RICE UNIVERSITY

Multiscale Analysis for Intensity and Density Estimation

by

Rebecca M. Willett

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

MASTER OF SCIENCE

APPROVED, THESIS COMMITTEE:

Robert Nowak, Chair
Assistant Professor of Electrical and Computer Engineering

Don H. Johnson
J. S. Abercrombie Professor of Electrical and Computer Engineering and of Statistics

Michael Orchard
Professor of Electrical and Computer Engineerng

Houston, Texas

April, 2002

ABSTRACT

Multiscale Analysis for Intensity and Density Estimation

by

Rebecca M. Willett

The nonparametric multiscale polynomial and platelet algorithms presented in this thesis are powerful new tools for signal and image denoising and reconstruction. Unlike traditional wavelet-based multiscale methods, these algorithms are both well suited to processing Poisson and multinomial data and capable of preserving image edges. At the heart of these new algorithms lie multiscale signal decompositions based on polynomials in one dimension and multiscale image decompositions based on platelets in two dimensions. This thesis introduces platelets, localized atoms at various locations, scales and orientations that can produce highly accurate, piecewise linear approximations to images consisting of smooth regions separated by smooth boundaries. Polynomial- and platelet-based maximum penalized likelihood methods for signal and image analysis are both tractable and computationally efficient. Simulations establish the practical effectiveness of these algorithms in applications such as medical and astronomical, density estimation, and networking; statistical risk bounds establish the theoretical near-optimality of these algorithms.

Acknowledgments

My thesis advisor, Dr. Robert Nowak, has been a tremendous source of encouragement and guidance during my graduate studies. His insight was invaluable whether we were discussing the big picture or the gritty details. Rob's energy motivated me to doggedly push past any stumbling blocks.

My committee members, Dr. Don Johnson and Dr. Michael Orchard, and Dr. Richard Baraniuk, all offered significant insights into my research and were always open for discussions. Dr. Johnson's obvious love for signal processing helped me recognize much of the beauty and fascination of the subject. Dr. Orchard never let me forget to carefully examine problems in search of a deeper understanding of not only how but also why solutions are effective. Dr. Baraniuk's optimism was always revitalizing.

I am also grateful to Dr. Eric Kolaczyk of Boston University for providing Gamma Ray Burst data. Additional thanks are due to Tycho Hoogland, Division of Neuroscience, Baylor College of Medicine, for providing confocal microscopy data, and to Dr. Reginald Dufour and Brent Buckalew of the Rice University Physics and Astronomy Department for providing astronomical imaging data. All of the above were a source of many valuable discussions.

I am further indebted to the DSP group for their comments and suggestions during discussions of my research and for their companionship over the past two years.

I am very grateful to my family for their love, support, frequent visits, and care packages.

Contents

Abstract ii
Acknowledgments iii
List of Figures vi
List of Tables vii

1 Analysis of Poisson and Multinomial Processes 1
1.1 Wavelet and Multiresolution Methods . . . 3
1.2 Contribution . . . 5

2 Piecewise Polynomial and Platelet Approximations 7
2.1 Polynomial Approximation . . . 7
Theorem 2.1 . . . 7
Lemma 2.1 . . . 9
Theorem 2.2 . . . 10
2.2 Platelet Decompositions and Image Approximations . . . 10
2.2.1 Haar Multiscale Analysis and Image Partitioning . . . 11
2.2.2 Anecdotal Evidence of Platelet Approximation Capabilities . . . 14
2.2.3 Platelet Approximation Theory . . . 17
Theorem 2.3 . . . 19

3 Likelihood Factorizations 21
3.1 Polynomial Factorizations . . . 21
3.2 Platelet Factorizations . . . 23

4 Denoising 28
4.1 Maximum Penalized Likelihood Estimators . . . 28
4.2 MLE of Linear Signal Parameters . . . 30
4.3 Optimal Pruning Algorithms . . . 31
4.3.1 One Dimension . . . 33
4.3.2 Two Dimensions . . . 33
4.4 Polynomial Risk Analysis . . . 36
Theorem 4.1 . . . 37
Lemma 4.1 . . . 37
4.5 Platelet Analysis Computational Complexity . . . 39
Lemma 4.2 . . . 40
Theorem 4.2 . . . 40
4.6 Penalty Parameter Selection . . . 41
4.7 Applications and Experiments . . . 43
4.7.1 Photon-Limited Applications . . . 43
4.7.2 Density Estimation . . . 45

5 Deblurring and Reconstruction **53**
5.1 Poisson Inverse Problems . . . 55
5.2 EM-MLE Reconstruction . . . 57
5.3 EM-MPLE using Platelet Approximations . . . 57
5.4 Applications and Experiments . . . 60
5.4.1 Emission Computed Tomography . . . 60
5.4.2 Astronomical Image Reconstruction . . . 61
5.4.3 Confocal Microscopy . . . 65

6 Conclusions and Ongoing Work **69**

References **72**

A Proofs of Theorems and Lemmas **76**
A.1 Proof of Discrete Polynomial Approximation Lemma . . . 76
Proof of Lemma 2.1 . . . 76
A.2 Proof of Platelet Approximation Theorem . . . 79
Proof of Theorem 2.3 . . . 79
A.3 Proof of Risk Bound Theorem . . . 82
Proof of Lemma 4.1 . . . 82
A.4 Proof of Computational Complexity Theorem . . . 84
Proof of Lemma 4.2 . . . 84
Proof of Theorem 4.2 . . . 84

List of Figures

2.1 Haar and wedgelet partition spaces 13
2.2 Quadratic bowl 15
2.3 Quadratic bowl approximations 16
2.4 Approximation errors *vs.* number of parameters 17
2.5 Platelet Approximation Example 20

4.1 Optimal pruning for a piecewise constant signal. 32
4.2 Optimal pruning for a piecewise linear signal. 32
4.3 Variation of MSE with penalty parameter 42
4.4 Gamma Ray Burst Intensity Estimation 44
4.5 Denoising in nuclear medicine 48
4.6 Network Queue Density Estimation 49
4.7 Heavisine Density Estimation 50
4.8 Blocks Density Estimation 51
4.9 Bumps Density Estimation 52

5.1 SPECT simulation 63
5.2 Astronomical Imaging Results 64
5.3 Confocal microscopy simulation 67
5.4 Confocal microscopy application 68

A.1 Sequential calculation of wedgelet likelihoods 85

List of Tables

4.1 Polynomial Algorithm Pseudocode 34
4.2 Platelet Algorithm Pseudocode 35
4.3 $\frac{MSE}{SNR}$ for Denoising with Platelets and Wavelets 45
4.4 Density Estimation MSE, Normalized to MSE of Multiscale MPLE Estimate . . . 47

Chapter 1

Analysis of Poisson and Multinomial Processes

Poisson processes are a useful statistical model for a wide variety of counting problems in a number of different fields. For example, counting the photons being emitted from a radioactive source has applications in nuclear medicine, geochronology, and nuclear physics. Seismologists use Poisson processes to model earthquake occurrences, and failure analysts model failure rates. Recent work by the astronomical community on Gamma Ray Bursts has also attracted the attention of Poisson statisticians [1]. Many astronomical and medical imaging modalities involve the detection of (light or higher energy) photons, and often the random nature of photon emission and detection is the dominant source of noise in imaging systems. The data collected by these imaging systems are usually assumed to obey a spatial Poisson distribution involving a two-dimensional intensity image that describes the probability of photon emissions at different locations in space. Such cases are referred to as *photon-limited* imaging applications, since, in contrast to applications such as x-ray imaging, the relatively small number of detected photons is the factor limiting the signal-to-noise ratio. These applications include Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Confocal Microscopy, and Infrared (IR) imaging [2–4]. In these and other applications, intensity estimation can be a daunting but necessary task for successful data analysis. Closely linked to these applications is the important problem of probability density estimation. In this case the total aggregated intensity is known to be one, and so the density can be modeled by a Poisson process conditioned on its aggregate intensity, the equivalent of a multinomial distribution. Both intensity and density estimation are possible with the algorithms described in this thesis.

This thesis describes new multiscale methods for intensity and density estimation with piecewise polynomials in one dimension and platelets in two dimensions. A key feature of these approaches is that they are nonparametric, meaning that no *a priori* limit is placed on the degrees of freedom used to describe the observed data. These methods constitute a non-trivial extension of the work done by Kolaczyk and Nowak, in which minimax optimal Haar wavelet-based methods restricted the intensity estimate to a piecewise-constant signal [5]. Piecewise polynomials are well-known for their ability to approximate one-dimensional signals more efficiently than piecewise constants. The algorithms developed here take advantage of these capabilities in the context of multiscale analysis. Efficient piecewise polynomial approximation on recursive dyadic partitions leads to fast, maximum penalized likelihood methods for intensity and density estimation. These methods are both theoretically and practically effective. The statistical risk associated with multiscale polynomial estimation is analyzed and bounded, and the algorithms are applied to both intensity and density estimation problems.

These concepts are then extended into two dimensions with platelet image representations. Platelets are localized functions at various scales, locations, and orientations that produce piecewise linear image approximations. Platelets are well suited for approximating images consisting of smooth regions separated by smooth boundaries. For smoothness measured in certain Hölder classes, it is shown that the error of m-term platelet approximations can decay significantly faster than that of m-term approximations in terms of sinusoids, wavelets, or wedgelets. This suggests that platelets may outperform existing techniques for image denoising and reconstruction. Moreover, the platelet decomposition is based on a recursive image partitioning scheme which, unlike conventional wavelet decompositions, is very well suited to photon-limited imaging applications involving Poisson distributed data. Fast, platelet-based, maximum penalized likelihood methods

for photon-limited image denoising, deblurring and tomographic reconstruction problems are developed. Because platelet decompositions of Poisson distributed images are tractable and computationally efficient, existing image reconstruction methods based on expectation-maximization type algorithms can be easily enhanced with platelet techniques. Experimental results demonstrate that platelet-based methods can outperform standard reconstruction methods currently in use in confocal microscopy, image restoration and emission tomography.

1.1 Wavelet and Multiresolution Methods

Many investigators have considered the use of wavelet representations for signal and image denoising, deblurring, and image reconstruction; for examples, see [6–14]. However, in the context of photon-limited imaging most wavelet-based approaches are based on Gaussian or other simplifying approximations to the Poisson or multinomial likelihood. This is due to the fact that it is very difficult in general to apply wavelets (as well as more recent innovations such as complex wavelets [15] and curvelets [16]) to Poisson data, but wavelets and related representations are easy to use in the Gaussian case. The Haar wavelet system is the only exception [17]; it does provide a tractable multiscale analysis framework for Poisson data, and this research builds on the Haar-based multiscale likelihood factorizations developed in [18, 19]. There are several reasons why Gaussian approximations are undesirable. First, the approximations are usually only reasonable if the numbers of detected photons (or other events occurring) in a discrete interval are sufficiently large, so that the Poisson data, possibly after a suitable transformation, is roughly Gaussian distributed. To insure that the photon count levels are large, the detections must be binned or aggregated over regions (usually intervals/pixels/voxels) of sufficiently large length/area/volume. Thus, one must immediately sacrifice spatial resolution in order to accommodate the approximations. This runs counter to the entire

philosophy of wavelet and multiscale methods, which attempt to achieve some degree of resolution or spatial adaptivity in order to recover as much of the signal or image detail and structural nuances as possible from the data. Secondly, taking advantage of the wealth of theoretical, algorithmic, experimental and clinical expertise developed in photon-limited imaging in the past two decades, we observe that methods which retain the Poisson likelihood criterion as the fundamental tool for statistical inquiry are quite advantageous. This paper describes new multiscale methods for photon-limited image denoising, deblurring, and reconstruction that are based on the Poisson likelihood and the classical EM algorithm.

Multiscale and wavelet methods fall under the broad heading of *computational harmonic analysis* (CHA). In denoising and reconstruction problems, the basic approach pursued by CHA is to (i) define a sufficiently rich class of functions that reasonably captures the characteristics of the signals or images under study, (ii) find a basis or similar representation consisting of simple, rapidly computable, space-scale localized elements that is capable of approximating all functions in the class with a relatively small number of terms, (iii) employ a coefficient thresholding or pruning criterion to remove terms with small (presumably "noisy") coefficients and reconstruct an estimate of the underlying signal or image, and (iv) prove that the estimator is optimal or near optimal in some sense. The basic idea is that because all functions in the class can be represented by a small number of simple elements, it is possible to transform the raw data into the alternate representation (e.g., wavelet) and then remove most of the terms (which in turn eliminates most of the noise) without losing much signal. The result can be a very good estimate of the underlying signal or image. The conventional wisdom is that a representation that provides good approximations will also provide good estimations [6].

1.2 Contribution

This thesis focuses on the four steps of the CHA program: (i) defining a class of functions (multiscale piecewise polynomials and platelets) suitable for describing the signals or images commonly encountered in Poisson or multinomial process applications, (ii) studying the approximation capabilities of piecewise polynomials and platelets, (iii) devising MPLEs based on piecewise polynomials and platelets, and (iv) demonstrating the near-optimality of piecewise polynomial estimators for Poisson, multinomial, or Gaussian data. Piecewise polynomials are efficient approximators for a wide range of smooth functions. Likewise, platelets generalize Donoho's wedgelets [20], and like complex wavelets and curvelets, platelets are capable of concisely representing edges at various scales, locations and *orientations*. It is shown that platelet approximations can significantly outperform conventional wavelet and wedgelet approximations (in terms of the number of terms required to achieve a given approximation error). Moreover, polynomial and platelet representations are especially well-suited to the analysis of Poisson data, unlike most other multiscale signal and image representations, and they can be rapidly computed.

Piecewise polynomial signal and platelet image representations are central to my maximum penalized likelihood criterion that encompasses denoising, deblurring, and tomographic reconstruction problems. The criterion can be incorporated very easily into the classical EM algorithm for image reconstruction and the overall computational burden is nearly the same as that of the basic EM-MLE algorithm [21]. The performance of polynomial- and platelet-based MPLEs is explored in this paper through a comprehensive set of realistic simulations. Simulation results demonstrate that platelet-based MPLEs can outperform conventional EM-MLE algorithms. The performance of polynomial- and platelet-based MPLEs is also explored with real data from astronomy, networking, density estimation, nuclear medicine and confocal microscopy experiments. Finally, it is shown that

piecewise polynomial signal estimates exhibit near-optimal statistical characteristics.

This thesis is laid out as follows. Chapter 2 introduces the platelet representation and quantifies the approximation capabilities of piecewise polynomials and platelets. I will demonstrate that piecewise polynomials can approximate signals in certain smoothness classes much more efficiently than Fourier or wavelet bases. Similarly, a fundamental class of images composed of smooth regions separated by smooth boundaries is defined, and it is shown that platelets provide dramatically superior approximations to images in the class compared to Fourier, wavelet and other existing multiscale methods. Chapter 3 discusses multiscale analysis methods for Poisson data and the notion of multiscale likelihood factorizations. I show that a Poisson likelihood, parameterized by a polynomial or platelet representation, admits a multiscale likelihood factorization, which is a probabilistic analog of classical multiresolution analysis. Chapter 4 proposes a new penalized likelihood criterion based on polynomials or platelets for "denoising" Poisson or multinomial process observations. The multiscale likelihood factorization enables a fast, globally optimal algorithm that computes the MPLE. Chapter 5 studies more challenging (inverse) problems including photon-limited image deblurring (restoration) and tomographic reconstruction. It is shown that MPLEs can be computed very rapidly using an EM algorithm in which the E-Step is identical to that of the classical EM-MLE algorithm and the M-Step is quickly computed using the "denoising" algorithm developed in the previous section. Examples from confocal microscopy are examined through simulation and with real data. Chapter 6 summarizes the new methodology and algorithms and discusses ongoing and future research directions.

Chapter 2

Piecewise Polynomial and Platelet Approximations

As discussed in the previous section, CHA necessitates a function representation which allows all functions in the class under consideration to be accurately approximated by the combination of just a few simple elements. Piecewise polynomials in one dimension and platelets in two dimensions meet this criterion, and we will explore their approximation capabilities in this chapter.

2.1 Polynomial Approximation

The approximation of one-dimensional continuous functions using piecewise polynomials is a well-studied problem. In this section I will review continuous approximation error decay rates and then use these rates to bound the error between a sampled version of the object function and a sampled version of the piecewise polynomial function. This bound will become a key component of the derivation of statistical risk bounds in Chapter 4.

Suppose $\mu(\cdot)$ is in the Besov space $B^r_\tau(L_\tau(\Omega))$, where $1/\tau = r + 1/2$ and $\Omega = [0, 1]$. Further assume that $\mu(t) \in [c, C]$, $\forall t \in \Omega$. Let $\tilde{\mu}_f(\cdot)$ be the best d-piece free-knot piecewise polynomial of order r. That is, $\|\theta - \tilde{\theta}_f\|_{L_2(\Omega)} \leq \|\theta - g\|_{L_2(\Omega)}$ for any $g(\cdot)$, where g is any d-piece free-knot piecewise polynomial of order r. DeVore [22] characterizes the approximation capabilities of free-knot polynomials of order r with d pieces on the space Ω as follows:

Theorem 2.1 *The squared L_2 approximation error decays exponentially with the number of polynomial pieces, d, where the rate of decay is proportional to the polynomial degree, r. That is,*

$$\|\mu - \tilde{\mu}_f\|^2_{L_2(\Omega)} \leq O(d^{-2r}). \tag{2.1}$$

This means that as the number of polynomials used to approximate the function increases, the approximation error will decrease exponentially, and the rate at which the error decreases is faster when using polynomials of higher degree. Compare this with approximation errors in Fourier or wavelet analyses. The best dr-term Fourier approximation error decays like $O((dr)^{-1/2})$, which is significantly slower than the polynomial approximation error of $O(d^{-2r})$ for $r \geq 1$. We consider dr terms in the Fourier approximation for a fair comparison because each of the d polynomials in the piecewise polynomial approximation can be represented by r basis vectors. Similarly, wavelet errors decay like $O((dr)^{-1})$, much more slowly than polynomial errors decay. Refer to [16, 20, 23] for the Fourier and wavelet error rates. This analysis foreshadows the significant denoising and reconstruction capability gains we will observe when applying piecewise polynomial estimation algorithms to noisy data.

We are primarily concerned with our ability to approximate a sampled version of $\mu(\cdot)$, which will play a key role in demonstrating the near-optimality of these methods in Chapter 4. We now use DeVore's $L_2(\Omega)$ bound in the continuous domain to establish a discrete bound in the ℓ_2 space. Use integration to generate the discrete vector $\boldsymbol{\mu}$ as follows: $\mu_i \equiv \int_{I_i} \mu(t)dt$ for $i = 0 \ldots N-1$ and $I_i \equiv [i/N, (i+1)/N)$. Likewise, define the elements of the free-knot approximation vector $\tilde{\boldsymbol{\mu}}_f$ as $\tilde{\mu}_{f,i} \equiv \int_{I_i} \tilde{\mu}_f(t)dt$. Next, consider the fact that the polynomial breakpoints in $\tilde{\mu}_f(\cdot)$ will not normally lie on partition breakpoints, but rather inside some interval I_i. This means that for each of the $d-1$ breakpoints, the corresponding vector element will be the result of integrating two distinct polynomials. This difficulty is overcome by defining a second approximation function $\tilde{\mu}_p(\cdot)$ which is equivalent to $\tilde{\mu}_f(\cdot)$ except that its $d-1$ breakpoints are all shifted to the nearest partition edge. $\tilde{\boldsymbol{\mu}}_p$ is then the integrated over intervals version of $\tilde{\mu}_p(\cdot)$, and $\tilde{\mu}_{f,i} = \tilde{\mu}_{p,i}$ for all but $d-1$ vector elements. Finally, $\tilde{\boldsymbol{\mu}}_p$ must be discretized for computer storage. This is accomplished by quantizing

each of the $r+1$ polynomial coefficients in each of the d polynomials to yield the vector $\tilde{\boldsymbol{\mu}}'$.

These definitions are central to bounding the ℓ_2 error $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}$. Using the construction of $\tilde{\boldsymbol{\mu}}'$ outlined above and the triangle inequality, we obtain

$$\begin{aligned}
\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2 &= \|(\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f) + (\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p) + (\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}')\|_{\ell_2}^2 \\
&\leq \underbrace{\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}^2}_{a} + \underbrace{\|\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p\|_{\ell_2}^2}_{b} + \underbrace{\|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2}_{c} \\
&\quad +2\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}\|\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p\|_{\ell_2} + 2\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}\|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_2} \\
&\quad +2\|\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p\|_{\ell_2}\|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}. \qquad (2.2)
\end{aligned}$$

Note that the last three terms are combinations of the square roots of the first three terms. Thus, after bounding the first three terms (marked a, b, and c), a total bound can be easily calculated. The terms a, b, and c of (2.2) can each be bounded according to the following lemma:

Lemma 2.1 *Let $\boldsymbol{\mu}$, $\tilde{\boldsymbol{\mu}}_f$, $\tilde{\boldsymbol{\mu}}_p$ and $\tilde{\boldsymbol{\mu}}'$ be defined as above. Then*

$$\begin{aligned}
\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}^2 &= O(\frac{d^{-2r}}{N}), \\
\|\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p\|_{\ell_2}^2 &\leq \frac{(C-c)^2 d}{N^2}, \text{ and} \\
\|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2 &\leq \frac{C'^2 r^3}{N^2}.
\end{aligned}$$

See Appendix A.1 for a proof. This lemma demonstrates that each of the first three terms in (2.2) are bounded. We now can use the three bounds in the lemma to bound the remaining terms in (2.2). For example, the fourth term is simply twice the square root of term a times the square root

of term b, etc. Specifically,

$$\begin{aligned} 2\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}\|\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p\|_{\ell_2} &= O\left(\frac{d^{-r+1/2}}{N^{3/2}}\right) \\ 2\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}\|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_2} &= O\left(\frac{d^{-r}r^{3/2}}{N^{3/2}}\right) \\ 2\|\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p\|_{\ell_2}\|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_2} &= O\left(\frac{d^{1/2}r^{3/2}}{N^2}\right) \end{aligned}$$

We have now bounded each term in (2.2); combining them, we can bound the entire quantity $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2$ by sum of the six bounds just derived for each term to yield the following theorem:

Theorem 2.2 *Let $\boldsymbol{\mu}$, $\tilde{\boldsymbol{\mu}}_f$, $\tilde{\boldsymbol{\mu}}_p$ and $\tilde{\boldsymbol{\mu}}'$ be defined as above. Then, using Lemma 2.1 we obtain the bound*

$$\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2 = O\left(\left(\frac{d^{-r}}{N^{1/2}} + \frac{d^{1/2}}{N} + \frac{r^{3/2}}{N}\right)^2\right). \tag{2.3}$$

In other words, our discrete error decays more slowly than the continuous error, and this "slow-down" decreases as we refine our sampling and N increases. We will re-address this bound in Chapter 4 as we bound the statistical risk associated with using discrete piecewise polynomials in our Poisson or multinomial estimation algorithms.

