when the exact noise level ($\epsilon = 0\%$) is used (black curve in the center of Figure 4.14). The lowest SNRI among the three examples was obtained with $\hat{\sigma} = 0.12$ ($\epsilon = +20\%$), corresponding to the black curve at the bottom of Figure 4.14. In this instance, the overestimation of σ leads to the most abrupt elimination of all sharp spikes present in the original signal (grey curve).

Note that in expressions Eq. 4.16 and 4.22 for the information criterion of a PSM, the data-fidelity term is scaled by $1/\sigma^2$. Since $0 < \sigma < 1$, underestimating σ assigns a larger relative weight to data fidelity over model complexity. Consequently, using a slightly underestimated $\hat{\sigma}$ makes the model more *conservative* by reducing the suppression of high-frequency signal features. This mitigation of over-smoothing leads to superior signal reconstruction compared to using the exact noise level, explaining the SNRI gains observed in Figure 4.13 when

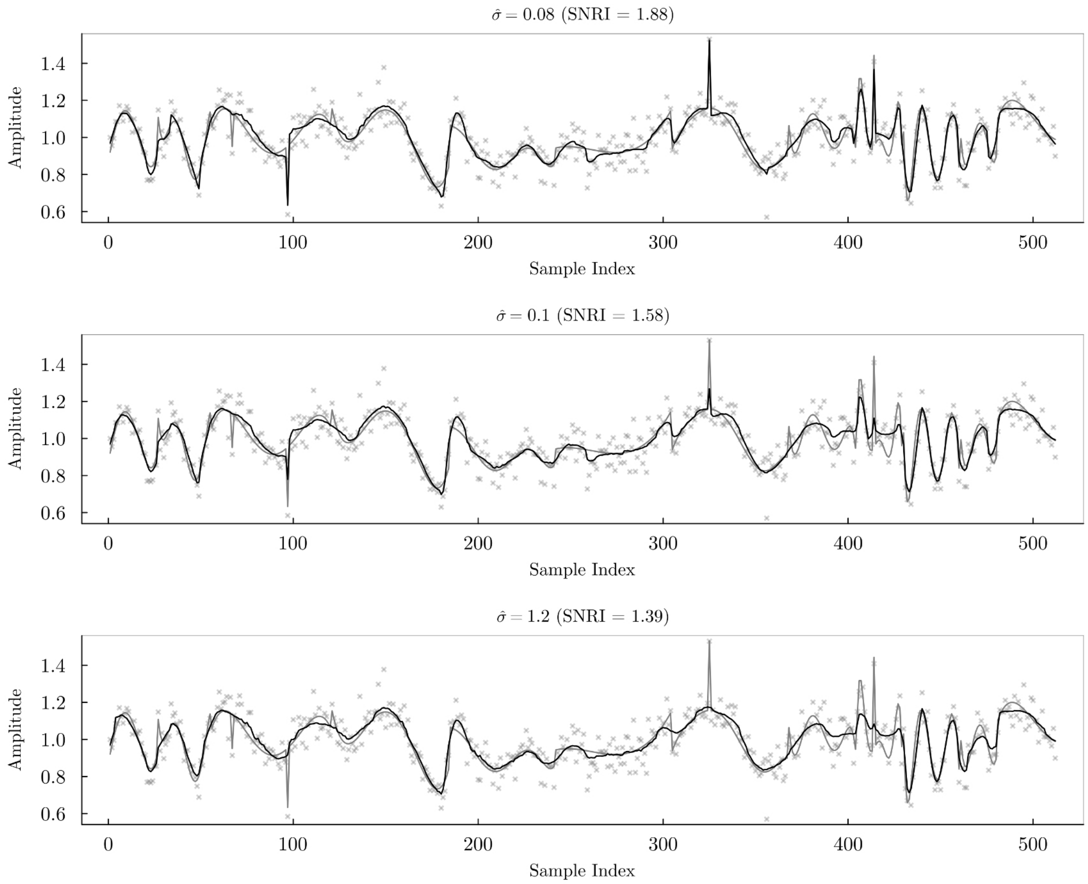

Figure 4.14 Signals: y ($\sigma = 0.1$, grey $\times$'s), x (grey curves) and SW-PSM signal estimates (black curves) obtained using $\hat{\sigma} = 0.08$ (top), $\hat{\sigma} = 0.1$ (center), and $\hat{\sigma} = 1.2$ (bottom).

$-20\% < \epsilon < -5\%$.

Conversely, overestimating σ reduces the weight of the data-fidelity term relative to the complexity penalty in the AIC and BIC expressions. This shift biases the model selection procedure toward simpler models, which leads to the over-smoothing of high-frequency components and a subsequent reduction in SNRI (Figure 4.14, bottom).

Noise-level estimation via high-order differencing

We introduce a noise-level estimation method for piecewise smooth signals that yields accurate or slightly-underestimated values of σ, suitable for the SW-PSM method. We derive this estimate by analyzing local variations in $y \in \mathbb{R}^N$. Specifically, we utilize a high-order differencing approach designed to vanish on piecewise smooth regions of x, and isolating the noise component.

For any sample index i, we denote a contiguous 5-point segment as $y_{i,i+4} = (y[i], y[i+1], y[i+2], y[i+3], y[i+4])^\top \in \mathbb{R}^5$. Following the noise model, each segment is composed of a clean signal component $x_{i,i+4} = (x[i], x[i+1], x[i+2], x[i+3], x[i+4])^\top \in \mathbb{R}^5$, and a noise component $w_{i,i+4} = (w[i], w[i+1], w[i+2], w[i+3], w[i+4])^\top \in \mathbb{R}^5$. Following the piecewise smooth assumption, for most i's, the segment x_i is well-approximated by a cubic polynomial

$$x[i+m] = a_0 + a_1 m + a_2 m^2 + a_3 m^3, \quad m \in 0, \ldots, 4, \tag{4.28}$$

where the coefficients a_k depend on i. The clean segment can be expressed as $x_{i,i+4} = \sum_{k=0}^{3} a_k P_k$, where $P_k := (0^k, 1^k, 2^k, 3^k, 4^k)^\top \in \mathbb{R}^5$ for $k \in \{0, 1, 2, 3\}$.

A unit-norm vector in $\mathbb{R}^5$ that is orthogonal to P_k, $k = 0, 1, 2, 3$ is $v := \frac{1}{\sqrt{70}}(1, -4, 6, -4, 1)$, i.e., $v^\top P_k = 0$ for all $k \in \{0, 1, 2, 3\}$. Multiplying this vector by $y_{i,i+4}$ yields $v^\top y_{i,i+4} = v^\top x_{i,i+4} + v^\top w_{i,i+4}$. Where, if (4.28) holds, $v^\top x_{i,i+4} = 0$, leaving $v^\top y_{i,i+4} = v^\top w_{i,i+4}$. Since the components of $w_{i,i+4}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ and $\|v\| = 1$, the projection $v^\top y_{i,i+4}$ follows a Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

The sequence of all such projections can be computed efficiently via the discrete convolution product $v * y$. This leads to the preliminary noise-level estimate

$$\hat{s} = \sqrt{\frac{1}{M} |v * y|^2}, \tag{4.29}$$

where M is the dimension of the resulting convolution vector.

If the signal considered is piecewise continuous, then the quantities $v^\top x_{i,i+4}$ may not vanish if the interval $[i, i+4]$ contains a singularity, e.g. a discontinuity or an angular point. Such singularities can bias the estimate (4.29).

To mitigate this bias, we restrict the application of (4.29) to locally smooth regions. We implement a procedure that scans the signal using a sliding window of length ℓ and fits a third-degree polynomial to each segment. We then evaluate the residuals of these fits using the Ljung-Box test to verify their independence. A p-value exceeding 0.05 suggests the polynomial successfully captured the local signal, leaving only noise. and because x is piecewise smooth, multiple such segments are typically found. These validated segments are collected as candidate regions. For regions of size ℓ, the final noise level estimate $\hat{\sigma}_\ell$ is calculated as

$$\hat{\sigma}_\ell = \sqrt{\frac{1}{R} \sum_{i=1}^{\nu} \hat{s}_i^2}, \tag{4.30}$$

where $R > 0 \in \mathbb{N}$ is the number of candidate regions, and $\hat{s}_i$ is the noise level estimate corresponding to the i^{th} candidate region.

The results in Figure 4.15 indicate that larger window lengths ℓ tend to introduce a positive bias ($\epsilon > 0$), particularly in the low-noise regime ($\sigma \leq 0.025$). Conversely, as σ increases, the estimates generally trend toward underestimation. Based on these distributions, the estimator using $\ell = 8$ yields errors near zero or slightly negative across the tested range of σ. By avoiding significant overestimation, this choice prevents the SW-PSM method from being biased toward over-smoothing. As discussed previously, the tendency toward slight underestimation is beneficial, as it preserves fine-scale signal features and results in higher SNRI values.

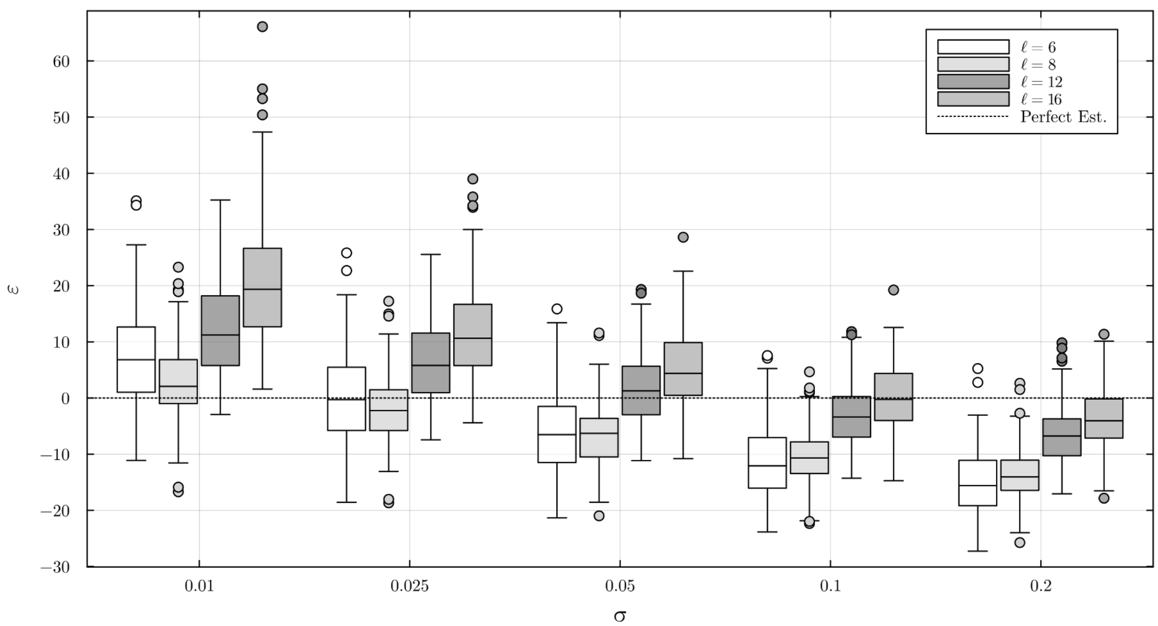

Figure 4.15 Relative estimation error ϵ obtained via 4.29 across five values of σ. For each σ, boxplots represent the distribution of ϵ for $\ell \in \{6, 8, 12, 16\}$ (300 realizations per box). The dashed line at $\epsilon = 0$ indicates the perfect estimate.

4.10.1 Statistical Analysis of the Noise-Level Estimator

The bias of $\hat{s}$ arises from two distinct phenomena: structural model mismatch and the statistical transformation from power to standard deviation. Let $\delta_i = v^\top x_{i,i+4}$ denote the residual of the clean signal. The expectation of $\hat{s}^2$ is given by

$$\mathbb{E}[\hat{s}^2] = \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}[(\delta_i + \mathbf{v}^\top \mathbf{w}_{i,i+4})^2] = \sigma^2 + \frac{1}{M} \sum_{i=1}^{M} \delta_i^2. \tag{4.31}$$

The term $\frac{1}{M}\sum \delta_i^2 \geq 0$ represents a positive structural bias. In regions where the signal is not perfectly cubic (e.g., near singularities or high-order curvatures), this term inflates the noise estimate. The candidate region selection procedure via the Ljung-Box test is specifically designed to minimize this bias. By discarding segments that do not locally conform to the third-degree polynomial model, it filters out regions where the deviation δ_i would otherwise inflate the estimate.

Because the projections z_i are computed via a sliding window convolution, they are not independent. The covariance between projections at lag k is $\mathrm{Cov}(z_i, z_{i+k}) = \sigma^2 C(k)$, where $C(k)$ is the lag-k autocorrelation of the vector v. For the chosen v, the squared autocorrelations are $C(0)^2 = 1$, $R(1)^2 = 0.64$, $C(2)^2 = 0.16$, $C(3)^2 = 64/4900$, and $C(4)^2 = 1/4900$. The variance of the sum of squared correlated Gaussian variables is [112]:

$$\mathrm{Var}(\hat{s}^2) = \frac{2\sigma^4}{M^2}\left[M \cdot C(0)^2 + 2\sum_{k=1}^{4}(M-k)C(k)^2\right]. \tag{4.32}$$

For large $M >> k$, this simplifies to

$$\mathrm{Var}(\hat{s}^2) \approx \frac{2\sigma^4}{M}\left(1 + 2\sum_{k=1}^{4} C(k)^2\right) \approx \frac{5.253\sigma^4}{M}. \tag{4.33}$$

This result shows that for a sufficiently large M, the variance of $\hat{s}^2$ scales linearly with σ^4 and is inversely proportional to the number of samples M.

4.11 Comparison of the SW-PSM method with other NR methods

We now examine the SNRI performance of SW-PSM under minimal-AIC and minimal-BIC selection, in comparison with other noise-reduction methods reported in the literature. Specifically, SW-PSM is evaluated against the following approaches, which are implemented using standard, publicly available software libraries. These methods were selected because they are both widely used and publicly available: *Total Variation* (TV) regularization [30], which defines the signal estimate by minimizing the L^1 norm of the gradient of $\hat{x}$ to preserve sharp discontinuities while suppressing noise, and which is implemented using the `proxtv` Python package [113]; *Wiener filtering* (WF) [100], that utilizes a frequency-dependent filter based on the signal-to-noise power spectral density ratio to achieve a minimum mean square error estimate, implemented using the `scipy` [114] library in Python; *Wavelet thresholding* (WT), employing Daubechies wavelets with universal thresholding [45], implemented using the `PyWavelets` library [115]; and *Variational Mode Decomposition* (VMD) [79], which estimates x by decomposing y into data-adaptive additive components. This method solves a constrained variational problem to identify the reconstruction components and their respective center frequencies and is implemented using the `vmdpy` Python package [116].

Parameters for all non-SW-PSM methods were selected by minimizing the mean squared error against the exact signal x over a predefined grid. (Supplementary Material S2). Although such oracle optimization not feasible in practice, it establishes a performance upper bound, ensuring that results reflect fundamental algorithmic capabilities rather than suboptimal tuning. Following standard practice in signal denoising and model selection [43, 45], this benchmark provides a near-optimal reference for evaluating practical, data-driven approaches.

In contrast, the SW-PSM parameters (L^* and Δ) were estimated directly from the noisy signal y. The comparison therefore evaluates how effectively the proposed SW-PSM performs relative to classical denoising methods whose parameters are selected in a near-optimal (oracle) manner.

Table 4.2 Mean SNRI for all methods across SNR levels and signal classes. Errors correspond to two standard deviations of each mean. For each SNR-signal pair, the two highest SNRI values are highlighted in bold.

Signal	Method	SNR=5	SNR=10	SNR=20	SNR=40	SNR=100
PP	SW-PSM_AIC	2.02 ± 0.02	$\mathbf{1.79 \pm 0.01}$	$\mathbf{1.79 \pm 0.01}$	$\mathbf{1.76 \pm 0.01}$	$\mathbf{1.65 \pm 0.01}$
	SW-PSM_BIC	1.81 ± 0.01	1.50 ± 0.01	$\mathbf{1.57 \pm 0.02}$	$\mathbf{1.63 \pm 0.01}$	$\mathbf{1.59 \pm 0.01}$
	TV	2.09 ± 0.01	1.71 ± 0.01	1.47 ± 0.01	1.28 ± 0.01	1.12 ± 0.01
	VMD	$\mathbf{2.13 \pm 0.02}$	1.47 ± 0.01	1.15 ± 0.01	1.04 ± 0.02	0.99 ± 0.01
	WT	1.74 ± 0.01	1.48 ± 0.01	1.33 ± 0.01	1.24 ± 0.01	1.18 ± 0.01
	WF	$\mathbf{2.10 \pm 0.01}$	$\mathbf{1.81 \pm 0.01}$	1.56 ± 0.01	1.34 ± 0.01	0.82 ± 0.01
Doppler	SW-PSM_AIC	$\mathbf{2.13 \pm 0.01}$	$\mathbf{1.98 \pm 0.01}$	$\mathbf{1.87 \pm 0.01}$	$\mathbf{1.85 \pm 0.01}$	$\mathbf{1.72 \pm 0.01}$
	SW-PSM_BIC	$\mathbf{2.28 \pm 0.01}$	$\mathbf{2.13 \pm 0.01}$	$\mathbf{1.99 \pm 0.01}$	$\mathbf{1.97 \pm 0.01}$	$\mathbf{1.80 \pm 0.01}$
	TV	1.92 ± 0.01	1.55 ± 0.01	1.36 ± 0.01	1.23 ± 0.01	1.11 ± 0.01
	VMD	1.98 ± 0.01	1.51 ± 0.01	1.25 ± 0.00	1.10 ± 0.01	1.02 ± 0.10
	WT	1.88 ± 0.01	1.66 ± 0.01	1.54 ± 0.01	1.45 ± 0.01	1.36 ± 0.01
	WF	1.95 ± 0.01	1.68 ± 0.01	1.45 ± 0.01	1.24 ± 0.01	0.72 ± 0.01
PR	SW-PSM_AIC	2.64 ± 0.01	$\mathbf{2.49 \pm 0.02}$	$\mathbf{2.45 \pm 0.01}$	$\mathbf{2.35 \pm 0.01}$	$\mathbf{2.16 \pm 0.01}$
	SW-PSM_BIC	2.94 ± 0.01	$\mathbf{2.71 \pm 0.02}$	$\mathbf{2.81 \pm 0.01}$	$\mathbf{2.66 \pm 0.01}$	$\mathbf{2.38 \pm 0.02}$
	TV	$\mathbf{2.98 \pm 0.02}$	2.41 ± 0.02	2.00 ± 0.01	1.69 ± 0.01	1.41 ± 0.01
	VMD	$\mathbf{3.09 \pm 0.01}$	2.29 ± 0.01	1.72 ± 0.01	1.33 ± 0.01	1.07 ± 0.01
	WT	2.19 ± 0.01	1.82 ± 0.01	1.57 ± 0.01	1.43 ± 0.01	1.33 ± 0.01
	WF	2.63 ± 0.01	2.26 ± 0.01	2.05 ± 0.01	1.85 ± 0.01	1.56 ± 0.01

We evaluated all methods on the Piece-Regular (PR) and Doppler test signals [43] (see Figure 4.16 for an SW-PSM example), as well as on our reference Piecewise Polynomial (PP) signals. The PR and Doppler signals were chosen as standard NR benchmarks. The PR signal combines sharp transients with smooth regions, testing edge preservation under nonstationary conditions, while the Doppler signal features spatially varying frequency content, assessing adaptability to local smoothness. For each signal and SNR value in $5, 10, 20, 40, 100$, 500 independent realizations were generated. Each realization was processed with all NR methods, and the resulting SNRI was computed.

