Platelets: A Multiscale Approach for Recovering Edges and Surfaces in Photon-Limited Medical Imaging

Rebecca M. Willett, *Student Member, IEEE,* and Robert D. Nowak, *Member, IEEE* *

March 12, 2004

Abstract

The nonparametric multiscale platelet algorithms presented in this paper, unlike traditional wavelet-based methods, are both well suited to photon-limited medical imaging applications involving Poisson data and capable of better approximating edge contours. This paper introduces platelets, localized functions at various scales, locations, and orientations that produce piecewise linear image approximations, and a new multiscale image decomposition based on these functions. 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. 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 suggest that platelet-based methods can outperform standard reconstruction methods currently in use in confocal microscopy, image restoration, and emission tomography.

Keywords: wavelets, multiresolution, denoising, tomography

1 Photon-Limited Medical Imaging

Many 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. Such cases are referred to as *photon-limited* imaging applications, since 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 [1–3]. 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. The mean and variance of a

*The authors are with the Department of Electrical and Computer Engineering, Rice University MS 366, Houston, TX 77251-1892 USA (e-mail: nowak@ece.rice.edu, phone: 713 348 6318, fax: 713 348 6196). R. Nowak was partially supported by the National Science Foundation, grant no. MIP–9701692, the Army Research Office, grant no. DAAD19-99-1-0349, the Office of Naval Research, grant no. N00014-00-1-0390. R. Willett was partially supported by the National Science Foundation Graduate Student Fellowship.

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.

1.1 Statistical Methods for Restoration and Reconstruction

Statistical methods for photon-limited image deblurring and reconstruction are especially effective since they can account for the special properties of the Poisson distribution. 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 [4–6] (also known as the Richardson-Lucy algorithm in the context of Poisson data [7]). 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 this problem the maximum likelihood criterion can be replaced with maximum penalized likelihood criteria [8–11]. 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 [8, 9], or a complexity measure [12, 13]. 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 [14, 15]; the more terms (basis functions) used to represent the image, the higher the complexity measure. 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 [16].

An alternative to MPLE-based methods is the "stopped" EM-MLE solution [17]. 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. This paper hopes to convince the reader that 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.

1.2 Wavelet and Multiresolution Methods

While, as mentioned above, images of normal homogeneous tissue exhibit smoothly varying intensities, the goal of medical imaging is often to detect boundaries of organs or abnormalities such as lesions or tumors. Such regions would be indicated by sharp changes in the uptake of the radioactive pharmaceuticals, *i.e.* edges in the intensity

image. Wavelets and multiresolution analysis are especially well-suited for the detection of discontinuities and edges, which in medical images are critical for diagnostics.

Many investigators have considered the use of wavelet representations for image denoising, deblurring, and tomographic image reconstruction; for examples, see [18–25]. However, in the context of photon-limited imaging most wavelet-based approaches are based on Gaussian or other simplifying approximations to the Poisson 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 [26], steerable pyramids [27], and curvelets [28]) to Poisson data, but wavelets and related representations are easy to use in the Gaussian case. The Haar wavelet system is the only exception [29]; it does provide a tractable multiscale analysis framework for Poisson data, and this paper builds on the Haar-based multiscale likelihood factorizations we developed in [30, 31]. There are several reasons why Gaussian approximations are undesirable. First, the approximations are usually only reasonable if the numbers of detected photons 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 enough for Gaussian-based reconstruction algorithms to be effective, the detections must be binned or aggregated over regions (usually pixels/voxels) of sufficiently large 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 spatial adaptivity in order to recover as much of the 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 image 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 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 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 image. The conventional wisdom is that a representation that provides good approximations will also provide good estimations [18].

1.3 Contribution

This paper introduces a new multiscale image representation based on atoms called platelets. Platelets are 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. Platelets generalize Donoho's wedgelets [32], and like complex wavelets, steerable pyramids, 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 (with respect to the number of terms required to achieve a given approximation error). Moreover, platelet representations are especially well-suited to the analysis of Poisson data, unlike most other multiscale image representations, and they can be rapidly computed. We propose a platelet-based maximum penalized likelihood criterion that encompasses denoising, deblurring, and tomographic reconstruction problems. The criterion can be incorporated very easily into the classical EM algorithm and the overall computational burden is

nearly the same as that of the basic EM-MLE algorithm [5]. This paper focuses on the first three steps of the CHA program: (i) defining a class of functions (platelets) suitable for describing the images commonly encountered in medical imaging applications, (ii) studying the approximation capabilities of platelets, and (iii) devising MPLEs based on platelets. The performance of 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 platelet-based MPLEs is also explored with real data from nuclear medicine and confocal microscopy experiments.