2.2 Platelet Decompositions and Image Approximations

Platelet approximation capabilities in two dimensions are more challenging to analyze than those of piecewise polynomials in one dimension, primarily due to the wedgelet partition space I employ in platelet analysis. I will therefore begin by reviewing Haar multiscale image analysis and its connection to recursive partitions. This sets the stage for developing multiscale platelet representations. Next I will formally define platelets, provide a qualitative explanation of why they

are an effective image representation, and conclude with an analysis of platelet approximation. The approximation capabilities of platelets are explored and contrasted with Fourier-, wavelet-, and wedgelet-based approximations.

2.2.1 Haar Multiscale Analysis and Image Partitioning

Image partitions are central to the approximation capabilities of platelets. In one dimension, we split the entire domain $[0, 1]$ into dyadic intervals; a similar mechanism with dyadic squares can be used in two dimensions, as will be discussed. We will see that piecewise constant image estimation on dyadic squares is analogous to the Haar wavelet analysis, and that greater efficiency can be achieved by employing wedgelet partitions.

Consider an image $x(u, v)$ on $[0, 1] \times [0, 1]$. A J-scale, Haar multiscale analysis of the image is achieved by defining the dyadic squares

$$S_{m,n,j} \equiv [m/2^j, (m+1)/2^j) \times [n/2^j, (n+1)/2^j),$$

for $j = 0, \ldots, J-1$, $m, n = 0, \ldots, 2^j - 1$, where J dictates the size of squares at the finest scale of analysis. Each dyadic square is associated with a coefficient $x_{S_{m,n,j}} \equiv \int_{S_{m,n,j}} x(u, v)$. That is, we define an analysis separating the information in x into its components at various combinations of position and scale (m, n, j). This strategy underlies the analysis of x with respect to an orthonormal basis of dyadic Haar wavelets. Note that each dyadic square $S_{m,n,j}$ splits into four smaller dyadic squares of equal size. These four squares are called the "children" of $S_{m,n,j}$ and are denoted by $\{ch(S^i_{m,n,j})\}_{i=1}^4$. In the following, the index i will be suppressed to keep the notation cleaner. The coefficients associated with the four children squares, denoted $\{x_{ch(S_{m,n,j})}\}$, will be referred to as

the children of $x_{S_{m,n,j}}$.

The relationship between "parent" and "children" dyadic squares suggests the notion of a recursive partition. The sequence of dyadic squares (from coarse-to-fine) can be interpreted as a recursive dyadic partition of $[0,1]^2$. Consider a sequence of nested partitions $\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_M$ of the unit square, where $\mathcal{P}_1 = [0,1]^2$, $\mathcal{P}_M = \cup_{m,n=0}^{2^J-1} S_{m,n,J}$, and each partition $\mathcal{P}_\ell$, $\ell = 2, 3, \ldots, M$, results from splitting one and only one of the dyadic squares in $\mathcal{P}_{\ell-1}$ into four smaller dyadic squares. Hence the collection $\mathcal{P}^* \equiv \{\mathcal{P}_\ell\}_{\ell=1}^M$ is sometimes called a *complete recursive partition (C-RP)* of the square $[0,1]^2$.

We can examine these recursive partitions in a wavelet context for increased clarity. Note that a C-RP can be associated with a "quad-tree" graph. This tree can be adaptively pruned to produce an incomplete RP with different size squares at different spatial locations, which will be central to the estimation algorithm described in Chapter 4. In the dyadic square C-RP, the pruning process is quite similar to thresholding (setting to zero) Haar wavelet coefficients except that the thresholding is performed with a hereditary constraint (i.e., Haar wavelet coefficients at a given scale may be kept only if all "parent" coefficients at coarser scales are retained in the representation) [24]. This hereditary constraint ensures that every pruning is an RP with a tree structure; in general, thresholding Haar wavelet coefficients does not correspond to an RP.

Recursive partitions (and hereditary Haar analysis) are important because they allow for efficient extensions of classical Haar multiscale analysis. In particular, one need not restrict the analysis to dyadic square partitions. The wedgelet partition is a dyadic, square recursive partition which allows for non-square, "wedge-shaped" partitions only at the final level of the partition [20]. That is, a wedgelet partition is based on a recursive dyadic square partition of the image in which the final nodes are allowed to terminate with a wedge instead of a square. Consider Figure 2.1 as a simple

illustration of the efficiency of the wedgelet partition space. Figures 2.1(b) and (c) contain rough approximations of the Shepp-Logan phantom to within the same error using the piecewise constant and wedgelet analyses. Notice how many fewer partitions are required for wedgelet approximation.

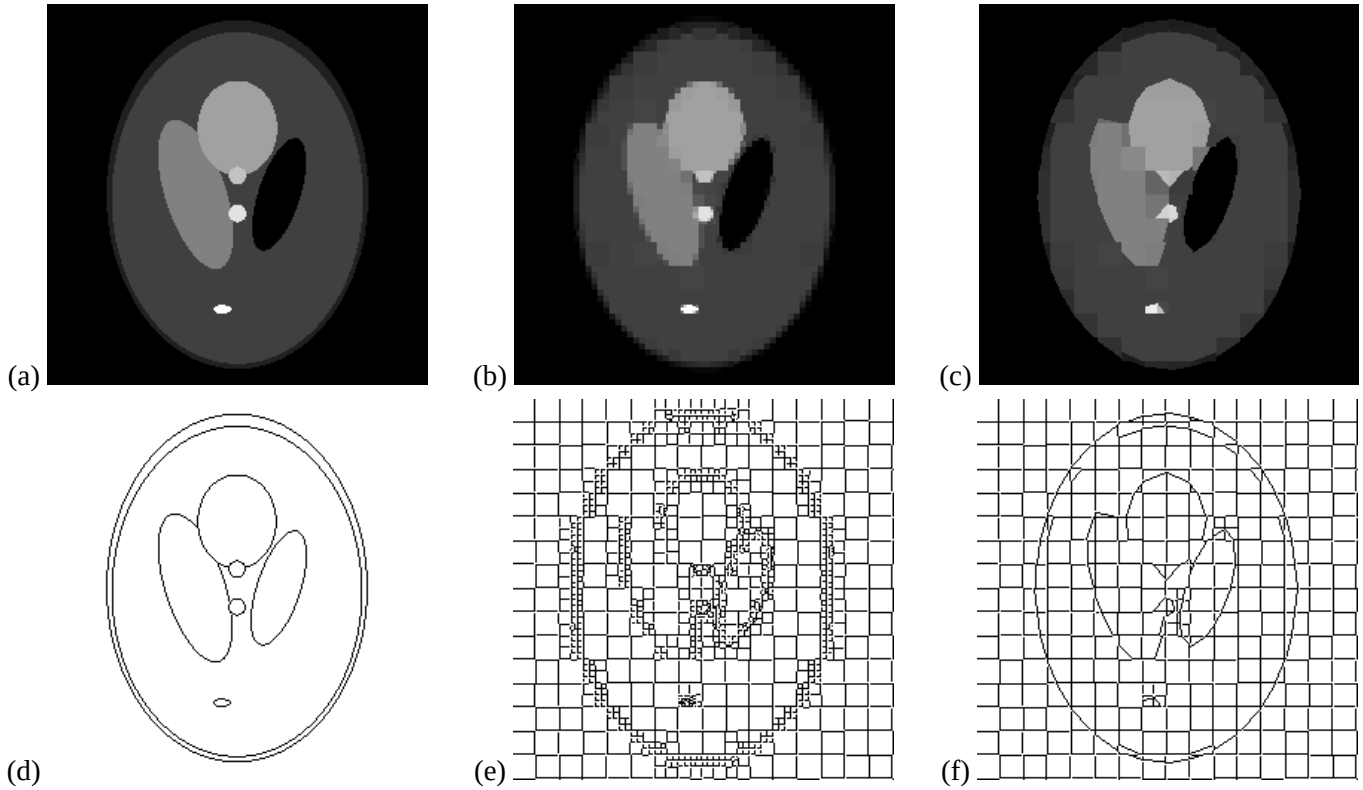


Figure 2.1 Image partitions. (a) Original image. (b) Haar approximation (error = 0.0155). (c) Wedgelet approximation (error = 0.0158). (d) True edges. (e) Haar partition. (f) Wedgelet partition. Note that while the two approximations are about equally accurate, the wedgelet partition is much more efficient in that many fewer partitions are required to approximate the image.

The wedge split is defined by a line connecting two points on the sides of the square. The points are not arbitrary; rather, they are chosen from a finite set of vertices spaced evenly δ apart around the perimeter of the square. This restriction is crucial because it means that the resulting "dictionary" of wedgelet elements is finite and easily computed. The spacing δ is referred to as the *wedgelet resolution*, a key parameter of the wedgelet analysis. Following the development in Donoho's original wedgelet paper, it is assumed that $\delta = 2^{-J-K}$ with $K \geq 0$ [20].

The power of wavelets and wedgelets is realized in connection with m-term approximations. *An m-term approximation to an image is a superposition of any m representation elements* (e.g., m wavelet functions or m wedgelets). It is important to note that one can select the m elements that provide the best approximation, where the selection is unconstrained in the case of wavelet approximations and is constrained only by the hereditary conditions dictated by the partition in hereditary Haar and wedgelet approximations. This is sometimes referred to as *nonlinear* approximation because the selection will depend on the image under consideration (in contrast to linear approximation in which the terms used in the approximation are selected without consideration of the image, e.g., the first (low frequency) m-terms in a Fourier series).

2.2.2 Anecdotal Evidence of Platelet Approximation Capabilities

Recall that in the standard Haar and wedgelet partitions, the image is modeled as piecewise constant. Instead of approximating the image on each piece of the partition by a constant, we can approximate it with a planar surface to produce image gradients. In many applications it is beneficial to have this added flexibility. Image gradients, or smooth transitions between regions of varied intensities, encode information about light emission or reflection as well as surface geometry. We define a platelet $f_S(x, y)$ to be a function of the form

$$f_S(x, y) = (A_S x + B_S y + C_S)\ I_S(x, y), \tag{2.4}$$

where $A_S, B_S, C_S \in \mathbb{R}$, S is a dyadic square or wedge associated with a terminal node of an RP, and I_S denotes the indicator function on S. Each platelet requires three coefficients, compared with the one coefficient for piecewise constant approximation. Although each platelet has two more

parameters per term, for images of sufficient smoothness many fewer platelets than constant blocks are needed to approximate an image to within a certain error. Thus, platelet approximations may require far fewer coefficients for a given level of accuracy.

Before proceeding to a more formal analysis of platelet approximation error decay rates, consider the example image in Figure 2.2. The surface plot of the test image reveals it to be a quadratic "bowl" with a depressed quadratic "bump" in its center. This image provides anecdotal evidence of platelet superiority over piecewise constant and wedgelet approximations.

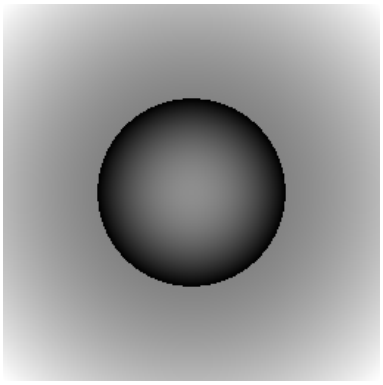
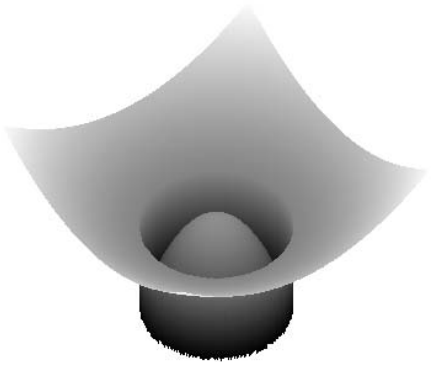
(a)  (b) 

Figure 2.2 Original quadratic test image. (a) Image. (b) Surface plot.

Approximations with similar L_2 percent errors were chosen for Figure 2.3. Figure 2.3(a) is a hereditary piecewise constant approximation, and contains 2068 constant squares of varying sizes. Figure 2.3(b) is a wedgelet approximation. Note that this image has an error similar to that of the image obtained with constants, yet it contains only 56% as many terms, where a term is a constant block or a constant wedge; *i.e.*, each block divided into two wedgelets is represented by two terms. Next consider the platelet approximation in Figure 2.3(c). For each platelet, the coefficients A_S, B_S, C_S are chosen to minimize the squared error of the fit to the image. Again witness comparable approximation error with significantly fewer parameters. Here each dyadic square is fitted with a platelet defined by three coefficients. Finally, Figure 2.3(d) is an approximation with a platelet

fitted to each square or wedge region. In this case a block may be diagonally split into two different gradients. For such a block, we need to store six parameters – three coefficients for each term, or region. In Figure 2.3(d), only 744 parameters are stored, in contrast to the 256×256 different pixel values in the image. Clearly platelets provide an accurate, sparse representation of smooth images containing smooth boundaries.

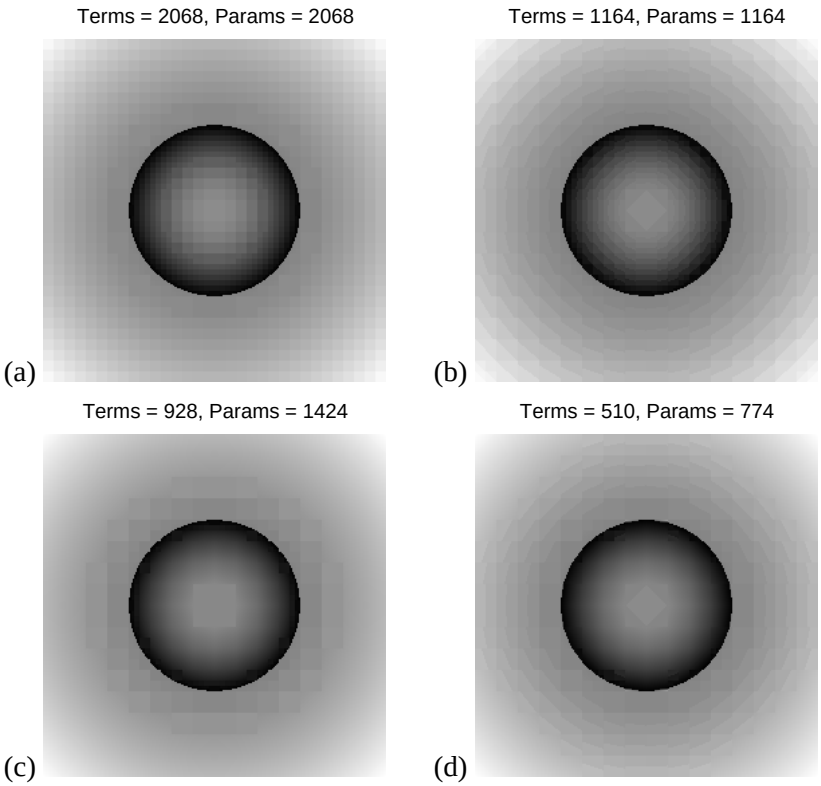


Figure 2.3 Approximations. (a) Haar. (b) Wedgelets. (c) Platelets. (d) Combined platelets and wedgelets. Each approximation has error $\approx 3 \times 10^{-4}$.

In fact, Figure 2.4 shows the decay in approximation error as the number of terms decreases. The platelet representation not only is more accurate with fewer terms, but also exhibits a rate of error decay faster than that of the constant and wedgelets representations. What follows is an analytical characterization of these rates.

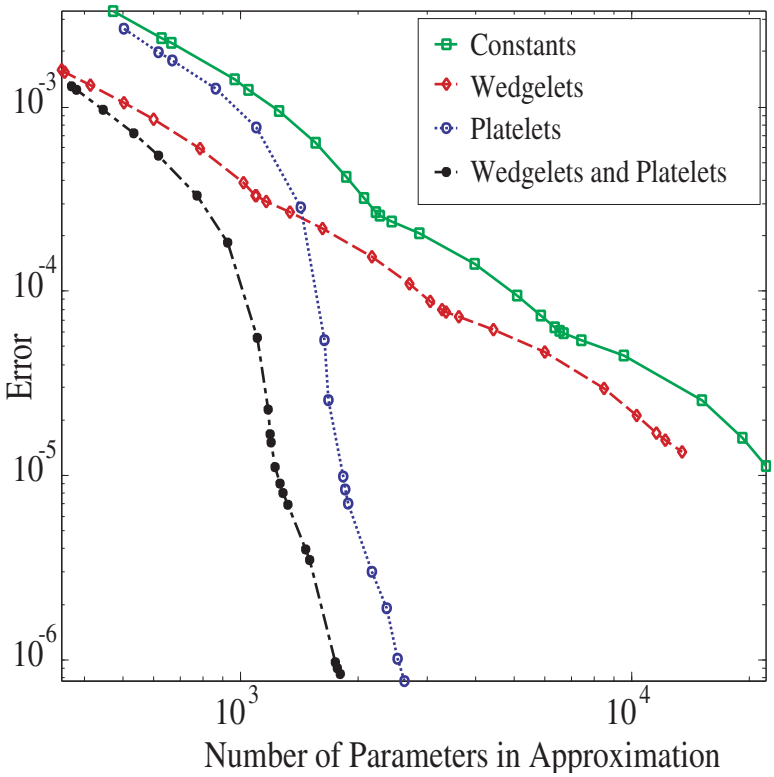


Figure 2.4 Approximation errors *vs.* number of parameters for bowl image

2.2.3 Platelet Approximation Theory

For this analysis, consider images which are Hölder smooth apart from a Hölder smooth boundary over $[0,1]^2$. Images of this form can be modeled by fusing two (everywhere) smooth images f_1 and f_2 into one single image according to

$$f(x,y) = f_1(x,y) \cdot I_{\{y \geq H(x)\}} + f_2(x,y) \cdot \left(1 - I_{\{y \geq H(x)\}}\right), \quad \forall (x,y) \in [0,1]^2, \tag{2.5}$$

where $I_{\{y \geq H(x)\}} = 1$ if $y \geq H(x)$ and 0 otherwise, and the function $H(x)$ describes a smooth boundary between a piece of f_1 and a piece of f_2. This is a generalization of the "Horizon" image model proposed in [20], which consisted of two constant regions separated by a Hölder smooth boundary.

To be more specific, the boundary is described by $y = H(x)$, where

$$H \in \text{Hölder}^{\alpha,1}(C_\alpha), \;\; \alpha > 1,$$

where $\text{Hölder}^{\alpha,1}(C_\alpha)$ is the set of functions satisfying

$$\left| \frac{\partial}{\partial x} H(x_1) - \frac{\partial}{\partial x} H(x_0) \right| \leq C_\alpha |x_1 - x_0|^{\alpha - 1}, \;\; \text{for all } x_0, x_1 \in [0,1].$$

Similarly, the smoothness of the images f_1 and f_2 is characterized by a two-dimensional Hölder surface condition

$$f_i \in \text{Hölder}^{\beta,2}(C_\beta), \;\; \beta > 1, \;\; i = 1, 2,$$

where $\text{Hölder}^{\beta,2}(C_\beta)$ is the set of functions satisfying

$$\begin{aligned} \left| \frac{\partial}{\partial x} f_i(x_1, y_1) - \frac{\partial}{\partial x} f_i(x_0, y_0) \right| &\leq C_\beta \left| \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \right|^{\beta - 1}, \text{ and} \\ \left| \frac{\partial}{\partial y} f_i(x_1, y_1) - \frac{\partial}{\partial y} f_i(x_0, y_0) \right| &\leq C_\beta \left| \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \right|^{\beta - 1} \end{aligned}$$

for all $(x_0, y_0), (x_1, y_1) \in [0,1]^2$. For more information on Hölder spaces see [25].

In other words, the model in (2.5) describes a image composed of two Hölder smooth pieces separated by a Hölder smooth boundary. The boundary of the model is specified as a function of x (hence the name "Horizon"), but we could have just as easily specified it as a function of y. Furthermore, more complicated boundaries (which are not functions of x or y) can be constructed with compositions of two or more Horizon-type boundaries.

The squared L_2 error of m-term platelet approximations for images of this form is bounded in

the following theorem.

Theorem 2.3 *Consider the class of images*

$$f(x,y) = f_1(x,y) \cdot I_{\{y \geq H(x)\}} + f_2(x,y) \cdot \left(1 - I_{\{y \geq H(x)\}}\right) \quad \forall (x,y) \in [0,1]^2$$

where $f_i \in$ *Hölder*$^{\beta,2}(C_\beta)$, $i = 1, 2$, *and* $H \in$ *Hölder*$^{\alpha,1}(C_\alpha)$ *with* $\alpha, \beta > 1$. *Suppose that* $2 \leq m \leq 2^J$, *with* $J > 1$. *The squared* L_2 *error of* m*-term,* J*-scale, resolution* δ *platelet approximation to images in this class is less than or equal to* $K_{\alpha,\beta} m^{-\min(\alpha,\beta)} + \delta$, *where* $K_{\alpha,\beta}$ *depends on* C_α *and* C_β.

Theorem 2.3 shows that for images consisting of smooth regions ($\beta > 1$) separated by smooth boundaries ($\alpha > 1$), m-term platelet approximations may significantly outperform Fourier, wavelet, or wedgelet approximations. For example, if the derivatives in the regions and along the boundary are Lipschitz ($\alpha, \beta = 2$, *i.e.*, smooth derivatives), then the m-term platelet approximation error behaves like $O(m^{-2}) + \delta$, whereas the corresponding Fourier error behaves like $O(m^{-1/2})$ and the wavelet and wedgelet errors behave like $O(m^{-1})$ at best. For very large m, the δ term will dominate the the platelet approximation error. However, for the modest values of m of the most practical interest, the $O(m^{-2})$ can be most significant, and the platelet approximation may then be significantly better than the other approximations. (To more precisely compare the approximation errors one needs to specify the constants C_α and C_β, as well as δ.) Wavelets and Fourier approximations do not perform well on this class of images due to the boundary. Refer to [16, 20, 23] for the Fourier and wavelet error rates. Wedgelets can handle boundaries of this type, but they produce piecewise constant approximations and perform poorly in the smoother (but non-constant) regions of images.

For the wedgelet case, consider an image of the form $f(x, y) = Ax$, a linear gradient image. This image is also in the class under consideration. Wedgelet approximations suffer due to their piecewise constant nature. Because the gradient is constant, the best dyadic m-term wedgelet approximation in this case partitions $[0, 1]^2$ into roughly equal area regions, each of side length $O(1/\sqrt{m})$. The L_∞ error in each region is $O(1/\sqrt{m})$ because the gradient is a fixed constant. Thus the L_2^2 error in each region is $O(m^{-2})$ (squared L_∞ error $\times$ area) and the total L_2^2 error of the best m-term wedgelet approximation is $O(m^{-1})$.