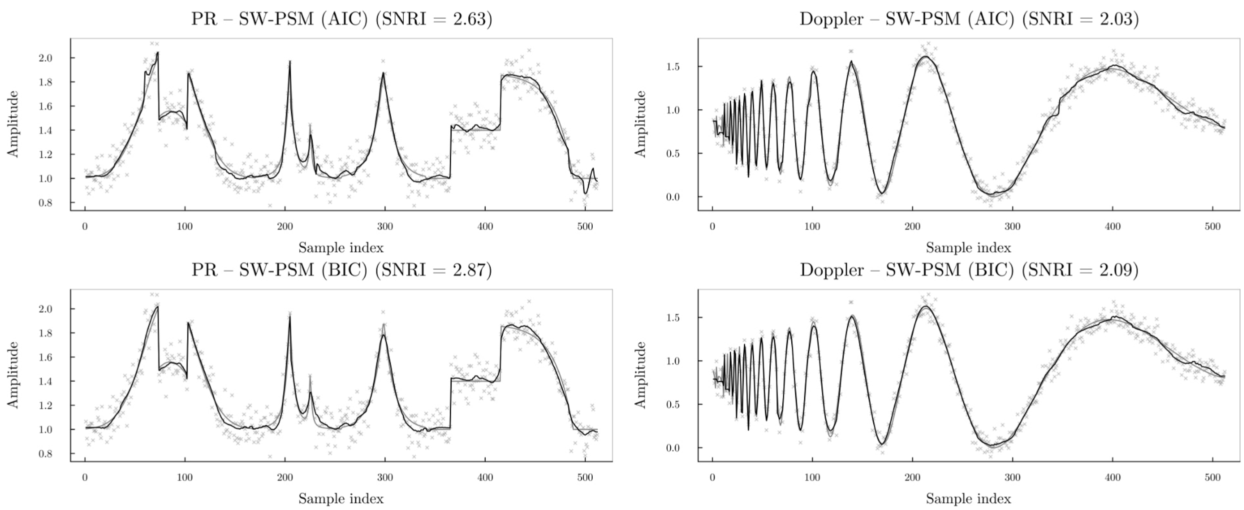

Figure 4.16 Signals: x (grey curve), y with SNR= 10 (grey ×'s), and signal estimate $\hat{x}$ (black curve) obtained by our proposed NR methods for the 'Piece-Regular' (left), and 'Doppler' (right) signals. The SNRI for each example is displayed at the top of each panel.

The average SNRI across realizations is reported in Table 4.2. Overall, SW-PSM consistently outperforms near-ideal implementations of competing methods in low-to-moderate noise regimes (SNR $\in [10, 100]$), while maintaining comparable performance at SNR $= 5$. For the Doppler and PR signals, the BIC-optimal SW-PSM achieves higher SNRI than its AIC-optimal counterpart. In contrast, for PP signals, which contain more discontinuities and rapid transitions, the AIC-optimal variant performs better. These results align with previous observations on the relative behavior of AIC- and BIC-optimal PSM.

Typical estimates for the PR and Doppler signals are presented in Figure 4.16. The left panels show that while AIC and BIC versions yield comparable results, SW-PSM-AIC can introduce minor oscillatory artifacts, notably around indices 60, 225, and 500. In the right panels, both variants fail to track the highest-frequency components accurately.

SW-PSM achieves particularly strong performance on PP signals (Table 4.2). For example, at SNR $= 100$, most competing methods provide little noise reduction (SNRI ≈ 1), whereas the AIC-optimal SW-PSM attains a mean SNRI of 1.65, outperforming oracle-optimized WT and VMD by factors of approximately 1.47 and 1.67, respectively. Substantial gains are also observed at SNR = 20 and SNR = 40.

Figure 4.17 illustrates the estimates (black curves) obtained for a noisy PP signal y at SNR = 20 ($\sigma = 0.05$). Among the competing methods, AIC-optimal SW-PSM provides the most accurate reconstruction of x, effectively preserving isolated spikes and smooth regions. In contrast, the other methods fail to adequately suppress residual noise, notably around indices 250 and 480.

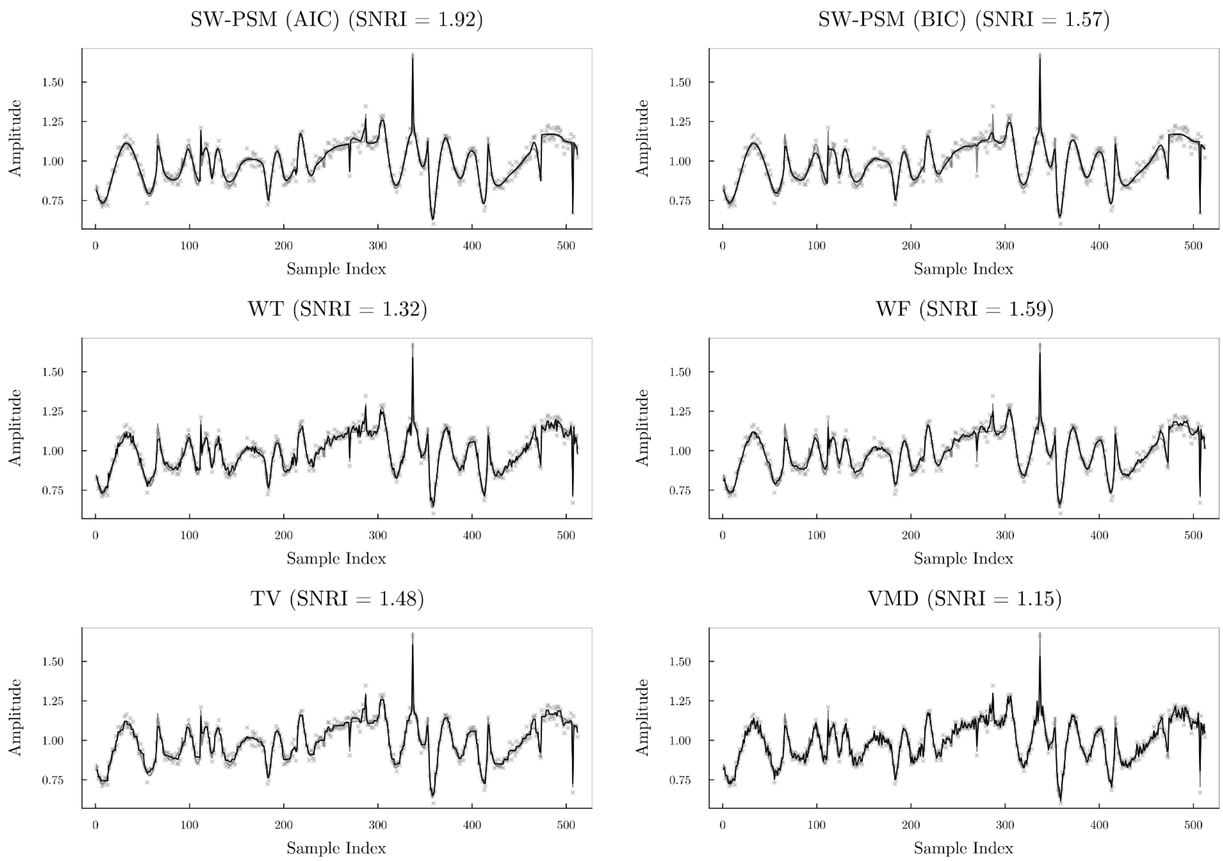

Figure 4.17 Piecewise polynomial signal x (grey curve), y with SNR= 20 (grey $\times$'s), and $\hat{x}$ (black curve) obtained by each of the NR methods.

In summary, SW-PSM consistently delivers superior or competitive SNRI across noise levels and signal types, despite not relying on oracle tuning. For PP signals at very low SNR, oracle-optimized VMD and WF may achieve comparable or slightly better performance; however, SW-PSM dominates in moderate to high SNR regimes. By aggregating information-criterion-minimal local models to represent the underlying signal, SW-PSM provides a robust and flexible alternative to existing NR methods.

Finally, we present an application of the SW-PSM noise reduction method for a real ECG signal obtained from the MIT-BIH arrhythmia database [117]. We estimated the noise level $\hat{\sigma} = 0.011$ using Eq. (4.30) with smooth regions of size $\ell = 8$, and applied the SW-PSM method with BIC minimization to the raw signal (grey curve in Figure 4.18). The SW-PSM estimate (black curve) produced a smooth signal that preserves both the sharp spikes and the smooth waveform components. As shown in the center panel, our approach avoids removing clinical features while eliminating high-frequency artifacts. The standardized residuals (right panel) were compared to a Gaussian distribution $\mathcal{N} \sim (0, 1)$ using a Kolmogorov-Smirnov statistical test. The test passed at a 95% significance level ($p > 0.05$), suggesting that the discarded component is statistically consistent with white Gaussian noise.

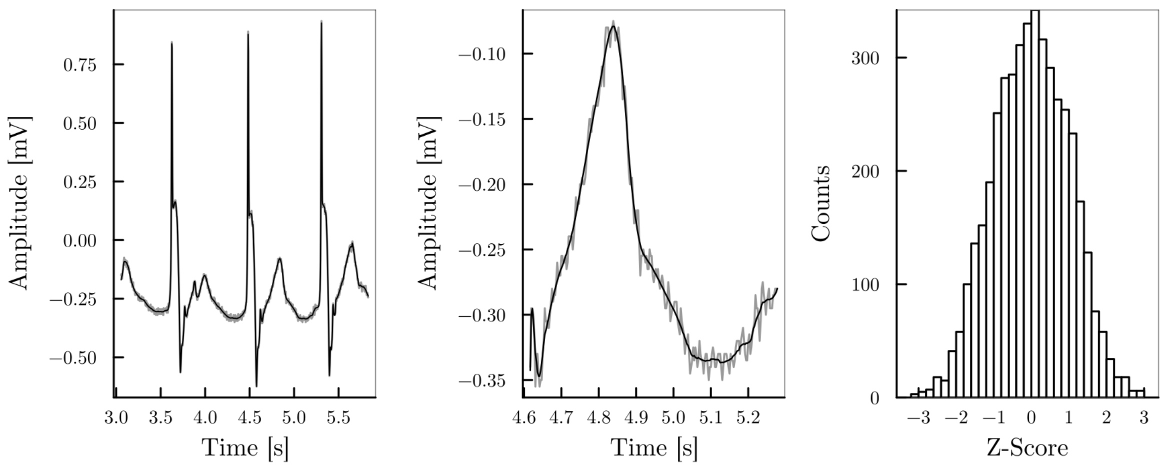

Figure 4.18 **Left:** Raw signal y (grey, $\hat{\sigma} = 0.011$), and SW-PSM (BIC) estimate $\hat{x}$ (black). **Center:** Detail of y and $\hat{x}$. **Right:** Histogram of standardized residuals $y - \hat{x}$.

4.12 Concluding remarks

In this work, we introduced the Piecewise Smooth Model (PSM) and the Sliding-Window Piecewise Smooth Model (SW-PSM) as novel model-based strategies for reducing Gaussian noise in piecewise smooth signals.

The PSM method provides an adaptive framework that combines fast curve-fitting with shortest-path optimization to select an optimal combination of constant, linear, and parabolic segments using an information criterion. BIC-minimal PSM is particularly effective for signals dominated by large smooth segments, whereas AIC-minimal PSM better captures fast-varying or fine-scale features. PSM's main advantages include fully data-driven parameter estimation and precise identification of smooth-segment supports. Its limitations arise from reliance on polynomial segments, which restrict modeling of non-polynomial components, and from the non-overlapping segment constraint, which can introduce discontinuity artifacts in the signal estimate.

The Sliding-Window Piecewise-Smooth Model (SW-PSM) extends PSM by performing local PSM estimation within overlapping windows and aggregating these estimates into a global reconstruction. This enables low-order polynomial fits to capture higher-order or non-polynomial features and mitigates boundary discontinuities. Data-driven selection of window size and step size provides a principled balance between local adaptability, accuracy, and computational efficiency. Experiments show that SW-PSM consistently outperforms PSM on triangular and sinusoidal waves, as well as on piecewise polynomial signals of degree greater than two, while providing comparable results on square waves. Moreover, SW-PSM surpasses near-optimal versions of established noise-reduction methods across multiple noise levels, particularly at low to medium noise. Computationally, it replaces a single large optimization with multiple smaller ones, achieving at least a twofold speed-up compared to standard PSM.

Overall, SW-PSM offers a robust, interpretable, and flexible framework that outperforms alternative methods in capturing both smooth and rapidly varying signal components. Potential challenges for our method should be acknowledged. The shortest-path optimization strategy assumes that the information criterion associated with a PSM can be decomposed into additive contributions corresponding to each smooth segment. We show that this assumption is valid for AIC and BIC (Equations (4.16) and (4.22)). Nonetheless, for information criteria that do not verify this assumption, e.g. in the case of correlated noise, a different optimization strategy should be employed. More generally, the additivity assumption is closely tied to the use of non-overlapping segment supports in the PSM formulation. Allowing overlapping supports within a single global PSM would break the independence between consecutive segment fits and invalidate the derived information criterion expressions.

Several directions remain for future research. Developing optimization strategies that jointly accommodate both non-overlapping and overlapping supports within a unified global model, as well as allowing adaptive (non-constant) window sizes, could further improve modeling flexibility. The PSM framework also shows promise for piecewise regression problems. Extending it to two-dimensional signals would enable applications in image denoising, enhancement, and compression.

Declaration of Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgments: A. Saucier acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grants RGPIN-115965.

4.13 Supplementary Material for Adaptive Noise Reduction for 1-Dimensional Signals via Multiscale Piecewise Smooth Modeling

4.13.1 Curve-fitting on covering sets via cumulants

The PSM method uses as input the results of all relevant fits on the support S using constant, linear and quadratic models. For a support size L, the set of L-sized intervals is $I_L := \{[i, i+L-1],\ i = 1,\, 2,\, ...,\, N-(L-1)\}$. Fits are performed for $L \in [2,\, L_{\max}]$, where $2 \leq L_{\max} \leq N$. Using $L_{\max} = N$ is always possible but it can lead to unnecessary computations unless the signal is globally smooth, e.g. a straight line or a parabola. In practice, the length $L_{\max}$ is the maximal interval size for which an acceptable fit to a straight line is likely, so $L_{\max}$ is essentially determined by the signal smoothness. It follows that the set of intervals considered in practice is $I := \cup_{k=1}^{L_{\max}} I_k$.

Fitting a new model independently for each window generates significant computation redundancies that can result in excessive computing time. Considering that the maximal number of fits for a given regression model is $N(N+1)/2$ for a signal $y \in \mathbb{R}^N$, i.e. increases quadratically as N increases, it is preferable to take advantage of information obtained from previous fits to perform new fits. In this section, we introduce fast algorithms to compute these regressions without losing accuracy.

Performing linear regressions requires the evaluation of several summations of the form $S_u(i,j) := \sum_{n=i}^{j} u[n]$, where $u \in \mathbb{R}^N$ is a discrete signal. These summations will be evaluated efficiently using

$$S_u(i,j) := C_u[j] - C_u[i-1], \tag{4.34}$$

where the *cumulant* $C_u[j]$ associated with the signal u is defined by

$$C_u[j] := \sum_{n=0}^{j} u[n],\ j = 0,\, 1,\, ...,\, N-1. \tag{4.35}$$

A fast computation of C_u can be done with the iteration

$$\begin{cases} C_u[j] & := & C_u[j-1] + u[j],\ j = 1,\, 2, ...,\, N-1, \\ C_u[0] & = & u[0]. \end{cases} \tag{4.36}$$

To perform all the fits required by PSM, the three cumulants C_y, C_{ny} and C_{n^2y} are needed. Since fits must be performed in many intervals $[i, j]$, we will use the following notations for the arrays restricted to this range:

$$\begin{aligned} x_{i,j} &:= (x[i],\, x[i+1],\, ...,\, x[j])^T \in \mathbb{R}^{j-i+1}, \\ y_{i,j} &:= (y[i],\, y[i+1],\, ...,\, y[j])^T \in \mathbb{R}^{j-i+1}, \\ w_{i,j} &:= (w[i],\, w[i+1],\, ...,\, w[j])^T \in \mathbb{R}^{j-i+1}. \end{aligned} \tag{4.37}$$

We will also define $N_{i,j} := j - i + 1$ for the cardinality of the set $\{i,\, i+1,\, ...,\, j\}$.

4.13.2 Fast computation of regressions for the constant model

The constant model of y in the range $[i, j]$ is

$$y[i+m] = a_{i,j} + w[i+m], \;\; m \in [0, j-i],$$

where $a_{i,j} \in \mathbb{R}$ is a constant and w is the noise. Since the noise variance is constant, the maximum likelihood estimate of the parameter $a_{i,j}$ coincides with the Least Squares Estimate (LSE), hence curve-fitting to the constant model consists in minimizing $F(a_{i,j}) := \sum_{m=0}^{j-1}(y[i+m] - a_{i,j})^2$ with respect to $a_{i,j}$. The first-order optimality condition $F'(a_{i,j}) = 0$ yields the parameter estimate $\hat{a}_{i,j} = (\sum_{n=i}^{j} y[n])/(j-i+1)$, which is the arithmetic average of the components of $y_{i,j}$. The estimate $\hat{x}_{i,j}$ of $x_{i,j}$ is the constant signal $\hat{x}_{i,j} = \hat{a}_{i,j}\, u_0$, where $u_0 := (1, 1, ..., 1)^T \in \mathbb{R}^{N_{i,j}}$. The fast computation of $\hat{a}_{i,j}$ is performed via the cumulant of y with

$$\hat{a}_{i,j} = S_y(i, j)/(i - j + 1), \tag{4.38}$$

where S_y is defined and computed with (4.34), (4.35) and (4.36) using $u = y$. The variance of the components of $\hat{x}_{i,j}$ is constant and given by

$$\mathrm{Var}\,(\hat{x}_{i,j}[n]) = \frac{\sigma^2}{j-i+1}, \;\; n \in [i, j]. \tag{4.39}$$

4.13.3 Fast computation of regressions for the linear and quadratic models

The linear model of y in the range $[i, j]$ is

$$y[i+m] = a_{i,j} + b_{i,j}\, m + w[i+m], \; m \in [0, j-i], \tag{4.40}$$

where $a_{i,j}$ and $b_{i,j}$ are real parameters and w is the noise signal. Similarly, The quadratic model of y in the range $[i, j]$ is

$$y[i+m] = a_{i,j} + b_{i,j}\, m + c_{i,j}\, m^2 + w[i+m], \; m \in [0, j-i], \tag{4.41}$$

where $a_{i,j}$, $b_{i,j}$ and $c_{i,j}$ are real parameters and w is the noise signal. The models (4.40) and (4.41) take the vector form

$$\begin{aligned} y_{i,j} &= a_{i,j}\; u_0 + b_{i,j}\; u_1 + w_{i,j}, \\ y_{i,j} &= a_{i,j}\; u_0 + b_{i,j}\; u_1 + c_{i,j}\; u_2 + w_{i,j}, \end{aligned} \tag{4.42}$$

respectively, where

$$\begin{aligned} u_0 &:= (1, 1, 1, ..., 1)^T \in \mathbb{R}^{N_{i,j}}, \\ u_1 &:= (0, 1, 2, ..., N_{i,j} - 1)^T \in \mathbb{R}^{N_{i,j}}, \\ u_2 &:= (0^2, 1^2, 2^2, ..., (N_{i,j} - 1)^2)^T \in \mathbb{R}^{N_{i,j}}. \end{aligned} \tag{4.43}$$

Both the models in (4.42) can be written in the matrix form

$$y_{i,j} = H\; \theta_{i,j} + w_{i,j}, \tag{4.44}$$

where the form of H and $\theta_{i,j}$ depends on the model. For the linear model, they are given by

$$\begin{aligned} \theta_{i,j} &:= (a_{i,j},\, b_{i,j})^T \in \mathbb{R}^2, \\ H &:= (u_0,\, u_1) \in \mathbb{R}^{N_{i,j}\times 2}. \end{aligned} \tag{4.45}$$

For the quadratic model, they are given by

$$\begin{aligned} \theta_{i,j} &:= (a_{i,j},\, b_{i,j},\, c_{i,j})^T \in \mathbb{R}^3, \\ H &:= (u_0,\, u_1,\, u_2) \in \mathbb{R}^{N_{i,j}\times 3}. \end{aligned} \tag{4.46}$$

The LSE estimate of the parameter vector $\theta_{i,j}$ is obtained by minimizing the function $F(\theta_{i,j}) := \frac{1}{2}||y_{i,j} - H\theta_{i,j}||^2$ with respect to $\theta_{i,j}$, where $||...||$ denotes the euclidian norm. The first-order optimality condition $\nabla F(\theta_{i,j}) = 0$ takes the form of the linear system of equations

$$(H^T H)\; \theta_{i,j} = H^T y_{i,j}, \tag{4.47}$$

which is known as the normal equations of the least squares problem. The square symmetric matrix $H^T H$ has dimensions $d \times d$, where $d = 2$ for the linear model and $d = 3$ for the quadratic model. Denoting the solution of (4.47) by $\hat{\theta}_{i,j}$, the estimate $\hat{x}_{i,j}$ of $x_{i,j}$ is

$$\hat{x}_{i,j} := H\; \hat{\theta}_{i,j} = H(H^T H)^{-1} H^T y_{i,j}. \tag{4.48}$$

The variance of the components of $\hat{x}_{i,j}$ can be obtained as follows. Using the covariance property $\mathrm{cov}(A\, y_{i,j}, A\, y_{i,j}) = A\, \mathrm{cov}(y_{i,j},\, y_{i,j})\; A^T$ with $A := H(H^T H)^{-1} H^T$ and $\mathrm{cov}(y_{i,j}, y_{i,j}) =$

$\sigma^2\ I$, where $I \in \mathbb{R}^{d\times d}$ is the identity matrix, we obtain successively

$$\begin{aligned}
\operatorname{cov}(\hat{x}_{i,j},\, \hat{x}_{i,j}) &= \operatorname{cov}(A\, y_{i,j},\, A\, y_{i,j}), \\
&= \sigma^2 A\, A^T, \\
&= \sigma^2\ H\ (H^T H)^{-1}\ H^T.
\end{aligned} \tag{4.49}$$

Equation (4.49) is used to compute the variances of the components of $\hat{x}_{i,j}$, which are located on the diagonal of the matrix $\operatorname{cov}(\hat{x}_{i,j}, \hat{x}_{i,j})$. We use these variances to aggregate multiple point-estimates via inverse variance weighting.