Deriving explicit mathematical bounds on the performance of the MPLEs is beyond the scope of this paper due the the diverse nature of the problems under consideration. However, the impressive approximation capabilities of platelets suggest that the MPLEs may indeed be close to optimal (in a minimax risk sense). Recently, minimax risk bounds for Haar wavelet-based multiscale MPLEs of Poisson intensities have been derived [29] which should be extendable to platelet-based denoising, deblurring and reconstruction. This is currently under investigation and we plan to report on this in a follow-up paper.

The paper is laid out as follows. Section 2 introduces the platelet representation. 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. Section 3 discusses multiscale analysis methods for Poisson data and the notion of multiscale likelihood factorizations. We show that a Poisson likelihood, parameterized by a platelet representation, admits a multiscale likelihood factorization, which is a probabilistic analog of classical multiresolution analysis. Section 4 proposes a new penalized likelihood criterion based on platelets for "denoising" photon-limited images. The multiscale likelihood factorization enables a fast, globally optimal algorithm that computes the MPLE. Section 5 studies more challenging (inverse) problems including photon-limited image deblurring (restoration) and tomographic reconstruction. It is shown that MPLEs can be computed 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 simulations and with real data. Section 6 summarizes the new methodology and algorithms and discusses ongoing and future research directions.

2 Platelet Decompositions and Image Approximations

We begin by reviewing Haar multiscale image analysis and its connection to recursive partitions. This sets the stage for developing multiscale platelet representations. The approximation capabilities of platelets are explored and contrasted with Fourier, wavelet, and wedgelet based approximations.

2.1 Haar Multiscale Analysis and Image Partitioning

Image partitions are central to the approximation capabilities of platelets. 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 $j = 0, \ldots, J-1$ and $m, n = 0, \ldots, 2^j - 1$, 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 $[0,1]^2$.

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. 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 [33] (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). 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 important extensions of classical Haar multiscale analysis. In particular, one need not restrict the analysis to dyadic square partitions. The wedgelet partition [32] is a dyadic, square recursive partition which allows for non-square, "wedge-shaped" partitions only at the final level of the partition. 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 1 as a simple illustration of the efficiency of the wedgelet partition space. Figures 1(c) and (e) contain rough approximations of the Shepp-Logan phantom to within the same error using the hereditary Haar and wedgelet analyses. Notice how many fewer partitions are required for wedgelet approximation.

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* and it is a key parameter of the wedgelet analysis. Following the development in [32], it is assumed that $\delta = 2^{-J-K}$ with $K \geq 0$.

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 only constrained 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).

Figure 2 shows the decay in hereditary Haar and wedgelet approximation errors as the number of terms decreases. The wedgelet representation not only is more accurate with fewer terms, but also exhibits a rate of error decay faster than that of the hereditary Haar representation. This demonstrates the sparseness of the wedgelet partition space for multiple levels of approximation and supports the theoretical results in [32].