The conclusions of Theorem 2.3 as well as the error rates of Fourier/wavelet/wedgelet approximations above also carry over to more complicated images composed of several smooth regions and boundaries. Each additional boundary increases the m-term approximation errors by integer multiples (two boundaries roughly double the error from the case of a single boundary). Figure 2.5 displays such an image (a linear top and quadratic bottom separated by a cubic boundary) and an approximation of it with platelets.

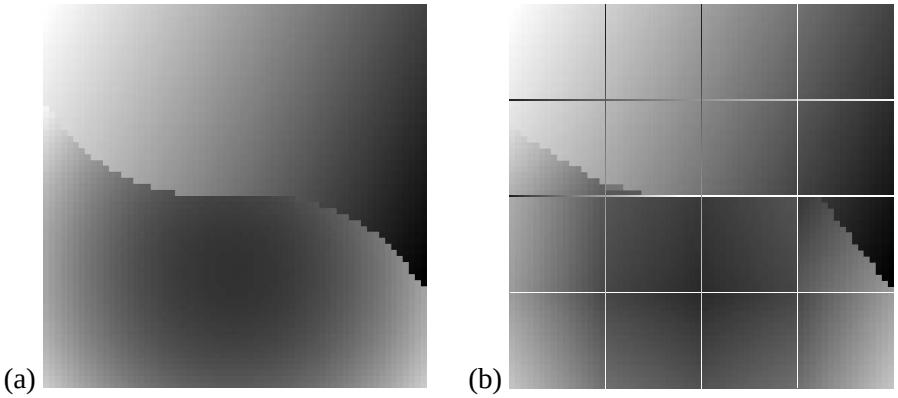


Figure 2.5 Platelet Approximation Example. (a) Image of smooth regions separated by a smooth boundary. (b) Rough approximation of (a) with platelets (lines drawn to illustrate platelet boundaries).

As was the case with piecewise polynomial analysis, the discrete ℓ_2 approximation error must be characterized to bound the statistical risk associated with platelet approximation. This is an area for future exploration.

Chapter 3

Likelihood Factorizations

In the previous chapter, we determined that piecewise polynomials defined on dyadic intervals and platelets can accurately and efficiently approximate functions in certain smoothness classes. Now suppose that we observe a one- or two-dimensional realization of a Poisson or multinomial process, and wish to exploit the approximation capabilities of polynomials or platelets in estimating the underlying Poisson intensity function or multinomial probabilities. I describe here likelihood factorizations based on multiscale polynomials and platelets, a key factor in the optimality of the estimation algorithm described in the following chapter. This extends the work done by Kolaczyk and Nowak for piecewise constant likelihood factorizations [17]. The polynomial and platelet factorizations described here provide an alternative probabilistic representation of the information in an observation, in a manner indexed by the various location/scale combinations offered by a given recursive partition. A likelihood factorization allows the likelihood of the entire image to be represented in a tree structure in which both likelihoods and parameters of children are inherited by parents. Thus, a likelihood factorization serves as a probabilistic analogue of an orthonormal wavelet decomposition of a function. The parameters of the conditional likelihoods play the same role as wavelet coefficients in a conventional wavelet-based multiscale analysis. As in the previous chapter, I first examine multiscale polynomials for signal analysis, followed by platelets for image analysis.

3.1 Polynomial Factorizations

We first consider a realization of a one-dimensional Poisson or multinomial process and its likelihood factorization in terms of approximating polynomials. Specifically, suppose that $x(u)$

is a realization of a Poisson or multinomial process. Underlying this process is an continuous intensity or density function $\mu(u)$, $(u) \in [0,1]$. Assume that either by choice or perhaps by the limitations of measuring instruments, $x(u)$ is observed only discretely on the measurement intervals I_n, $n = 0, \ldots, N-1$. It is assumed that the effect of the discretization is to yield a vector of count measurements $\boldsymbol{x} \equiv \{x_n\}_{n=0}^{N-1}$, associated with an array of intensity parameters or multinomial probabilities $\boldsymbol{\mu} \equiv \{\mu_n\}_{n=0}^{N-1}$. Each x_n is simply the number of events in the interval I_n and $\mu_n \equiv \int_{I_n} \mu(u)$ for multinomial processes or $\mu_n \equiv \int_{I_n} N\mu(u)$ for Poisson processes. The counts are conditionally independent; given $\{\mu_n\}$, $x_n \sim$ Multinomial$(\sum x_n, \mu_n)$ or $x_n \sim$ Poisson(μ_n). The likelihood of $\boldsymbol{x}$, given the intensities or probabilities $\boldsymbol{\mu}$, is denoted by $p(\boldsymbol{x}|\boldsymbol{\mu})$.

A Haar multiscale analysis of the count data is obtained by associating a count statistic $x_{I_{n,j}} \equiv \sum_{k:(\frac{k}{N}) \in I_{n,j}} x_k$ with each dyadic interval $I_{n,j} \equiv [n/2^j, (n+1)/2^j)$, $j = 0, \ldots, J-1$, $n = 0, \ldots, 2^j - 1$, and $J = \log_2(N)$. The set of all dyadic intervals $\{I_{n,j}\}$ corresponds to a *complete* recursive dyadic partition (RDP) of $[0,1]$. There also exists a Haar multiscale analysis of the intensity/density function which is defined analogously on dyadic intervals. This RDP is called complete because all terminal nodes in the partition are intervals of width $1/N$ at the finest scale. An incomplete RDP would contain larger terminal intervals which could correspond to intervals of homogeneous or smoothly varying intensities. In earlier work, Nowak and Kolaczyk introduced in this context a class of *multiscale likelihood factorizations* that provide an alternative probabilistic representation (i.e., in addition to that of the original likelihood) of the information in $\boldsymbol{x}$, in a manner indexed by the various time/scale combinations offered by a given RDP [5]. The keys to the likelihood factorization are (1) that sums of Poisson variates are Poisson and (2) that the conditional distribution of a collection of Poisson variates given their sum is multinomial. For a given RDP $\mathcal{P}$

the likelihood $p(\boldsymbol{x}|\boldsymbol{\mu})$ may be factored as

$$p(\boldsymbol{x}|\boldsymbol{\mu}(\mathcal{P},\boldsymbol{\theta})) = \\ p(x_{I_0}|\mu_{I_0}) \times \prod_{I\in NT(\mathcal{P})} p(\{x_{ch(I)}\}|x_I,\theta_I) \\ \times \prod_{I\in T(\mathcal{P})} p(\{x_n\}_{n/N\in I}|x_I,\theta_I), \tag{3.1}$$

where $I_0 \equiv [0,1]$, $NT(\mathcal{P})$ is the set of all non-terminal intervals in $\mathcal{P}$, and $T(\mathcal{P})$ is the set of all terminal intervals in $\mathcal{P}$. The terminal node likelihood factors $p(\{x_n\}_{n/N\in I}|x_I,\theta_I)$ are the multinomial likelihoods of the data in I given a polynomial model θ_I of the intensity on I. Note that in this factorization the intensity is constrained to be piecewise polynomial on each interval in the partition, as indicated by the notation $\boldsymbol{\mu}(\mathcal{P},\boldsymbol{\theta})$. For readability, however, this function will simply be referred to as $\boldsymbol{\mu}$ for the remainder of this thesis. In Poisson processes, μ_{I_0} is a parameter to be estimated, while in multinomial processes, μ_{I_0} is known to be one. The factorization in (3.1) can be shown to follow from a set of sufficient conditions whose form and function are remarkably similar to those of a Haar wavelet analysis – effectively a multiresolution analysis of the likelihood function. This is important because it allows a simple framework for the multiscale analysis of non-Gaussian data–a non-trivial task for wavelets alone. Details may be found in [5].

3.2 Platelet Factorizations

These likelihood factorization concepts are extendable to platelets in two dimensions. Now suppose that $x(u,v)$ is a realization of a Poisson process. Underlying this process is a continuous intensity or density function $\mu(u,v),(u,v)\in[0,1]^2$. Assume that either by choice or perhaps by the limitations of measuring instruments, $x(u,v)$ is observed only discretely on the squares

(pixels) $S_{m,n}$, $m, n = 0, \ldots, N-1$. It is assumed that the effect of the discretization is to yield an array of count measurements $\boldsymbol{x} \equiv \{x_{m,n}\}_{m,n=0}^{N-1}$, associated with an array of intensity parameters $\boldsymbol{\mu} \equiv \{\mu_{m,n}\}_{m,n=0}^{N-1}$. Each $x_{m,n}$ is simply the number of events in the square $S_{m,n}$ and $\mu_{m,n} \equiv \int_{S_{m,n}} \mu(u,v)$ for multinomial processes or $\mu_{m,n} \equiv \int_{S_{m,n}} N\mu(u,v)$ for Poisson processes. The counts are conditionally independent; given $\{\mu_{m,n}\}$, $x_{m,n} \sim$ Poisson$(\mu_{m,n})$ or $x_{m,n} \sim$ Multinomial$(\sum x_{m,n}, \mu_{m,n})$. The Poisson likelihood of $\boldsymbol{x}$, given the intensities $\boldsymbol{\mu}$, is denoted by $p(\boldsymbol{x}|\boldsymbol{\mu})$.

As in the one-dimensional case, a Haar multiscale analysis of the count data is obtained by associating a count statistic $x_{S_{m,n,j}} \equiv \sum_{k,l:(\frac{k}{N},\frac{l}{N}) \in S_{m,n,j}} x_{k,l}$ with each dyadic square $S_{m,n,j}$, $j = 0, \ldots, J-1$, $m, n = 0, \ldots, 2^j - 1$, and $J = \log_2(N)$. There also exists a Haar multiscale analysis of the intensity function which is defined analogously on dyadic squares. For a given C-RP $\mathcal{P}^*$ the likelihood $p(\boldsymbol{x}|\boldsymbol{\mu})$ is factored as

$$p(\boldsymbol{x} \,|\, \boldsymbol{\mu}) = p(x_{S_0} \,|\, \mu_{S_0}) \prod_{S \in NT(\mathcal{P}^*)} p(\{x_{ch(S)}\}|x_S, \theta_S) \ , \tag{3.2}$$

where $S_0 \equiv [0,1]^2$, and $NT(\mathcal{P}^*)$ is the set of all non-terminal squares in $\mathcal{P}^*$ (i.e., excluding individual pixels $S_{m,n}$). θ_S denotes the three parameters of the multinomial conditional likelihood of $\{x_{ch(S)}\}$ given x_S, which consist of the ratios of the child intensities to the parent intensities. The child intensities are then $\mu_S\theta_S^1$, $\mu_S\theta_S^2$, $\mu_S\theta_S^3$, and $\mu_S(1 - \theta_S^1 - \theta_S^2 - \theta_S^3)$.

For a general (incomplete) RP, say $\mathcal{P}$, certain terminal squares may include several pixels. The

multiscale factorization in this general case takes the form

$$p(\boldsymbol{x} \,|\, \boldsymbol{\mu}(\mathcal{P}, \boldsymbol{\theta})) \;=\; p(x_{S_0} \,|\, \mu_{S_0}) \prod_{S \in NT(\mathcal{P})} p(\{x_{ch(S)}\} | x_S, \theta_S) \times \prod_{ch(S) \in T(\mathcal{P})} p(\{x_{m,n}\}_{(m/N, n/N) \in ch(S)} | x_{ch(S)}), \tag{3.3}$$

where $\boldsymbol{\theta} \equiv \mu_{S0} \cup \{\theta_S\}_{S \in (\mathcal{P})}$ and $T(\mathcal{P})$ is the set of all terminal squares in $\mathcal{P}$ and the conditional likelihood corresponding to each terminal square $p(\{x_{m,n}\}_{(m/N,n/N) \in ch(S)} | x_{ch(S)})$ is multinomial with equal probabilities (since the intensity in the square is modeled as constant). Note that in this case the intensity is constrained to be piecewise constant on each piece of the partition, as indicated by the notation $\boldsymbol{\mu}(\mathcal{P}, \boldsymbol{\theta})$. As in the one-dimensional case, this function will simply be referred to as $\boldsymbol{\mu}$ for the remainder of this thesis. The fact that a partition of the image space underlies the multiscale likelihood factorization is key. For example, it is not possible to obtain a similar factorization with conventional smooth wavelets. This is simply because the distribution of an arbitrary linear combination (e.g., inner product or filtering) of Poisson variables does not have a simple, easily expressed likelihood function; in fact, a closed-form expression does not exist. Only unweighted summations of Poisson variables possess a simple expression; namely the sum of independent Poisson variables is Poisson distributed. Thus, Haar multiscale analysis and its generalizations are especially well suited to the study of Poisson data.

Equation (3.3) holds for a general recursive partition of the data, that is, when terminal nodes are restricted to dyadic squares. Recall, however, that a key feature of platelets is the underlying wedgelet partition, in which a terminal node of the partition can denote wedge-shaped regions in addition to squares; that is, $S \in \mathcal{P}$ may be a dyadic square or wedge. Because the spatial resolution of the acquired image data is limited to pixel size squares ($2^{-J} \times 2^{-J}$), the continuous wedgelet

partitions discussed in the previous section are replaced by "digital" wedgelets. The splitting "line" defining a digital wedgelet is a pixel-scale approximant of the ideal line; i.e., the digital wedgelet split boundary follows the boundaries of pixels, producing a staircase-like approximation to a line. A digital wedgelet splits a dyadic square into two pieces, and the two pieces contain disjoint sets of pixels. Note that this also implies that the wedgelet resolution (spacing of vertices) is $\delta = 2^{-J}$, the side length of a pixel. Additionally, instead of approximating the intensity function on each piece of the partition by a constant, we can approximate it with a planar surface (platelet). The likelihood factorization is still valid, but the terminal node likelihood functions may now be parameterized by a planar intensity function requiring two extra coefficients for the slope of the gradient in addition to a coefficient corresponding to the estimated total intensity of the associated square or wedge shaped region. In this case,

$$\begin{aligned} p(\boldsymbol{x} \,|\, \boldsymbol{\mu}) &= p(x_{S_0} \,|\, \mu_{S_0}) \prod_{S \in NT(\mathcal{P})} p(\{x_{ch(S)}\} | x_S, \theta_S) \\ &\quad \times \prod_{ch(S) \in T(\mathcal{P})} p(\{x_{m,n}\}_{(m/N,n/N) \in ch(S)} | x_{ch(S)}, \theta_{ch(S)}), \end{aligned} \tag{3.4}$$

where the terminal node likelihood factors $p(\{x_{m,n}\}_{(m/N,n/N)\in ch(S)}|x_{ch(S)}, \theta_{ch(S)})$ are the multinomial likelihoods of the data in $ch(S)$ given a planar model $\theta_{ch(S)}$ of the intensity on $ch(S)$. More specifically, $x_{ch(S)}$ is the total photon count in the region $ch(S)$. This leaves us with two degrees of freedom (embodied in $\theta_{ch(S)}$) that complete the description of a planar intensity model on $ch(S)$; *e.g.* $\theta_{ch(S)} = \{A_{ch(S)}, B_{ch(S)}\}$. It turns out that maximum likelihood estimates of the parameters of the linear surface $\theta_{ch(S)}$ are easy to compute because, as shown in the next section, the log likelihood function is concave.

Because polynomial and platelet parameters factor with the data in the likelihood factorizations,

it is possible to optimally prune one of the above recursive partitions of the data. Thus, these factorizations enable the development of fast algorithms for optimal intensity estimation.

Chapter 4

Denoising

The recursive partitions and likelihood factorizations discussed in the previous two chapters are critical to the optimal tree-pruning estimation/denoising algorithm detailed in this section. This relationship will be explored for first the one-dimensional case and then more complicated the two-dimensional case. We will see that the maximum penalized likelihood criterion coupled with the likelihood factorizations results in a globally optimal estimation/denoising algorithm. The effectiveness of these algorithms is demonstrated both theoretically, by bounding the statistical risk of piecewise polynomial estimation, and practically, by applying the algorithm to real and simulated data and comparing the results with those of wavelet denoising.

4.1 Maximum Penalized Likelihood Estimators

The multiscale likelihood factorizations in the previous chapter provide for a very simple framework for maximum penalized likelihood estimation, wherein the penalization is based on the complexity of the underlying estimate. The maximum penalized likelihood criterion we employ here is

$$L_\gamma(\boldsymbol{\mu}) \;\equiv\; \log p(\boldsymbol{x} \mid \boldsymbol{\mu}) \;-\; \gamma\,\{\#\boldsymbol{\theta}\}, \tag{4.1}$$

where $p(\boldsymbol{x} \mid \boldsymbol{\mu})$ denotes a likelihood (factorization) of the form (3.1), (3.2), (3.3), or (3.4), and $\{\#\boldsymbol{\theta}\}$ is the number of parameters in the vector $\boldsymbol{\theta}$. Note that the first term, the likelihood, is penalized, or *regularized* by the second term, proportional to the complexity of the estimate, hence the name complexity regularized penalized likelihood. In one-dimensional signal analysis, the vector $\boldsymbol{\theta}$ has

one element for each constant interval and r elements for each polynomial interval of degree r. In two-dimensional image analysis, $\boldsymbol{\theta}$ has one element for each constant square or wedge and three for each planar square or wedge. Recall $\boldsymbol{\mu}$ is a function of the multiscale parameters and partition; *i.e.* $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}, \mathcal{P})$. The constant $\gamma > 0$ is a weight that balances between fidelity to the data (likelihood) and complexity regularization (penalty), which effectively controls the bias-variance trade-off.

The solution of

$$\begin{aligned} (\widehat{\mathcal{P}}, \widehat{\boldsymbol{\theta}}) &\equiv \arg\max_{\mathcal{P},\boldsymbol{\theta}} L_\gamma(\boldsymbol{\mu}(\mathcal{P}, \boldsymbol{\theta})) \\ \widehat{\boldsymbol{\mu}} &\equiv \boldsymbol{\mu}(\widehat{\mathcal{P}}, \widehat{\boldsymbol{\theta}}) \end{aligned} \tag{4.2}$$

is called a maximum penalized likelihood estimator (MPLE). Larger values of γ produce smoother, less complex estimators; smaller values of γ produce more complicated estimators. Note that if $\gamma = 0$, then no penalty is assigned and a C-RP (with as many free parameters as measurements or pixels) maximizes the likelihood. Since a C-RP corresponds to a element- or pixel-based partition, in the $\gamma = 0$ case the MPLE reduces to the conventional maximum likelihood estimator (MLE).

The approximation-theoretic results of Theorems 2.1 and 2.3 have important implications for the performance of the platelet-based MPLE. The mean square error (MSE) of the MPLE can be separated into squared bias and variance. The variance of the MPLE is proportional to the number of terms $d \propto \{\#\boldsymbol{\theta}\}$ in a polynomial or platelet approximation. The squared bias is proportional to the error of a d-term polynomial or platelet approximation to the underlying true intensity. Theorem 2.1 tells us that for functions in Besov spaces piecewise polynomial approximations with a small number of polynomial pieces (relative to the number of measurements) can be quite accurate. This means that if the underlying intensity belongs to this class (or if it can be closely approximated by

a member of the class), then Theorem 2.1 shows that for a small value of d there exists an accurate d-term piecewise polynomial approximation to the true intensity. If $d \ll N$, then the reduction in variance will be dramatic, and the polynomial-based MPLE will tend to have a small MSE. An analogous statement holds for platelet approximations. We have seen through Theorem 2.3 that for a class of images characterized by regions of smooth intensity separated by smooth boundaries platelet approximations with a small number of terms (relative to the number of pixels) can be quite accurate. If $d \ll N^2$, we will again observe a significant reduction in variance, and the platelet-based MPLE will tend to have a small MSE.

Maximizing (4.1) involves adaptively pruning the C-RP based on the data. This pruning can be performed optimally and very efficiently using bottom-up, CART-like algorithms [24]. A solution to (4.2) using platelets can be computed in $O(N^2)$ operations (i.e., number of operations is proportional to number of pixels), as demonstrated below. The pruning process is akin to a "keep or kill" wavelet thresholding rule. An MPLE provides higher resolution and detail in areas of the signal or image with higher count levels (higher SNR) and/or where strong discontinuities are present. The partition underlying the MPLE is pruned to a coarser scale (lower resolution) in areas with lower count levels (low SNR) and where the data suggest that the intensity is fairly smooth. Moreover, in image analysis the possibility of wedgelet partitions and planar fits allows the MPLE to adapt to the contours of smooth boundaries in the image and smooth (but non-constant) variation in the image intensity.

4.2 MLE of Linear Signal Parameters

Before detailing my multiscale intensity estimation algorithm, I describe here my method of calculating the maximum likelihood estimate of a polynomial (one interval) fit to Poisson and multi-

nomial data. This technique extends the work of Unser and Eden and will be applied to every dyadic interval in the multiresolution algorithm [26].

The model for the intensity of the process is constrained such that $\boldsymbol{\mu} = T \cdot \boldsymbol{\theta}$, where T is a known Vandermonde matrix transforming the vector of polynomial coefficients to be estimated, $\boldsymbol{\theta}$, to the polynomial signal, $\boldsymbol{\mu}$. In the Poisson case, $\boldsymbol{\mu}$ would represent the Poisson intensity vector, while in the multinomial case, $\boldsymbol{\mu}$ would represent the probability vector and be constrained such that $\sum \boldsymbol{\mu} = 1$. There exists no closed form solution for the MLE of the parameter vector $\boldsymbol{\theta}$; therefore a numerical solution must be calculated using a gradient or steepest descent algorithm. We will see in Section 4.5 that this is a convex minimization problem, which demonstrates that an optimization algorithm will correctly identify a global maximum. The convexity of this optimization problem plays a key role in computational efficiency of the algorithm described below.

4.3 Optimal Pruning Algorithms

Observe that the structure of the penalized likelihood criterion stated in (4.1) and the likelihood factorization described in Chapter 3 allow an optimal intensity estimate to be computed quickly. The likelihood factorization allows the likelihood of the entire signal or image to be represented by a tree structure in which both likelihoods and parameter penalties of children are inherited by parents. Using this, it is possible to optimally prune an RDP of the data using a fast algorithm reminiscent of dynamic programming and the CART algorithm [5]. For the one dimensional case, consider Figures 4.1 and 4.2; in these cases, the binary trees have been optimally pruned to efficiently represent the piecewise constant and linear natures of the underlying signal intensities. The algorithm will now be detailed for one- and two-dimensional denoising.