The evaluation of $\hat{x}_{i,j}$ and $\operatorname{cov}(\hat{x}_{i,j}, \hat{x}_{i,j})$ with (4.48) and (4.49) requires the computation of the square matrix $H^T H$ and the vector $H^T y_{i,j}$. For the linear model,

$$H^T H = \begin{pmatrix} u_0^T u_0 & u_1^T u_0 \\ u_0^T u_1 & u_1^T u_1 \end{pmatrix} \text{ and } H^T y_{i,j} = \begin{pmatrix} u_0^T y_{i,j} \\ u_1^T y_{i,j} \end{pmatrix}.$$

For the quadratic model,

$$H^T H = \begin{pmatrix} u_0^T u_0 & u_1^T u_0 & u_2^T u_0 \\ u_0^T u_1 & u_1^T u_1 & u_2^T u_1 \\ u_0^T u_2 & u_1^T u_2 & u_2^T u_2 \end{pmatrix} \text{ and } H^T y_{i,j} = \begin{pmatrix} u_0^T y_{i,j} \\ u_1^T y_{i,j} \\ u_2^T y_{i,j} \end{pmatrix}.$$

The elements of the matrix $H^T H$ depend only on $N_{i,j}$ and can be evaluated analytically as follows:

$$\begin{aligned}
u_0^T u_0 &:= \textstyle\sum_{m=0}^{N_{i,j}-1} 1 &&= N_{i,j}, \\
u_0^T u_1 &:= \textstyle\sum_{m=0}^{N_{i,j}-1} m &&= \tfrac{1}{2} N_{i,j}\ (N_{i,j}-1), \\
u_0^T u_2 &:= \textstyle\sum_{m=0}^{N_{i,j}-1} m^2 &&= \tfrac{1}{6}(N_{i,j}-1)\ N_{i,j}\ (2N_{i,j}-1), \\
u_1^T u_2 &:= \textstyle\sum_{m=0}^{N_{i,j}-1} m^3 &&= \tfrac{1}{4}(N_{i,j}-1)^2\ N_{i,j}^2, \\
u_2^T u_2 &:= \textstyle\sum_{m=0}^{N_{i,j}-1} m^4 &&= \tfrac{1}{30}(N_{i,j}-1)\ N_{i,j}\ (2N_{i,j}-1)\left(3N_{i,j}^2 - 3N_{i,j} - 1\right), \\
u_1^T u_1 &= \quad u_0^T u_2.
\end{aligned} \tag{4.50}$$

The evaluation of the quantities $u_k^T y$ for $k \in \{0,1,2\}^2$ can be done efficiently using the three cumulants C_y, $C_{n\,y}$ and $C_{n^2\,y}$, as follows:

$$u_0^T y_{i,j} := \sum_{m=0}^{N_{i,j}-1} y[i+m] = \sum_{n=i}^{j} y[n] = S_y(i,j); \tag{4.51}$$

$$\begin{aligned}
u_1^T y_{i,j} &:= \textstyle\sum_{m=0}^{N_{i,j}-1} m\, y[i+m], \\
&= \textstyle\sum_{m=0}^{N_{i,j}-1} ((i+m)-i)\, y[i+m], \\
&= \textstyle\sum_{m=0}^{N_{i,j}-1} (i+m)\, y[i+m] - i\ \textstyle\sum_{m=0}^{N_{i,j}-1} y[i+m], \\
&= S_{ny}(i,j) - i\ S_y(i,j);
\end{aligned} \tag{4.52}$$

$$\begin{aligned}
u_2^T y_{i,j} &:= \textstyle\sum_{m=0}^{N_{i,j}-1} m^2\, y[i+m], \\
&= \textstyle\sum_{m=0}^{N_{i,j}-1} (i+m-i)^2\, y[i+m], \\
&= \textstyle\sum_{m=0}^{N_{i,j}-1} ((i+m)^2 + i^2 - 2i(i+m))\ y[i+m], \\
&= S_{n^2y}(i,j) + i^2\ S_y(i,j) - 2\ i\ S_{ny}(i,j).
\end{aligned} \tag{4.53}$$

4.13.4 Oracle Optimization for Competing Noise Reduction Methods

To provide an informative point of reference, we evaluate competing noise reduction (NR) methods under an oracle optimization strategy. For each method, parameters are selected to minimize the mean squared error (MSE) with respect to the ground-truth signal.

Wavelet Thresholding. NR via wavelet thresholding is implemented using the discrete wavelet transform with periodic boundary conditions. A fixed Daubechies-4 wavelet is employed, providing a good trade-off between computational efficiency and localization of rapid variations. Soft-thresholding is applied to all wavelet coefficients. The threshold parameter t is selected via an exhaustive search over 600 uniformly spaced values in the interval $[0, 3\sigma]$, where σ denotes the noise standard deviation. For each threshold, the signal is reconstructed and its MSE is computed. The threshold yielding the minimum MSE is retained.

Wiener Filtering. An oracle search is performed over integer window lengths $w \in \{2, \ldots, 72\}$.

Total Variation Filter. Total variation (TV) filtering is implemented using a proximal formulation. The regularization parameter λ is selected by exhaustive search over a logarithmically spaced grid of 100 values, $\lambda \in \{10^{-10}, \ldots, 10^0\}$. This logarithmic distribution is adopted because the reconstruction is characterized by scale-dependent properties, where the sensitivity to edge preservation and noise removal is highest at lower parameter values [118]. This approach ensures high resolution in the high-sensitivity regions of the parameter space where λ is small. For each candidate λ, the signal estimate is computed and its MSE with respect to x, is calculated. The estimate achieving the lowest MSE is retained.

Variational Mode Decomposition (VMD). For VMD-based NR, oracle optimization is conducted over multiple parameters. The number of modes is selected from $K \in \{2, 3, \ldots, 10\}$, and the quadratic penalty parameter is selected from a logarithmically spaced grid of 40 values ranging from $\alpha \in \{10^1, \ldots, 10^4\}$. A logarithmic scale is used because the inverse relationship between α and the bandwidth of the resulting modes [79], which makes the decomposition highly sensitive to small numerical changes in the lower range of α, and less so on larger values. Furthermore, the logarithmic scaling circumvents the computational complexity of a fine-grained linear search while maintaining high precision in the high-sensitivity region. For each parameter pair (K, α), the signal is decomposed into K intrinsic mode functions. The modes are ordered from high to low center frequency, and NR is achieved by reconstructing the signal after discarding the first n high-frequency modes, with $n \in \{0, 1, 2\}$, $n < K$. The combination of (K, α, n) minimizing the Mean Squared Error (MSE) is selected.

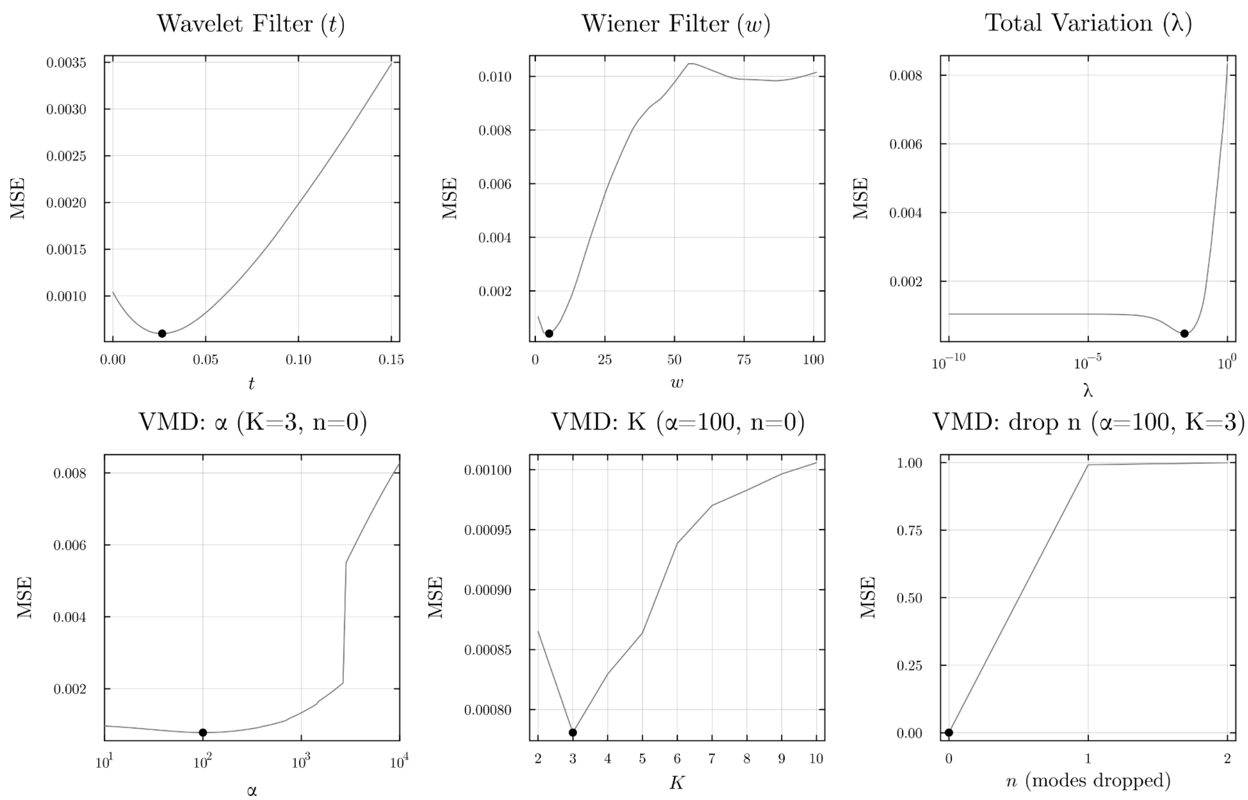

Figure 4.19 MSE profiles for oracle optimization across different NR methods. The plots validate that the search ranges for the wavelet threshold (t), Wiener window (w), TV regularization (λ), and VMD parameters (α, K, n) are correctly centered to capture the global minimum reconstruction error. Optimal values are highlighted with black dots.

Figure 4.19 provides an example of the Mean Squared Error (MSE) profiles between the estimated and pure piece-polynomial signals for each noise reduction (NR) method. The black dots denote the optimal parameters identified through the grid search oracle optimization methods described previously. These plots confirm that the defined search ranges are well-posed, as they successfully capture the global minimum reconstruction error for each model.

Similar trends were observed across different scenarios, providing confidence in the sensitivity analysis and the chosen grids in this study.

CHAPTER 5 ARTICLE 2: ADAPTIVE IMAGE NOISE REDUCTION VIA AGGREGATED ONE-DIMENSIONAL SMOOTH MODELING

Submitted on July 10, 2025 to "Digital Signal Processing". Authors: Noel Plascencia-Diaz and Antoine Saucier.

Introduction to Article 2

The Smoothness and Self-Similarity Fusion (S3F) model is a training-free framework for image denoising that integrates piecewise-smooth and self-similarity priors in a principled and computationally efficient manner. The results presented here strongly delineate multiple interesting research avenues. In particular, we provide evidence that S3F outperforms both the two-dimensional extension of the Sliding-Window Piecewise-Smooth Model (SW-PSM) and BM3D (which remains the state-of-the-art among non-training-based, data-driven image noise reduction methods), while requiring approximately half the execution time of the latter. The scope and depth of these findings suggest that at least two independent research articles may stem from this work: one focused on the piecewise-smooth modeling framework and its 2D extension, and one centered on the self-similarity component and its integration with SW-PSM using an improved model weighting scheme. We present this manuscript as a comprehensive record of the methodology, experiments, and conclusions, noting that it has not been submitted for publication.

5.1 Abstract

We propose the Smoothness and Self-Similarity Fusion (S3F) model, a novel framework for additive white Gaussian noise reduction in images. S3F explicitly integrates the complementary strengths of piecewise-smooth and self-similarity priors, adaptively applying the appropriate noise reduction (NR) strategy based on the local textural context. The framework utilizes the Sliding-Window Piecewise-Smooth Model (SW-PSM) via directional vectorizations for homogeneous regions, and Reduced-Rank Regression (RR) with AIC-based rank selection for textured areas. Additionally, we introduce a robust noise-level estimation method for Gaussian noise. Experimental results demonstrate that S3F delivers effective NR across varying noise intensities in natural images, achieving performance that is superior or comparable to the BM3D method while requiring approximately 50% of the execution time. S3F provides a transparent and practical alternative to black-box, learning-based methods,

particularly in resource-constrained environments.

5.2 Keyword

Image noise reduction, noise level estimation, multi-model inference, reduced rank regression

5.3 Introduction

Image noise reduction (INR) is a fundamental task in image processing, aimed at estimating a noise-free image from a noisy observation. This computational challenge is particularly critical in applications spanning medical imaging, computer vision, and remote sensing, where noise can degrade the quality of posterior analyses [119, 120].

The ubiquity of noise in imaging systems arises from various sources, including thermal fluctuations in sensor electronics, photon shot noise, and quantization errors during analog-to-digital conversion.Thermal noise can be well-approximated by a Gaussian distribution, is uncorrelated, and under suitable acquisition conditions, the constant noise-variance assumption holds [18]. Shot noise, which arises from the discrete nature of charge or energy carriers, such as electrons in electronic circuits or photons in optical systems, is usually modeled as a Poisson distribution, but when the number of electrons/photons is large, the Poisson process can be suitably approximated by a Gaussian distribution [19, 20].

5.3.1 Problem statement

We consider the additive noise model where the observed noisy image $Y \in \mathbb{R}^{M \times N}$, is generated through the degradation of a clean image $X \in \mathbb{R}^{M \times N}$ by Additive White Gaussian (AWG) Noise with zero-mean and noise level $\sigma \in \mathbb{R}$, i.e., $Y = X + W$, where $W \in \mathbb{R}^{M \times N}$ represents the noise matrix whose entries are independently and identically (iid) sampled from a zero-mean Gaussian distribution with standard deviation σ.

The noise reduction (NR) problem is formulated as an inverse problem that seeks to recover an accurate estimate $\hat{X} \in \mathbb{R}^{M \times N}$ of the noise-free image X. Typically, the quality of image NR algorithms is quantified using the Peak Signal-to-Noise Ratio (PSNR), defined as PSNR $= 10 \log_{10} \frac{\max(X)^2}{\text{MSE}}$, where MSE denotes the Mean Squared Error, that quantifies the average squared deviation between the original and reconstructed images, and higher PSNR values indicate better reconstruction quality.

5.3.2 Literature review

Image Noise Reduction (INR) methods include transform domain filtering techniques including Wavelet and Fourier transforms, spatial domain filters such as mean, median, and Gaussian filters, dictionary-based approaches like K-SVD [84], and block matching algorithms such as Non-local Means [102] and BM3D [83].

Methods that rely on Supervised Learning (SL) gained significant attention due to their superior performance in specific Gaussian noise reduction tasks [57, 58]. Learning-based INR methods predominantly rely on supervised learning techniques, which assume identical distributions between training and testing data [61–63]. However, this assumption frequently fails in real-world scenarios, where noise characteristics and image exhibit substantial variability. Furthermore, the availability of noise-free images is often limited, posing a constraint in the generalization capabilities of SL-based NR methods [70, 121, 122]. The limited generalization of supervised learning-based INR methods represents a methodological limitation that has motivated the development of Self-Supervised Learning (SSL) INR methods, these techniques operate without requiring clean training data. Empirical studies have demonstrated that SSL INR methods are capable of achieving NR performance comparable to their supervised counterparts [64] [65]. Nonetheless, SSL methods tend to be computationally intensive [63], as they require large datasets to match or improve the performance of supervised methods [68], which limits their use in real-time applications [69] [70] [71]. These constraints reduce the practicality of learning-based models in many real-world settings, including those where data sources and/or computational resources are limited.

In these scenarios, training-free image NR techniques maintain their relevance by providing a balance between implementation ease, performance, interpretability, and computational efficiency.

Paper Organization

The remainder of this paper is organized as follows: Section 5.4 details the procedure for estimating image noise level, Section 5.5 introduces an extension of the Sliding-Window Piecewise Smooth Model (SW-PSM) for images. This model functions as an initial NR pass for piecewise smooth regions in Y and utilizes point-estimate variance to identify non-smooth regions where the SW-PSM method is less suitable. Section 5.6 proposes a Reduced-Rank Regression model designed to suppress noise where local smoothness assumptions fail. In section 5.7, we define the final image estimate by aggregating the outputs of both the SW-PSM and Reduced-Rank Regression frameworks. Section 5.8 presents the results and discussion of the proposed methods. Conclusions are presented in Section 5.9.

This work presents the following contributions: 1) the application of the SW-PSM method for image noise reduction, 2) the introduction of a data-driven reduced rank regression approach to reduce noise from non-smooth regions, and 3) a classification strategy to distinguish between smooth and non-smooth regions in Y, ensuring that each local region is processed using the most suitable noise reduction framework.

5.4 Noise-level estimation

Noise-level estimation on 1D vectors

The proposed framework requires an estimate of the noise standard deviation σ, which is typically unknown. We derive this estimate by analyzing local variations in the observed signal $y \in \mathbb{R}^N$. We assume that y is corrupted by additive white Gaussian noise (AWGN), such that $y = x + w$, where x is the clean signal and the components of w are independent and identically distributed random variables (i.i.d.) following a Gaussian distribution $\mathcal{N} \sim (0, \sigma^2)$. Specifically, we utilize a high-order differencing approach proposed in [123].

$$\hat{s} = \sqrt{\frac{1}{M}|v * y|^2}, \tag{5.1}$$

where v is a vector designed to isolate the noise by vanishing the smooth components in y and M is the dimension of the resulting convolution vector.

With the purpose of reducing execution time while maintaining an accurate estimation of σ, we propose to estimate σ in an image based on the average of multiple noise-level estimates, each derived by applying Eq. (5.1) to suitable pixel sequences in Y. We preserve the accuracy of this procedure by restricting the application of Eq. (5.1) to sequences found in a relatively homogeneous square patch $H \in \mathbb{R}^{m \times m}$ within Y, i.e., $H \subset Y$.

Where H is defined as the patch that minimizes the residual sum of squares (RSS) under a two-dimensional (2D) linear regression of its pixel intensities. Figure 5.1 displays examples of H of size $m \times m$ ($m = 50$) for the test images *Barbara*, *Boats*, *Couple*, *House*, *Man*, and *Plane* (obtained from [124,125]) corrupted with AWGN at $\sigma = 0.05$. The test original images are presented in Figure 5.4.