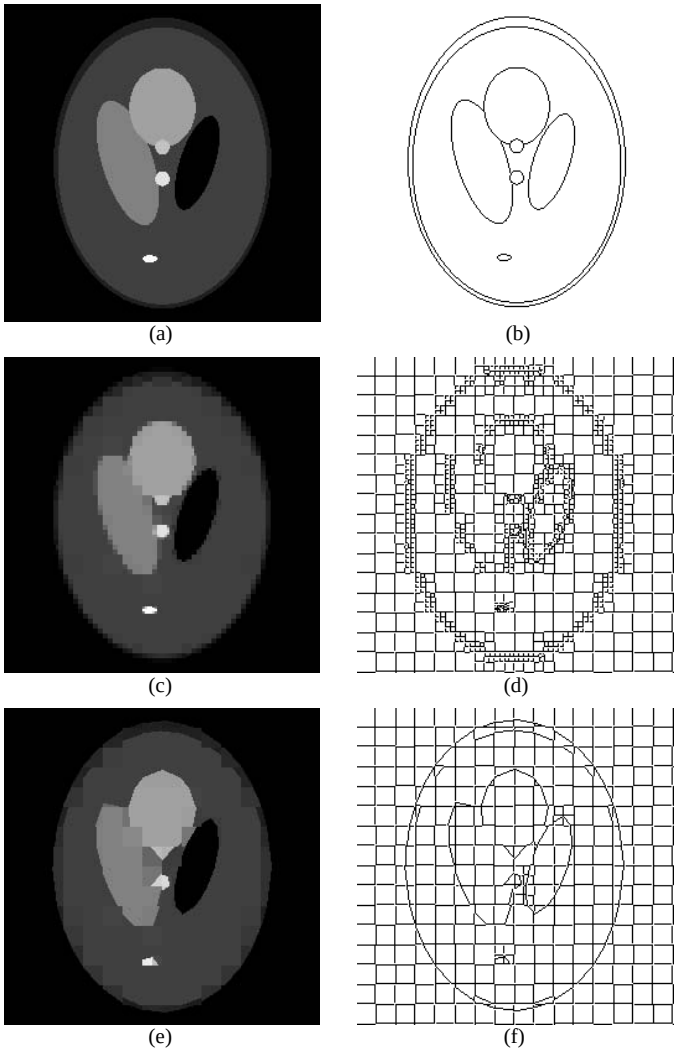

Figure 1: Image Partitions. (a) Original image. (b) True edges. (c) Haar approximation (error = 0.0155). (d) Haar partition - over 850 wavelet coefficients. (e) Wedgelet approximation (error = 0.0158). (f) Wedgelet partition - less than 370 wedgelets.

2.2 Platelet Analysis

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. 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

$$f_S(x, y) = (A_S x + B_S y + C_S)\ I_S(x, y), \tag{1}$$

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 parameters, compared with the one parameter 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

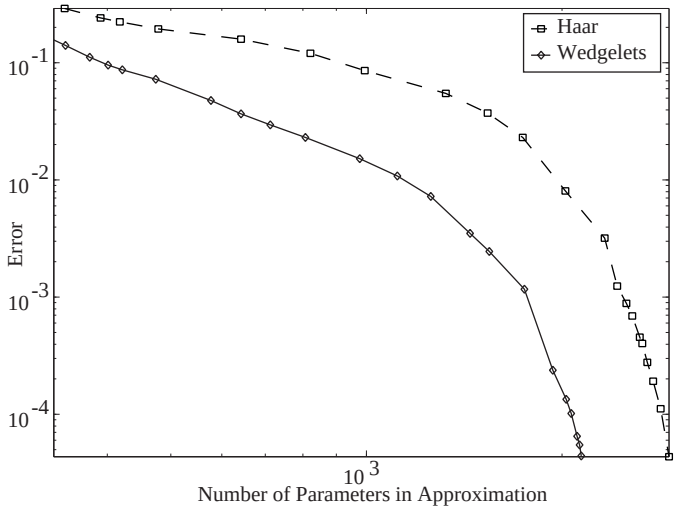

Figure 2: Approximation errors *vs.* number of parameters for Shepp-Logan phantom

approximations may require far fewer parameters 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 3. 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 hereditary Haar and wedgelet approximations.

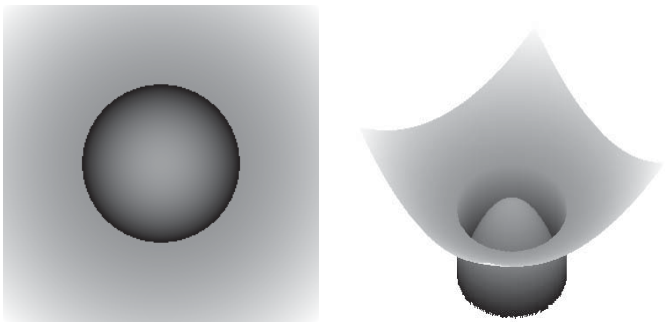

Figure 3: Quadratic test image. (a) Image. (b) Surface plot.

Approximations with similar L_2 percent errors were chosen for Figure 4. Figure 4(a) is a hereditary Haar approximation, and contains 2068 constant squares of varying sizes. Figure 4(b) is a wedgelet approximation. Note that this image has an error similar to that of the image obtained with Haar approximation, 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 rectangular platelet approximation (computed on a rectangular partition instead of wedgelet partition) in Figure 4(c). For each platelet, the parameters A_S, B_S, C_S are chosen to minimize the total 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 parameters. Finally, Figure 4(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 parameters for each term, or region. In the image below, 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.