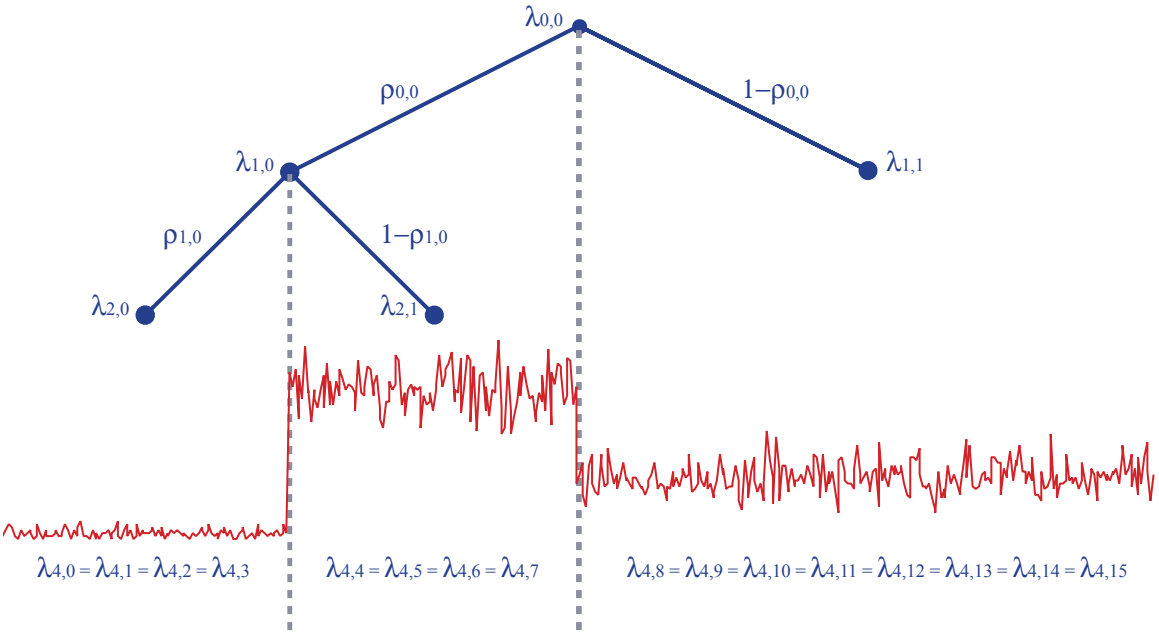


Figure 4.1 Example of optimal pruning for a piecewise constant signal. The binary tree has been pruned so each leaf node represents an interval of estimated constant intensity.

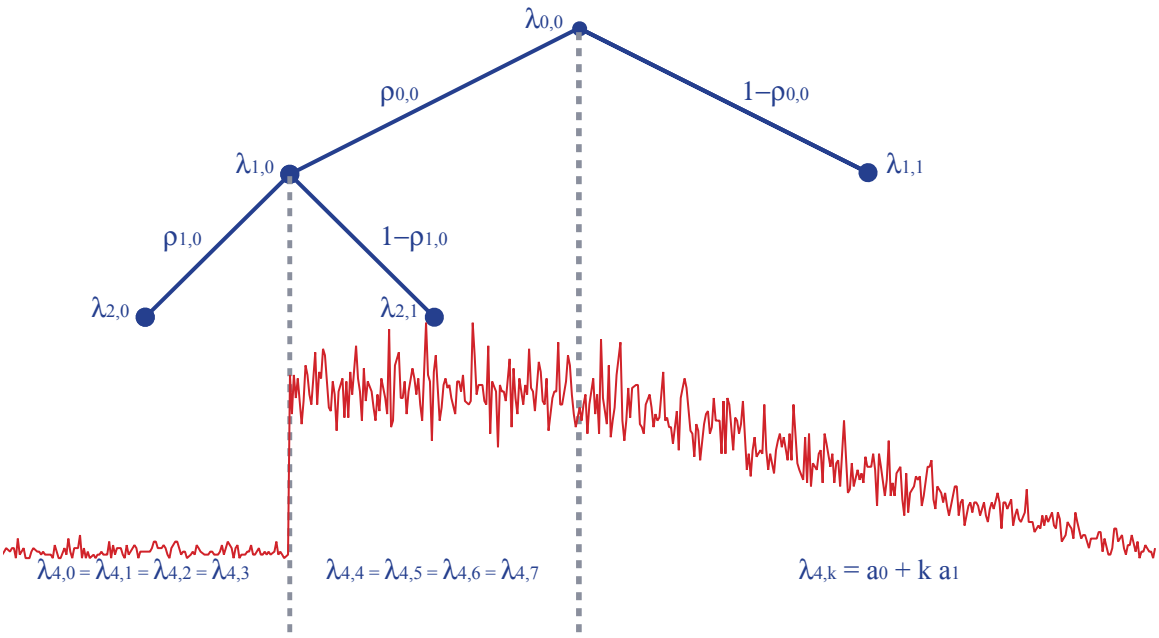


Figure 4.2 Example of optimal pruning for a piecewise linear signal. The binary tree has been pruned so each leaf node represents an interval of estimated linear or constant intensity.

4.3.1 One Dimension

The goal of the algorithm is to estimate the intensity $\boldsymbol{\mu}$ according to (4.2). In order to perform the estimation, the algorithm considers each dyadic interval in the partition of the observation interval and performs an M-ary hypothesis test. The hypotheses for each dyadic interval are as follows:

- $\mathbf{H_m}$: Degree m ($m = 0, 1, 2, \ldots$,M) polynomially varying intensity segment (terminal node)
- $\mathbf{H_{M+1}}$: Inherit from children (non-terminal node)

When the maximum polynomial degree $M = 0$, the algorithm coincides with Haar analysis. It is also possible to consider a subset of $\{\mathbf{H}_i\}$; *e.g.* if $M = 2$, one might use only $\mathbf{H_2}$ and $\mathbf{H_3}$ to restrict the set of estimates to piecewise quadratic intensities. The algorithm begins one scale above the leaf nodes in the binary tree and traverses upwards, performing a tree-pruning operation at each stage. For each node (i.e., dyadic interval) at a particular scale, the maximum likelihood parameter vector is determined for each hypothesis as described in 4.2 and the penalized log likelihoods for each hypothesis are calculated. In particular, the penalized log likelihood for the split is computed using the optimal penalized log likelihoods computed at the previous, finer scale for both of the two children. The algorithm pseudocode is in Table 4.1. In the table, $L_\gamma(\boldsymbol{\theta}_{H_i}; I_{n,j})$ denotes the penalized log likelihood term for segment $I_{n,j}$ under hypothesis H_i.

4.3.2 Two Dimensions

Equation (4.1) may be expanded using (3.4) to yield:

$$\begin{aligned} L_\gamma(\boldsymbol{\mu}) &= \log p(x_{S_0} \,|\, \mu_{S_0}) + \text{pen}\,(\mu_{S_0}) + \sum_{S \in NT(\mathcal{P})} \log p(\{x_{ch(S)}\}|x_S, \theta_S) + \text{pen}\,(\theta_S) \\ &\quad + \sum_{ch(S) \in T(\mathcal{P})} \log p(\{x_{m,n}\}_{(m/N,n/N) \in ch(S)}|x_{ch(S)}, \theta_{ch(S)}) + \text{pen}\,(\theta_{ch(S)}). \end{aligned}$$

Initialize:	$j = J - 1$
Loop:	for each node $I_{n,j}$ at level j
Calculate:	$L_\gamma(\boldsymbol{\theta}_{H_i}; I_{n,j})$ for $0 \leq i \leq M$
	$L_\gamma(\boldsymbol{\theta}_{H_{M+1}}; I_{n,j}) = \sum_{I' \in ch(I_{n,j})} L_{min}(I')$
Save:	$L_{min}(I_{n,j}) = \min_{0 \leq i \leq M+1} L_\gamma(\boldsymbol{\theta}_{H_i}; I_{n,j})$
	$\boldsymbol{\theta}_{min}(I_{n,j}) = \arg \min_{0 \leq i \leq M+1} L_\gamma(\boldsymbol{\theta}_{H_i}; I_{n,j})$
Coarsen:	Scale $j = j - 1$
Goto Loop:	if $j \geq 0$
Prune:	Perform a depth first search for terminal nodes. When a terminal node is found, record the MPLE for each terminal interval descending from the current node.

Table 4.1 Polynomial Algorithm Pseudocode

The penalty of the total image intensity parameter, μ_{S_0}, is γ because of the singular dimension of the parameter. Each nonterminal node's penalty, pen (θ_S), is either γ or $3\,\gamma$ because once the value of the parent node is known, there are only one or three free parameters necessary to describe how the intensity is distributed among the children in a wedge split or a quad split, respectively. Finally, each terminal node's penalty, pen $(\theta_{ch(S)})$, is either 0 or $2\,\gamma$; once the value of the parent node is known, no parameter is needed to represent a constant-valued region, and two parameters are needed to to represent the gradient of a planar fit to the data.

Similar to the one-dimensional algorithm, the platelet algorithm considers each dyadic square in the partition of the observed image and performs an M-ary hypothesis test. The hypotheses for each dyadic square are as follows:

- $\mathbf{H_0}$: Constant = homogeneous square (terminal node)
- $\mathbf{H_1}$: Wedgelet = two homogeneous wedges (terminal node)
- $\mathbf{H_2}$: Platelet = dyadic linear gradient on square (terminal node)
- $\mathbf{H_3}$: Wedged Platelet = linear gradients on two wedges (terminal node)

Initialize:	$j = J - 1$
Loop:	for each node $S_{j,m,n}$ at level j
Calculate:	$L_\gamma(\boldsymbol{\theta}_{H_i}; S_{j,m,n})$ for $0 \leq i \leq 3$
	$L_\gamma(\boldsymbol{\theta}_{H_4}; S_{j,m,n}) = \sum_{S' \in ch(S_{j,m,n})} L_{min}(S')$
Save:	$L_{min}(S_{j,m,n}) = \min_{0 \leq i \leq 4} L_\gamma(\boldsymbol{\theta}_{H_i}; S_{j,m,n})$
	$\boldsymbol{\theta}_{min}(S_{j,m,n}) = \arg\min_{0 \leq i \leq 4} L_\gamma(\boldsymbol{\theta}_{H_i}; S_{j,m,n})$
Coarsen Scale:	$j = j - 1$
Goto Loop if $j \geq 0$	
Prune:	Perform a depth first search for terminal nodes. When a terminal node is found, record the MPLE for each pixel descending from the current node.

Table 4.2 Platelet Algorithm Pseudocode

- $\mathbf{H_4}$: Quad split = inherit from children (non-terminal node)

It is also possible to consider a subset of $\{\mathbf{H}_i\}$; *e.g.* using only $\mathbf{H_0}$ and $\mathbf{H_4}$ coincides with the hereditary Haar analysis (with no wedgelets or platelets).

For each node (i.e., dyadic square) at a particular scale, the penalized log likelihoods for each hypothesis are calculated. In particular, the penalized log likelihood for the quad split is computed using the optimal penalized log likelihoods computed at the previous, finer scale for each of the four children. As demonstrated earlier, these four child penalized log likelihoods add to yield the penalized log likelihood of the parent node, and then this log likelihood is compared with those for the other four hypotheses. If, for a given node, the maximum penalized log likelihood is associated with a hypothesis other than a quad split, then that node is made a terminal node with parameters appropriate to the said hypothesis; its children are then pruned from the RP. The maximum log likelihood for this newly terminal node will be used for all quad split log likelihood computations for all of the node's ancestors unless it, too, is pruned during analysis on a higher scale. The algorithm pseudocode is in Table 4.2. In the table, $L_\gamma(\boldsymbol{\theta}_{H_i}; S_{j,m,n})$ denotes the penalized log likelihood term for square $S_{j,m,n}$ under hypothesis H_i.

4.4 Polynomial Risk Analysis

While the polynomial pruning algorithm described here yields the optimal MPLE, the analysis up to this point does not quantify the effectiveness of this optimal algorithm. Nowak and Kolaczyk established statistical risk bounds associated with estimating an intensity or density with piecewise constant Haar functions [5]. This analysis provided a bound on the expected error between $\boldsymbol{\mu}$ and $\tilde{\boldsymbol{\mu}}$. I demonstrate here that similar near-optimal bounds exist for piecewise polynomial approximations. In this thesis I define risk to be proportional to the expected squared Hellinger distance between two densities; that is,

$$R(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) \equiv \frac{1}{N}\mathbb{E}_{\boldsymbol{\mu}}\left[L(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu})\right] \tag{4.3}$$

where

$$L(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) \equiv H^2(p_{\hat{\boldsymbol{\mu}}}, p_{\boldsymbol{\mu}}) = \int \left[\sqrt{p(\boldsymbol{x}|\hat{\boldsymbol{\mu}})} - \sqrt{p(\boldsymbol{x}|\boldsymbol{\mu})}\right]^2 \nu(\boldsymbol{x}) \tag{4.4}$$

is the squared Hellinger distance between the true and estimated intensities, where ν is the dominating measure.

The squared Hellinger distance is an appropriate error metric here for several reasons. First, it is expressed in terms of the densities themselves instead of the parameters of any particular density, which is beneficial here because my algorithm is not limited to any one noise model. Secondly, it provides an upper bound for the affinity between two densities, a squared-error like measure tailored to the given density functions. Finally, using the squared Hellinger distance allows me to take advantage of a key inequality derived by Li and Barron in the analysis below [27, 28].

The expected loss can be bounded using the Kullback-Leibler (KL) divergence. This was accomplished in Theorem 4 of Kolaczyk and Nowak's analysis without any assumptions about the nature of the underlying signal, and hence we can use it here without alteration. For clarity, I restate

it now:

Theorem 4.1 *Let Γ_N be a finite collection of estimators $\boldsymbol{\mu}'$ for $\boldsymbol{\mu}$, and pen$(\cdot)$ a function on Γ_N satisfying the condition*

$$\sum_{\boldsymbol{\mu}' \in \Gamma_N} e^{-pen(\boldsymbol{\mu}')} \leq 1 \quad . \tag{4.5}$$

Let $\hat{\boldsymbol{\mu}}$ be a penalized maximum likelihood estimator of the form

$$\hat{\boldsymbol{\mu}}(\boldsymbol{X}) \equiv \arg \min_{\boldsymbol{\mu}' \in \Gamma_N} \left\{ -\log p\left(\boldsymbol{X} \mid \boldsymbol{\mu}'\right) + 2\, pen\left(\boldsymbol{\mu}'\right) \right\} \quad . \tag{4.6}$$

Then

$$E\left[H^2\left(p_{\hat{\boldsymbol{\mu}}}, p_{\boldsymbol{\mu}}\right)\right] \leq \min_{\boldsymbol{\mu}' \in \Gamma_N} \left\{ K(p_{\boldsymbol{\mu}}, p_{\boldsymbol{\mu}'}) + 2\, pen(\boldsymbol{\mu}') \right\} \quad . \tag{4.7}$$

In other words, assume we have an intensity $\boldsymbol{\mu}$ that we wish to estimate, and that we want to pick the very best estimate, $\hat{\boldsymbol{\mu}}$, out of a collection of estimates, Γ_N. Each of these estimates has a penalty associated with it. Γ_N may be a very large collection, but it must be finite and satisfy (4.5). We decide which estimate is "best" using the criterion in (4.6). If all these criteria are met, then the expected value of the squared Hellinger error is bounded above by the lowest possible sum of the KL divergence and penalty.

In the following Lemma, I define this collection of estimators, Γ_N, and demonstrate that it meets the criterion in (4.5). See Appendix A.3 for a proof.

Lemma 4.1 *Let Γ_N be the collection of all N-length vectors $\boldsymbol{\mu}'$ with coefficients $a'_j \in D_N[R_{j,1}, R_{j,2}]$, for some $R_{j,1} < R_{j,2}$ for $j = 0, 1, \ldots, r$ (polynomials of degree r), where $D_N[R_1, R_2]$ denotes a uniform discretization of the interval $[R_1, R_2]$ into $N^{1/2}$ equispaced values. Let $\#(\boldsymbol{\mu}')$ count the number of polynomially-varying sequences in the vector $\boldsymbol{\mu}'$, i.e., in analogy to the number of pieces*

of a piecewise polynomial function. Then

$$\sum_{\boldsymbol{\mu}' \in \Gamma_N} e^{-\gamma \log_e(N) \#(\boldsymbol{\mu}')} \leq 1 \tag{4.8}$$

for $\gamma \geq \frac{3}{2} + \frac{r}{2}$ *and* $N \geq 3$.

Thus if we define $\text{pen}(\boldsymbol{\mu}') \equiv -\gamma \log_e(N) \#(\boldsymbol{\mu}')$, we will have satisfied (4.5). The next step in bounding the risk is then to bound the KL divergence in (4.7). Kolaczyk and Nowak demonstrated that if $\boldsymbol{\mu}$ is a vector of multinomial probabilities (that is, restricted so $\sum_i \mu_i = 1$), then

$$(1/N) K(p_{\boldsymbol{\mu}}, p_{\tilde{\boldsymbol{\mu}}'}) \leq \frac{2n}{c} \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2. \tag{4.9}$$

Similarly, if $\boldsymbol{\mu}$ is a vector of Poisson intensities (that is, the sum of the elements if unknown), then

$$(1/N) K(p_{\boldsymbol{\mu}}, p_{\tilde{\boldsymbol{\mu}}'}) \leq \frac{1}{Nc} \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2. \tag{4.10}$$

Recall that we bounded $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2$ in (2.3) in Chapter 2. We will now restrict our attention to the case where $\boldsymbol{\mu}$ is a vector of multinomial probabilities. The risk bounds for a Poisson intensity vector can be derived in an analogous manner. Because the pruning algorithm described earlier estimates $\boldsymbol{\mu}$ on recursive dyadic partitions, any of the d polynomial segments represented by $\tilde{\boldsymbol{\theta}}'$ that do not lie on a dyadic partition need to be repartitioned a maximum of $\log_2(N)$ times. Hence, the number of polynomial segments in $\tilde{\boldsymbol{\theta}}'$ which lie on dyadic partition intervals is bounded as follows: $\#(\tilde{\boldsymbol{\theta}}') = O(d \log_2(N))$. These bounds can then be combined with (4.7) to produce the following bound on the expected squared Hellinger error for estimates with d polynomial pieces:

$$\min_{\boldsymbol{\mu}' \in \Gamma_N^{(d)}} \left\{ \frac{1}{N} \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2 + \frac{2\gamma \log_2(N)}{N} \#(\tilde{\boldsymbol{\mu}}') \right\} \leq O\left(d^{-2r}\right) + O\left(\frac{d}{N}\right) + O\left(\frac{1}{N}\right) + O\left(d^{-r}\left(\frac{d}{N}\right)^{1/2}\right) + O\left(\frac{d^{-r}}{\sqrt{N}}\right) + O\left(\frac{d^{\frac{1}{2}}}{N}\right) + \frac{2\gamma \log_2^2(N)}{N} d$$

This expression is minimized for $d \sim \left(\frac{\log^2(N)}{N}\right)^{\frac{-1}{2r+1}}$. Substitution then yields that $R(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu})$ is bounded above by a quantity of the order $O\left(\left(\frac{\log^2(N)}{N}\right)^{\frac{2r}{2r+1}}\right)$. For $r = 1$, that is for piecewise constant estimators, the bound is $O((\log^2(N)/N)^{2/3})$, which equals the bound derived directly for piecewise constant estimators. The penalization structure employed here minimizes the upped bound on the risk. Furthermore, this is within a logarithmic factor of the lower bound on the minimax risk, demonstrating the near-optimality of my algorithm [5].

4.5 Platelet Analysis Computational Complexity

While the previous section established the near-optimality of the pruning algorithm, a more practical concern is the computational complexity of the platelet estimation algorithm. Before examining this complexity, let us first establish the concavity of the log likelihood function, which will allow us to perform the optimization rapidly. There is no closed-form solution to the MLE of the plate parameters with respect to the Poisson or multinomial likelihood; however, they can be computed numerically, as described for the Poisson case by Unser [26]. In this case, the likelihood in the plate factorization is concave in the plate parameters, which means that a numerical optimization technique such as Newton's method or gradient descent can find the optimal parameter values. In addition, since there are only two parameters per platelet, the optimization is over a two-

dimensional space and may be solved with relatively few iterations of the optimization algorithm of choice. The parameters in the case of constant regions or wedgelets are just the count averages, and in that case no such numerical optimization technique is necessary, although the concavity result still holds. The following lemma, proved in Appendix A.4, establishes the concavity. Consider a terminal node log likelihood term for a platelet $\log p(\{x_{m,n}\}_{(m/N,n/N)\in ch(S)}|x_{ch(S)}, \theta_{ch(S)})$. The subscripts $ch(S)$ will be dropped to simplify the notation at this point. The log likelihood is multinomial and the multinomial parameters, denoted ρ, are not equal (as in the case of a constant region) but instead obey a linear model of the form $\rho = T\theta$, where θ is a two-parameter vector describing the gradient of the intensity (and hence the linear relationship between the multinomial parameters ρ) and T is a matrix relating the parameters to the gradient field.

Lemma 4.2 *The log (multinomial) likelihood function of a platelet is concave in θ.*

This lemma makes the following bounds of the computational complexity of this algorithm possible. We bound the computational complexity of performing either an "approximate" or exact MPLE, where the "approximate" estimate uses a (suboptimal) least-squares platelet fit and the exact estimate is obtained by numerically optimizing the (concave) log likelihood function for the most likely platelet fit. The theorem below is also proved in Appendix A.4.

Theorem 4.2 *(MPLE): A Haar MPLE can be computed in $O(N^2)$ operations, where N^2 is the number of pixels in the image. A wedgelet or "approximate" platelet MPLE can be computed in $O(N^3)$ operations. An exact platelet MPLE can be computed in $O(N^4)$ operations.*

The "approximate" platelet fit is used in all experiments discussed in this paper.

4.6 Penalty Parameter Selection

The selection of the penalty parameter γ in the penalized likelihood (4.1) plays a significant role in the performance of the MPLE. Large values of γ favor variance reduction and may introduce a non-negligible bias. Small values of γ result in a nearly unbiased estimator which may have a large variance. The best overall performance (as measured by MSE) depends on the choice of γ. In this section we study the performance of the MPLE in simulated denoising experiments and investigate the effect of γ. We give a simple rule for setting γ which provides very good performance over a broad range of intensity levels.

There has been a significant amount of theoretical work regarding the choice of γ in the context of wavelet (Gaussian) denoising methods as well as for Haar-based MPLE Poisson denoising [5, 24, 29]. In those works, as in Section 4.4, it is shown that near minimax optimal results are obtained with $\gamma = c\log(N)$, where N is the number of pixels and c is a constant such that (4.5) is satisfied. I choose $\gamma = c\log(N)$ because it guarantees that I meet the conditions of Lemma 4.1, which is central to the statistical risk bounds I derived in Section 4.4. This result is similar to the penalty derived for Minimum Description Length (MDL) estimation with a few distinctions [30]. In MDL, the estimator complexity is defined very precisely in terms of the minimum code length, while I measure complexity in terms of the number or partitions and parameters to be stored. In addition, MDL theory does not allow for the added flexibility of the tuning parameter c.