Although H is selected for its global linearity, it may still contain localized non-smooth structures that could bias the estimation of σ [123]. We refine the pixel sequence selection by fitting a third-degree polynomial model using a minimum mean-squared error estimator

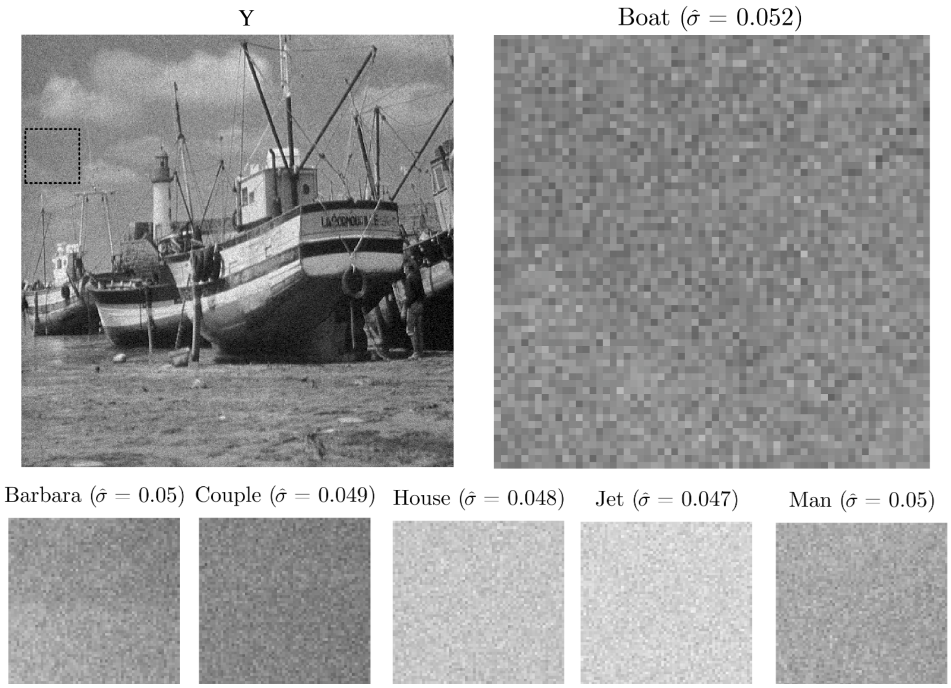

Figure 5.1 **Top-left**: Noisy *Boats* image Y and the corresponding homogeneous patch H (square delimited by dashed lines). **Top-right:** Smooth patch Y_s (zoomed-in) obtained from Y. The noise-level estimate is $\hat{\sigma} = 0.052$. **Bottom:** H for to different noisy ($\sigma = 0.05$) test images .

to each of the rows and columns in H and validate the model fit employing the Ljung-Box autocorrelation test performed across all lags from 1 to $m - 1$. Using a significance level of $\alpha = 0.05$, we reject the null hypothesis of uncorrelated residuals for any sequence exhibiting significant structural dependencies. Sequences that pass this criterion are designated as *smooth sequences*. For the 2d case, our noise level estimate is defined as

$$\hat{\sigma} = \sqrt{\frac{1}{m} \sum_{i=0}^{m} \hat{s}_i^2}, \tag{5.2}$$

m is the number of smooth sequences that are kept.

5.5 SW-PSM for Image Noise Reduction

Overview

The Sliding-Window Piecewise-Smooth Model (SW-PSM) [123] builds upon the Piecewise-Smooth Modeling (PSM) framework, a statistical NR method originally developed for 1D signals corrupted by AWGN. The PSM method employs an optimized sequence of polynomials to model the signal, utilizing the Akaike Information Criterion (AIC) to balance goodness-of-fit against model complexity. This approach is particularly effective at preserving signal discontinuities and angular points while maintaining high fidelity in smooth regions. SW-PSM extends this framework by analyzing overlapping segments through a sliding window. Within each window, local PSM estimates and their corresponding uncertainties are generated. These local estimates are aggregated using inverse-variance weighting (IVW) to produce a global reconstruction along with a point wise variance of the final estimate.

This section introduces the SW-PSM method in the context of 2D signal processing. To apply the SW-PSM method to 2D data, we employ vectorization operators that map the 2D input Y to a collection of 1D vectors, which are then processed individually using the SW-PSM method.

Image vectorization

We define a vectorization $\text{vec}(Y)$ of Y as a mapping of Y to a collection of 1D vectors y_i, i.e. $\text{vec}(Y) = \{y_1, y_2, ..., y_\ell\}$, where the number of vectors $\ell \in \mathbb{N}$ depends on the vectorization scheme. Within each vectorization scheme, the supports of any pair of vectors (y_i, y_j), $i \neq j$, are non-overlapping. Our method typically combines four SW-PSM estimates obtained from four different vectorizations. Specifically, we use the four vectorizations shown in Figure 5.2, where the gray lines (horizontal, vertical, and the two diagonals with slopes 1 and -1) represent the supports of the vectors y_i associated with each vectorization scheme. Note that the dimension of y_i depends on i for the two diagonal vectorizations. Using multiple vectorizations increases the likelihood that a pixel belongs to a sufficiently large smooth segment in one of the 1D signals y_i, leading to more accurate estimates.

The horizontal and vertical vectorizations (Figure 5.2, panels 1-2) cover the image Y entirely; every pixel is included in exactly one vector y_i. In contrast, the diagonal vectorizations (panels 3–4) exclude certain vectors in $V(Y)$ from the estimation process— specifically those near the image corners with lengths ≤ 4.

SW-PSM Image Estimate

For each pixel-value $Y[p] \in \mathbb{R}$ at position $p := (i, j) \in \{1, 2, ..., M\} \times \{1, 2, ..., N\}$, we denote by $V_p[Y]$ the subset of vectorized signals whose supports include the pixel at position p. Each vector in $V_p[Y]$ is used to estimate $X[p]$ and its variance. Note that the number of vectors $\ell_p \in \mathbb{N}$ may be smaller than n when p is located near an image corner.

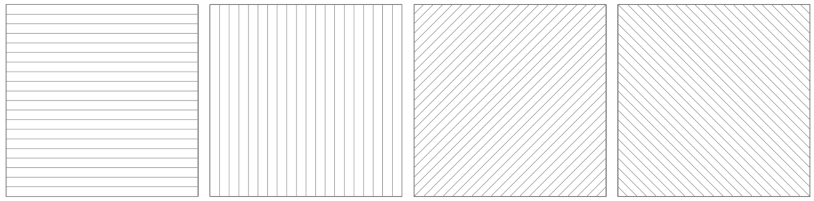

Figure 5.2 Illustration of the four vectorization schemes (black squares) used in SW-PSM. The support of each vector is indicated by gray lines within the corresponding black square. From left to right: horizontal, vertical, slope = 1 and slope = -1 vectorizations.

The integer indices of the vectors in $V_p[Y]$ are collected in the set $I_p = \{1, 2, ..., \ell_p\}$. For each index $k \in I_p$, the corresponding estimate of $X[p]$ derived from the k-th vector is denoted by $\hat{X}_k[p]$. The final pixel-value estimate $\hat{X}[p]$ is computed as a weighted average of these individual estimates with

$$\hat{X}_{\mathrm{PS}}[p] := \frac{\sum_{k \in I_p} \hat{X}_k[p]/(\sigma_k[p])^2}{\sum_{k \in I_p} 1/(\sigma_k[p])^2}, \tag{5.3}$$

where $(\sigma_k[p])^2$ denotes the variance of $\hat{X}_k[p]$. The standard deviation ($\sigma[\hat{X}_{\mathrm{PS}}[p]]$) of $\hat{X}_{\mathrm{PS}}[p]$ is given by

$$\sigma[\hat{X}_{\mathrm{PS}}[p]] := \frac{1}{\sqrt{\sum_{k \in I_p} 1/(\sigma_k[p])^2}}. \tag{5.4}$$

It may happen that a given pixel does not belong to any smooth signal piece, i.e. it is an outlier. If this pixel is classified as an outlier for a majority of the estimates $\hat{X}_k[p]$, $p \in I_p[Y]$, then we use $\hat{X}[p] = Y[p]$ and $\sigma[\hat{X}[p]] = \hat{\sigma}$. By applying (5.3) to each pixel in Y, we produce the SW-PSM image estimate $\hat{X}_{\mathrm{PS}} \in \mathbb{R}^{n \times n}$.

5.6 NR via Reduced-Rank Regression

5.6.1 Motivation and overview

Natural images typically exhibit both piecewise smooth regions and heterogeneous structures, the latter manifesting as highly textured regions and fine-scale details. In textured regions, high frequency intensity variations occur across small pixel neighborhoods, causing the signal to deviate significantly from the continuous, piecewise-smooth structure required for SW-PSM to perform optimally.

To overcome this problem, we exploit the Self-Similarity (SS) property of natural images, i.e., the tendency for structural patterns to repeat throughout a scene [126]. Building on this principle, we propose to use a Reduced-Rank (RR) Regression approach to complement SW-

PSM in non-smooth regions. For a noisy image Y, we consider overlapping $L \times L$ patches, where $L \in \mathbb{N}$. We denote $\mathbf{y}_i \in \mathbb{R}^d$ (with $d = L^2$) as the column-vectorized version of the patch centered at the i^{th} pixel. The RR approach is motivated by the fact that similar image patches often represent the same underlying local structure, differing only by minor spatial shifts or intensity fluctuations.

Mathematically, self-similarity implies that a reference patch in an image can be matched with a collection of highly similar patches. Due to structural redundancy, this collection does not span $\mathbb{R}^d$; instead, these vectors cluster within a local signal subspace $\mathcal{S}_i \subset \mathbb{R}^d$ of low rank, defined by a small number of basis vectors [127, 128]. To suppress noise in the non-smooth regions of Y, we treat noise reduction as the estimation of a basis for $\mathcal{S}_i$, which is then used to filter the reference patch $\mathbf{y}_i$.

5.6.2 Patch grouping

For each reference patch $\mathbf{y}_i$, we search for all similar candidates within a local neighborhood $\mathcal{N}_i$, defined as the intersection of the image Y and an $m \times m$ square centered at pixel i ($m \geq 31$ is an odd integer). All overlapping $L \times L$ patches within $\mathcal{N}_i$ are evaluated. A candidate patch $\mathbf{y}_j$ is included in the set $\mathcal{C}_i$ if it satisfies the distance threshold: $\| \mathbf{y}_i - \mathbf{y}_j \|^2 < q\,\sigma^2\,d$, where d is the patch dimension, σ^2 is the noise variance, and $q > 0 \in \mathbb{R}$ is a scalar controlling the selectivity of the search. In this context, we use $q = 6$, which, with high probability, ensures that two noisy versions of the same patch are selected as part of the same patch group without being excessively restrictive [129]. The k^{th} element of $\mathcal{C}_i$ is denoted by $\mathbf{y}_{i,k}$, so that $\mathcal{C}_i = \{\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, ..., \mathbf{y}_{i,n_i}\}$, where n_i represents the cardinality $|\mathcal{C}_i|$.

In scenarios where $n_i < d$, the search results in an underdetermined system, which may produces inaccurate estimates of the covariance matrix, limiting the model's ability to accurately capture the underlying low-rank structure. To mitigate this, we increase the search space of similar patches using a data augmentation step on the candidate patches within $\mathcal{N}_i$. This involves applying the D_8 transformation group (comprising rotations and reflections) to the neighboring patches [130]. Each transformed version is subsequently evaluated against the similarity threshold; if a transformed patch satisfies the distance criterion, it is appended to $\mathcal{C}_i$, thereby increasing the sample size and improving the robustness of the subsequent rank-reduction process.

We then construct the observation matrix $\mathbf{Y}_i = [\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, \ldots, \mathbf{y}_{i,n_i}]$, $\mathbf{Y}_i \in \mathbb{R}^{d \times n_i}$ by stacking the similar patches as columns where $\mathbf{y}_{i,1} = \mathbf{y}_i$, and $\mathbf{y}_{i,j}$ with $j = 2, 3, ..., n_i$ is the j-th closest patch to $\mathbf{y}_i$ in Euclidean distance. Within this patch-based framework, the observation matrix

$\mathbf{Y}_i$ is modeled as

$$\mathbf{Y}_i = \mathbf{X}_i + \mathbf{W}_i, \tag{5.5}$$

where $\mathbf{X}_i \in \mathbb{R}^{d \times n_i}$ represents the unknown clean patch group and $\mathbf{W}_i \in \mathbb{R}^{d \times n_i}$ is a matrix with white Gaussian noise components. Although the use of overlapping patches introduces correlation between the noise components of adjacent patches, we follow the standard convention in non-local processing by treating $\mathbf{W}_i$ as a random matrix [62, 82, 102] with locally independent entries distributed as $\mathcal{N}(0, \sigma^2)$.

5.6.3 Reduced-Rank Representation and Estimation

Under the Self-Similarity (SS) hypothesis, the high degree of redundancy in natural images induces strong linear dependencies between the columns of the clean patch group $\mathbf{X}_i$, implying a low-rank structure. We assume $\mathbf{X}_i$ can be written as $\mathbf{X}_i = U_i\, Z_i$, where $U_i \in \mathbb{R}^{d \times d}$ is an orthonormal basis spanning the subspace $\mathcal{S}_i \subset \mathbb{R}^d$ and $Z_i \in \mathbb{R}^{d \times n_i}$ represents the coordinates of $\mathbf{X}_i$ in the space spanned by U_i.

In the reduced-rank (RR) representation framework, it is assumed that $\mathbf{X}_i$ has a low-rank structure, meaning that a basis of $\mathbf{X}_i$ contains $k < n_i$ elements and that the extra $n_i - k$ dimensions are typically composed by noise. Therefore, the goal is to obtain a reduced-rank estimate of U_i that will be employed to reconstruct a noise-reduced version of $\mathbf{x}_i$.

To estimate U_i, from the noisy patch group $\mathbf{Y}_i$, we assume

$$\mathbf{Y}_i = U_i\, Z_i + \mathbf{W}_i, \tag{5.6}$$

where $\mathbf{W}_i \in \mathbb{R}^{d \times n_i}$ represents a noise matrix [131]. The optimal rank-k estimate of $\mathbf{X}_i = U_i\, Z_i$ in the Frobenius norm is obtained via the truncated Singular Value Decomposition (SVD) of $\mathbf{Y}_i$ [132]. The SVD of $\mathbf{Y}_i$ is $\mathbf{Y}_i = \hat{U}_i \hat{\Sigma}_i \hat{V}_i^\top$, where $\hat{U}_i \in \mathbb{R}^{d \times d}$ and $\hat{V}_i \in \mathbb{R}^{n_i \times n_i}$ are orthogonal matrices. The diagonal entries of $\hat{\Sigma}_i \in \mathbb{R}^{d \times n_i}$ are the singular values $\hat{\lambda}_{i,j}$, which are non-negative real numbers arranged in non-increasing order ($\hat{\lambda}_{i,1} \geq \hat{\lambda}_{i,2} \geq \cdots \geq \hat{\lambda}_{i,n_i} \geq 0$). The columns of $\hat{U}_i$ and $\hat{V}_i$ are the corresponding singular vectors, ordered to match this sequence. The truncated SVD approximation of rank-k is given by

$$\hat{\mathbf{X}}_i = \hat{U}_{i,k} \hat{\Sigma}_{i,k} \hat{V}_{i,k}^\top \tag{5.7}$$

where $\hat{U}_{i,k} \in \mathbb{R}^{d \times k}$ and $\hat{V}_{i,k} \in \mathbb{R}^{n_i \times k}$ denote the first k columns of U_i and V_i, and $\hat{\Sigma}_{i,k} \in \mathbb{R}^{k \times k}$ is a diagonal matrix containing the k largest singular values in $\hat{\Sigma}_i$. The residual error of this approximation is given by the sum of the discarded squared singular values, $\sum_{j=k+1}^{n_i} \hat{\lambda}_{i,j}^2$.

Intuitively, the RR model preserves the k most significant elements in the basis, and discards the low-variance elements that are typically dominated by noise.

This approximation is numerically equivalent to performing Principal Component Analysis (PCA) on $\mathbf{Y}_i$. In practice, we perform centered PCA by first computing the empirical mean patch of the group, $\bar{\mathbf{y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{y}_{i,j}$. The SVD is then applied to the centered matrix $\tilde{\mathbf{Y}}_i = \mathbf{Y}_i - \bar{\mathbf{y}}_i \mathbf{1}^\top$, where $\mathbf{1}^\top = [1, 1, \ldots, 1] \in \mathbb{R}^{1\times n_i}$.

Given the basis $\hat{U}_{i,k}$, the clean patch estimate $\hat{\mathbf{x}}_i$ is obtained by projecting the noisy reference patch onto the k-dimensional subspace and restoring the mean, i.e.,

$$\hat{x}_i = \bar{\mathbf{y}}_i + \hat{U}_{i,k}\hat{U}_{i,k}^\top(\mathbf{y}_i - \bar{\mathbf{y}}_i). \tag{5.8}$$

In this framework, the choice of k is critical: an excessively small k may lead to oversmoothing, while an excessively large k may retain too much noise. Rather than using arbitrary thresholds, we propose an objective selection of k via Akaike Information Criterion (AIC) [10]. In the following section, we approximate the likelihood function necessary to compute the AIC for a rank-k RR model.

5.6.4 Surrogate Likelihood of the RR model

To select the rank k, we employ an information-theoretic approach based on the Akaike Information Criterion (AIC). While overlapping in $\mathbf{Y}_i$ introduces statistical dependencies in the noise, the independent AWGN likelihood serves as a robust and computationally efficient surrogate likelihood for the model selection framework. We model the clean patch group as $\mathbf{X}_i = U_{i,k}\, Z_i$, where $U_{i,k} \in \mathbb{R}^{d\times k}$ is the orthonormal basis spanning the k-dimensional subspace of similar patches, and $Z_i \in \mathbb{R}^{k\times n_i}$ is the matrix of representation coefficients. Each column of Z_i contains the weights required to reconstruct a clean patch as a linear combination of the basis vectors in $U_{i,k}$.

Under the independent AWGN assumption, the joint probability density function for $\mathbf{Y}_i$ is constructed as the product of the probabilities of its $d \times n_i$ individual entries, given by

$$L(\mathbf{Y}_i|U_{i,k}, Z_i) = \prod_{j=1}^{d}\prod_{m=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}}\, \exp\left(-\frac{(\mathbf{Y}_i[j,m] - (U_{i,k}Z_i)[j,m])^2}{2\sigma^2}\right),$$

where $\mathbf{Y}_i[j,m]$ and $(U_{i,k}Z_i)[j,m]$ are the pixel values at the j^{th} row and m^{th} column of $\mathbf{Y}_i$

and $U_{i,k}Z_i$, respectively. Applying the natural logarithm yields the log-likelihood function

$$\ell_i(\mathbf{Y}_i|U_{i,k}, Z_i) = \sum_{j=1}^{d}\sum_{m=1}^{n_i}\left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(\mathbf{Y}_i[j,m] - (U_{i,k}Z_i)[j,m])^2}{2\sigma^2}\right].$$

The double sum over the second term in the brackets represents the sum of squared residuals between $\mathbf{Y}_i$ and the rank-k approximation, which is equivalent to the squared Frobenius norm $\|\mathbf{Y}_i - U_{i,k}Z_i\|_F^2$. Simplifying the equation above yields the expression

$$\ell_i(\mathbf{Y}_i|U_{i,k}, Z_i) = -\frac{d\,n_i}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\parallel \mathbf{Y}_i - U_{i,k}Z_i \parallel_F^2 . \tag{5.9}$$

This function is maximized when $\parallel \mathbf{Y}_i - U_{i,k}\,Z_i \parallel_F^2$, is minimal. As established by the Eckart-Young-Mirsky theorem, this error is minimized when $U_{i,k}$ consists of the first k left singular vectors of $\mathbf{Y}_i$ derived via SVD. At this optimum, the residual sum of squares is $\sum_{j=k+1}^{n_i}\hat{\lambda}_{i,j}^2$. The resulting maximum surrogate log-likelihood for a rank-k RR model is

$$\ell_i(\mathbf{Y}_i|\hat{U}_{i,k}, \hat{Z}_i) = -\frac{d\,n_i}{2}\,\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=k+1}^{n_i}\hat{\lambda}_{i,j}^2. \tag{5.10}$$

This expression serves as the data-fidelity component in the following subsection, where it is combined with the model complexity penalty to derive the AIC for the RR model.