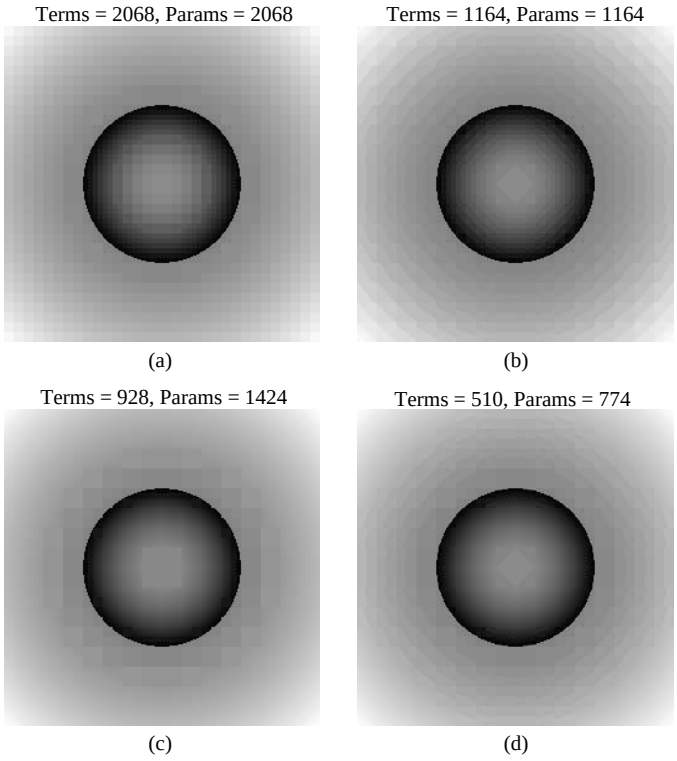

Figure 4: Approximations. (a) Haar. (b) Wedgelets. (c) Rectangular platelets. (d) Combined platelets and wedgelets. Each approximation has error $\approx 3 \times 10^{-4}$.

In fact, Figure 5 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 Haar and wedgelets representations. What follows is an analytical characterization of these rates.

2.3 Platelet Approximation Theory

For this analysis, consider images which are 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

$$\begin{aligned} f(x,y) &= f_1(x,y) \cdot I_{\{y \geq H(x)\}} + \\ & \quad f_2(x,y) \cdot \left(1 - I_{\{y \geq H(x)\}}\right), \end{aligned}$$

for all $(x,y) \in [0,1]^2$, 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 [32], which consisted of two constant regions separated by a Hölder smooth boundary.

The boundary is described by $y = H(x)$, where

$$H \in \text{Hölder}^{\alpha}(C_{\alpha}), \;\; \alpha \in (1,2],$$

where $\text{Hölder}^{\alpha}(C_{\alpha})$ for $\alpha \in (1,2]$ is the set of functions satisfying

$$\left| \frac{\partial}{\partial x} H(x_1) - \frac{\partial}{\partial x} H(x_0) \right| \leq C_{\alpha} |x_1 - x_0|^{\alpha - 1},$$

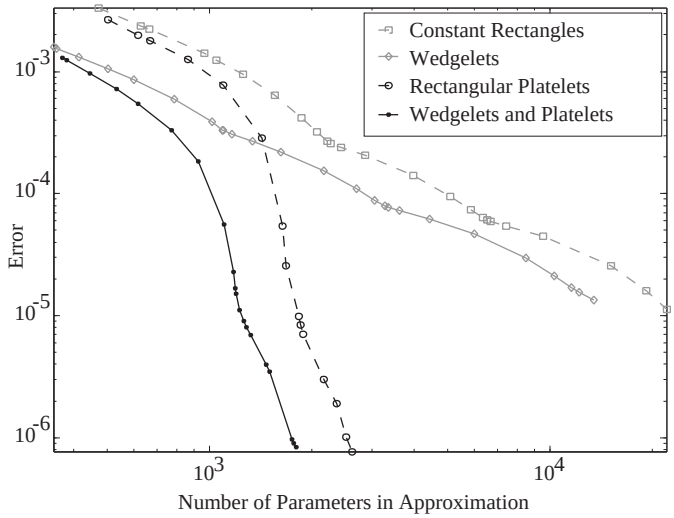

Figure 5: Approximation errors *vs.* number of parameters for bowl image

for all $x_0, x_1 \in [0,1]$. For more information on Hölder spaces see [34].

The smoothness of the images f_1 and f_2 is characterized by a two-dimensional surface smoothness condition defined in [35]

$$f_i \in \Sigma_2(\beta, C_\beta), \;\; \beta \in (1,2], \;\; L > 0, \;\; i = 1,2,$$

where $\Sigma_2(\beta, C_\beta)$ is the set of functions $f : [0,1]^2 \to \mathbb{R}^1$ with $k = \lfloor \beta \rfloor = 1$ continuous partial derivatives satisfying

$$|f(y_1) - p_{y_0}(y_1)| \leq C_\beta \, |y_1 - y_0|^\beta ,$$

for all $y_0, y_1 \in [0,1]^2$, where $p_{y_0}(y_1)$ is the Taylor polynomial of order k for $f(y_1)$ at point y_0.

The model describes an image composed of two smooth surfaces separated by a Hölder smooth boundary. This is similar to the "grey-scale boundary fragments" class of images defined in [35], but our model has fewer restrictions on boundary, consistent with the boundaries in [32]. 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 global squared L_2 error of m-term platelet approximations for images of this form is bounded in the following theorem (proved in Appendix A).