In this section we use the above form for γ and determine a "good" setting for the constant c. Consider the graphs in Figure 4.3. The MSE performance of platelet-based denoising of the "Bowl" test intensity and wedgelet-based denoising of the Shepp-Logan intensity is examined over a range of count levels (SNRs) and γ settings. For each case (SNR and γ) I compute the MPLE for 10 independent realizations (Poisson count images generated from each intensity image, scaled

to produce the desired SNR) and display the MSE (averaged squared error) in the plots. Note that $\gamma = \frac{1}{5}\log(\#\text{counts})$ consistently results in a low MSE, nearly the minimum MSE in all cases examined. This setting for γ is used in all denoising and deblurring experiments described in this paper. In the tomography problem (described in the next section), the nature of the projection process changes the relationship between γ and the MSE. Through similar experiments (not reported here) we have found that $\gamma = \frac{1}{50}\log(\#\text{counts})$ provides low MSEs over a range of count levels.

Similar experiments were performed for polynomial signal analysis. I have studied the performance of the MPLE in simulated intensity and density estimation experiments and investigated the effect of γ; my experiments reveal that $\gamma = \frac{1}{5}\log(\#\text{counts})$ consistently results in a low MSE over a broad range of intensity levels. This setting for γ is used in all one-dimensional experiments described in this paper. The fact that γ is proportional to $\log(\#\text{counts})$ provides a built-in adaptivity to the underlying SNR level.

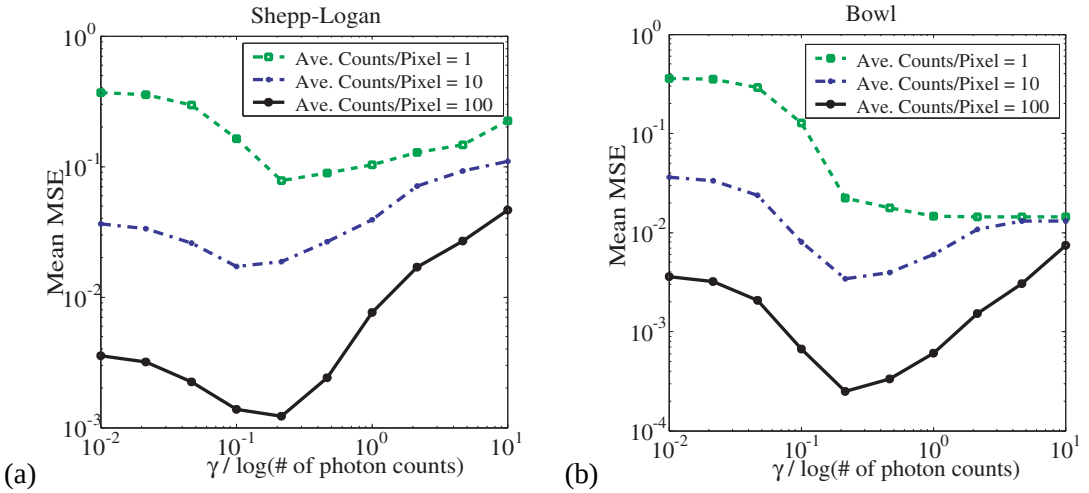


Figure 4.3 Variation of MSE with penalty parameter γ for different SNRs. (a) Wedgelet denoising on Shepp-Logan phantom. (b) Platelet denoising on quadratic bowl image.

4.7 Applications and Experiments

In a variety of applications, data is acquired by observing the times or locations at which events occur. Frequently, either by choice or due to limitations of the measuring device, events are only observed on discrete intervals, resulting in a discrete signal representing the number of events occurring within each interval. Examples include the arrival of photons at a detector or observations of a random variable; we can estimate the intensity of photon emission or the probability density function by modeling these processes as Poisson and multinomial processes, respectively.

Two applications of multiscale MPLE are now explored, and density estimation capabilities are compared with those of wavelet-based methods and kernel methods.

4.7.1 Photon-Limited Applications

The results of the denoising techniques described above are demonstrated in this section with a one-dimensional gamma ray burst example and a two-dimensional nuclear medicine image example.

One of the most fascinating problems in astrophysics today is the nature and origin of gamma ray bursts (GRBs) – quick, extremely intense bursts of gamma rays commonly associated with star formation and supernovae. Photon-counting measuring devices obtain GRB signals, indicating the presence of Poissonian noise. Kolaczyk has done preliminary work in GRB intensity estimation using Haar wavelets, an approach similar to the one described in this paper with the restriction that the intensity be piecewise constant [1]. Figure 4.4 demonstrates our algorithm's ability to produce an intensity estimate balanced between fidelity to the data and the underlying complexity. This estimate is much more faithful to the data than the piecewise constant Haar estimate in [1].

Figure 4.5(a) depicts an image of a heart obtained from a nuclear medicine study. The image was

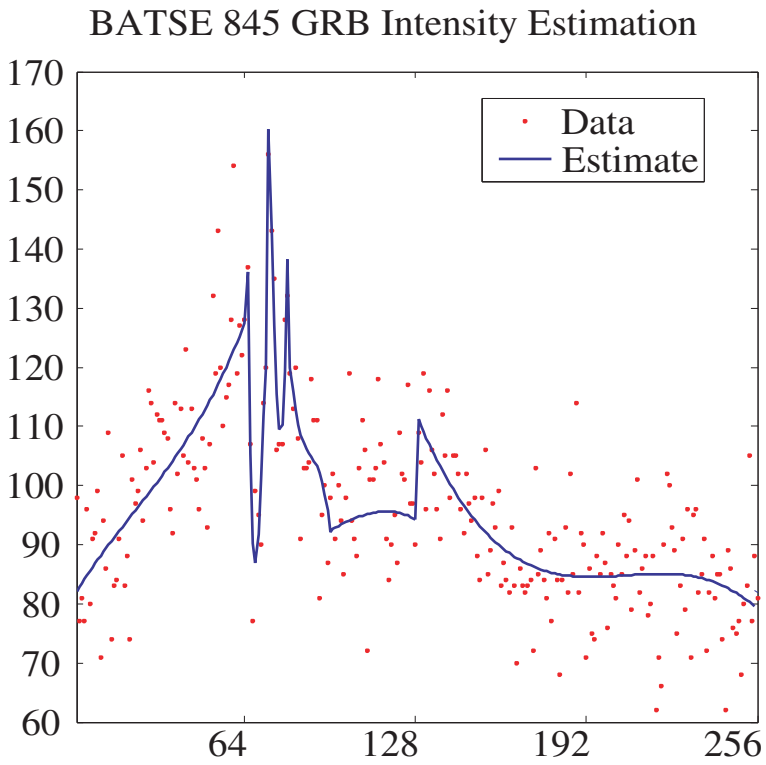


Figure 4.4 Gamma Ray Burst Intensity Estimation

obtained using the radiopharmaceutical Thallium-201. In this type of study, the radiopharmaceutical is injected into the bloodstream of the patient and moves into the heart wall in proportion to the local degree of blood perfusion. The purpose of the procedure is to determine if there is decreased blood flow to the heart muscle. Figure 4.5(f) depicts an image of the spine obtained from a nuclear medicine study. The radiopharmaceutical used here is Technetium-99m labeled diphosphonate. In bone studies such as this, brighter areas indicate increased uptake of blood in areas where bone growth or repair is occurring. Functional changes in the bone can be detected using nuclear medicine image before they will show up in X-ray images. More information on this data can be found in [31].

The hereditary Haar and platelet based MPLEs are shown in Figure 4.5. The first set of estimates (Figures 4.5 (b),(c),(g), and (h)) were generated by one pass through the denoising algorithm. For the second set (Figures 4.5 (d),(e),(i), and (j)), I applied a technique called “averaging over shifts”

SNR	Platelets	D4 Wavelets	D6 Wavelets
1	0.038	0.041	0.048
10	0.029	0.101	0.130
100	0.039	0.159	0.198

Table 4.3 $\frac{MSE}{SNR}$ for Denoising with Platelets and Wavelets

or "cycle-spinning" [32, 33]. This entails circularly shifting the raw data by a few pixels, denoising, and then shifting the estimate back to its original position. Five shifts in each direction (horizontal and vertical) yielded a total of twenty-five estimates, which are then averaged. This technique often improves denoising and reconstruction results because it reduces the dependence of the estimator on the dyadic partition. I employ this technique in all subsequent experiments.

Comparison with Wavelets

In order to compare the image denoising performance of platelets and wavelets, I generated Poisson data using Figure 2.5(a) as a scaled version of the true intensity at resolution 128×128 pixels. For SNRs (average intensities per pixel) of one, ten, and one hundred, seventy-five realizations were generated and denoised with platelets and D4 and D6 wavelets. The wavelet hard threshold level was chosen clairvoyantly for each noisy image to yield the minimum mean square error, and the MPLE penalty $\gamma = \frac{1}{2}\log(\#\text{counts})$ was used for the platelet estimator. The MSEs (averaged over all trials, and normalized by the total intensity) for each SNR are displayed in Table 4.3. Clearly platelets have a significant advantage over conventional wavelet denoising methods.

4.7.2 Density Estimation

Queuing delays are one of the most critical performance metrics in data networks. Accurate estimates of the queuing delay distribution can aid in optimizing communication network routing and service strategies. Here I apply my density estimation method to this problem. Queuing delay

measurements were generated using the *ns*-2 network simulator. In the simulation, the queue buffer size was 35 packets and the traffic was a mixture of TCP and UDP flows. The estimate in Figure 4.6 is plotted above a histogram of the measurements. Observe that the density estimate is smoothly varying and yet detects sharp peaks and discontinuities in the density.

Comparison with Wavelets

The final denoising experiment consisted of comparing the polynomial density estimation technique presented here with the commonly used normal kernel and wavelet methods. I consider three densities, generated from the well known test functions 'HeaviSine' (Figure 4.7), 'Blocks' (Figure 4.8), and 'Bumps' (Figure 4.9) [34]. Probability mass functions (pmfs) of length 1024 were constructed from these test signals by shifting them to be strictly positive and normalizing them to one. This mimics a density estimation problem in which the measurement system has an accuracy of 10 bits. To simulate a set of observations from each density, approximately 1024 iid samples from each pmf were generated by a random number generator. Notice that the total number of samples is approximately the same as the dimension of the pmfs, simulating the ideal situation in which the data are not binned.

The densities were estimated with the MPLE algorithm described in this paper, the normal kernel method described in [35] (p. 56) and the wavelet hard-thresholding method described in [7]. Ten estimations were performed using the kernel method (with adaptively chosen optimal bandwidths for each observation), the wavelet thresholding method (using D6 wavelets with threshold levels chosen clairvoyantly in each simulation to obtain the best MSE), and our MPLE piecewise quadratic polynomial fits. In the wavelet thresholding case, the clairvoyant thresholds could not be obtained in practice, but here they provide a lower bound on the achievable MSE performance for

	HeaviSine	Bumps	Blocks
Normal Kernel	1.24	1.98	1.23
Clairvoyant D6 Wavelet	2.06	1.71	2.02
Multiscale MPLE	1	1	1

Table 4.4 Density Estimation MSE, Normalized to MSE of Multiscale MPLE Estimate

any practical hard-thresholding scheme. The MSE of these estimates normalized to the MSE of the MPLE piecewise polynomial estimates are displayed in Table 4.4. Clearly, even without the benefit of setting the penalization factor clairvoyantly or data adaptively, the multiscale MPLE yields much smaller errors than the kernel and wavelet techniques for both smooth and spiky densities. Notably, unlike wavelet-based techniques, the polynomial technique is guaranteed to result in a non-negative density estimate.

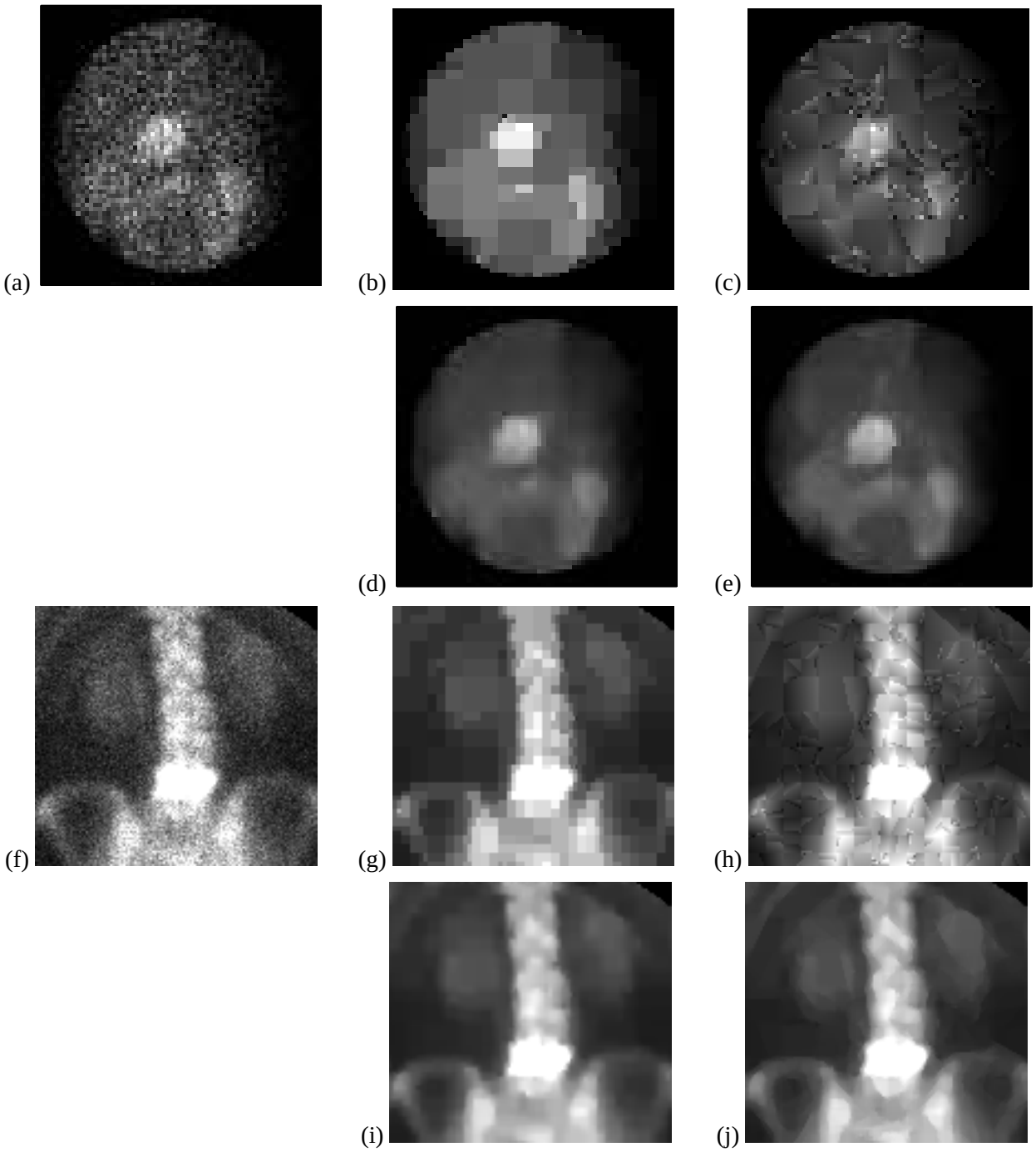


Figure 4.5 Denoising in nuclear medicine. (a) "Raw" nuclear medicine cardiac image (64×64 pixels). (b) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (c) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (d) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts. (e) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts. (f) "Raw" nuclear medicine spine image (256×256 pixels). (g) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (h) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (i) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts. (j) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts.

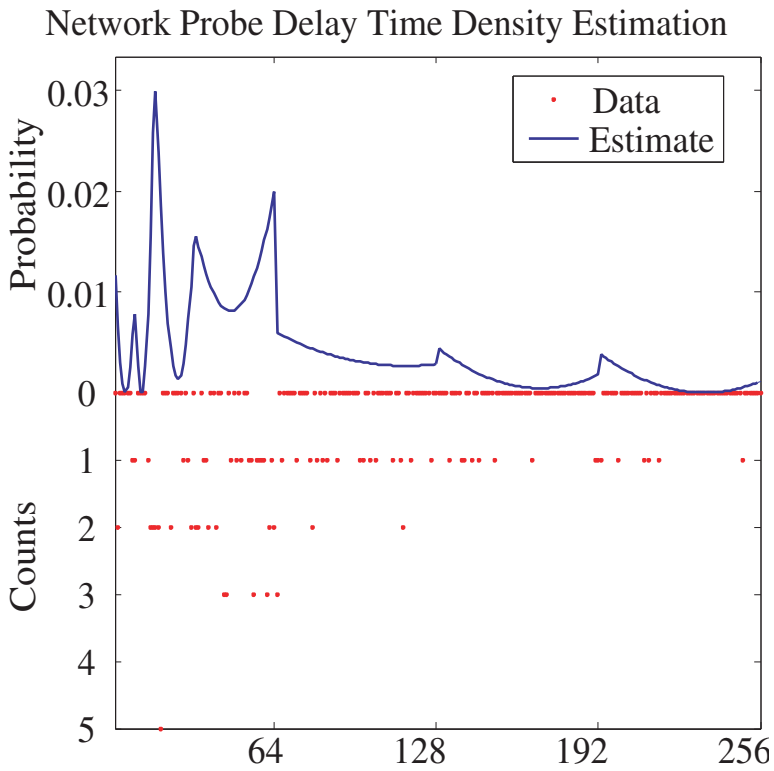


Figure 4.6 Network Queue Density Estimation

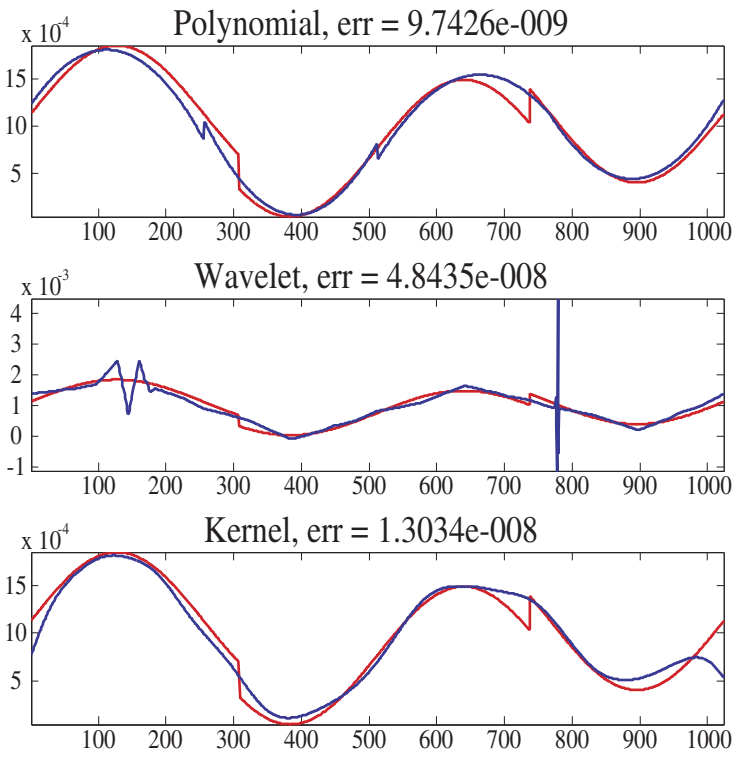


Figure 4.7 Heavisine Density Estimation Performance

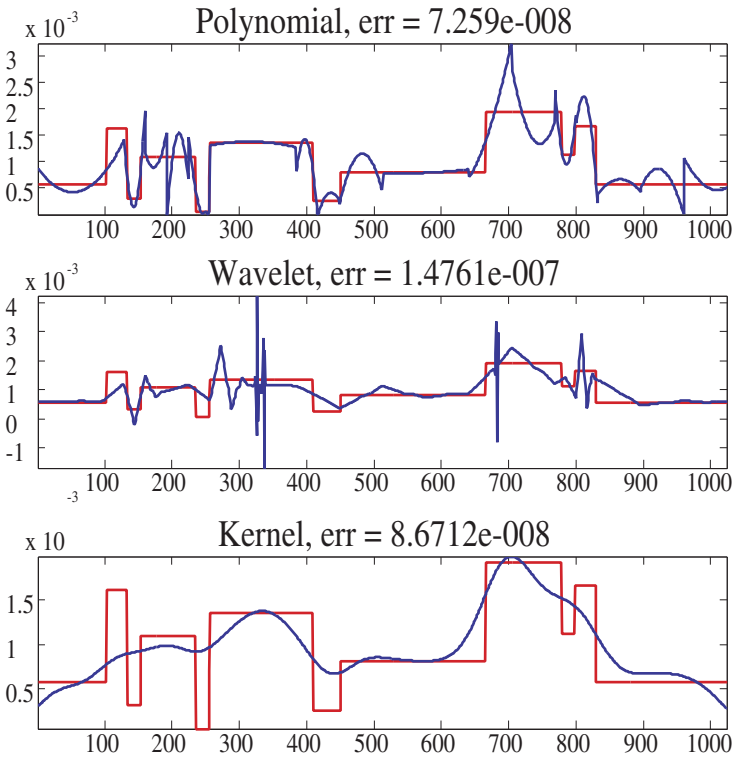


Figure 4.8 Blocks Density Estimation Performance

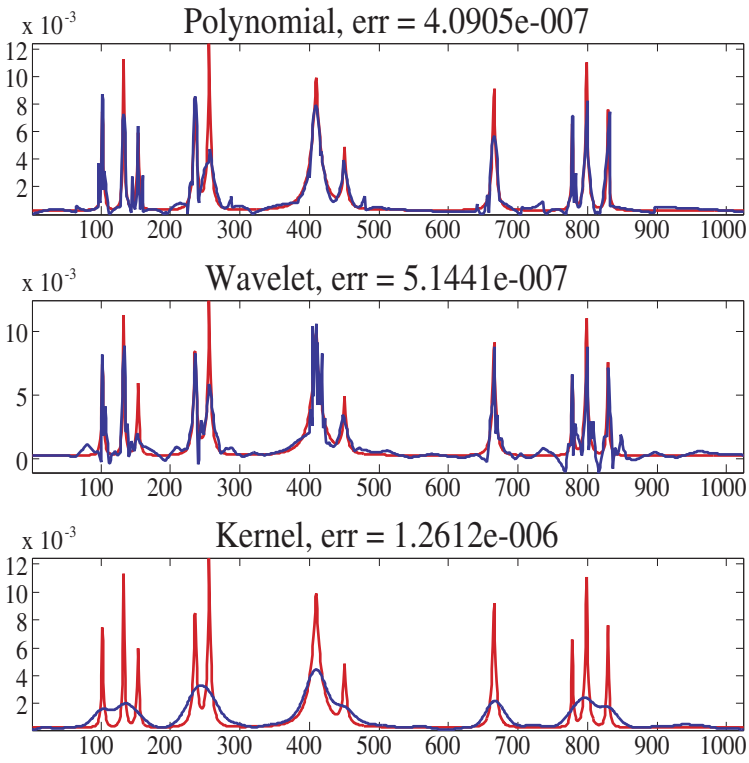


Figure 4.9 Bumps Density Estimation Performance

Chapter 5

Deblurring and Reconstruction

Statistical methods for photon-limited signal and image deblurring and reconstruction are especially effective since they can account for the special properties of the Poisson distribution. The mean and variance of a Poisson process are equal to the intensity. The intensity/mean is the "signal" of interest and the variability of the data about the mean can be interpreted as "noise." Thus, as the intensity varies spatially as a function of anatomy, structure, or function, so does the signal-to-noise ratio. In this sense it could be said that the noise in photon-limited imaging is signal-dependent.