5.6.5 AIC for Rank Selection and Model Aggregation

The selection of the rank k requires balancing data fidelity and model complexity. Selecting a value that is too small, i.e., a small number of components, may omit subtle image features, producing an oversmoothed estimate $\hat{x}_i$ that cannot capture fine-scale textures. Conversely, choosing k too large may include noise-dominated singular vectors in the basis, leading to insufficient noise suppression.

To achieve a balance between under- and oversmoothing, we utilize the Akaike Information Criterion (AIC) rather than the Bayesian Information Criterion (BIC). Since we are mainly interested in using the RR approach to reduce noise from non-smooth signal regions, BIC was not adopted here for two reasons: theoretically, it is better suited for large sample sizes [11], a condition not guaranteed for the small local patches considered here; and empirically, it was observed to oversmooth fast-varying structures, whereas the AIC better preserves such features in noisy signals [123].

To achieve a balance between under- and oversmoothing, we utilize the Akaike Information

Criterion (AIC), defined as

$$\mathrm{AIC} = -2\ell(z|\hat{\theta}) + 2p, \tag{5.11}$$

where $\ell(z|\hat{\theta})$ is the maximized log-likelihood of the data z given the parameter vector $\hat{\theta}$ which represents a data-fidelity term, and $p \in \mathbb{N}$ is the number of parameters defining the model [10]. For problems requiring the selection of a single model among multiple candidates, selecting the model that minimizes Eq. (5.11) yields a principled compromise between data fidelity and model complexity.

The total number of parameters in the RR model includes d parameters for the empirical means of the d coordinates of each patch vector, $k(d - \frac{k+1}{2})$ parameters required to specify the orthonormal vectors in $\hat{U}_{i,k}$ [133], and the k singular values. Furthermore, the product $\hat{\Sigma}_{i,k}\hat{V}_{i,k}^{\top} \in \mathbb{R}^{k\times n_i}$ in Eq. (5.7) represents the projection coefficients of the patches onto the subspace spanned by $\hat{U}_{i,k}$. Following a fixed-effects approach to PCA [134], we explicitly include the $n_i\,k$ projection coefficients as incidental parameters in our AIC penalty. If the noise level is unknown, one additional parameter is included for its estimation. The number of free parameters for a rank-k RR model associated with the i-th pixel in Y is therefore

$$p_i(k) = d + k\left(d - \frac{k+1}{2}\right) + k + n_i\,k + 1 \tag{5.12}$$

We utilize Eq. (5.10) as a surrogate of the likelihood term in Eq. (5.11). By substituting $p_i(k)$ and the surrogate log-likelihood from Eq. (5.10) into Eq. (5.11) and discarding constant terms that do not depend on k, the effective AIC of the rank-k RR model associated with i-th noisy patch in Y is

$$\mathrm{AIC}_i(k) = \frac{1}{\sigma^2}\sum_{j=k+1}^{\min(d,n_i)} \hat{\lambda}_{i,j}^2 + 2k\left(d - \frac{k-1}{2} + n_i\right). \tag{5.13}$$

The selected rank k^* is subsequently chosen as the value that minimizes $\mathrm{AIC}_i(k)$ over the range $k \in \{1, \ldots, n_i - 1\}$.

5.6.6 Reduced Rank Multi-Model Inference

While minimizing Eq. (5.13) provides a principled mechanism for selecting k, relying on a single model can lead to instability when multiple ranks yield comparable AIC values. To account for this model uncertainty, we adopt a Multi-Model Inference framework [135].

Let $\mathcal{K}_i = \{1, 2, \ldots, \min(d, n_i)\}$ denote the set of candidate ranks. For each $k \in \mathcal{K}_i$, we define the AIC difference relative to the AIC-minimal model as: $\Delta_i(k) := \mathrm{AIC}_i(k) - \mathrm{AIC}_i(k^*)$.

The relative likelihood of the k-th model, given the data, is proportional to $\exp(-\frac{1}{2}\Delta_i(k))$. To focus the estimation on the most plausible candidates, we consider only the subset of models with substantial empirical support, defined as $\mathcal{K}_i^* = \{k \in \mathcal{K}_i : \Delta_i(k) < 2\}$ [135]. By normalizing the likelihoods over this subset, we obtain the AIC weights $w_i(k)$ [135]

$$w_i(k) = \frac{\exp\left(-\frac{1}{2}\Delta_i(k)\right)}{\sum_{m\in\mathcal{K}_i^*}\exp\left(-\frac{1}{2}\Delta_i(m)\right)}, \tag{5.14}$$

where $0 \leq w_i(k) \leq 1$ and $\sum_{k\in\mathcal{K}_i^*} w_i(k) = 1$ We define the final estimate of $\mathbf{x}_i$ as

$$\hat{\mathbf{x}}_i^* = \sum_{k\in\mathcal{K}_i^*} w_i(k)\,\hat{\mathbf{x}}_{i,k}. \tag{5.15}$$

This procedure mitigates the ambiguity of selecting a single rank. The covariance matrix of $\hat{\mathbf{x}}_i^*$ is given by [135]:

$$\mathrm{cov}(\hat{\mathbf{x}}_i^*, \hat{\mathbf{x}}_i^*) = \sum_{k\in\mathcal{K}_i^*} w_i(k)\left[\sigma^2\hat{U}_{i,k}\hat{U}_{i,k}^\intercal + (\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i^*)(\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i^*)^\intercal\right], \tag{5.16}$$

where the variances of the $\hat{\mathbf{x}}_i^*$s are located in the diagonal of $\mathrm{cov}(\hat{\mathbf{x}}_i^*, \hat{\mathbf{x}}_i^*)$.

Each pixel $p \in \{1, 2, ..., M\}\times\{1, 2, ..., N\}$ is typically contained in several overlapping patches. We denote by $\hat{X}^*[p, k] \in \mathbb{R}$ the point-estimate of pixel p associated with the patch index k, and we collect these estimates in the list $\mathcal{E}_p = \{\hat{X}^*[p, 1], \hat{X}^*[p, 2], ..., \hat{X}^*[p, |\mathcal{E}_p|]$. The final image estimate $\hat{X}_{\mathrm{RR}} \in \mathbb{R}^{M\times N}$ is obtained using a weighted average of the point-estimates $\hat{X}^*[p, k]$ for each pixel position p across the entire image:

$$\hat{X}_{\mathrm{RR}}[p] = \frac{\sum_{k=1}^{|\mathcal{E}[p]|}\hat{X}^*[p, k]/(\sigma[\hat{X}^*[p, k])^2}{\sum_{k=1}^{|\mathcal{E}[p]|} 1/(\sigma[\hat{X}^*[p, k])^2}, \tag{5.17}$$

where $\sigma[\hat{X}^*[p, k])^2$ denotes the estimated variance of $\hat{X}^*[p, k]$.

The inverse-variance weighting in Eq. (5.17) is motivated by the classical best linear unbiased estimate (BLUE) under independent, heteroscedastic errors [111]. However, since each pixel p may appear in multiple spatially overlapping patches, the estimates $\{\hat{X}^*[p, k]\}$ are not mutually independent. Specifically, since patches sharing pixels generate correlated residuals, the independence assumption required for BLUE optimality is violated. This correlation is a well-known limitation of patch-based aggregation, shared by methods such as NLM [102] and BM3D [82], and is generally accepted as a tractable approximation. The justification is that even correlated averaging reduces variance relative to any single patch estimate [102].

5.7 Hybrid SW-PSM and RR method

While self-similarity (SS) methods can effectively reduce noise in redundant, non-smooth regions, the requisite patch-grouping and transformation steps result in significantly higher computational costs than piecewise-smooth (PS) approaches. However, in regions with low self-similarity, such as non-repeating structures or stochastic textures, SS methods may be insufficient for robust noise reduction.

For these reasons, we propose a hybrid framework that integrates the PS and SS hypotheses into a single estimator, we call this method Smoothness and Self-Similarity Fusion (S3F). In the first stage of the S3F method, we use SW-PSM model with a dual purpose: 1) providing the final estimate for smooth regions where the PS assumption holds, and 2) acting as a baseline to identify non-smooth regions that may benefit from SS-based processing via the reduced-rank regression (RR) method presented in section 5.6. This texture-guided method is motivated by the fact that for homogeneous areas, SW-PSM's local polynomial support is a superior model precisely because it is optimized for smooth representations. In textured regions where both the PS and RR estimates are available, a final estimate is obtained through an averaging scheme.

Within the SW-PSM framework, the variance of a point estimate is inversely proportional to the spatial support size of the estimated smooth component [123]. While extensive supports in homogeneous regions yield low estimation variance, the SW-PSM estimate exhibits significantly higher variance in regions characterized by large variations over small supports, e.g. high-frequency textures.

Building on this property of the SW-PSM estimate variance, we partition the image domain $\mathcal{D} = \{1, \ldots, N\} \times \{1, \ldots, M\}$ into two disjoint sets based on the SW-PSM estimation variance. The set of textured (non-smooth) pixel positions is defined as

$$\mathcal{D}_T = \{p \in \mathcal{D} : \mathrm{Var}(\hat{X}_{\mathrm{PS}}[p]) > \tau\}. \tag{5.18}$$

To ensure that the variance threshold τ effectively separates smooth and textured regions, we define τ as the typical value of the SW-PSM pointwise variance estimate in the homogeneous region H defined in Section 5.4, i.e., $\tau := \mathrm{median}(\mathrm{Var}(\hat{X}_{\mathrm{PS}}(H)))$, where $\mathrm{Var}(\hat{X}_{\mathrm{PS}}(H))$ denotes the collection of SW-PSM point-wise variance estimates associated with H. Using the median allows τ to reflect the typical estimation variance while remaining robust to atypically high values caused by isolated discontinuities or residual edges that may be found in H. Conversely, the set of smooth pixel positions is $\mathcal{D}_S = \{p \in \mathcal{D} : \mathrm{Var}(\hat{X}_{\mathrm{PS}}[p]) \leq \tau\}$.

The S3-F estimate $\hat{X}_H$ at pixel p is defined by a spatially adaptive selection rule that switches between the SW-PSM and RR models. Depending on whether a pixel belongs to the smooth or textured partition, the estimate is given by:

$$\hat{X}_H[p] = \begin{cases} \hat{X}_{\mathrm{PS}}[p] & \text{if } p \in \mathcal{D}_S, \\ \frac{\omega_{\mathrm{PS}}[p]\hat{X}_{\mathrm{PS}}[p]+\omega_{\mathrm{RR}}[p]\hat{X}_{\mathrm{RR}}[p]}{\omega_{\mathrm{PS}}[p]+\omega_{\mathrm{RR}}[p]} & \text{if } p \in \mathcal{D}_T, \end{cases} \tag{5.19}$$

where $\omega_{\mathrm{PS}}[p]$ and $\omega_{\mathrm{RR}}[p]$ are the inverse variance weights associated with the estimates $\hat{X}_{\mathrm{PS}}[p]$ and $\hat{X}_{\mathrm{RR}}[p]$, respectively. This formulation ensures that smooth regions benefit from the piecewise-smooth model, while textured regions are processed using a linear combination of the SW-PSM and RR estimates.

To account for the possible bias in the IVW weighting scheme, we introduce a simpler, uniformly weighted hybrid estimate $\hat{X}_U$, defined as

$$\hat{X}_U[p] = \begin{cases} \hat{X}_{\mathrm{PS}}[p] & \text{if } p \in \mathcal{D}_S, \\ \frac{\hat{X}_{\mathrm{PS}}[p]+\hat{X}_{\mathrm{RR}}[p]}{2} & \text{if } p \in \mathcal{D}_T. \end{cases} \tag{5.20}$$

Figure 5.3 Left to right: Noisy ($\sigma = 0.05$) sample, SW-PSM , RR, and S3F (uniform weighting) estimates.

A comparison of the image estimates provided by SW-PSM, RR with $q = 8$ and patch size 9×9, and the proposed S3F framework is presented in Figure 5.3. The leftmost panel shows a zoomed-in region of the *Barbara* image at $\sigma = 0.05$. The SW-PSM estimate (second panel) reduces noise in homogeneous regions, such as the face and hand, yet remains less effective in the high-frequency "striped" textures. Conversely, while the RR estimate (third panel) successfully reconstructs the redundant textures of the clothing, it introduces noticeable artifacts in the smoother regions. The S3F estimate with uniform weighting aggregation (rightmost panel) successfully integrates the advantages of both methods; by using only the SW-PSM estimate on the smooth region and combining both RR and SW-PSM estimates on the textured regions, S3F achieves a PSNR improvement of more than 1 dB over the individual SW-PSM and RR models.

5.8 Results and Discussion

5.8.1 Image Noise Reduction with synthetic AWGN

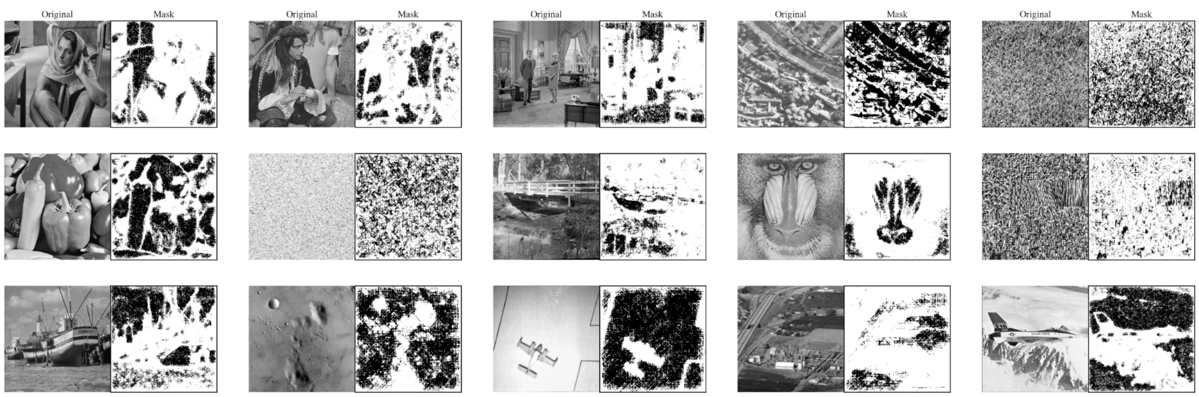

Figure 5.4 Test images used in the experiments. Each pair corresponds to the original image (left) and its texture segmentation when $\sigma = 0.05$ (right), depicting smooth (black) and textured (white) regions.

We evaluated the performance of the proposed noise reduction (NR) methods against the benchmark BM3D algorithm as implemented in [85]. Two variants of the S3F method were tested: one utilizing inverse variance weighting (S3F-IVW) and uniform weighting (S3F-UW) for aggregation. The evaluation was conducted on grayscale images corrupted with synthetic AWGN at three noise levels: $\sigma \in \{0.025, 0.05, 0.1\}$.

The test images are displayed in Figure 5.4 along with an example of the textural segmentation ($\sigma = 0.05$) described in Section 5.7, with smooth and textured regions represented by black and white pixels, respectively. We chose an evaluation set combining standard natural scenes (e.g. *couple, house, plane, boat, and peppers*) with images characterized by dense texture, including *grass, sand, and texture_mosaic*. All test images are accessible in [124].

All NR methods utilized the noise level estimation procedure described in Section 5.4. Per-

Table 5.1 PSNR results for various denoising methods at Noise Level = 0.05.

Signal	SW-PSM	S3F-UW	S3F-IVW	BM3D
Airplane	34.84	35.51	35.68	**36.88**
Brodatz_Sand	28.51	**29.45**	28.79	28.96
Chemical_Plant	29.48	30.42	29.92	**30.97**
aerial	28.32	**28.91**	28.56	28.19
grass	26.10	**26.85**	26.39	26.75
baboon	27.93	**28.67**	28.27	28.52
Barbara	29.86	31.55	30.96	**32.30**
boat	31.05	32.01	31.54	**32.09**
bridge	28.69	**29.27**	28.96	29.04
couple	30.87	31.79	31.41	**31.98**
house	34.96	36.13	36.03	**38.29**
man	30.98	31.69	31.36	**31.82**
moonsurface	30.71	**31.09**	30.91	30.38
peppers	32.46	**33.08**	32.86	32.78
plane	32.39	32.87	32.78	**33.91**
texmos1	24.31	**25.09**	24.62	23.28

formance was quantified using the Peak Signal-to-Noise Ratio (PSNR), where higher values indicate superior noise suppression. For the self-similarity component of our Hybrid method we use a patch size of 9×9, which is the typical patch size employed by self-similarity-based NR methods [82, 102, 128]. The resulting PSNR values for $\sigma = 0.05$ are presented in Table 5.1, tables for other noise levels can be consulted in the Appendix.

Table 5.1 shows that S3F-UW typically outperforms S3F-IVW and SW-PSM. The most significant performance differences, with PSNR gains reaching 0.5 dB, are observed in the *baboon, grass, and texture_mosaic* images, which are characterized by predominantly textured content. We attribute the performance degradation of the IVW scheme relative to uniform aggregation to the assumptions of the IVW design. Specifically, by assuming independence and neglecting the correlation between estimates, the SW-PSM framework tends to underestimate the true variance [136]. This leads the IVW to erroneously overweight the smooth component of the S3F model, resulting in a reduction in NR performance. Since similar observations were drawn for other noise levels (see Appendix 5.10.3), in subsequent analysis we adhere to the uniformly weighted variant of S3F.

Figure 5.5 illustrates the PSNR performance of the SW-PSM, BM3D, and S3F-UW methods across four noise levels ($\sigma \in \{0.01, 0.025, 0.05, 0.1\}$) and all the test images. At $\sigma = 0.01$, S3F-UW exhibits a higher median PSNR and a notably tighter interquartile range compared to both the standalone SW-PSM and BM3D. As σ increases, the performance margin between

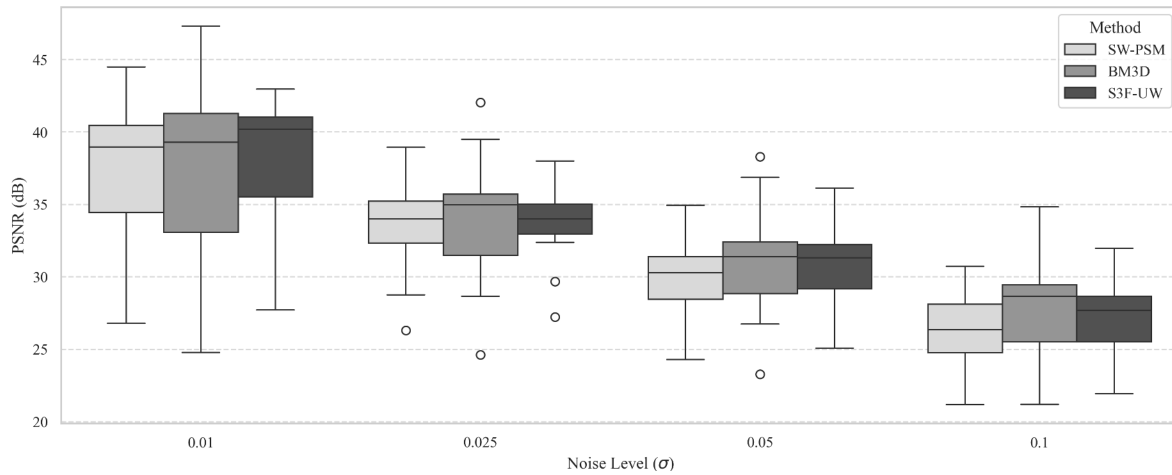

Figure 5.5 Comparative PSNR analysis for $\sigma \in \{0.01, 0.025, 0.05, 0.1\}$. Boxes represent the performance of SW-PSM, BM3D, and the proposed S3F-UW across the image dataset.

the S3F-UW and BM3D narrows, with both methods maintaining similar median values at $\sigma = 0.05$. While the PSNR gains are less pronounced at the highest noise level ($\sigma = 0.1$), S3F-UW maintains a higher lower-quartile and higher median than SW-PSM.

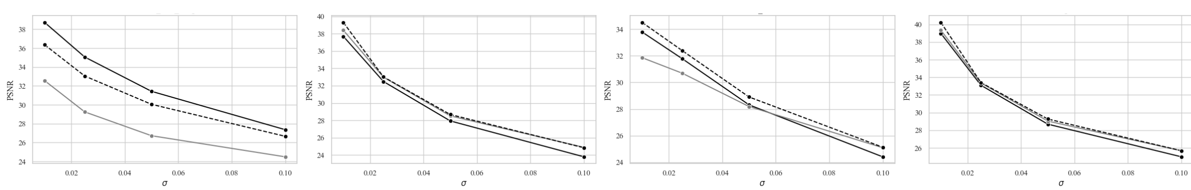

Figure 5.6 PSNR obtained for the SW-PSM (black, solid curve), S3F-UW (black, dashed), and BM3D (grey) models for the (left to right) *disks, baboon, aerial*, and *bridge* images at varying σ.