Theorem 1 *Consider the class of images*

$$f(x,y) = f_1(x,y) \cdot I_{\{y \geq H(x)\}} + f_2(x,y) \cdot \left(1 - I_{\{y \geq H(x)\}}\right)$$

for all $(x,y) \in [0,1]^2$ *where* $f_i \in \Sigma_2(\beta, C_\beta)$, $i = 1,2$, *and* $H \in$ *Hölder*$^\alpha(C_\alpha)$ *with* $\alpha, \beta \in (1,2]$. *Suppose that* $2 \leq m \leq 2^J$, *with* $J > 1$. *The squared* L_2 *error of an* 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 1 shows that for images consisting of smooth regions ($\beta \in (1,2]$) separated by smooth boundaries ($\alpha \in (1,2]$) 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 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. The reader is referred to [28, 32, 36] for the Fourier and wavelet error rates. Wedgelets can handle boundaries of this type, but produce piecewise 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 sidelength $O(1/\sqrt{m})$. 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 1 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 doubles the error from the case of a single boundary).

We do not consider images with higher orders of smoothness (*i.e.* $\alpha, \beta > 2$) in this analysis, although proving an analog to Theorem 1 for a more general case would be a simple theoretical problem. In order to achieve the approximation error decay rates derived in Theorem 1, it would be necessary to use higher order approximating functions. For example, in approximating the class of images where $\alpha, \beta \in (2, 3]$, one would need to use quadratic atoms separated by quadratic boundaries to achieve the desired rates. While the theoretical aspects of such an approximation are straightforward, the implementation is intractable. Currently, optimal wedgelets are determined through an exhaustive search of each slope-intercept combination. Computing optimal quadratic boundaries would mean introducing an additional degree of freedom and significantly increasing the computational complexity of the algorithm. Quadratic surface fits on a wedgelet partition space would not be difficult to compute, but the approximation error decay rate would still be fundamentally limited by the wedgelets.

3 Platelet Analysis of Poisson Processes

In the previous section, we determined that platelets can accurately and efficiently approximate functions in certain smoothness classes. Now suppose that we observe a two-dimensional realization of a Poisson process, and wish to exploit the approximation capabilities of platelets in estimating the underlying Poisson intensity function. We describe here likelihood factorizations based on platelets, a key factor in the optimality of the estimation algorithm described in the following section. This extends the work done by Kolaczyk and Nowak for piecewise constant *multiscale 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.

Suppose that $x(u, v)$ is a realization of a Poisson process. Underlying this process is a continuous intensity function $\lambda(u, v), (u, v) \in [0, 1]^2$. Assume that either by choice or perhaps 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 $\boldsymbol{\lambda} \equiv \{\lambda_{m,n}\}_{m,n=0}^{N-1}$. Each $x_{m,n}$ is simply the number of events in the square $S_{m,n}$ and $\lambda_{m,n} \equiv \int_{S_{m,n}} \lambda(u,v)$. The counts are conditionally independent; given $\{\lambda_{m,n}\}$, $x_{m,n} \sim \text{Poisson}(\lambda_{m,n})$. The Poisson likelihood of $\boldsymbol{x}$, given the intensities $\boldsymbol{\lambda}$, is denoted by $p(\boldsymbol{x}|\boldsymbol{\lambda})$.

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. The multiscale likelihood factorizations 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 position/scale combinations offered by a given C-RP $\mathcal{P}^*$ (in this case the partition underlying the dyadic Haar analysis) [29–31]. The keys to the likelihood factorization are the facts that sums of Poisson variates are Poisson and that the conditional distribution of a collection of Poisson variates given their sum is multinomial. Therefore, the conditional distribution of four children $\{x_{ch(S)}\}$ given the parent x_S is multinomial. For a given C-RP $\mathcal{P}^*$ the likelihood $p(\boldsymbol{x}|\boldsymbol{\lambda})$ is factorized as

$$p(\boldsymbol{x}\,|\,\boldsymbol{\lambda}) = p(x_{S_0}\,|\,\lambda_{S_0}) \prod_{S\in NT(\mathcal{P}^*)} p(\{x_{ch(S)}\}|x_S,\theta_S) \ , \tag{2}$$

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 $\lambda_S\theta_S^1$, $\lambda_S\theta_S^2$, $\lambda_S\theta_S^3$, and $\lambda_S(1-\theta_S^1-\theta_S^2-\theta_S^3)$. The expression in (2) 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. In fact, there is a one-to-one mapping between the multinomial parameters and the Haar wavelet and scaling coefficients. The factorization in (2) 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 [29].