The maximum likelihood estimator (MLE) is the most popular statistical tool and is routinely applied in scientific and clinical practice. In most cases the MLE must be computed numerically, and the most common method for this purpose is the expectation-maximization (EM) algorithm [21, 36, 37] (also known as the Richardson-Lucy algorithm in the context of Poisson data [38]). EM algorithms have been widely studied and applied and provide a very simple means to compute the MLE. However, the maximum likelihood criterion is not always useful. For example, in PET and SPECT the resulting system of equations is very ill-posed, and often the MLE is extremely noisy (highly variable). The EM algorithm may even diverge.

To combat these problems the maximum likelihood criterion can be replaced with maximum penalized likelihood criteria [39–42]. These criteria are constructed by adding a penalizing function to the Poisson log-likelihood function. The penalizing function measures the smoothness of the intensity image and assigns weaker penalties to smoother intensities and stronger penalties to more irregular intensities. Assumptions of smoothness are not unreasonable in practice. For example, radioactive pharmaceuticals diffuse smoothly in regions of homogeneous tissue, resulting in smoothly

varying intensities within organs. Basic anatomical and physiological considerations suggest that extremely irregular intensities are unnatural.

The penalizing function can be specified by an ad hoc smoothness measure, a Bayesian prior distribution for the intensity image or a complexity measure [39, 40, 43, 44]. Smoothness measures include simple quadratic functions that measure the similarity between the intensity values of neighboring pixels, as well as non-quadratic measures that better preserve edges. Similar penalty functions result from Markov Random Field (MRF) priors. Complexity measures are usually associated with an expansion of the intensity image with respect to a set of basis functions (*e.g.* Fourier or wavelet) and count the number of terms retained in a truncated or pruned series; the more terms (basis functions) used to represent the image, the higher the complexity measure [45, 46]. An intensity that maximizes a penalized likelihood criterion is called a maximum penalized likelihood estimator (MPLE). Many algorithms (*e.g.* EM algorithms or close relatives) have been developed to compute MPLEs under various observation models and penalization schemes [47].

An alternative to MPLE-based methods is the "stopped" EM-MLE solution [48]. The idea here is to stop the iterations of the EM-MLE algorithm at a suitable point, before it converges to the undesirable MLE or diverges. The stopped EM-MLE algorithm implicitly produces a smooth solution and is perhaps the most widely applied approach in practice. The popularity of this procedure is probably due to its simplicity and the availability of very fast EM-type algorithms for the basic maximum likelihood criterion. However, MPLEs based on complexity penalized multiscale (or "multiresolution") image representations not only compare favorably with EM-MLE reconstructions in terms of computational speed, but also can provide reconstructions that are significantly superior to the best stopped EM-MLE solution.

5.1 Poisson Inverse Problems

In many medical imaging applications, the detected photons are indirectly related to the object of interest. For example, confocal imaging systems may involve a blurring process, and SPECT and PET require the measurement of tomographic projections. Blurring and tomographic projections can be described statistically as follows. Photons are emitted (from the emission space) according to an intensity $\boldsymbol{\mu}$. Those photons emitted from location (k, l) are detected (in the detection space) at position (m, n) with transition probability $p_{k,l,m,n}$. In such cases, the measured data are distributed according to

$$x_{m,n} \sim \text{Poisson}\left(\sum_{k,l} p_{k,l,m,n}\, \mu_{k,l}\right). \tag{5.1}$$

Notice that the intensity of the observation is $\sum_{k,l} p_{k,l,m,n}\mu_{k,l}$, rather than $\mu_{m,n}$ as in the direct case. The transition probabilities $p_{k,l,m,n}$ represent the blurring or projection process. The recovery $\boldsymbol{\mu}$ from $\boldsymbol{x}$ is an *inverse problem*; one must invert the effects of $\boldsymbol{p} \equiv \{p_{k,l,m,n}\}$.

The Poisson likelihood function of $\boldsymbol{x}$ given $\boldsymbol{\mu}$ is denoted $p(\boldsymbol{x}|\boldsymbol{\mu})$. The log-likelihood is

$$\log\, p(\boldsymbol{x}|\boldsymbol{\mu}) \;=\; \sum_{m=0}^{M_1}\sum_{n=0}^{M_2}\left(-\sum_{k,l=0}^{N-1} p_{k,l,m,n}\mu_{k,l} + x_{m,n}\log\left(\sum_{k,l=0}^{N-1} p_{k,l,m,n}\mu_{k,l}\right) - \log x_{m,n}!\right), \tag{5.2}$$

where $M_1 \times M_2$ is the dimension of the detection space and $N \times N$ is the dimension of the emission (or image) space. M_1 and M_2 are arbitrary, but for convenience assume that N is a power of two. The likelihood function here is much more complicated due the the presence of $\boldsymbol{p}$, and a multiscale likelihood factorization is not possible. Therefore the inverse problem faced in deblurring (restoration) and tomographic reconstruction cannot be solved by simple tree-pruning methods. The multiscale approach developed in the previous chapters can however be applied within the context

of an EM algorithm. The key idea in the EM algorithm (as it generally applies to photon-limited imaging problems) is that the indirect (inverse) problem can be broken into two subproblems; one which involves computing the expectation of the unobserved direct data (as though no blurring or projection took place) and one which entails estimating the underlying image from this expectation.

Central to the EM algorithm is the notion of "complete" data, defined as $\boldsymbol{z} = \{z_{k,l,m,n}\}$, where $z_{k,l,m,n}$ denotes the number of photons emitted from (k, l) and detected at (m, n) [47]. The complete data are Poisson distributed according to

$$z_{k,l,m,n} \sim \text{Poisson}(\mu_{k,l}\, p_{k,l,m,n}) \quad . \tag{5.3}$$

Hence the observed data $\boldsymbol{x}$ in (5.1) are given by $x_{m,n} = \sum_{k,l} z_{k,l,m,n}$. Additionally, were we able to observe $\boldsymbol{z} = \{z_{k,l,m,n}\}$, the direct emission data for each location (k, l) is given by sums of the form $y_{k,l} \equiv \sum_{m,n} z_{k,l,m,n}$, from which it follows that $y_{k,l} \sim \text{Poisson}(\mu_{k,l})$. Therefore, if $\boldsymbol{z}$ were known, we could avoid the inverse problem altogether and simply deal with the issue of estimating a Poisson intensity given direct observations.

In earlier work, Nowak and Kolaczyk developed an EM algorithm that can be adapted to the MPLE considered here [49]. The approach developed in this earlier work employed a Haar-based Bayesian estimator. The priors of the estimator were quite different from the complexity penalty used here. In addition, that approach only applies to the Haar case and is not applicable to wedgelet and platelet-based MPLEs. Furthermore, tomography experiments with the multiscale Bayesian method showed that its performance was competitive with (but could be slightly inferior to) stopped EM-MLE reconstruction methods [49]. Here we demonstrate that the Haar and wedgelet based MPLEs perform *better* than even the best possible stopped EM-MLE procedure.

5.2 EM-MLE Reconstruction

It is well-known that the maximizer of (5.2) cannot be expressed in closed-form, but the concavity of the log-likelihood allows a numerical determination. While in principle any numerical optimization method could be used, the iterative EM algorithm, as first proposed for this problem in Shepp and Vardi's work, has a number of features that make it especially desirable, most notably its natural, probabilistic formulation, computationally straightforward calculations at each iteration step, and numerical stability [47, 50]. Moreover, it can be shown that the EM algorithm monotonically increases the log-likelihood at each iteration and converges to a global (not necessarily unique) point of maximum for (5.2) [21].

Unfortunately, due to the ill-posed nature of the likelihood equations, the variance of the MLE can be quite high, particularly for applications involving low counts. In fact, in many cases the MLE is practically useless. A popular remedy is to stop the EM algorithm prior to convergence (e.g., [48]). Stopping the algorithm acts implicitly as a smoothing operation and can produce acceptable results. However, it may be preferable to abandon the strictly likelihood-based perspective altogether, and approach the inverse problem with a different criterion, one that smoothes through a well-defined optimal solution, while still providing useful and meaningful results.

5.3 EM-MPLE using Platelet Approximations

The maximum penalized likelihood function employed here is

$$L_\gamma(\boldsymbol{\mu}) \;\equiv\; \log p(\boldsymbol{x} \,|\, \boldsymbol{\mu}) \;-\; \gamma \,\{\#\boldsymbol{\theta}\}, \tag{5.4}$$

where $\{\#\boldsymbol{\theta}\}$ is the number of parameters in the vector $\boldsymbol{\theta}$. Again, keep in mind that the intensity is a function of the partition and the multiscale parameters; i.e., $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}, \mathcal{P})$. The constant $\gamma > 0$ is again a weight that balances between fidelity to the data (likelihood) and complexity penalty. In maximizing this function, the resulting reconstruction will be one that has a relatively high likelihood value as well as a relatively low complexity Haar, wedgelet, or platelet representation. The EM algorithm can be easily modified to produce a sequence of reconstructions that monotonically increase this function. To simplify the notation in the derivation of the EM algorithm, the intensity functions will be denoted by $\boldsymbol{\mu}$ (without explicit indication of the dependence on $\mathcal{P}$ and $\boldsymbol{\theta}$).

The EM algorithm is based on consideration of the following alternative or surrogate function:

$$L_\gamma^c(\boldsymbol{\mu}) \;\equiv\; \log p(\boldsymbol{z} \,|\, \boldsymbol{\mu}) \;-\; \gamma\,\{\#\boldsymbol{\theta}\}, \tag{5.5}$$

where the likelihood of the observed data $\boldsymbol{x}$ is replaced by the likelihood of the (unobserved) complete data $\boldsymbol{z}$. The E-Step of the algorithm computes the conditional expectation of $L_\gamma^c(\boldsymbol{\mu})$ given the observed data. This expectation is computed using the Poisson distribution corresponding to the previous iterate of the algorithm. That is, the E-Step at the $i+1$-th iteration computes

$$Q(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}) \;\equiv\; E_{\boldsymbol{\mu}^{(i)}}\left[L_\gamma^c(\boldsymbol{\mu})|\boldsymbol{x}\right], \tag{5.6}$$

where the subscript on the expectation is the previous iterate $\boldsymbol{\mu}^{(i)}$ of the algorithm and indicates that the expectation is computed using the Poisson distribution of that intensity. Note that the conditional expectation is

$$E_{\boldsymbol{\mu}^{(i)}}\left[L_\gamma^c(\boldsymbol{\mu})|\boldsymbol{x}\right] = E_{\boldsymbol{\mu}^{(i)}}\left[\log p(\boldsymbol{z} \,|\, \boldsymbol{\mu})|\boldsymbol{x}\right] \;-\; \gamma\,\{\#\boldsymbol{\theta}\}. \tag{5.7}$$

The penalty term does not depend on $\boldsymbol{x}$ and so it is simply a constant in this step. Therefore, the E-Step here is equivalent to the E-Step of the conventional EM-MLE algorithm; it computes $E_{\boldsymbol{\mu}^{(i)}}\left[\log p(\boldsymbol{z} \mid \boldsymbol{\mu})|\boldsymbol{x}\right]$. The complete data log likelihood $\log p(\boldsymbol{z} \mid \boldsymbol{\mu})$ happens to be a linear function of $\boldsymbol{z}$ and so this calculation simplifies to computing the $\boldsymbol{z}^{(i)} \equiv E_{\boldsymbol{\mu}^{(i)}}[\boldsymbol{z}|\boldsymbol{x}]$. A closed-form expression for this calculation can be found in [21]. In general each E-Step requires $O(M_1 M_2 N^2)$ operations, but often the structure of $\boldsymbol{p}$ can be exploited to simplify the calculation (e.g., if $\boldsymbol{p}$ corresponds to a convolution and $M_1 = M_2 = N$, then the E-Step can be calculated in $O(N^2)$ operations).

The M-Step of the algorithm is the maximization of $Q(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu})$ over $\boldsymbol{\mu}$. The penalty, which is a function of $\boldsymbol{\mu}$, plays a key role in this step. Based on the derivation of the EM algorithm in Nowak and Kolaczyk's work, $Q(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu})$ can have the form

$$Q(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}) \;=\; \log p(\boldsymbol{y}^{(i)} \mid \boldsymbol{\mu}) \;-\; \gamma\,\{\#\boldsymbol{\theta}\} \;+\; C(\boldsymbol{x}), \tag{5.8}$$

where $\boldsymbol{y}^{(i)}$ is the unobserved direct data (computed from $\boldsymbol{z}^{(i)}$) and $C(\boldsymbol{x})$ is a constant depending on $\boldsymbol{x}$ but not $\boldsymbol{\mu}$. Thus, the M-Step is equivalent to maximizing $\log p(\boldsymbol{y}^{(i)} \mid \boldsymbol{\mu}) \;-\; \gamma\,\{\#\boldsymbol{\theta}\}$. Remarkably, because $\boldsymbol{y}$ is the direct data, the M-Step is equivalent to the denoising calculation discussed in the previous section with $\boldsymbol{y}^{(i)}$ in place of $\boldsymbol{x}$. To compute the M-Step, first compute $y^{(i)}_{k,l} = \sum_{m,n} z^{(i)}_{k,l,m,n}$, where $\boldsymbol{z}^{(i)}$ is computed in the E-Step. Then carry out the pruning algorithm described in the previous section to compute the vector $\boldsymbol{\theta}$ (and from $\boldsymbol{\theta}$ construct $\boldsymbol{\mu}$) that maximizes $Q(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu})$. This results in the next iterate of the algorithm, $\boldsymbol{\mu}^{(i+1)}$.

The EM algorithm must be initialized, and in all examples considered here the initialization is a constant image of total intensity equal to the total number of counts detected. The penalized log likelihood of each iterate is equal to or greater than that of the previous iterate (monotonicity

of EM), however the algorithm is not guaranteed to converge to a global maximum of (5.4). Two applications of the MPLE and EM algorithm are examined next.

5.4 Applications and Experiments

The ill-posed inverse problems discussed above encompass confocal microscopy image deblurring and tomographic reconstruction. In the following sections, I present simulation results and comparisons with results of more traditional techniques for both of these applications.

5.4.1 Emission Computed Tomography

Here I consider the application of my framework to emission computed tomography (ECT). In medical ECT, a human subject is injected with a radioactive pharmaceutical specifically designed for absorption in certain bodily organs or tissues. The distribution of this pharmaceutical within the subject can provide functional and/or anatomical diagnostic information. To obtain a mapping of pharmaceutical uptake, data are collected by detecting beta-ray photons that are emitted from within the subject as the pharmaceutical decays. From these *projection* data (the indirect data $\boldsymbol{y}$ in our problem), we wish to estimate the underlying pharmaceutical distribution (intensity $\boldsymbol{\mu}$). The probability transition matrix $\boldsymbol{p}$ is derived from the physics and geometry of the detection device and data collection process [21].

In Figure 5.1 we illustrate the application of our multiscale framework to a simulated single photon ECT problem. The underlying 2-d intensity in our simulation is the common Shepp-Logan model, a standard benchmark in SPECT. The intensity $\boldsymbol{\mu}$ is a 64×64 square image shown in Figure 5.1(a). The transition probability matrix $\boldsymbol{p}$, corresponding to a *parallel strip-integral geometry* with 80 radial samples and 60 angular samples distributed uniformly over 180°, was generated by

the *ASPIRE* software system [51]. $\boldsymbol{p}$ was applied to $\boldsymbol{\mu}$, and I used a standard Poisson random number generator to synthesize the projection data $\boldsymbol{y}$. The Shepp-Logan model is piecewise constant, and so I employ a wedgelet-based MPLE instead of using platelets.

For comparison, Figure 5.1(b) shows the very best likelihood-based reconstruction obtained by stopping the likelihood-based EM algorithm at the reconstruction having the smallest squared error, which is impossible to determine in practice since the true intensity is unknown. The Haar and wedgelet based MPLE algorithms converge to good reconstructions, better in quality than the best possible reconstruction obtained by the stopped likelihood-based EM algorithm. The wedgelet based MPLE performs best overall. Note that, unlike the EM-MLE algorithm for which the error diverges, the MPLE errors settle down towards minimums as we continue to iterate.

5.4.2 Astronomical Image Reconstruction

The computed tomography example demonstrates the edge-preserving capabilities of our wedgelet-based reconstruction algorithm. Wedgelets allow extra flexibility in defining edges in the recursive dyadic partition framework, which leads to a decrease in the blocking artifacts resulting from the Haar-based MPLE. Alternatively, the averaging-over-shifts technique described in Chapter 4 can result in a reasonable reduction of blocking artifacts when a large number of shifts are used. This is demonstrated here in the context of astronomical imaging.

Astronomical images taken by CCD cameras in ground telescopes exhibit a combination of blurring, induced by the atmosphere and telescope point spread function (PSF), and Poisson noise associated with the CCD mechanism. Figure 5.2(a) contains an original noisy and blurred image of the irregular galaxy NGC 1569. In this example I employ the hereditary Haar-based MPLE and average over 256 shifts at each iteration. The result of the stopped EM-MLE at iteration four is

displayed in 5.2(b). The final EM-MLE and Haar-based EM-MPLE images are displayed in Figures 5.2(c) and (d), respectively. Clearly the EM-MLE algorithm does not converge to a useful estimate of the intensity. Even the stopped EM-MLE image is not as smooth and has not accomplished as much deblurring as the Haar-based algorithm. The averaging-over-shifts significantly limits the blocking artifacts in the Haar-based result, highlighting the need for further investigation in to shift-invariant extensions of this work.

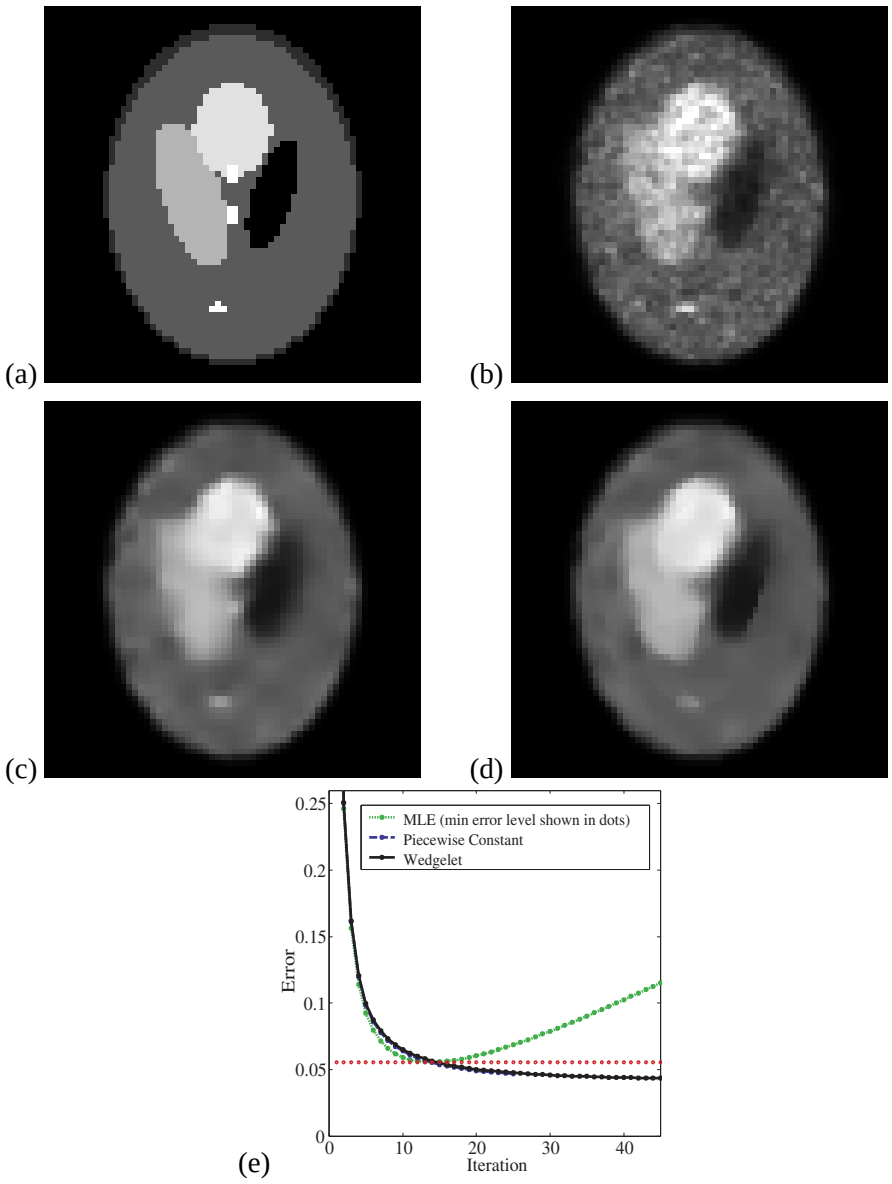


Figure 5.1 SPECT simulation. (a) Shepp-Logan phantom (64×64). (b) Best EM-MLE reconstruction (stopped at 14 iterations which gave the minimum squared error reconstruction). (c) Non-hereditary Haar-based MPLE (averaged over 5×5 shifts). (d) Wedgelet-based MPLE (averaged over 5×5 shifts). (e) Error vs. iteration for all three methods. In all cases, $\gamma = \frac{1}{50}\log(\#\text{counts})$ and convergence was declared when $||\boldsymbol{\mu}^{(i+1)} - \boldsymbol{\mu}^{(i)}||_2 / ||\boldsymbol{\mu}^{(i)}||_2 < 10^{-5}$ (roughly 25 iterations in these cases.) Note that some edges inside the Shepp-Logan phantom are sharper and more pronounced in the wedgelet reconstruction than in the Haar or MLE reconstructions.