We present image-specific examples of PSNR comparisons across noise levels in Figure 5.6. The leftmost panel corresponds to *disks*, a synthetic image which adheres strictly to a piecewise-smooth prior but is designed to have low self-similarity (see Appendix 5.10.2 for more details). Here the SW-PSM (solid black) outperforms BM3D (gray), which suffers from the absence of self-similar patches. Conversely, in the heavily textured natural images *baboon*, *aerial*, and *bridge* (panels 2-4), BM3D frequently surpasses the standalone SW-PSM, as the latter struggles to differentiate high-frequency textural details from noise. Across both categories, the S3F model (dashed line) consistently appears as the most robust approach; it avoids the significant performance drops of BM3D in non-redundant smooth regions while matching or exceeding BM3D's performance in textured images. This stability across diverse

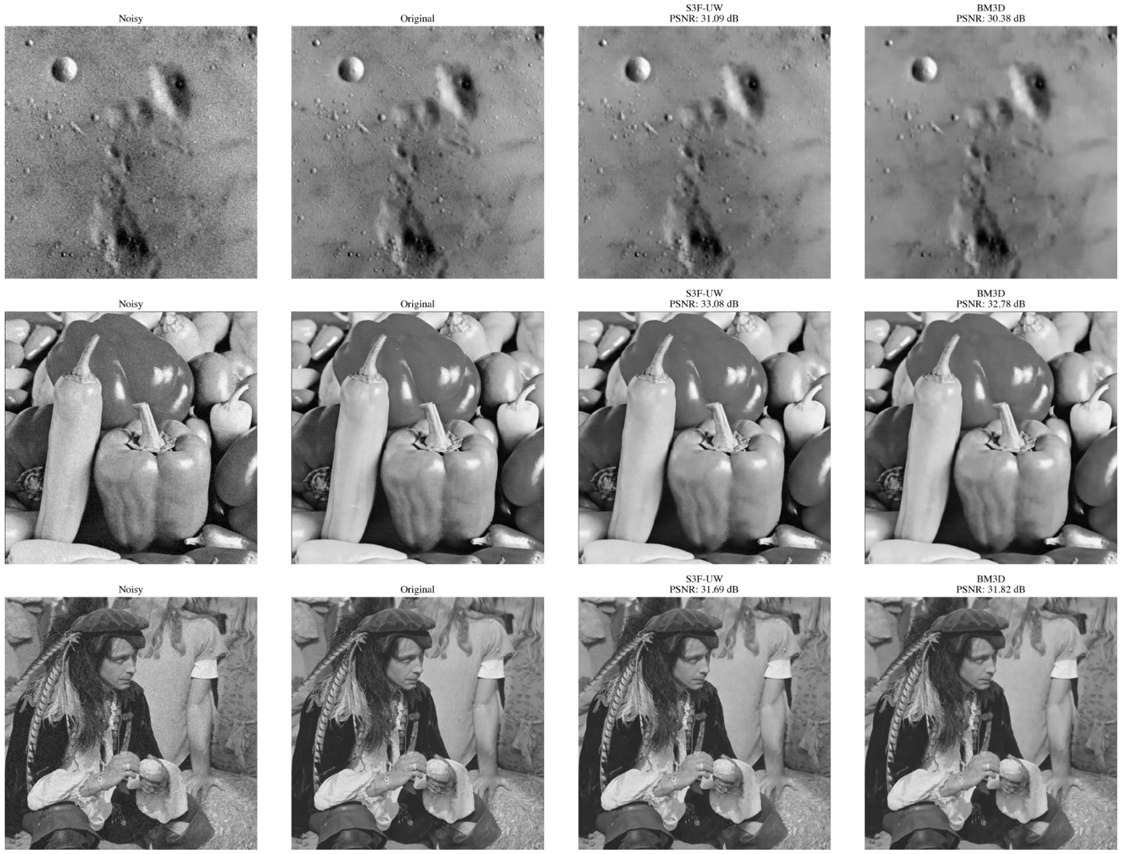

Figure 5.7 From left to right: Noisy image ($\sigma = 0.05$), original image, S3F-UW estimate, and BM3D estimate for *moon* (top, $\hat{\sigma} = 0.053$) *peppers* (center, $\hat{\sigma} = 0.052$), and *man* (bottom, $\hat{\sigma} = 0.048$). The actual noise level is $\sigma = 0.05$.

image characteristics validates the effectiveness of the texture-guided fusion approach.

We encountered multiple instances in which S3F-UW preserves fine-scale features more faithfully than BM3D. Examples of this are provided in the reconstruction of *moon* at $\sigma = 0.05$ shown in the top panel of Figure 5.7; while the BM3D estimate (right) tends to oversmooth small crater textures, the S3F-UW approach (left) retains them.

A detailed illustration of superior texture preservation via the S3F-UW approach is provided for the *peppers* image reconstruction at $\sigma = 0.05$ (Figure 5.7, bottom), with zoomed-in regions displayed in Figure 5.8. Notably, the S3F-UW estimate (Figure 5.8, top row) recovers fine-scale details of the original image (center row) with greater precision than BM3D (bottom row). Specifically, the BM3D estimate exhibits pronounced oversmoothing artifacts in the first four panels from the left; in contrast, these fine-scale details are preserved with greater fidelity in the S3F-UW result. Furthermore, BM3D tends to suppress the white spots in the original image and introduces checkered artifacts (Figure 5.7, rightmost panel); conversely,

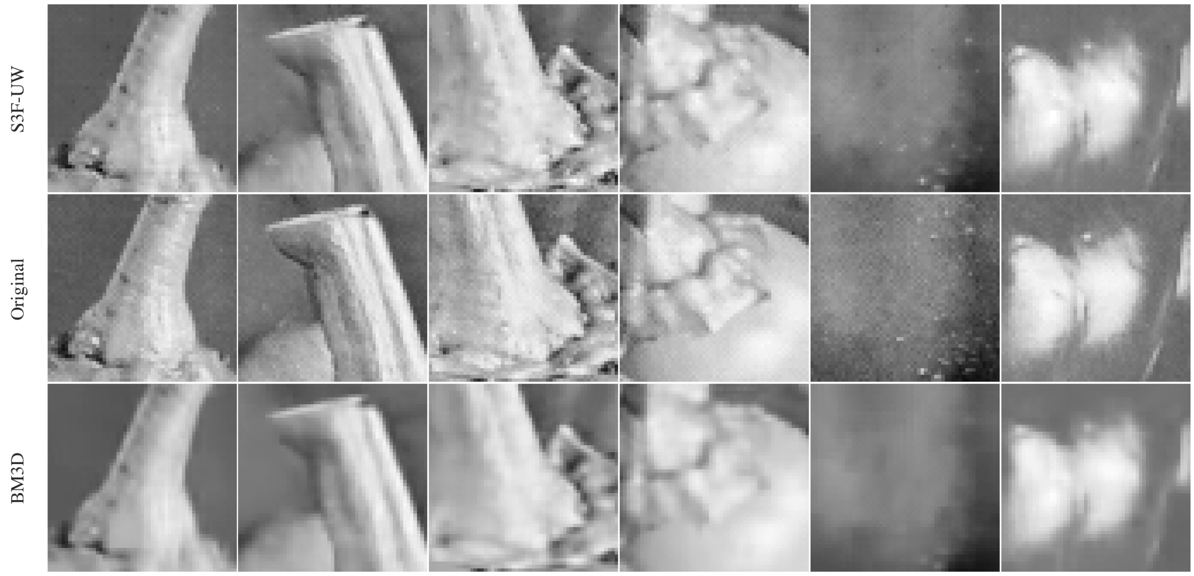

Figure 5.8 Zoomed fragments of *peppers*. Rows: original image (center), S3F-UW estimate (top), and BM3D estimate (bottom); the latter two were obtained from a noisy observation at $\sigma = 0.05$ with $\hat{\sigma} = 0.052$.

the S3F-UW method better preserves these features without introducing discernible visual artifacts.

Even when BM3D marginally outperforms S3F-UW in global PSNR (e.g., Figure 5.7, bottom), the latter often preserves fine-scale details with greater fidelity. As demonstrated in Figure 5.9, textures in the rings, hair, and grass (panels 1 - 3) are more accurately recovered by S3F-UW. Furthermore, unlike BM3D, the S3F-UW estimate remains free of characteristic checkered artifacts (rightmost panel).

As shown in Figure 5.10, which illustrates the execution time across the test images at $\sigma = 0.05$, the S3F-UW model reduces processing time by approximately 50% relative to BM3D. We attribute the reduced execution to the fast optimization techniques in the SW-PSM framework, specifically, sliding-window processing, shortest-path optimization and fast curve-fitting. Furthermore, by restricting the low-rank estimation to a subset of the image, the S3F-UW approach benefits from an additional reduction in execution time in comparison to patch based methods that are applied to the entire image domain.

5.8.2 Noise Reduction on Raw Images

We evaluated the NR performance of the S3F-UW method on raw natural images. Unlike the synthetic AWG noise reduction scenario presented in the previous section, we treat the

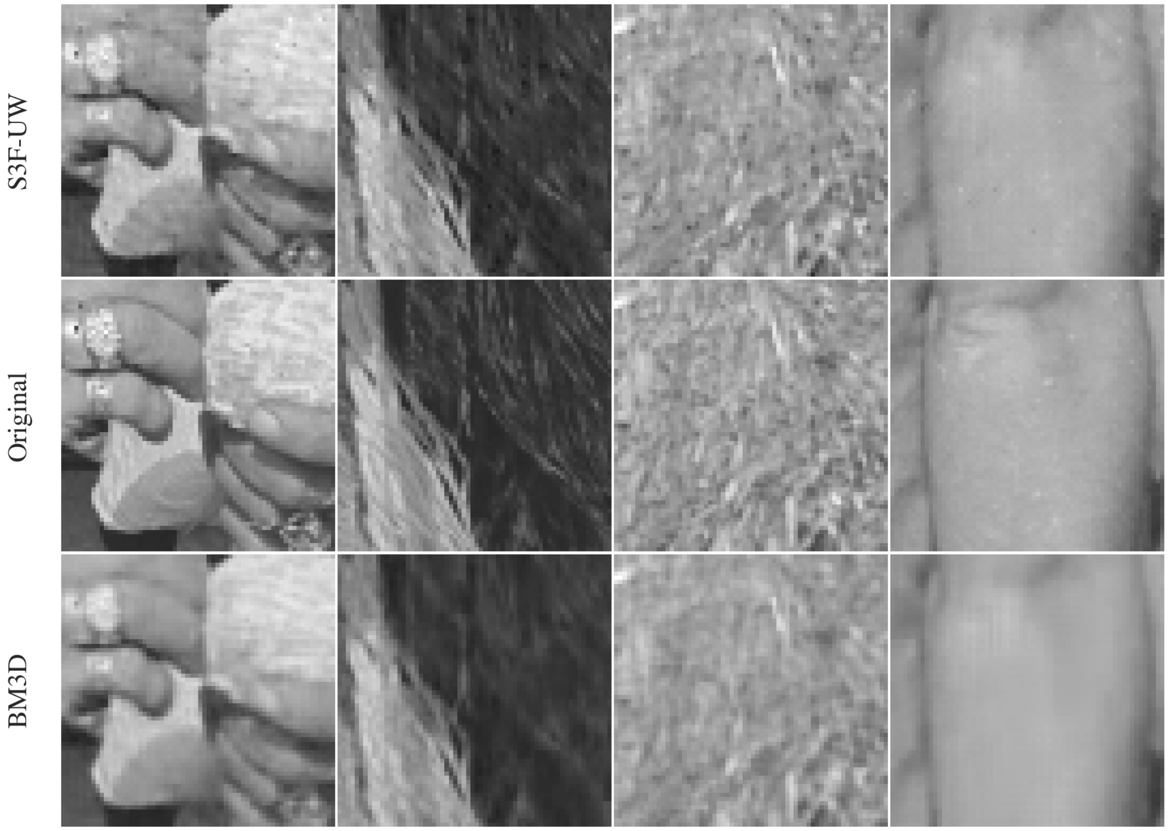

Figure 5.9 Zoomed fragments of *man*. Rows: original image (center), S3F-UW estimate (top), and BM3D estimate (bottom); the latter two were obtained from a noisy observation at $\sigma = 0.05$ with $\hat{\sigma} = 0.048$.

raw images as inherently noisy. We process each raw image using the S3F-UW method with the noise-level estimate in Section 5.4.

Figure 5.11 shows a comparison of the S3F-UW method applied to raw image fragments extracted from *baboon* ($\hat{\sigma} = 0.0085$), *Barbara* ($\hat{\sigma} = 0.0084$), *couple* ($\hat{\sigma} = 0.0075$), and *peppers* ($\hat{\sigma} = 0.0073$). The top row displays the original signal, while the bottom row shows the corresponding S3F-UW estimate. Notably, the S3F-UW estimate reduces noise from smooth regions (white texture in *baboon*, the skin in *Barbara*, and the smooth regions in *couple*, and *peppers*) while preserving the detail of textured regions.

The preservation of fine-scale details alongside smooth textures is further illustrated by the zoomed-in regions of the *Barbara* image in Figure 5.12. corresponding to the S3F-UW (top row) image estimate of a raw version of the *Barbara* image (bottom row). Comparing the raw image (bottom row) with the S3F-UW reconstruction (top row) highlights the model's ability to recover the characteristic striped textures. Simultaneously, the noise within the

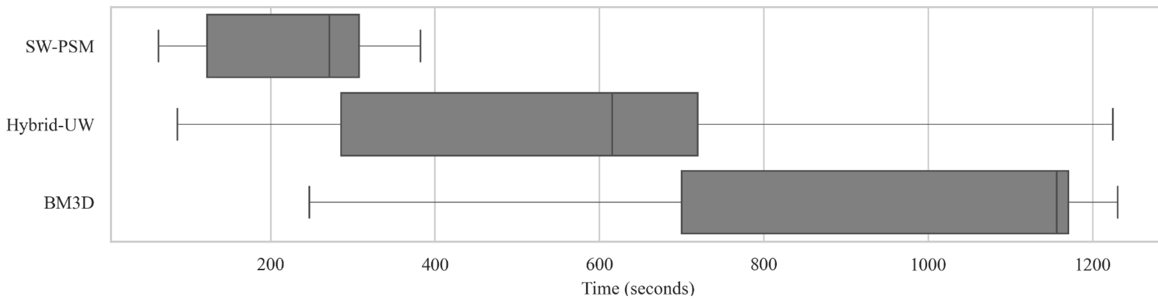

Figure 5.10 Execution time in seconds for SW-PSM, S3F-UW, and BM3D at a noise level of $\sigma = 0.05$. The box plots illustrate the distribution of computational time across the test images.

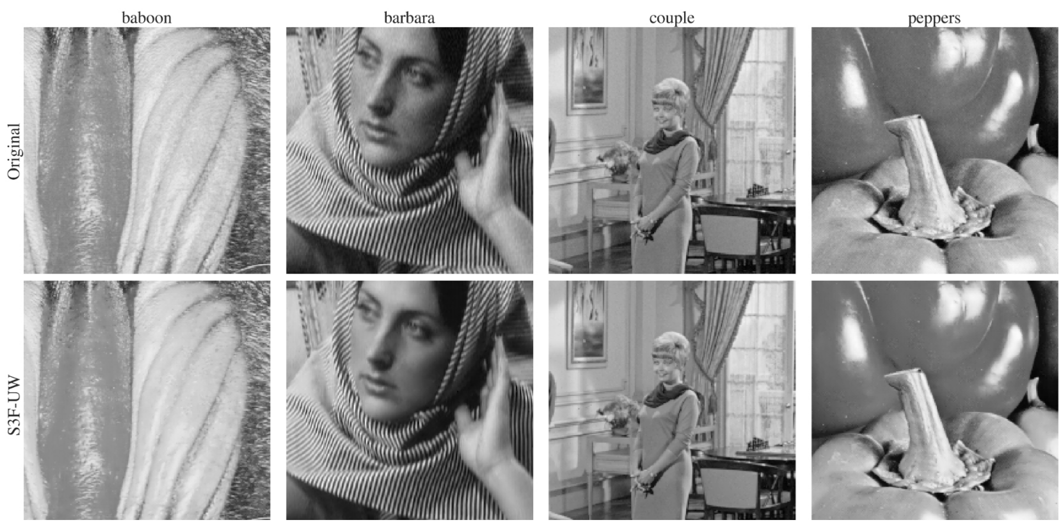

Figure 5.11 **Top**: Fragments of raw test images. **Bottom:** Corresponding S3F-UW image estimates.

skin tones is significantly attenuated We may therefore conclude that the S3F-UW estimator effectively reduces noise in smooth regions without blurring edges, and also performs well in complex textured regions.

5.9 Concluding Remarks

We introduced a novel image noise reduction framework that adaptively integrates piecewise smooth and self-similarity-based approaches by leveraging local textural content. By dynamically balancing these two strategies, the proposed method enhances noise reduction across diverse image regions, using the SW-PSM framework not only for initial estimation but as a tool to identify non-smooth regions. Furthermore, we refined the self-similarity

Figure 5.12 **Top**: Details of the S3F-UW image estimate. **Bottom:** Corresponding fragments of the raw image.

component through a principled AIC-based selection of regression components and the application of Inverse Variance Weighting to combine multiple reduced-rank models. Finally, the global image estimate is computed as the mean of the piecewise smooth and self-similarity estimates, ensuring a balanced reconstruction that preserves high-frequency details without compromising the integrity of homogeneous regions.

Self-similarity methods often face a trade-off between under-smoothing and over-smoothing, driven by patch size [137]. Small patches recover fine details but overfit in smooth regions, while large patches offer robust noise reduction at the expense of structural features. Multiscale variants address this but introduce significantly parameter-heavy models. Our method circumvents these problems by using a multiscale piecewise smooth model for homogeneous regions, reserving a single-scale self-similarity model to refine fine-textured details.

Compared to learning-based methods which require intensive offline training and massive datasets, our method only requires the noisy image to operate, making it suitable for applications where data variability is high or training data is scarce.

It was shown that S3F-UW performs consistently across different noise levels and images with particularly strong performance at smaller noise levels, where it was observed that S3F-UW outperforms BM3D in PSNR in multiple instances while providing a reduction in execution time. Qualitative assessments showed an improved preservation of detailed textures. The performance of the proposed method across different image types demonstrates its robustness and suggests that it may be useful for other real-world applications.

Future developments could explore alternative data-driven weighting schemes. Specifically, Stein's Unbiased Risk Estimate (SURE) could be employed as a self-supervised tuning strategy to optimize the mixing weights without requiring noise-free data [61]. Additional research could explore extensions to other noise distributions and non-stationary noise.

The results obtained in this work reinforce the idea that principled statistical methods, when properly extended and engineered, can remain competitive, and even preferable, in domains where data variability, interpretability, and availability are ongoing challenges.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

The author A. Saucier acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grants RGPIN-115965.

5.10 Appendix

5.10.1 Stability of the Noise Level Estimate

We evaluate the performance of (5.1) on noisy linear signals y with length $N \in \{2^k, k = 6, 7, ..., 11\}$. These signals were corrupted with AWGN with $\sigma \in \{0.01, 0.02, ..., 0.3\}$, spanning low-to-moderately large noise level regimes. To quantify the degree in which the noise corruption affects the pure signal x, we compute the signal-to-noise ratio (SNR), given by $\mathrm{SNR} = \frac{1}{\sigma}\left(\sum_{i=0}^{N} x[i]^2\right)^{1/2}$. For each pair (N, σ), we generate 300 noisy linear signals denoted by y. Each y is characterized by a random slope sampled uniformly from $[-100, 100] \subset \mathbb{R}$, and an intercept set to zero without loss of generality. Subsequently, we compute the noise level estimate $\hat{\sigma}_0$ for each y. The estimator's performance was assessed using the absolute percentage error $\epsilon = 100\frac{|\hat{\sigma}_0 - \sigma|}{\sigma}$. The results of this experiment are displayed in Figure 5.13 in terms of the average ϵ across experimental realizations, denoted by $\bar{\epsilon}$. Each gray curve in the left panel of Figure 5.13 corresponds to $\bar{\epsilon}$ as a function of N for a fixed σ. The black curve reports the average ϵ across noise levels as a function of N. The right panel displays the standard deviation $\mathrm{STD}(\bar{\epsilon}) = \mathrm{STD}(\epsilon)/\sqrt{300}$.