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

$$\begin{aligned} p(\boldsymbol{x}\,|\,\boldsymbol{\lambda}(\mathcal{P},\boldsymbol{\theta})) = {} & p(x_{S_0}\,|\,\lambda_{S_0})\times \\ & \prod_{S\in NT(\mathcal{P})} p(\{x_{ch(S)}\}|x_S,\theta_S)\times \\ & \prod_{S\in T(\mathcal{P})} p(\{x_{m,n}\}_{(\frac{m}{N},\frac{n}{N})\in S}|x_S), \end{aligned} \tag{3}$$

where $\boldsymbol{\theta} \equiv \lambda_{S0} \cup \{\theta_S\}_{S\in(\mathcal{P})}$ and $T(\mathcal{P})$ is the set of all terminal squares in $\mathcal{P}$ and the conditional likelihood corresponding to each terminal square $p(\{x_{m,n}\}_{(\frac{m}{N},\frac{n}{N})\in S}|x_S)$ is multinomial with equal probabilities (since the intensity in the square is modeled as constant). Note that the intensity is constrained to be defined with respect to the partition, as indicated by the notation $\boldsymbol{\lambda}(\mathcal{P},\boldsymbol{\theta})$. For readability, however, this function will simply be referred to as $\boldsymbol{\lambda}$ for the remainder of this paper. 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 due to the fact that 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 (e.g., arising when summing over a piece of the partition) is Poisson distributed. Thus, Haar multiscale analysis and its generalizations are especially well suited to the study of Poisson data.

Wedgelet partitions produce a multiscale likelihood factorization of the form (3), in which case the terminal nodes of the partition can denote wedge-shaped regions in addition to squares; that is, $S \in T(\mathcal{P})$ 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, 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 sidelength 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 parameters for the slope of the gradient in addition to a parameter corresponding to the estimated total intensity of the associated square or wedge shaped region. In this case,

$$\begin{aligned} p(\boldsymbol{x} \,|\, \boldsymbol{\lambda}) = p(x_{S_0} \,|\, \lambda_{S_0}) \times & \\ \prod_{S \in NT(\mathcal{P})} p(\{x_{ch(S)}\} | x_S, \theta_S) \times & \\ \prod_{S \in T(\mathcal{P})} p(\{x_{m,n}\}_{(\frac{m}{N}, \frac{n}{N}) \in S} | x_S, \theta_S), & \end{aligned} \tag{4}$$

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

4 Photon-Limited Image Denoising

The recursive partitions and likelihood factorizations discussed in the previous two sections are critical to the optimal tree-pruning estimation/denoising algorithm detailed in this section. We will see that the maximum penalized likelihood criterion coupled with the likelihood factorizations results in a globally optimal estimation/denoising algorithm. The potential effectiveness of this algorithms is demonstrated by applying the algorithm to real and simulated data.

4.1 Maximum Penalized Likelihood Estimators

The multiscale likelihood factorizations above provide for a very simple framework for maximum penalized likelihood estimation, wherein the penalization is based on the complexity of the underlying partition. While this paper focuses on complexity regularization, a variety of penalization schemes could be effective, including the roughness penalties described in [37]. A key feature of our penalization scheme is that the penalty of each partition cell is inherited along with the pixel estimates for that cell, as will be described shortly. This inheritance is possible because the penalties are separable and additive, which ensures the optimality of the tree-pruning algorithm proposed in Section 4.2. Any alternative penalization scheme must exhibit this separability and additivity.

In our complexity regularization scheme, the complexity of a given partition is proportional to the total number of regions. For example, a C-RP partitions the image into each individual pixel, and thus has complexity $O(N^2)$. An (incomplete) RP with R regions has complexity $O(R)$. The maximum penalized likelihood criterion we employ here is

$$L_\gamma(\boldsymbol{\lambda}) \;\equiv\; \log p(\boldsymbol{x} \,|\, \boldsymbol{\lambda}) \;-\; \gamma \{\#\boldsymbol{\theta}\}, \tag{5}$$

where $p(\boldsymbol{x} \mid \boldsymbol{\lambda})$ denotes a likelihood (factorization) of the form (2), (3), or (4), and $\{\#\boldsymbol{\theta}\}$ is the number of parameters in the vector $\boldsymbol{\theta}$ (one for each constant region, three for each planar region). (Recall $\boldsymbol{\lambda}$ is a function of the multiscale parameters and partition; *i.e.* $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta}, \mathcal{P})$.) 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

$$\begin{aligned} (\widehat{\mathcal{P}}, \widehat{\boldsymbol{\theta}}) &\equiv \arg\max_{\mathcal{P},\boldsymbol{\theta}} L_\gamma(\boldsymbol{\lambda}(\mathcal{P}, \boldsymbol{\theta})) \\ \widehat{\boldsymbol{\lambda}} &\equiv \boldsymbol{\lambda}(\widehat{\mathcal{P}}, \widehat{\boldsymbol{\theta}}) \end{aligned} \tag{6}$$

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 N^2 free parameters) maximizes the likelihood. Since a C-RP corresponds to a pixel-based partition, in the $\gamma = 0$ case the MPLE reduces to the conventional maximum likelihood estimator (MLE).