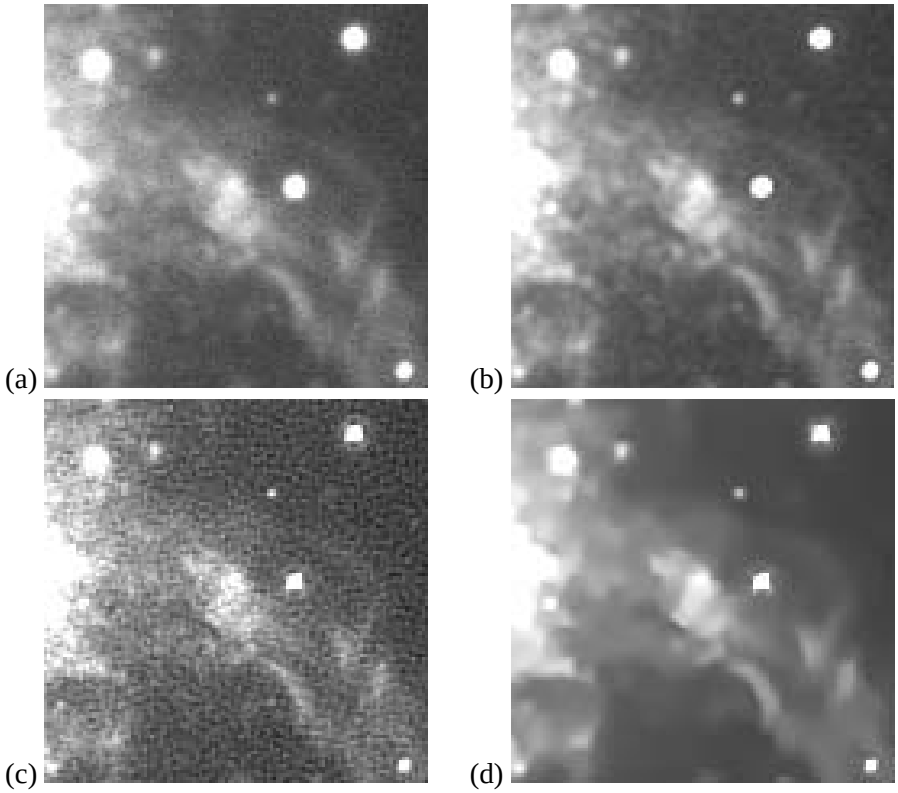


Figure 5.2 Astronomical Imaging Results. (a) Original data. (b) EM-MLE result, stopped after four iterations. (c) EM-MLE result after convergence. (d) Haar-based EM-MLE result after convergence. In this simulation, $\gamma = \frac{1}{10}\log(\#\text{counts})$ and convergence was declared when $||\boldsymbol{\mu}^{(i+1)} - \boldsymbol{\mu}^{(i)}||_2 / ||\boldsymbol{\mu}^{(i)}||_2 < 10^{-5}$ (roughly 30 iterations in these cases.)

5.4.3 Confocal Microscopy

The combination of wedgelets, platelets, and averaging-over-shifts can result in high-quality image reconstruction, as demonstrated here in the context of confocal microscopy. Confocal microscopy is used to obtain volume images of small fluorescent objects with high spatial resolution [4]. To generate a confocal fluorescence microscope (CFM) image, the microscope performs a 3D scan of the object. At each point in the scan, a photo multiplier tube measures the emission of fluorescence light from the object, essentially acting as a photon counter.

Due to the geometry of these microscopes, a "blurring" is introduced into the measurement process. This distortion of the image is commonly modeled by the convolution of the true image with the point spread function of the microscope. Since the arrival of fluorescence light at the photo multiplier tube can be modeled as a Poisson process, the "de-blurring" and estimation process may be viewed as a Poisson inverse problem well suited to the application of iterative estimation using the EM algorithm, as detailed in [37]. The E-step of the algorithm is rapidly computed using the FFT to perform the convolution. In practice, the M-step commonly consists of performing maximum likelihood estimation, but these estimates are known to diverge from the true image after several iterations. A common use for CFMs is the imaging the dendritic spines of neurons. In order to demonstrate the capabilities of the platelet-based MPLE in confocal microscopy, I have created a scale phantom of an image of a dendrite segment with thin, stubby, and mushroom spines. See [52] for descriptions of dendrites and the various types of spines. Some regions of the object will be closer to the detector or exhibit more fluorescence than others, resulting in image gradients and making this an excellent candidate for platelet analysis.

Figure 5.3(a) contains the 128×128 phantom developed for this experiment, and Figure 5.3(d) contains the blurred and noisy data. The best MLE estimate in Figure 5.3(b) is the MLE image at

the iteration when its L_2 error was smallest. In practice it is not possible to know which iteration yielded the best ML estimate, but it is included here to demonstrate that MPLE algorithms converge to a point with lower error than the best possible image obtainable with the commonly used stopped-MLE technique [48]. Figure 5.3(c) is a closeup of a small region of Figure 5.3(b). The averaged-over-shifts Haar-based MPLE appears in Figure 5.3(e), and its closeup in Figure 5.3(f). Likewise, the averaged-over-shifts platelet-based MPLE appears in Figure 5.3(h) and its closeup in Figure 5.3(i). Finally, Figure 5.3(g) plots the L_2 error of each of the three estimates at each iteration. After several iterations the EM-ML estimate worsens considerably with each subsequent iteration. In contrast, both MPLEs converge eliminating the need to choose which iteration is the best stopping point, as done with the EM-MLE. Furthermore, the converged MPLEs exhibit significantly less L_2 error than the best MLE.

Figure 5.4 demonstrates the capabilities of platelet analysis on a real confocal microscopy image of a dendritic spine. In this case the structure of the dendritic spines is of critical interest to researchers. As demonstrated, the platelet-based MPLE can perform better than both the stopped MLE and the Haar-based MPLE. The platelet algorithm effectively extracts image gradients and edges better than the Haar-based algorithm, but it converges with fewer artifacts than the EM-MLE algorithm.

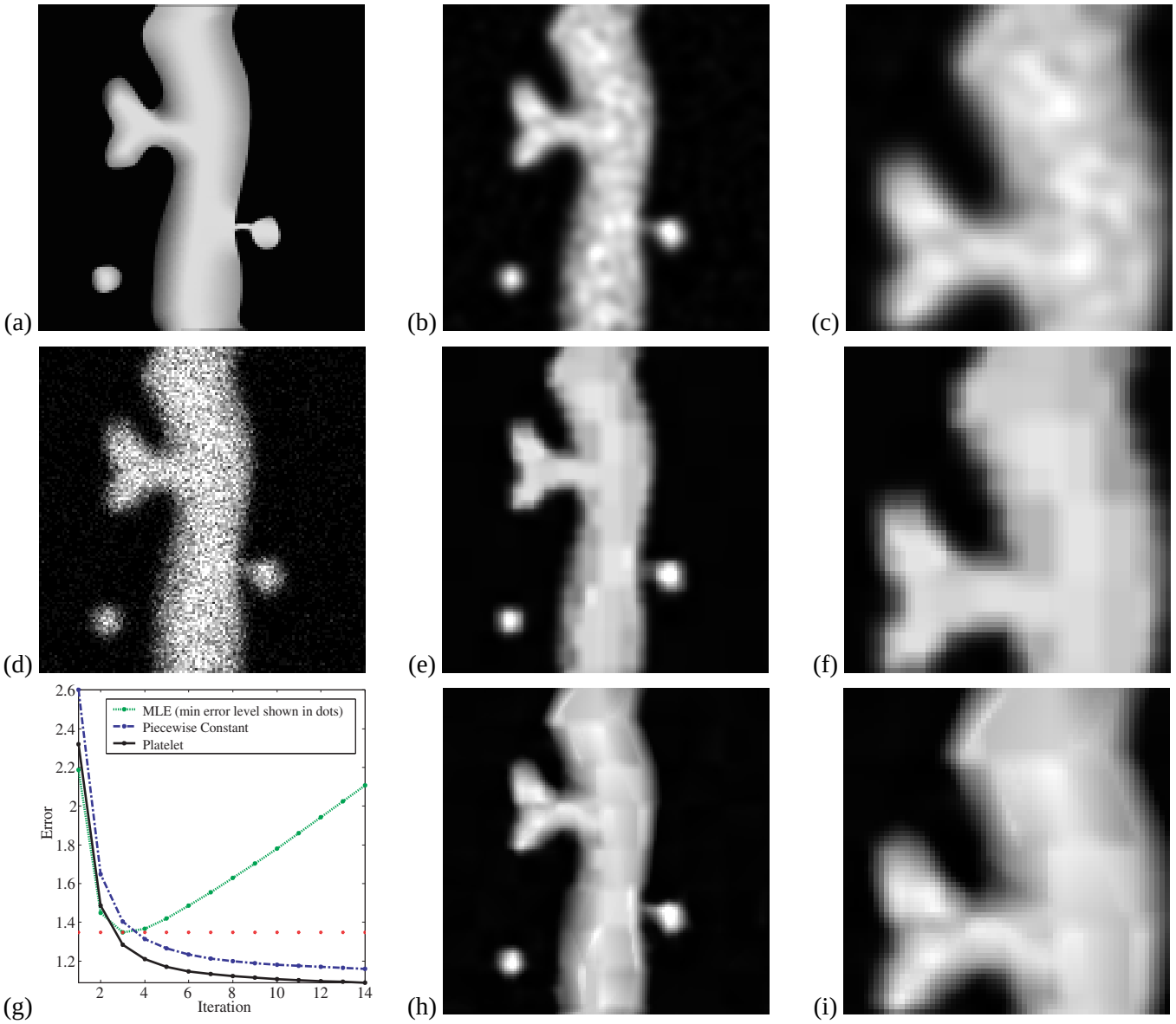


Figure 5.3 Confocal microscopy simulation. (a) Phantom (128×128 pixels). (b) Best EM-MLE restoration. (c) Best EM-MLE (zoomed). (d) Blurred and noisy phantom. (e) Hereditary Haar MPLE (averaged over 3×3 shifts). (f) Hereditary Haar MPLE (zoomed). (g) Error decay by iteration. (h) Platelet MPLE (averaged over 3×3 shifts), (i) Platelet MPLE (zoomed). In all cases, $\gamma = \frac{1}{5}\log(\#\text{counts})$ and convergence was declared when $||\boldsymbol{\mu}^{(i+1)} - \boldsymbol{\mu}^{(i)}||_2 / ||\boldsymbol{\mu}^{(i)}||_2 < 10^{-5}$ (14 iterations in this case).

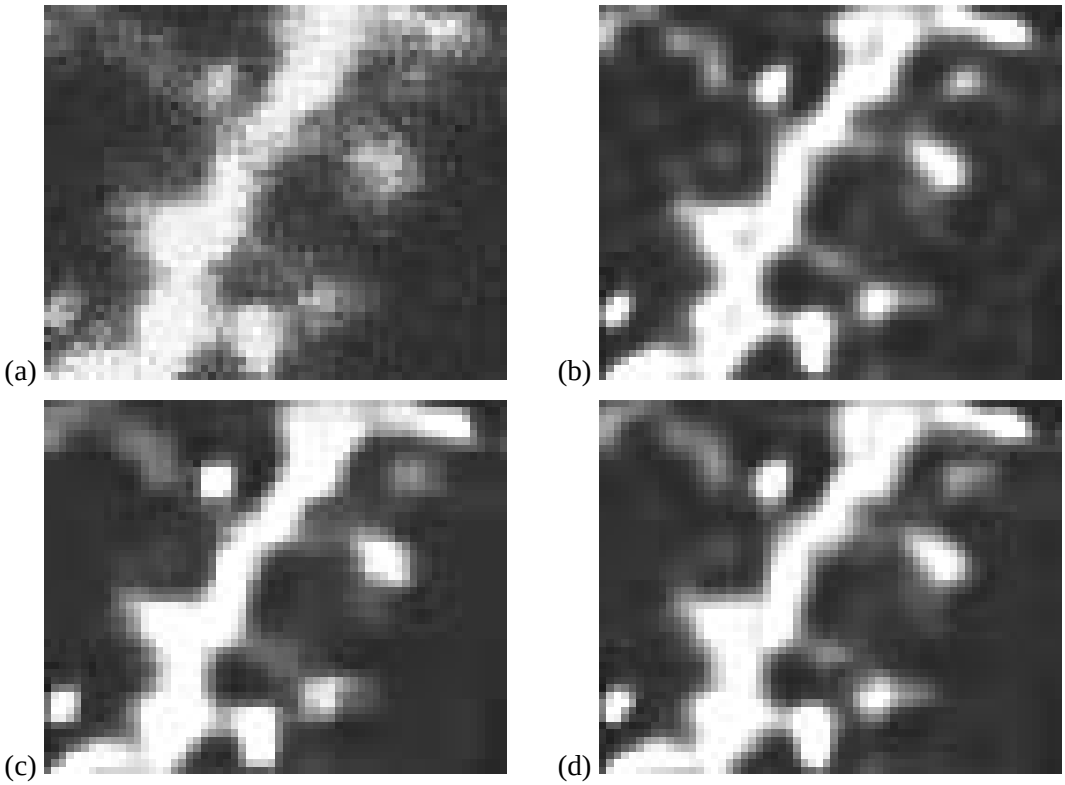


Figure 5.4 Confocal microscopy application. (a) Raw data (64×64 pixels). (b) EM-MLE restoration after 30 iterations. (c) Haar-based MPLE (averaged over 3×3 shifts). (d) Platelet-based MPLE (averaged over 3×3 shifts). In these cases, $\gamma = \frac{1}{5}\log(\#\text{counts})$ and convergence was declared when $||\boldsymbol{\mu}^{(i+1)} - \boldsymbol{\mu}^{(i)}||_2/||\boldsymbol{\mu}^{(i)}||_2 < 10^{-4}$ (30 iterations in this case).

Chapter 6

Conclusions and Ongoing Work

This thesis discussed piecewise polynomial- and platelet-based methods for the analysis, denoising, and reconstruction of Poisson or multinomial process observations in one and two dimensions. Both methods outperform conventional wavelets because they take into account the special properties of the Poisson or multinomial distributions. Platelets in particular outperform conventional wavelet representations because of their ability approximate smooth boundaries more efficiently than wavelets. Moreover, because the platelet analysis of Poisson distributed images is tractable and computationally efficient, existing image deblurring and tomographic reconstruction methods based on expectation-maximization algorithms can be easily enhanced with platelet-based complexity penalties with only a modest increase in computational complexity. Experimental results with real and simulated data from astronomy, networking, density estimation, confocal microscopy and nuclear medicine demonstrate that polynomial- and platelet-based methods can outperform the popular and widely used wavelet, kernel, and EM-MLE methods.

The piecewise polynomial-based MPLE algorithm makes an important contribution to the field of multiscale analysis. Like wavelet-based methods, my method relies on localized approximation at multiple resolutions, but unlike wavelet-based methods, my method is not restricted to an orthonormal basis representation. Furthermore, the statistical risk bounds derived in Chapter 4 for accurate noise models are within a logarithmic factor of of the risk associated with wavelet-based methods under the (over-simplifying) assumption of Gaussian noise. Thus my analysis demonstrates that there do exist methods for multiscale signal analysis with performance and computational complexity comparable to those of wavelets which exhibit superior statistical characteristics.

Similarly, the platelet-based MPLE has an important theoretical strength over more conventional approaches like the stopped EM-MLE and methods based on quadratic or non-quadratic roughness penalties or smoothness priors. As discussed in Section 4, because platelets can accurately approximate images with a very small number of terms, the platelet-based MPLE can have a very small bias (approximation error) and variance (proportional to the number of terms in the platelet representation). Although the conventional methods mentioned above have been studied extensively with experiments on simulated and real data and have been demonstrated to provide high quality image reconstructions, I am not aware of any comparable theoretical results for these methods. Certainly, very little is known about the theoretical error performance of MPLEs based on standard quadratic or even non-quadratic edge preserving roughness penalties for the class of images we considered in our platelet analysis. Even less is known for the stopped EM approaches. It may be possible to compare MPLEs using quadratic penalties with those based on platelets. Quadratic penalties can be interpreted as a penalty weighting applied to a Fourier expansion of the intensity. I have demonstrated theoretically that platelet approximations to piecewise smooth intensity functions can significantly outperform Fourier approximations. Thus, we may expect that the error performance of the platelet-based MPLE will be considerably better than that of quadratically penalized MPLEs.

In contrast, it should be possible to quantify the theoretical error performance of the platelet-based MPLE very precisely. Kolaczyk and Nowak show that the Haar-based MPLE is near minimax optimal when the underlying Poisson intensity belongs to Bounded Variation or Besov function spaces [5, 25]. These spaces are characterized by mostly smooth images with isolated point singularities. My interest in this work is in images with singularities along smooth curves (edges and boundaries) rather than at points, and I expect that improved minimax bounds can be obtained for my platelet-based MPLE in such cases. I am currently pursuing this work. Minimax error bounds

should be obtained relatively easily in the denoising context (applications without a blurring or projection operator). Inverse problems like the tomography case are more challenging to analyze, but in related problems with Gaussian noise, wavelet-based approaches have been shown to be near minimax optimal [53]. I am investigating similar approaches to the Poisson deblurring and tomography problems examined in this paper. The platelet approximation results of this paper are a key first step in this direction.

In addition to the ongoing theoretical analysis of the polynomial- and platelet-based MPLE, further experimental comparisons between it and conventional methods are necessary. Furthermore, polynomial and platelet analyses can be extended to signal processing and imaging problems involving other noise distributions. The multiscale likelihood factorizations that underpin the platelet-based MPLE also exist for Gaussian, multinomial and other data types, and only difference in the piecewise polynomial and platelet analyses in these cases will be the parametric form of the conditional likelihood factors involved [17]. In fact, in the Gaussian case all the conditional likelihood factors are Gaussian themselves and the maximum likelihood criterion is equivalent to the conventional least squares criterion.

References

1. Eric D. Kolaczyk, "Nonparametric estimation of gamma-ray burst intensities using haar wavelets," *The Astrophysical Journal*, vol. 483, pp. 340–349, 1997.

2. Z. H. Cho, J. P. Jones, and M. Singh, *Foundations of Medical Imaging*, Wiley-Interscience, New York, 1993.

3. D. Synder and M. Miller, *Random Point Processes in Time and Space*, New York: Springer-Verlag, 1991.

4. J. B. Pawley, Ed., *Handbook of Biological Confocal Microscopy*, Plenum Press, New York, 1995.

5. E. Kolaczyk and R. Nowak, "Risk analysis for multiscale penalized maximum likelihood estimators," submitted to Annals of Stat. Available at `http://www.ece.rice.edu/~nowak/publications.html`.

6. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

7. D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, "Density estimation by wavelet thresholding," *Ann. Statist.*, vol. 24, pp. 508–539, 1996.

8. J. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge Univ. Press, 1998.

9. A. Aldroubi and M. Unser, "Wavelets in medicine and biology," CRC Pr., Boca Raton FL, 1996.

10. M. Bhatia, W. C. Karl, and A. S. Willsky, "A wavelet-based method for multiscale tomographic reconstruction," *IEEE Trans. Med. Imaging*, vol. 15, no. 1, pp. 92–101, 1996.

11. E. D. Kolaczyk, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statistica Sinica*, vol. 9, pp. 119–135, 1999.

12. N. Lee and B. J. Lucier, "Wavelet methods for inverting the radon transform with noisy data," *IEEE Trans. Image Proc.*, vol. 10, pp. 79 –94, 2001.

13. J. Lin, A. F. Laine, and S. R. Bergmann, "Improving pet-based physiological quantification through methods of wavelet denoising," *IEEE Trans. Bio. Eng.*, vol. 48, pp. 202 –212, 2001.

14. J. Weaver, Y. Xu, D. Healy, and J. Driscoll, "Filtering MR images in the wavelet transform domain," *Magn. Reson. Med.*, vol. 21, pp. 288–295, 1991.

15. N. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Royal Society London A*, vol. 357, pp. 2543–2560, 1999.

16. E. Candès and D. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," To appear in Curves and Surfaces, L. L. Schumaker et al. (eds), Vanderbilt University Press, Nashville, TN.

17. E. Kolaczyk and R. Nowak, "A multiresolution analysis for likelihoods: Theory and methods," submitted to Annals of Stat. Available at `http://www.ece.rice.edu/~nowak/publications.html`.

18. K. Timmermann and R. Nowak, "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, April, 1999.

19. E. Kolaczyk, "Bayesian multi-scale models for Poisson processes," *J. Amer. Statist. Assoc.*, vol. 94, pp. 920–933, 1999.

20. D. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Ann. Statist.*, vol. 27, pp. 859 – 897, 1999.

21. Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, pp. 8–37, 1985.

22. R. A. DeVore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.

23. D. Donoho, "Sparse components of images and optimal atomic decompositions," *Constr. Approx.*, vol. 17, pp. 353–382, 2001.

24. D. Donoho, "Cart and best-ortho-basis selection: A connection," *Annals of Stat.*, vol. 25, pp. 1870–1911, 1997.

25. H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*, North-Holland, Amsterdam, 1978.

26. M. Unser and M. Eden, "Maximum likelihood estimation of linear signal parameters for poisson processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 6, pp. 942–5, 1988.

27. Q. J. Li, *Estimation of Mixture Models*, Ph.D. thesis, Yale University, 1999.

28. Q. J. Li and A. R. Barron, *Advances in Neural Information Processing Systems 12*, chapter Mixture Density Estimation, MIT Press, 2000.

29. D. Donoho and I. Johnstone, "Ideal adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

30. J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, singapore, 1989.

31. R. Nowak and R. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *IEEE Transactions on Image Processing*, vol. 8, no. 5, 1999.

32. R. Coifman and D. Donoho, "Translation invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics*, vol. New York: Springer-Verlag, pp. 125–150, 1995.

33. M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.

34. David L. Donoho and Iain M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

35. B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

36. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, vol. 39, pp. 1–38, 1977.

37. G. M. P. van Kempen, H. T. M. van der Voort, J. G. J. Bauman, and K. C. Strasters, "Comparing maximum likelihood estimation and constrained tikhonov-miller restoration," *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, pp. 76 – 83, 1996.

38. W. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. of Am.*, vol. 62, pp. 55–59, 1972.

39. T. Hebert and R. Leahy, "A generalized EM algorithm for 3-d Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imaging*, vol. 8, no. 2, pp. 194–202, 1989.

40. P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. 9, no. 1, pp. 84–93, 1990.

41. J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1417–1429, 1995.

42. A. R. Depierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imaging*, pp. 132–137, 1995.

43. J. Liu and P. Moulin, "Complexity-regularized image denoising," *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 841 –851, 2001.

44. P. Moulin and J. Liu, "Statistical imaging and complexity regularization," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1762 –1777, 2000.

45. N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," *Wavelets in Geophysics*, Foufoula-Georgiou and Kumar (eds.), Academic Press, 1994.

46. H. Krim and I.C. Schick, "Minmax description length for signal denoising and optimal representation," *IEEE Trans. on Information Theory*, vol. 45, no. 3, April, 1999.

47. G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, New York, 1997.

48. J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Med. Imaging*, vol. 8, pp. 186–193, 1989.

49. R. Nowak and E. Kolaczyk, "A multiscale statistical framework for Poisson inverse problems," *IEEE Trans. Info. Theory*, vol. 46, pp. 1811–1825, 2000.

50. L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. 1, pp. 113–122, 1982.

51. J. A. Fessler, "Aspire 3.0 user's guide: A sparse iterative reconstruction library," Communication & Signal Processing Laboratory Technical Report No. 293, Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, 1998.

52. K. E. Sorra and K. M. Harris, "Overview on the structure, composition, function, development, and plaasticity of hippocampal dendritic spines," *Hippocampus*, vol. 10, pp. 501–511, 2000.