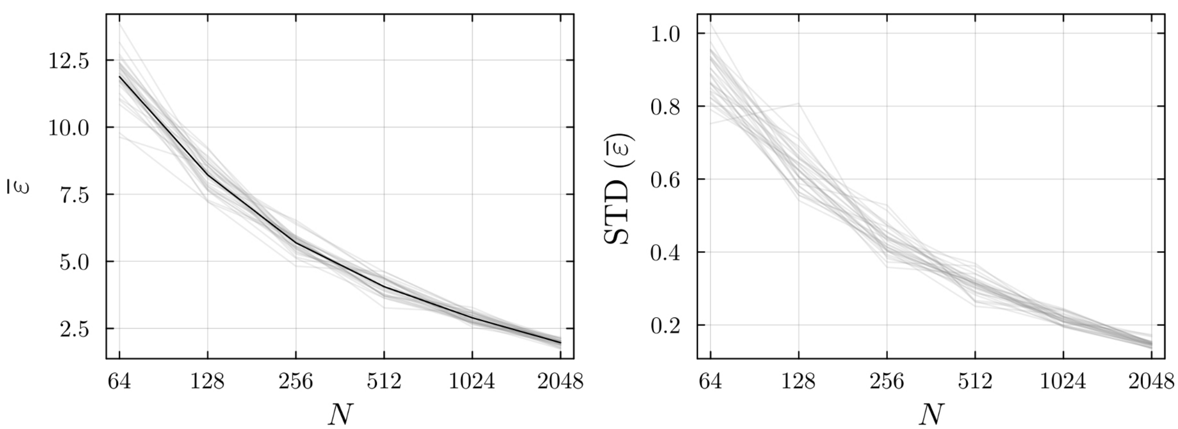

Figure 5.13 **Left:** Mean absolute percentage error $\bar{\epsilon}$ as a function of sample size N. Each gray curve corresponds to a fixed noise level $\sigma \in \{0.01, 0.02, \ldots, 0.30\}$. Black curve: Average $\bar{\epsilon}$ over noise levels. **Right:** Standard deviation of ϵ as a function of N, reported for the same set of noise levels.

Figure 5.13 demonstrates that accuracy and stability improve with increasing N : at $N = 64$, the estimator exhibits relatively large errors ($\bar{\epsilon} \approx 12\%$) and variability ($\mathrm{STD}(\bar{\epsilon}) \approx 8\%$), whereas at $N = 2048$, $\bar{\epsilon}$ falls below 3%. This trend confirms estimator consistency, as larger samples reduce both bias and variance. Moreover, convergence of the gray curves toward the mean trajectory indicates decreasing sensitivity to σ at larger N.

5.10.2 Disks Image

The synthetic test image is a 256×256 grayscale composition designed to evaluate piecewise-smooth modeling in the absence of self-similarity. The image is populated with a stochastic distribution of disks with radii R ranging from $R_{min} = 4$ to $R_{max} = 25.6$. The quantity of disks for a given radius follows a fractal power-law distribution, $N(R) = kR^{-d}$, where the constants k and d are calibrated to yield exactly $N_{min} = 1$ disk at the maximum radius and $N_{max} = 800$ disks at the minimum radius. Each disk is centered randomly within the image domain, with internal pixel intensities defined by a local linear gradient $A(i,j) = a(i-i_0)+b(j-j_0)$, where the slopes a and b are drawn from a uniform distribution $\mathcal{U}(-1,1)$. Finally, the composite image values are normalized to the unit interval $[0,1]$, resulting in a complex, overlapping piecewise-smooth structure that lacks the repeating patterns typical of natural textures. An example of this image and its image estimates is displayed in Figure 5.14.

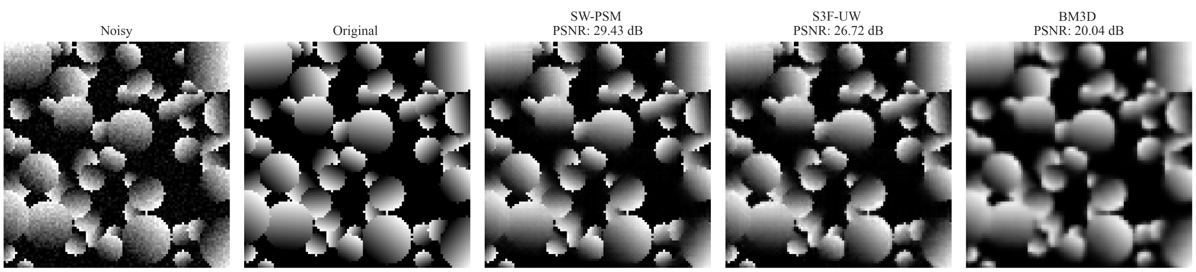

Figure 5.14 First and second panels: Noisy ($\sigma = 0.05$) and original *disks* image. Third to fifth panels: SW-PSM, S3F-(UW) with $q = 6$ and 9×9 patches, and BM3D estimates.

5.10.3 Additional PSNR Tables

Table 5.2 PSNR results for various denoising methods at Noise Level = 0.01.

Signal	SW-PSM	S3F-UW	S3F-IVW	BM3D
airplane	43.13	42.57	42.39	**43.62**
aerial	33.79	**34.50**	34.12	31.86
baboon	37.69	**39.30**	37.81	38.40
Barbara	39.59	40.63	39.86	**41.26**
boat	38.94	**40.16**	39.21	39.26
bridge	38.98	**40.23**	39.01	39.35
chemical_plant	39.25	40.27	39.33	**40.42**
couple	40.24	**40.98**	40.20	40.53
grass	31.32	**32.20**	31.69	30.84
house	44.49	42.97	43.41	**47.32**
man	41.03	41.21	40.91	**41.34**
moonsurface	34.23	**34.75**	34.30	31.94
peppers	38.57	**39.53**	38.77	37.97
plane	41.89	41.69	41.53	**42.56**
sand	34.51	**35.78**	34.98	33.46
texmos1	26.79	**27.73**	27.14	24.79

Table 5.3 PSNR results for various denoising methods at Noise Level = 0.025.

Signal	SW-PSM	S3F-UW	S3F-IVW	BM3D
airplane	38.71	38.00	37.51	**39.50**
aerial	31.78	**32.39**	32.14	30.69
baboon	32.48	**33.01**	32.46	32.99
Barbara	34.36	34.27	34.06	**35.96**
boat	34.76	34.83	34.53	**35.33**
bridge	33.10	**33.40**	32.97	33.36
chemical_plant	33.63	33.72	33.52	**34.93**
couple	34.90	34.83	34.55	**35.35**
grass	28.77	**29.69**	29.15	28.66
house	38.96	36.89	36.29	**42.04**
man	35.15	34.49	34.31	**35.62**
moonsurface	32.88	**33.18**	32.95	31.33
peppers	35.50	**35.60**	35.30	35.01
plane	36.46	35.75	35.57	**37.48**
sand	31.90	**32.85**	32.28	31.55
texmos1	26.32	**27.25**	26.67	24.64

Table 5.4 PSNR results for various denoising methods at Noise Level = 0.1.

Signal	SW-PSM	S3F-UW	S3F-IVW	BM3D
airplane	30.23	31.23	31.14	**33.65**
aerial	24.42	**25.14**	24.73	25.10
baboon	23.79	24.84	24.17	**24.88**
Barbara	25.23	27.38	26.13	**28.81**
boat	27.35	28.20	27.85	**29.13**
bridge	24.99	25.67	25.32	**25.68**
couple	27.03	28.00	27.52	**28.76**
chemical_plant	25.68	26.74	26.16	**27.45**
grass	21.90	22.80	22.19	**23.07**
house	30.71	31.98	31.91	**34.84**
man	27.29	28.06	27.75	**28.61**
moonsurface	28.04	28.47	28.42	**28.74**
peppers	28.73	29.54	29.32	**30.46**
plane	28.29	29.30	29.03	**30.42**
sand	24.88	25.64	25.15	**25.88**
texmos1	21.19	**21.96**	21.48	21.22

CHAPTER 6 GENERAL DISCUSSION

6.1 Gaussian noise reduction in one-dimensional piecewise smooth signals

This work addressed the long-standing problem of signal noise reduction. The proposed approach builds upon the PSM procedure, a novel formulation for piecewise regression that combines model selection and shortest path optimization. Compared to other piecewise regression methods in the literature [106, 138], the PSM method is designed to handle the case where multiple functional forms are considered for the individual regression candidates.

For one-dimensional signals, the methods and results presented in Chapter 4 provide strong evidence that the SW-PSM method successfully reduces noise across a wide variety of piecewise smooth signals. In particular, it outperforms oracle-optimized versions of classical noise reduction methods by preserving both smooth and fast-varying regions with greater detail. This performance gain is attributed to the accurate piecewise curve-fitting solution provided by locally optimal piecewise-smooth models, combined with the inverse variance weighting framework. The results on raw ECG signals further highlight the potential of the SW-PSM method for real-world applications, where the exact noise level is unknown. One noted limitation is that the SW-PSM method performs less well in regions where the noise-free signal exhibits fast variations, that is, where the signal deviates from the piecewise-smooth assumption. Interesting extensions to improve the SW-PSM method include modifying the PSM method to allow solutions that admit overlapping and non-overlapping supports, and integrating small-sample correction terms in the Information criterion cost function.

6.2 Gaussian noise reduction in natural images

For images, the S3F method proposed in Chapter 5 yields results that are comparable to, or exceed, those of the established BM3D method on several test cases. By restricting the reduced-rank regression model to a subset of pixels in the noisy image, the S3F method reduces computing time to roughly half that required by BM3D. The results in Chapter 5 further indicate that the uniform weighting scheme of SW-PSM combined with reduced-rank regression outperformed the inverse variance weighting scheme, suggesting that more refined model weighting strategies could lead to additional improvements in noise reduction quality.

Although recently developed image noise reduction methods are typically designed to handle both Gaussian and non-Gaussian noise, the S3F method offers a principled and interpretable alternative that can be extended to other noise distributions, avoids computationally demanding calculations, and delivers results comparable to those of classical Gaussian noise reduction benchmarks.

The general objectives guiding this thesis, i.e., designing high-quality, applicable, and generalizable Gaussian noise reduction frameworks, were successfully met. In particular, the SW-PSM method introduced in Chapter 4 yields excellent results for Gaussian noise reduction. Image noise reduction poses a greater challenge, as natural images typically contain regions that are not adequately captured by the piecewise-smooth hypothesis. To address this, the S3F methodology combines the piecewise-smooth and self-similarity hypotheses based on the textural context of the image, producing a model that balances image adaptation, computational efficiency, and high-quality noise reduction.

CHAPTER 7 CONCLUSION

7.1 Summary of Works

The proposed methodologies are motivated by the observation that methods that reduce Gaussian noise without introducing artifacts typically rely heavily on parameter tuning. Parameter tuning is a non-trivial task in the sense that it requires optimization procedures which not necessarily guarantee successful NR across signals coming from disparate domains. The work contained in this thesis represents an effort to reconcile high-quality noise reduction, applicability across domains, and efficient computation.

We proposed PSM as a theoretical framework for data-adaptive Gaussian noise reduction. For 1D signals, we extend PSM to SW-PSM, in which multiple local PSM estimates are combined to increase the accuracy of the signal estimate. The results of SW-PSM demonstrate substantially better performance than classical NR methods. Particularly, the SW-PSM method relying only on noisy data was shown to outperform oracle-optimized versions of classical and state-of-the-art NR methods for 1D signals at low-to-moderate noise levels, while maintaining a comparable at higher noise-levels. This was achieved for a large family of signals and with minimal manual parameter tuning. Furthermore, efficient computation was also achieved, taking less than one second to reduce noise from a signal of 512 points. This was done without relying on parallel computation, indicating that there is still room for improving computing speed. This factors show that SW-PSM represents a high-impact contribution to the field of AWGN reduction in 1D signals.

7.2 Concluding Remarks

We introduced the Smoothness and Self-Similarity Fusion (S3F) framework, a training-free method for additive white Gaussian noise reduction in images that adaptively integrates two complementary modeling strategies based on local textural content. The framework operates in three coupled stages: a noise-level estimator derives $\hat{\sigma}$ directly from the noisy image; the SW-PSM method is applied via four directional vectorizations, producing a pixel-wise estimate $\hat{X}_{\mathrm{PS}}$ and a pointwise variance map that simultaneously drives IVW aggregation and partitions the image into smooth ($\mathcal{D}_S$) and textured ($\mathcal{D}_T$) regions; and a Reduced-Rank Regression (RR) estimate $\hat{X}_{\mathrm{RR}}$ is computed for pixels in $\mathcal{D}_T$ via AIC-based rank selection and Multi-Model Inference over competitive rank candidates. The final estimate retains $\hat{X}_{\mathrm{PS}}$ in smooth regions and combines both estimates via uniform weighting in textured regions, a scheme found empirically superior to inverse-variance weighting.

Experimental results demonstrate that S3F-UW performs consistently across noise levels and image types, outperforming BM3D in PSNR across multiple images at lower noise levels while requiring approximately 50% of its execution time, and avoiding the checkered artifacts characteristic of BM3D. The method's advantage is especially pronounced in images with low self-similarity, where BM3D degrades significantly.

Unlike supervised learning approaches that depend on large, clean datasets, or classical filters requiring manual tuning, the proposed methods operate directly on noisy observations and remain competitive in terms of NR performance while introducing a minimal amount of artifacts.

7.3 Limitations

Despite its strengths, the proposed methodology is subject to a number of limitations. The current framework assumes additive Gaussian noise with constant variance. While this is a common assumption in many signal processing applications, it restricts applicability in scenarios where noise is signal-dependent or varies spatially, such as in computed tomography imaging or low-light photography.

Additionally, the application of SW-PSM to images relies on a predefined set of vectorization schemes. Although effective in practice, this strategy may inadequately capture local signal structures in images with highly-textured areas, potentially failing to reduce noise.

Another way to improve the PSM method would be to allow overlapping intervals in the covering definition. This has the potential of improving the signal estimation on angular points.

While the current implementation of S3F yields good results using a uniform weighting scheme for the SW-PSM and Reduced Rank Regression (RR) models, the development of an improved weighting scheme remains as an open challenge.

Finally, while the computational efficiency of the method is acceptable for most static image applications, scaling to large-scale image datasets or to real-time video processing environments may require further optimization.

7.4 Future Research

There are several promising directions for future investigation. Extending the current methods to handle non-stationary Gaussian noise, or other noise models such as Poisson and

Poisson-Gaussian, widen the applicability of the proposed methods.

Including data-adaptive vectorization strategies could enhance the ability of SW-PSM to represent and reduce noise from complex textures more effectively.

The optimal model fusion problem in S3F may be adequately tackled by minimizing the Stein's unbiased risk estimator (SURE) to select the model's weights. We also think that the S3F model may benefit from a more adaptive design, where the patch size and the parameter q controlling the number of similar patches could be selected according to the image structure.

In summary, this thesis lays a principled and practical foundation for adaptive statistical noise reduction. By offering adaptability, interpretability, and competitive performance, the proposed methods make a meaningful contribution to both the theoretical understanding and practical deployment of noise reduction techniques in signal and image processing.

REFERENCES

[1] M. Korki, N. Hosseinzadeh, and T. Moazzeni, “Performance evaluation of a narrowband power line communication for smart grid with noise reduction technique,” *IEEE Transactions on consumer electronics*, vol. 57, no. 4, pp. 1598–1606, 2012.

[2] P. Li and T.-M. Laleg-Kirati, “Signal denoising based on the schrödinger operator’s eigenspectrum and a curvature constraint,” *IET Signal Processing*, vol. 15, no. 3, pp. 195–206, 2021.

[3] G. Li, W. Bu, and H. Yang, “Noise reduction method for ship radiated noise signal based on modified uniform phase empirical mode decomposition,” *Measurement*, vol. 227, p. 114193, 2024.

[4] S. Kantamaneni, A. Charles, and T. R. Babu, “Speech enhancement with noise estimation and filtration using deep learning models,” *Theoretical Computer Science*, vol. 941, pp. 14–28, 2023.

[5] R. Serizel, M. Moonen, J. Wouters, and S. H. Jensen, “A speech distortion weighting based approach to integrated active noise control and noise reduction in hearing aids,” *Signal processing*, vol. 93, no. 9, pp. 2440–2452, 2013.

[6] M. Lozano, J. A. Fiz, and R. Jané, “Performance evaluation of the hilbert–huang transform for respiratory sound analysis and its application to continuous adventitious sound characterization,” *Signal Processing*, vol. 120, pp. 99–116, 2016.

[7] M. Z. U. Rahman, R. A. Shaik, and D. R. K. Reddy, “Efficient sign based normalized adaptive filtering techniques for cancelation of artifacts in ecg signals: Application to wireless biotelemetry,” *Signal processing*, vol. 91, no. 2, pp. 225–239, 2011.

[8] J. G. Proakis and M. Salehi, *Digital communications.* McGraw-hill, 2008.

[9] S. Konishi and G. Kitagawa, *Information criteria and statistical modeling.* Springer Science & Business Media, 2007.

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

[11] G. Schwarz, “Estimating the dimension of a model,” *The annals of statistics*, pp. 461–464, 1978.

[12] P. Jain and V. Tyagi, "A survey of edge-preserving image denoising methods," *Information Systems Frontiers*, vol. 18, pp. 159–170, 2016.

[13] C. Rohlfs, "Generalization in neural networks: A broad survey," *Neurocomputing*, vol. 611, p. 128701, 2025.

[14] J.-L. Starck and F. Murtagh, "Astronomical image and signal processing: looking at noise, information and scale," *IEEE Signal Processing Magazine*, vol. 18, no. 2, pp. 30–40, 2001.

[15] A. Vojtekova, M. Lieu, I. Valtchanov, B. Altieri, L. Old, Q. Chen, and F. Hroch, "Learning to denoise astronomical images with u-nets," *Monthly Notices of the Royal Astronomical Society*, vol. 503, no. 3, pp. 3204–3215, 2021.

[16] J. N. Hird and G. J. McDermid, "Noise reduction of ndvi time series: An empirical comparison of selected techniques," *Remote Sensing of Environment*, vol. 113, no. 1, pp. 248–258, 2009.

[17] L. A. Santos, K. R. Ferreira, G. Camara, M. C. Picoli, and R. E. Simoes, "Quality control and class noise reduction of satellite image time series," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 75–88, 2021.

[18] A. Van der Ziel, "History of noise research," *Advances in Electronics and Electron Physics*, vol. 50, pp. 351–409, 1980.

[19] C. A. Mack, "Shot noise: a 100-year history, with applications to lithography," *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 17, no. 4, pp. 041 002–041 002, 2018.

[20] M. Thompson, *Intuitive analog circuit design.* Elsevier, 2013.

[21] J. M. Sancho, M. San Miguel, S. Katz, and J. Gunton, "Analytical and numerical studies of multiplicative noise," *Physical Review A*, vol. 26, no. 3, p. 1589, 1982.

[22] X. Yu and J. Li, "Adaptive kalman filtering for recursive both additive noise and multiplicative noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 3, pp. 1634–1649, 2021.

[23] G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise," *SIAM journal on applied mathematics*, vol. 68, no. 4, pp. 925–946, 2008.

[24] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal processing*, vol. 81, no. 11, pp. 2403–2418, 2001.

[25] ——, "Microphone array post-filtering for non-stationary noise suppression," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 2002, pp. I–901.

[26] W. Thanhikam, Y. Kamamori, A. Kawamura, and Y. Iiguni, "Stationary and nonstationary wide-band noise reduction using zero phase signal," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 95, no. 5, pp. 843–852, 2012.

[27] Y. Zuo, W. Yao, Y. Zeng, J. Xie, Y. Fang, Y. Huang, and W. Jiang, "Cfnet: conditional filter learning with dynamic noise estimation for real image denoising," *Knowledge-Based Systems*, vol. 284, p. 111320, 2024.