The approximation-theoretic result of Theorem 1 has 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 $m \propto \{\#\boldsymbol{\theta}\}$ in a platelet approximation. The squared bias is proportional to the error of an m-term platelet approximation to the underlying true intensity. We have shown 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. 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 1 shows that for a small value of m there exists an accurate m-term platelet approximation to the true intensity. If $m \ll N^2$, then the reduction in variance will be dramatic, and the platelet-based MPLE will tend to have a small MSE. Our ongoing work includes a careful analysis of the minimax bounds on the error.

Maximizing (5) 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 [33]. A solution to (6) 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 image with higher count levels (higher SNR) and/or where strong edges 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, 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 Optimal Pruning Algorithm

Observe that the structure of the penalized likelihood criterion stated in (5) and the likelihood factorization described in Section 3 allow an optimal intensity estimate to be computed quickly. Equation (5) may be expanded using (4) to yield:

$$\begin{aligned} L_\gamma(\boldsymbol{\lambda}) = \log p(x_{S_0} \mid \lambda_{S_0}) \; + \text{pen}\,(\lambda_{S_0}) + \\ \sum_{S \in NT(\mathcal{P})} \log p(\{x_{ch(S)}\} | x_S, \theta_S) + \text{pen}\,(\theta_S) + \\ \sum_{S \in T(\mathcal{P})} \log p(\{x_{m,n}\}_{(\frac{m}{N}, \frac{n}{N}) \in S} | x_S, \theta_S) + \text{pen}\,(\theta_S). \end{aligned} \tag{7}$$

The penalty of the total image intensity parameter, λ_{S_0}, is γ because of the singular dimension of the parameter. Each nonterminal node's penalty, pen (θ_S), $S \in NT(\mathcal{P})$, is either γ or $3\,\gamma$ because once the value of the parent node is

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 $0 \le i \le 3$
	$L_\gamma(\boldsymbol{\theta}_{H_4}; S_{j,m,n}) = \sum_{S' \in ch(S_{j,m,n})} L_{min}(S')$
Save:	$L_{min}(S_{j,m,n}) = \min_{0 \le i \le 4} L_\gamma(\boldsymbol{\theta}_{H_i}; S_{j,m,n})$
	$\boldsymbol{\theta}_{min}(S_{j,m,n}) = \arg\min_{0 \le i \le 4} L_\gamma(\boldsymbol{\theta}_{H_i}; S_{j,m,n})$
Coarsen Scale:	$j = j - 1$
Goto Loop if $j \ge 0$	
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 1: Algorithm Pseudocode

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 (θ_S), $S \in T(\mathcal{P})$, 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. The likelihood factorization allows the likelihood of the entire image to be represented in 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 RP of the data using a fast algorithm reminiscent of dynamic programming, Coifman and Wickerhauser's "Best-Ortho-Basis" algorithm [38], or the CART algorithm [33].

The goal is to estimate the intensity $\boldsymbol{\theta}$ according to (6). In order to perform the estimation, the 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:

- $\mathbf{H}_0$: Square Constant = homogeneous square (terminal node)
- $\mathbf{H}_1$: Wedgelet = two homogeneous wedges (terminal node)
- $\mathbf{H}_2$: Square Platelet = dyadic linear gradient on square (terminal node)
- $\mathbf{H}_3$: Wedged Platelet = linear gradients on two wedges (terminal node)
- $\mathbf{H}_4$: 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). The algorithm begins one scale above the leaf nodes in the quad tree and traverses upwards, performing a tree-pruning operation at each stage. 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 above, 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 1. 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.

4.3 Computational Complexity

Before examining the computational complexity of this algorithm, we 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 [39]. In this case, the likelihood in the platelet factorization is concave in the platelet 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 B, establishes the concavity. Consider a terminal node log likelihood term for a platelet: $\log p(\{x_{m,n}\}_{(\frac{m}{N},\frac{n}{N})\in S}|x_S, \theta_S)$. The subscripts 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 1 *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 B.

Theorem 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.