53. D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *App. and Comp. Harmonic Analysis*, vol. 2, pp. 101–126, 1995.

54. M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, Dover Publications, Inc., New York, 1965.

55. T. R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1972.

Appendix A

Proofs of Theorems and Lemmas

A.1 Proof of Discrete Polynomial Approximation Lemma

Proof of Lemma 2.1 Consider each of the three bounds separately.

Term a: As in Kolaczyk and Nowak's work, we use the Haar basis to bound $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}^2$ [5]. Specifically, let $h_{j,k}(i)$ be the (j,k)-th Haar function on the discrete space $\{0, 1, \ldots, N-1\}$; in other words, $h_{j,k}(i) \equiv (\chi_{j+1,2k+1}(i) - \chi_{j+1,2k}(i))/N_{j,k}^{1/2}$, where $\chi_{j,k}$ is the characteristic function for the discrete analogue of the interval $I_{j,k} \equiv [k/2^j, (k+1)/2^j)$, for $j = 0, \ldots, J-1, k = 0, \ldots, 2^j - 1$, and $J = log_2(N)$, and $N_{j,k} = N/2^j$ is the cardinality of this set. Similarly, let $h_{j,k}^c(t)$ be the continuous analog of $h_{j,k}(i)$ on the interval $[0,1]$, or $h_{j,k}^c(t) \equiv 2^{j/2}(\chi_{j+1,2k+1}^c(t) - \chi_{j_1,2k}^c(t))$. It then follows that $\langle \boldsymbol{\mu}, h_{j,k} \rangle_{\ell_2} = N^{-1/2} \langle \mu, j_{j,k}^c \rangle_{L_2}$; combining this with Parseval's relation we have

$$\begin{aligned} \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_f\|_{\ell_2}^2 &= \sum_{(j,k)\in\mathcal{J}} \left(\langle \boldsymbol{\mu}, h_{j,k} \rangle_{\ell_2} - \langle \tilde{\boldsymbol{\mu}}_f, h_{j,k} \rangle_{\ell_2} \right)^2 \\ &= \frac{1}{N} \sum_{(j,k)\in\mathcal{J}} \left(\langle \mu, h_{j,k}^c \rangle_{L_2} - \langle \tilde{\mu}_f, h_{j,k}^c \rangle_{L_2} \right)^2 \end{aligned} \tag{A.1}$$

where $\mathcal{J}$ is the set of all (j,k) with $j = 0, 1, \ldots, J-1$ and $k = 0, 1, \ldots, 2^j - 1$. The expression (A.1) is bounded above by a similar sum over all (j,k), which is equal to the squared L_2 approximation error, $\|\mu - \tilde{\mu}_f\|_{L_2(\Omega)}^2$. From (2.1), we have $\|\mu - \tilde{\mu}_f\|_{L_2}^2 = O(d^{-2r})$, where d is the number of polynomially varying pieces in the function $\mu(\cdot)$, and r is the maximum degree of each polynomial.

Term *b*: By construction, $\tilde{\mu}_f(\cdot)$ has $d-1$ breakpoints. In order to form $\tilde{\boldsymbol{\mu}}_f$ and $\tilde{\boldsymbol{\mu}}_p$ by integration, the function domain $[0,1]$ is partitioned into N intervals of length $1/N$, and then the functions are integrated over each partition to form each vector element. Thus for all but $d-1$ of these intervals, $\tilde{\mu}_{f,i} = \tilde{\mu}_{p,i}$. For the remaining $d-1$ intervals, using the integration construction, we obtain

$$\begin{aligned} |\tilde{\mu}_{f,i} - \tilde{\mu}_{p,i}| &= \left| \int_{I_i} \tilde{\mu}_f(t) - \tilde{\mu}_p(t) dt \right| \\ &\leq \int_{I_i} |\tilde{\mu}_f(t) - \tilde{\mu}_p(t)| \, dt \\ &\leq \frac{C-c}{N} \end{aligned}$$

which leads to the final bound

$$\|\tilde{\boldsymbol{\mu}}_f - \tilde{\boldsymbol{\mu}}_p\|_{\ell_2}^2 \leq \frac{(C-c)^2 d}{N^2}. \tag{A.2}$$

Term *c*: Quantization produces the final error term. Instead of quantizing vector elements, we quantize the coefficients of an orthonormal polynomial basis representation of each polynomial piece. This is possible because the magnitude of each of these coefficients is bounded above. Consider a polynomial $p(t)$ defined on $t \in [0,1]$. This polynomial can be expressed in terms of the normalized shifted Legendre polynomial orthogonal basis, also defined on $[0,1]$ [54]. That is,

$$p(t) = \sum_{k=0}^{r} \langle p, \sqrt{2k+1}\bar{P}_k \rangle \sqrt{2k+1}\bar{P}_k$$

where $\bar{P}_k(t) = P_k(2t-1)$, P_k is the k^{th} Legendre polynomial defined on $[-1,1]$. The $\sqrt{2k+1}$ term is used because $\|\bar{P}_k\|^2 = \frac{1}{2k+1}$ and so each polynomial must be normalized.

The magnitude of each coefficient in the above orthonormal basis expansion is bounded above using an $L_1 - L_\infty$ bound argument; specifically,

$$\begin{aligned}\left|\langle p, \sqrt{2k+1}\bar{P}_k\rangle\right| &= \left|\int_0^1 p(t)\bar{P}_k(t)\sqrt{2k+1}dt\right| \\ &\leq \sqrt{2k+1}\int_0^1 |p(t)|dt \\ &\leq \sqrt{2k+1}C\end{aligned}$$

where we have used the Legendre polynomial characteristic that $|P_k(t)| \leq 1$. Thus the magnitude of coefficient k is bounded by $\sqrt{2k+1}C$, and so it is now possible to quantize this coefficient to one of $N^{1/2}$ levels in $\left[-\sqrt{2k+1}C, \sqrt{2k+1}C\right]$. Let the quantized version of coefficient $a_k = \langle p, \sqrt{2k+1}\bar{P}_k\rangle$, $k = 0, \ldots, r$ be denoted $[a_k]$. This quantization induces the following error for a given $t \in [0, 1]$:

$$\begin{aligned}|\tilde{\mu}_p(t) - \tilde{\mu}'(t)| &= \left|\sum_{k=0}^{r}(a_k - [a_k])\bar{P}_k(t)\right| \\ &\leq \sum_{k=0}^{r}|a_k - [a_k]|\left|\bar{P}_k(t)\right| \\ &\leq \sum_{k=0}^{r}\sqrt{2k+1}\frac{C}{\sqrt{N}}\left|\bar{P}_k(t)\right| \\ &\leq \frac{C}{\sqrt{N}}\sum_{k=0}^{r}\sqrt{2k+1} \\ &\leq \frac{C'r^{3/2}}{\sqrt{N}}\end{aligned}$$

The above bound allows a bound on the difference between average-sampled vector elements:

$$\begin{aligned} |\tilde{\mu}_{p,i} - \tilde{\mu}'_i| &\leq \int_{I_i} |\tilde{\mu}_p(t) - \tilde{\mu}'(t)| dt \\ &\leq \frac{C'^2 r^{3/2}}{N^{3/2}} \end{aligned}$$

This yields the final bound:

$$\begin{aligned} \|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_2}^2 &\leq \|\tilde{\boldsymbol{\mu}}_p - \tilde{\boldsymbol{\mu}}'\|_{\ell_\infty}^2 \\ &= \frac{C'^2 r^3}{N^2} \end{aligned}$$

■

A.2 Proof of Platelet Approximation Theorem

Proof of Theorem 2.3

It suffices to verify the theorem for dyadic m (powers of two); an error bound in the general case follows in a similar manner. The proof is constructive. For the most part, we will not specify the constants underlying the bounds discussed below and only speak in terms of orders of magnitude. It is possible to keep careful track of the constants, but this would make the proof less transparent.

First consider approximating the boundary with a J-scale (i.e., dyadic squares have side length greater than or equal to 2^{-J}) wedgelet partition. From [20] it is known that one can construct an $O(m)$-term dyadic wedgelet partition such that the boundary is completely contained in disjoint dyadic squares of equal side length $1/m$. To sketch the idea, imagine tiling $[0,1]^2$ with m^2 dyadic squares of side length $1/m$. Because the boundary function is Hölder$^{\alpha,1}(C_\alpha)$, with $\alpha > 1$, it must

also be Lipschitz (i.e., Hölder$^{1,1}(C_1)$, where the Lipschitz constant C_1 may be smaller than C_α) it is easy to check that the boundary passes through only $O(m)$ of these squares. Merge all squares not containing the boundary into larger dyadic squares (according to the RP associated with Haar analysis). It turns out that after merging there are only $O(m)$ squares in total. That is, there exists a constant $C > 0$ such that the total number of squares is less than or equal to $C\,m$. It can be shown that $C' = 8(C_1 + 2)$ will work [20], where C_1 is the Lipschitz constant above. This result holds for $2 \leq m \leq 2^J$.

Now consider the approximation of a wedgelet to the true boundary in one of the dyadic squares above. Each such square can be broken into three regions, two regions in which the true boundary and wedge boundary agree and one region where they disagree (area "between" the true boundary and the wedge boundary). The area between the true boundary and the wedgelet boundary in a square is $O(m^{-(\alpha+1)})$ at most. This is a simple consequence of the fact that the Hölder$^{\alpha,1}(C_\alpha)$ can be used to bound the L_∞ approximation error between the boundary function $H(x)$ and a line connecting the points $(i/m, H(i/m))$ and $((i+1)/m, H((i+1)/m))$, where i/m and $(i+1)/m$ refer to the horizontal boundaries of a given square. The L_∞ error bounds the area of the region in question using this line fit; the error is $O(m^{-(\alpha+1)})$. Wedgelets do not use arbitrary vertices (e.g., $H(i/m)$), but rather their vertices are restricted to equispaced points δ apart along the boundary of each square. This "quantization" effect adds an additional amount of area to the region. This additional amount is $O(\delta m^{-1})$.

Next generate a new partition by subdividing all squares of side length larger than $1/\sqrt{m}$ in the $O(m)$ wedgelet partition so that the new partition is a tiling of $[0,1]^2$ with dyadic squares having side length $1/\sqrt{m}$ or less. This partition also has $O(m)$ regions. To see this, note that tiling $[0,1]^2$ with squares of side length $1/\sqrt{m}$ requires m such squares, so the additional subdivision adds less

than m additional squares to the original $O(m)$ regions in the wedgelet partition. The L_2^2 error on each square of the partition just constructed above can be bounded above as follows.

First consider squares not containing the boundary (i.e., squares in smooth parts of the image). The first-order Taylor series expansion of f on such a square is

$$\widehat{f}(x,y) = f(x_0,y_0) + (x-x_0)\frac{\partial f}{\partial x}(x_0,y_0) + (y-y_0)\frac{\partial f}{\partial y}(x_0,y_0).$$

This expansion provides a planar approximation to f on that square. Because each square has side length of at most $1/\sqrt{m}$, the mean value theorem allows us to bound the the approximation error at any point (x,y) in the square:

$$\begin{aligned} |f(x,y) - \widehat{f}(x,y)| &= |f(x_0,y_0) + (x-x_0)\frac{\partial f}{\partial x}(x',y') + (y-y_0)\frac{\partial f}{\partial y}(x',y') \\ &\quad -f(x_0,y_0) - (x-x_0)\frac{\partial f}{\partial x}(x_0,y_0) - (y-y_0)\frac{\partial f}{\partial y}(x_0,y_0)| \\ &\leq \left|(x-x_0)\left[\frac{\partial f}{\partial x}(x',y') - \frac{\partial f}{\partial x}(x_0,y_0)\right]\right| + \left|(y-y_0)\left[\frac{\partial f}{\partial y}(x',y') - \frac{\partial f}{\partial y}(x_0,y_0)\right]\right|, \end{aligned}$$

where x' (y') is between x and x_0 (y and y_0). From here, the Hölder$^{\beta,2}(C_\beta)$ assumption gives the upper bound

$$\begin{aligned} |f(x,y) - \widehat{f}(x,y)| &\leq C_\beta \left|\sqrt{(x'-x_0)^2 + (y'-y_0)^2}\right|^{\beta-1} (|x-x_0| + |y-y_0|) \\ &\leq C_\beta \left(\frac{\sqrt{2}}{\sqrt{m}}\right)^{\beta-1} \cdot \frac{2}{\sqrt{m}} \\ &= O(m^{-\beta/2}) \end{aligned}$$

Since this holds for all (x,y) in the square, we have bounded the L_∞ error. Thus the L_2^2 error over

each such square is $O(m^{-(\beta+1)})$ (i.e., L_∞ error squared $\times$ area).

Next consider squares containing the boundary. Each such square has a side length of $1/m$. According to the wedgelet analysis above, the area between the true boundary and the wedge boundary in each square is $O(m^{-(\alpha+1)})$. (Since the image is Hölder$^{\beta,2}(C_\beta)$, $\beta > 1$, it is also Lipschitz and hence continuous and since the image has compact support, it must be bounded and therefore the L_2^2 error in this region is $O(m^{-(\alpha+1)} + \delta m^{-1})$). Select planar fits on the two wedges so that the L_2^2 errors in the other two smooth regions of the square are $O(m^{-2(\beta+1)})$ (the 2 in the exponent appears because the side length is $O(1/m)$ instead of $O(1/\sqrt{m})$). This gives a total L_2^2 error in each such square of $O(m^{-\min(\alpha+1,2(\beta+1))} + \delta m^{-1})$. Combining the planar approximations on dyadic squares of side length $1/\sqrt{m}$ or less in the smooth parts of the image with the platelet approximation of the boundary on dyadic squares of side length $1/m$ produces a total L_2^2 approximation error of $O(m^{-\min(\alpha,\beta)} + \delta)$. That is, the squared L_2 approximation error is bounded above by $K_{\alpha,\beta}\,(m^{-\min(\alpha,\beta)} + \delta)$, where the constant $K_{\alpha,\beta}$ depends on C_α, C_β, and C' defined above. In fact, this bound can be slightly improved since the squared error contributed by the wedgelet quantization can be bounded more tightly [20] to be less than or equal to δ, rather that the loose bound of $O(\delta)$ we used above. Thus, the total squared error is bounded above by $K'_{\alpha,\beta}\, m^{-\min(\alpha,\beta)} + \delta$. ■

A.3 Proof of Risk Bound Theorem

Proof of Lemma 4.1 Following Kolaczyk and Nowak's work, I begin by writing $\Gamma_N = \cup_{d=1}^{N} \Gamma_N^{(d)}$, where $\Gamma_N^{(d)}$ is the subset of values $\boldsymbol{\theta}'$ that are comprised of d polynomially-varying sequences [5]. Each of the members of $\Gamma_N^{(d)}$ has the same value for the summand in 4.8 because $\#(\boldsymbol{\theta}') = d$ for each $\boldsymbol{\theta} \in \Gamma_N^{(d)}$. Each of the $r+1$ quantized coefficients, a_j for $j = 0, \ldots, r$ in each polynomially-varying sequence can take one of $N^{1/2}$ values because of the coefficient quantization. There are d

such sequences in each $\boldsymbol{\theta}' \in \Gamma_N^{(d)}$, which means that given the locations of the $d-1$ breakpoints between polynomial sequences, there are $(N^{1/2})^{(r+1)d}$ possible values for $\boldsymbol{\theta}'$. Finally, the breakpoints between the d segments can occur in $N-1$ choose $d-1$ possible locations. Thus each $\Gamma_N^{(d)}$ contains $\binom{N-1}{d-1}(N^{1/2})^{(r+1)d}$ members. It then follows:

$$
\begin{aligned}
\sum_{\boldsymbol{\theta}' \in \Gamma_N} e^{-\gamma \log_e(N) \#(\boldsymbol{\theta}')} &= \sum_{d=1}^{N} \sum_{\boldsymbol{\theta}' \in \Gamma_N^{(d)}} e^{-\gamma d \log_e(N)} \\
&= \sum_{d=1}^{N} \binom{N-1}{d-1} (N^{1/2})^{(r+1)d} e^{-\gamma d \log_e(N)} \\
&= \sum_{d=1}^{N} \binom{N-1}{d-1} e^{-(\gamma - \frac{1}{2}(r+1)) d \log_e(N)} \\
&= \sum_{d'=0}^{N-1} \binom{N-1}{d'} e^{-(\gamma - \frac{1}{2}(r+1))(d'+1) \log_e(N)} \\
&= \sum_{d'=0}^{N-1} \binom{N-1}{d'} N^{-(\gamma - \frac{1}{2}(r+1))(d'+1)} \\
&\leq \sum_{d'=0}^{N-1} \frac{(N-1)^{d'}}{d'!} N^{-(\gamma - \frac{1}{2}(r+1))(d'+1)} \qquad \text{(A.3)} \\
&= N^{-(\gamma - \frac{1}{2}(r+1))} \sum_{d'=0}^{N-1} \frac{(N-1)^{d'}}{d'!} N^{-d'(\gamma - \frac{1}{2}(r+1))} \\
&\leq N^{-(\gamma - \frac{1}{2}(r+1))} \sum_{d'=0}^{N-1} \frac{N^{d'}}{d'!} N^{-d'(\gamma - \frac{1}{2}(r+1))} \\
&= N^{-(\gamma - \frac{1}{2}(r+1))} \sum_{d'=0}^{N-1} \frac{1}{d'!} N^{d'[1-(\gamma - \frac{1}{2}(r+1))]} \\
&\leq N^{-(\gamma - \frac{1}{2}(r+1))} \sum_{d'=0}^{N-1} \frac{1}{d'!} \qquad \text{(A.4)} \\
&\leq N^{-(\gamma - \frac{1}{2}(r+1))} \sum_{d'=0}^{\infty} \frac{1}{d'!} \\
&= N^{-(\gamma - \frac{1}{2}(r+1))} e \\
&\leq N^{1-(\gamma - \frac{1}{2}(r+1))} \leq 1 \qquad \text{(A.5)}
\end{aligned}
$$

The first inequality (A.3) is due to the fact that

$$\begin{aligned} \binom{a}{b} &= \frac{a!}{(a-b)!b!} \\ &\leq \frac{1}{b!}a^b \end{aligned}$$

The inequality (A.4) holds under our assumption that $\gamma \geq \frac{3}{2} + \frac{r}{2}$, or $\frac{3}{2} \leq \gamma - \frac{r}{2}$, which means N is raised to a negative power and hence $N^{d'[1-(\gamma-\frac{1}{2}(r+1))]} \leq 1$. The final result (A.5) is bounded by one, as desired, under our assumption that $N \geq 3$. ■

A.4 Proof of Computational Complexity Theorem

Proof of Lemma 4.2 It is easy to check that the log multinomial likelihood function is concave in its parameters ($\rho = T\,\theta$). To prove the lemma, we refer to Theorem 5.7 in [55], which states that if T is a linear transformation from $\mathbb{R}^N$ to $\mathbb{R}^M$, then, for each convex function g on $\mathbb{R}^M$, the function gT defined by

$$(gT)(\theta) \equiv g(T\theta)$$

is convex in θ on $\mathbb{R}^N$. This and the concavity of multinomial log likelihood shows that

$$L(x|\theta) \equiv \log \text{Multinomial } (x \,|\, T\,\theta)$$

is concave θ. ■

Proof of Theorem 4.2 In order to calculate a Haar MPLE, one may obtain the sums of the counts in each dyadic square by performing a Haar wavelet analysis, which requires $O(N^2)$ calculations. Once the tree has been built with these parameters, the tree-pruning takes only $O(N^2)$ operations,

yielding a total complexity of $O(N^2)$.

For wedgelets, note that there are $O(N^2)$ possible wedgelets to consider for an $N \times N$ pixel image. Direct calculation of the log likelihood for each wedgelet term requires $O(N^2)$ operations. The overall number of computations would then be $O(N^4)$. This number can be significantly reduced by carefully considering the impact of calculating the likelihood for each possible wedgelet in a sequential order. Consider two possible wedgelets, one dividing the image into regions A_1 and B_1, and the other dividing the image into regions A_2 and B_2, which are constructed so that the boundary separating A_1 and B_1 is different from the boundary separating A_2 and B_2 by only one pixel side length in one coordinate of one boundary endpoint. This is depicted in Figure A.1. The two wedgelets are very similar; in fact, the number of pixels which are common to both regions A_1 and B_2 or A_2 and B_1 may be bounded as follows: since the side length of the image is 1 and therefore the side length of each pixel is $1/N$, the image area between the two boundaries is upper bounded by $1/(2N)$. Since the area of each pixel is $1/N^2$, we have that the number of pixels which are in both regions A_1 and B_2 or A_2 and B_1 is of $O(N/2)$. Because the likelihood function is additive, once the likelihood of the image being separated into regions A_1 and B_1 has been calculated, we need only $O(N/2)$operations to calculate the likelihood of the image being separated into regions A_2 and B_2.

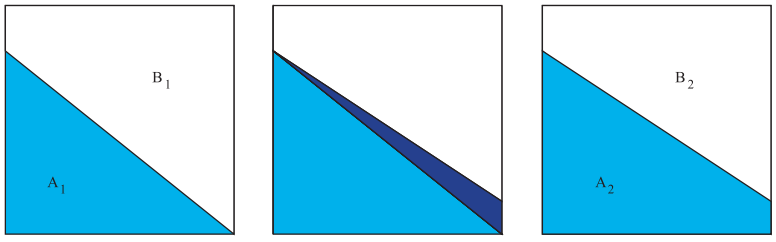


Figure A.1 Sequential calculation of wedgelet likelihoods

The complete complexity calculation must consider the number of likelihood calculations at

each level of the quad tree described above. At the coarsest scale, there is one block containing N^2 pixels and N^2 possible wedgelets, resulting in $O(N^3)$ operations for that level. On the second level, there are four blocks containing $N^2/4$ pixels each at $N^2/4$ possible wedgelets for each, resulting in $\frac{1}{2}O(N^3)$ operations for that level. This sequence continues, and is upper bounded by $O(N^3) \cdot \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i$, or $2\,O(N^3)$ for the entire image. Such a simplification is not possible when numerically searching for the maximum (multinomial) likelihood platelet fit, since all data in a given square are necessary to perform the search. Performing this search results in an $O(N^4)$ algorithm. An approximate platelet fit can be constructed using least-squares. Although this is suboptimal, our experimental results demonstrate it is a close approximation to the optimal platelet. Using the least squares approximation has the advantage of reducing the total number of operations to $O(N^3)$. The statistics of the data needed to calculate the least squares platelet fit are $\sum_{i,j} x_{i,j}$, $\sum_{i,j} i \cdot x_{i,j}$, $\sum_{i,j} j \cdot x_{i,j}$, where the sums are over the indices of pixels in the square or wedge under consideration (the individual data are not required). As in the wedgelet case, these statistics may be updated for each sequential platelet in $O(N)$ operations. As before, this coupled with the quad tree structure yields a total complexity of $O(N^3)$. ■