[28] D. Klionskiy, M. Kupriyanov, and D. Kaplun, "Signal denoising based on empirical mode decomposition," *Journal of Vibroengineering*, vol. 19, no. 7, pp. 5560–5570, 2017.

[29] Y. Zhang and M. Mondelli, "Matrix denoising with doubly heteroscedastic noise: Fundamental limits and optimal spectral methods," *arXiv preprint arXiv:2405.13912*, 2024.

[30] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

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

[32] A. Shah, J. I. Bangash, A. W. Khan, I. Ahmed, A. Khan, A. Khan, and A. Khan, "Comparative analysis of median filter and its variants for removal of impulse noise from gray scale images," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 3, pp. 505–519, 2022.

[33] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques," *Visual computing for industry, biomedicine, and art*, vol. 2, no. 1, p. 7, 2019.

[34] A. Dixit and P. Sharma, "A comparative study of wavelet thresholding for image denoising," *IJ Image, Graphics and Signal Processing*, vol. 12, pp. 39–46, 2014.

[35] A. Halidou, Y. Mohamadou, A. A. A. Ari, and E. J. G. Zacko, "Review of wavelet denoising algorithms," *Multimedia Tools and Applications*, vol. 82, no. 27, pp. 41 539–41 569, 2023.

[36] Z. Wang, F. Wan, C. M. Wong, and L. Zhang, "Adaptive fourier decomposition based ecg denoising," *Computers in Biology and Medicine*, vol. 77, pp. 195–205, 2016.

[37] M.-Y. Zhai, "Seismic data denoising based on the fractional fourier transformation," *Journal of Applied Geophysics*, vol. 109, pp. 62–70, 2014.

[38] A. C. To, J. R. Moore, and S. D. Glaser, "Wavelet denoising techniques with applications to experimental geophysical data," *Signal processing*, vol. 89, no. 2, pp. 144–160, 2009.

[39] M. A. Kutay and H. M. Ozaktas, "Optimal image restoration with the fractional fourier transform," *Journal of the Optical Society of America A*, vol. 15, no. 4, pp. 825–833, 1998.

[40] "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE transactions on medical imaging*, vol. 22, no. 3, pp. 323–331, 2003.

[41] X. Wang, "Moving window-based double haar wavelet transform for image processing," *IEEE Transactions on image processing*, vol. 15, no. 9, pp. 2771–2779, 2006.

[42] J. Varghese, S. Subash, and N. Tairan, "Fourier transform-based windowed adaptive switching minimum filter for reducing periodic noise from digital images," *IET image processing*, vol. 10, no. 9, pp. 646–656, 2016.

[43] S. Mallat, *A wavelet tour of signal processing.* Elsevier, 1999.

[44] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the american statistical association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[45] ——, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[46] I. Y. Novikov and S. B. Stechkin, "Basic wavelet theory," *Russian Mathematical Surveys*, vol. 53, no. 6, p. 1159, 1998.

[47] D. L. Donoho, "De-noising by soft-thresholding," *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 2002.

[48] F. M. Bayer, A. J. Kozakevicius, and R. J. Cintra, "An iterative wavelet threshold for signal denoising," *Signal Processing*, vol. 162, pp. 10–20, 2019.

[49] Y. Ding and I. W. Selesnick, "Artifact-free wavelet denoising: Non-convex sparse regularization, convex optimization," *IEEE signal processing letters*, vol. 22, no. 9, pp. 1364–1368, 2015.

[50] N. ur Rehman, S. Z. Abbas, A. Asif, A. Javed, K. Naveed, and D. P. Mandic, "Translation invariant multi-scale signal denoising based on goodness-of-fit tests," *Signal Processing*, vol. 131, pp. 220–234, 2017.

[51] V. Strela, P. N. Heller, G. Strang, P. Topiwala, and C. Heil, "The application of multiwavelet filterbanks to image processing," *IEEE Transactions on image processing*, vol. 8, no. 4, pp. 548–563, 1999.

[52] B. Subramanian, A. Ramasamy, K. Rangasamy *et al.*, "Performance comparison of wavelet and multiwavelet denoising methods for an electrocardiogram signal," *Journal of Applied Mathematics*, 2014.

[53] K. Naveed, M. T. Akhtar, M. F. Siddiqui, and N. ur Rehman, "A statistical approach to signal denoising based on data-driven multiscale representation," *Digital Signal Processing*, vol. 108, p. 102896, 2021.

[54] G. Tang, X. Yan, and X. Wang, "Chaotic signal denoising based on adaptive smoothing multiscale morphological filtering," *Complexity*, vol. 2020, pp. 1–14, 2020.

[55] L. Duan, L. Yang, and Y. Guo, "Paramps: Convolutional neural networks based on tensor decomposition for heart sound signal analysis and cardiovascular disease diagnosis," *Signal Processing*, vol. 227, p. 109716, 2025.

[56] W. Zhu, S. M. Mousavi, and G. C. Beroza, "Seismic signal denoising and decomposition using deep neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9476–9488, 2019.

[57] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[58] M. El Helou and S. Süsstrunk, "Blind universal bayesian image denoising with gaussian noise level learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 4885–4897, 2020.

[59] P. C. Tripathi and S. Bag, "Cnn-dmri: a convolutional neural network for denoising of magnetic resonance images," *Pattern Recognition Letters*, vol. 135, pp. 57–63, 2020.

[60] J. Schaefferkoetter, J. Yan, C. Ortega, A. Sertic, E. Lechtman, Y. Eshet, U. Metser, and P. Veit-Haibach, "Convolutional neural networks for improving image quality with noisy pet data," *EJNMMI research*, vol. 10, pp. 1–11, 2020.

[61] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *International conference on machine learning.* PMLR, 2019, pp. 524–533.

[62] M. Elad, B. Kawar, and G. Vaksman, "Image denoising: The deep learning revolution and beyond - a survey paper," *SIAM Journal on Imaging Sciences*, vol. 16, no. 3, pp. 1594–1654, 2023.

[63] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[64] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2043–2052.

[65] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2self with dropout: Learning self-supervised denoising from single image," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1890–1898.

[66] R. Wen, H. Yuan, D. Ni, W. Xiao, and Y. Wu, "From denoising training to test-time adaptation: Enhancing domain generalization for medical image segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 464–474.

[67] Y. Mansour, X. Zhong, S. Caglar, and R. Heckel, "Ttt-mim: Test-time training with masked image modeling for denoising distribution shifts," in *European Conference on Computer Vision.* Springer, 2024, pp. 341–357.

[68] T. Klug, D. Atik, and R. Heckel, "Analyzing the sample complexity of self-supervised image reconstruction methods," vol. 36, 2023, pp. 65 869–65 893.

[69] J. Lequyer, R. Philip, A. Sharma, W.-H. Hsu, and L. Pelletier, "A fast blind zero-shot denoiser," *Nature Machine Intelligence*, vol. 4, no. 11, pp. 953–963, 2022.

[70] C. Zhang and K. S. Yen, "An accelerated unsupervised denoising model based on risk estimation and sparsity variational techniques," *Digital Signal Processing*, p. 105317, 2025.

[71] H. Yang and X. Yang, "Two-stage zero-shot denoising and enhancement for structured illumination microscopy," *Digital Signal Processing*, p. 105345, 2025.

[72] T. Eriksen and N. u. Rehman, "Data-driven nonstationary signal decomposition approaches: a comparative analysis," *Scientific Reports*, vol. 13, no. 1, p. 1798, 2023.

[73] G. Yang, Y. Liu, Y. Wang, and Z. Zhu, "Emd interval thresholding denoising based on similarity measure to select relevant modes," *Signal Processing*, vol. 109, pp. 95–109, 2015.

[74] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046, pp. 1597–1611, 2004.

[75] M. Barbosh, P. Singh, and A. Sadhu, "Empirical mode decomposition and its variants: A review with applications in structural health monitoring," *Smart Materials and Structures*, vol. 29, no. 9, p. 093001, 2020.

[76] O. B. Fosso and M. Molinas, "Method for mode mixing separation in empirical mode decomposition," *arXiv preprint arXiv:1709.05547*, 2017.

[77] A. Tabrizi, L. Garibaldi, A. Fasana, and S. Marchesiello, "Influence of stopping criterion for sifting process of empirical mode decomposition (emd) on roller bearing fault diagnosis," in *Advances in Condition Monitoring of Machinery in Non-Stationary Operations: Proceedings of the third International Conference on Condition Monitoring of Machinery in Non-Stationary Operations CMMNO 2013*. Springer, 2014, pp. 389–398.

[78] H. Ahmadi and A. Ekhlasi, "Types of emd algorithms," in *2019 5th Iranian conference on signal processing and intelligent systems (ICSPIS)*. IEEE, 2019, pp. 1–5.

[79] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2013.

[80] J. Lian, Z. Liu, H. Wang, and X. Dong, "Adaptive variational mode decomposition method for signal processing based on mode characteristic," *Mechanical Systems and Signal Processing*, vol. 107, pp. 53–77, 2018.

[81] A. Saucier and C. Audet, “Construction of sparse signal representations with adaptive multiscale orthogonal bases,” *Signal Processing*, vol. 92, no. 6, pp. 1446–1457, 2012.

[82] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising with block-matching and 3d filtering,” in *Image processing: algorithms and systems, neural networks, and machine learning*, vol. 6064. SPIE, 2006, pp. 354–365.

[83] ——, “Bm3d image denoising with shape-adaptive principal component analysis,” in *SPARS’09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[84] M. Aharon, M. Elad, and A. Bruckstein, “K-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[85] M. Lebrun, “An analysis and implementation of the bm3d image denoising method,” *Image Processing On Line*, vol. 2, pp. 175–213, 2012.

[86] N. Mamaev, D. Yurin, and A. Krylov, “Choice of the parameter for bm3d denoising algorithm using no-reference metric,” in *2018 7th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2018, pp. 1–6.

[87] M. Scetbon, M. Elad, and P. Milanfar, “Deep k-svd denoising,” *IEEE Transactions on Image Processing*, vol. 30, pp. 5944–5955, 2021.

[88] C. Tian, M. Zheng, W. Zuo, B. Zhang, Y. Zhang, and D. Zhang, “Multi-stage image denoising with the wavelet transform,” *Pattern Recognition*, vol. 134, p. 109050, 2023.

[89] G. Vaksman, M. Elad, and P. Milanfar, “Lidia: Lightweight learned image denoising with instance adaptation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 524–525.

[90] Y. Zhao, Y. Li, and B. Yang, “Denoising of seismic data in desert environment based on a variational mode decomposition and a convolutional neural network,” *Geophysical Journal International*, vol. 221, no. 2, pp. 1211–1225, 2020.

[91] Z. Shamaee and M. Mivehchy, “Dominant noise-aided emd (demd): Extending empirical mode decomposition for noise reduction by incorporating dominant noise and deep classification,” *Biomedical Signal Processing and Control*, vol. 80, p. 104218, 2023.

[92] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, “Toward convolutional blind denoising of real photographs,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1712–1722.

[93] Z. Wang, J. Liu, G. Li, and H. Han, "Blind2unblind: Self-supervised image denoising with visible blind spots," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 2027–2036.

[94] D. M. Vo, D. M. Nguyen, T. P. Le, and S.-W. Lee, "Hi-gan: A hierarchical generative adversarial network for blind denoising of real photographs," *Information Sciences*, vol. 570, pp. 225–240, 2021.

[95] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[96] Y. Mansour and R. Heckel, "Zero-shot noise2noise: Efficient image denoising without any data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 018–14 027.

[97] C. Qiao, Y. Zeng, Q. Meng, X. Chen, H. Chen, T. Jiang, R. Wei, J. Guo, W. Fu, H. Lu *et al.*, "Zero-shot learning enables instant denoising and super-resolution in optical fluorescence microscopy," *Nature communications*, vol. 15, no. 1, p. 4180, 2024.

[98] Q. Ma, J. Jiang, X. Zhou, P. Liang, X. Liu, and J. Ma, "Pixel2pixel: A pixelwise approach for zero-shot single image denoising," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[99] C. Kim, J. Lee, and J. Shin, "Zero-shot blind image denoising via implicit neural representations," *arXiv preprint arXiv:2204.02405*, 2022.

[100] N. Wiener, "Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications," 1949.

[101] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique," *IEEE transactions on image processing*, vol. 3, no. 6, pp. 747–758, 1994.

[102] A. Buades, C. Bartomeu, and J.-M. Morel, "A non-local algorithm for image denoising," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 60–65, 2005.

[103] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE transactions on image processing*, vol. 17, no. 10, pp. 1737–1754, 2008.

[104] S. Aja-Fernández and G. Vegas-Sánchez-Ferrero, "Statistical analysis of noise in mri," *Switzerland: Springer International Publishing*, 2016.

[105] U. Satija, B. Ramkumar, and M. S. Manikandan, "A review of signal processing techniques for electrocardiogram signal quality assessment," *IEEE reviews in biomedical engineering*, vol. 11, pp. 36–52, 2018.

[106] I. Gkioulekas and L. G. Papageorgiou, "Piecewise regression analysis through information criteria using mathematical programming," *Expert Systems with Applications*, vol. 121, pp. 362–372, 2019.

[107] L. Yang, S. Liu, S. Tsoka, and L. G. Papageorgiou, "Mathematical programming for piecewise linear regression analysis," *Expert systems with applications*, vol. 44, pp. 156–167, 2016.

[108] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[109] P. S. Karlsson, L. Behrenz, and G. Shukur, "Performances of model selection criteria when variables are ill conditioned," *Computational Economics*, vol. 54, no. 1, pp. 77–98, 2019.

[110] K. E. Atkinson, *An introduction to numerical analysis.* John wiley & sons, 2008.

[111] L. Lyons, D. Gibaut, and P. Clifford, "How to combine correlated estimates of a single physical quantity," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 270, no. 1, pp. 110–117, 1988.

[112] P. J. Brockwell and R. A. Davis, *Time series: theory and methods.* Springer science & business media, 2009.

[113] A. Barbero and S. Sra, "proxtv: Fast proximal operators for total variation regularization," https://github.com/albarji/proxTV, 2014.

[114] P. e. a. Virtanen, "Scipy: Open source scientific tools for python," 2020.

[115] G. e. a. Lee, "Pywavelets: A python library for wavelet transform," https://pywavelets.readthedocs.io.

[116] K. Dragomiretskiy and D. Zosso, "vmdpy: Variational mode decomposition in python," https://github.com/vrcarva/vmdpy.

[117] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[118] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse problems*, vol. 19, no. 6, pp. S165–S187, 2003.

[119] Z. F. Mohd Apandi, R. Ikeura, S. Hayakawa, and S. Tsutsumi, "An analysis of the effects of noisy electrocardiogram signal on heartbeat detection performance," *Bioengineering*, vol. 7, no. 2, p. 53, 2020.

[120] S. A. Abdaljlil, O. Boiprav, A. R. Zerek, and A. Saleh, "Impact of thermal noise on communication system performance," in *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*. IEEE, 2022, pp. 569–573.

[121] D. Hendrycks, S. Basart, N. Mu, S. Kadavath, F. Wang, E. Dorundo, R. Desai, T. Zhu, S. Parajuli, M. Guo *et al.*, "The many faces of robustness: A critical analysis of out-of-distribution generalization," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8340–8349.

[122] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S. Y. Philip, "Generalizing to unseen domains: A survey on domain generalization," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 8, pp. 8052–8072, 2022.

[123] N. Plascencia-Diaz and A. Saucier, "Adaptive noise reduction via multiscale piecewise smooth modeling," **Submitted to Digital Signal Processing**, vol. This manuscript is currently under revision (Status: with Editor), manuscript number: DSP-D-25-02130.

[124] USC-SIPI. (1977) The usc-sipi image database, volume 3: Miscellaneous. [Online]. Available: https://sipi.usc.edu/database/database.php?volume=misc&image=23#top

[125] T. H. Johnny Chen. (2022) Juliaimages. [Online]. Available: https://testimages.juliaimages.org/stable/imagelist/

[126] D. L. Ruderman, "The statistics of natural images," *Network: computation in neural systems*, vol. 5, no. 4, p. 517, 1994.

[127] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, 2011.

[128] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2862–2869.

[129] A. Rajwade, A. Rangarajan, and A. Banerjee, "Image denoising using the higher order singular value decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 849–862, 2012.

[130] G. Chang, X. Ge, W. Yuan, Z. Li, Q. Song, L. Zhu, and S. Liu, "Beyond model design: Data-centric training and self-ensemble for gaussian color image denoising," *arXiv preprint arXiv:2604.11468*, 2026.

[131] G. C. Reinsel and R. P. Velu, *Multivariate Reduced-Rank Regression: Theory and Applications*, ser. Lecture Notes in Statistics. Springer New York, 1998, vol. 136.

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

[133] T. Minka, "Automatic choice of dimensionality for pca," *Advances in neural information processing systems*, vol. 13, 2000.

[134] I. Jolliffe, "Principal component analysis," *Encyclopedia of statistics in behavioral science*, 2005.

[135] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach.* Springer, 2002.

[136] H. Cooper, L. V. Hedges, and J. C. Valentine, *The handbook of research synthesis and meta-analysis.* Russell Sage Foundation, 2019.

[137] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *CVPR 2011.* IEEE, 2011, pp. 2833–2840.

[138] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

APPENDIX A GENERATION OF SYNTHETIC PIECEWISE-POLYNOMIAL SIGNALS

We generate synthetic piecewise smooth signals by concatenating M signal pieces $s_1, s_2, \ldots, s_M$, where $s_k \in \mathbb{R}^{M_k}$ and $M_k \leq N$. The piece lengths $M_k \in \mathbb{N}$ are independent and uniformly distributed (i.i.d.) random variables, using $M_k \sim \mathrm{U}[20, 80]$ [1] for $1 \leq k \leq M$. The total number of signal pieces M is chosen to be the smallest integer that satisfies $\sum_{k=1}^{M} M_k \geq N$ for a predefined signal length $N \in \mathbb{N}$.

The k^{th} signal piece is defined as a polynomial of degree d_k evaluated on the support $S_k := \{1,\, 2, \ldots,\, M_k\}$. The d_ks are independent integer random variables equidistributed on $\{3, 4, 5, 6\}$. These polynomials, chosen independently, are of the form

$$f_k(t) = \sum_{i=1}^{d_k+1} \alpha_i \;\; t^{i-1}. \tag{A.1}$$

To determine the polynomial coefficients $(\alpha_1, \alpha_2, \ldots, \alpha_{d+1})$, we generate a list with d_k+1 pairs $P_k = \{(t_i, p_i) \,|i = 1, 2, ..., d_k + 1\}$. To ensure signal values in $[1, 2] \subset \mathbb{R}$, we take t_i as equally spaced samples from the support S_k, and p_i as uniformly distributed random variables over the interval $[1, 2] \subset \mathbb{R}$. The list P_k combined with (A.1) yield the system of equations

$$\begin{bmatrix} t_1^0 & t_1^1 & t_1^2 & \ldots & t_1^d \\ t_2^0 & t_2^1 & t_2^2 & \ldots & t_2^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{d+1}^0 & t_{d+1} & t_{d+1}^2 & \ldots & t_{d+1}^d \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{d+1} \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{d+1} \end{bmatrix}, \tag{A.2}$$

which is solved to obtain $(\alpha_1, \alpha_2, ..., \alpha_{d+1})$, and compute the values signal piece by evaluating (A.1) on the points in S_k.

The piecewise polynomial signal x is obtained by juxtaposing M polynomial pieces generated using the method above. To ensure that the length of the concatenated signal matches N, the initial N points constituting the concatenation of pieces are selected.

To simulate the presence of signal outliers in the form of spikes, we generated a fixed number $N_s \in \mathbb{N}$ of random and independent spike locations $i_1, i_2, ..., i_{N_s}$, uniformly distributed on the support $\{0, 1, ..., N-1\}$. For each location i_k, $k \in [1, N_s]$, a random increment $\Delta x_{i_k} \sim$

[1] The notation $X \sim \mathrm{U}[a, b]$ indicates that a random variable X is uniformly distributed on the interval $[a, b]$.

$\mathrm{U}[-1, 1]$ was added to $x[i_k]$. The increments Δx_{i_k} are mutually independent. The value of N_s used here is the integer part of $0.01\,N$. Finally, a white Gaussian noise with standard deviation σ was added to x to produce the final noisy signal $y = x + w$.