4.4 Penalty Parameter Selection

The selection of the penalty parameter γ in the penalized likelihood (5) 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 [33, 40] as well as for Haar-based MPLE Poisson denoising [29]. In those works, it is shown that near minimax optimal results are obtained with $\gamma = c\log(n)$, where n is the number of pixels or the total number of counts and c is a constant. Here we use this form for γ and determine a "good" setting for the constant c.

Consider the graphs in Figures 6 and 7. 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) we 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

global squared error) in the plots. We see 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. The fact that γ is proportional to $\log(\#\text{counts})$ provides a built-in adaptivity to the underlying SNR level.

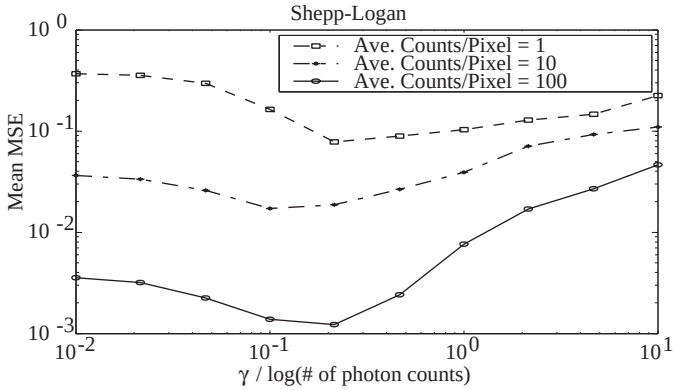

Figure 6: Wedgelet denoising on Shepp-Logan phantom. Variation of MSE with penalty parameter γ indicates that $\gamma = \frac{1}{5}\log(\#\text{counts})$ is adequate for wedgelet denoising for different SNRs.

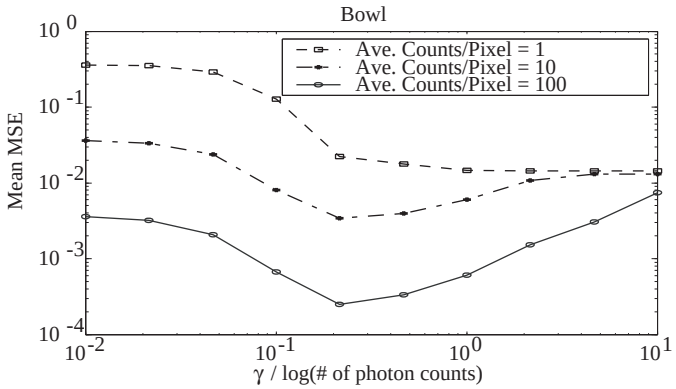

Figure 7: Platelet denoising on quadratic bowl image. Variation of MSE with penalty parameter γ indicates that $\gamma = \frac{1}{5}\log(\#\text{counts})$ is adequate for platelet denoising for different SNRs.

Here, as in all experiments described in this paper, we calculate errors using a global L_2 metric. This is an intuitive, straightforward metric commonly used to asses quality for general image processing tasks. There do exist a number of other metrics for assessing how well an algorithm accomplishes a specific task, but the MSE is an acceptable general measure of performance. This global measure may not be sensitive to some characteristics central to some medical imaging tasks; this warrants further investigation. We acknowledge that the experimental results presented in this paper do not completely characterize the numerical performance of platelet-based MPLE algorithms, but such evaluations are beyond the scope of this paper. The purpose of this paper is propose a new method for reconstructing medical images, prove theoretically that it has the potential to outperform some existing reconstruction algorithms, and provide some initial examples of the algorithms' effectiveness. A platelet-based algorithm could constitute the core of several different, sophisticated reconstruction algorithms, designed to compete with state-of-the-art algorithms

for specific medical imaging applications. Extensive numerical validation of the platelet-based MPLE would be best suited to this later stage.

4.5 Denoising Experiments and Applications

To demonstrate the effectiveness of platelet-based denoising, we have applied both the Haar-based and platelet-based MPLE to a confocal microscopy phantom with Poisson noise. This phantom is described in detail in Section 5.3. The 128×128 noisy data contained a total of 107,000 counts. The phantom and noisy data are displayed in Figures 8(a) and (b), respectively. Figures 8(c) and (d) show the Haar- and platelet-based MPLE results. We found that a penalty of $\gamma = \frac{1}{2}\log(\#\text{counts})$ yielded good results in this experiment. It was not chosen to minimize any error criterion, but because it resulted in an image of similar smoothness and quality to the images produced using real data below. Here the global MSE is lower for the platelet estimate, and the resulting image is visually more appealing than the Haar image due to the combination of smooth edges and smooth surfaces.

For the second set (Figures 8(e) and (f)), we applied a technique called "averaging over shifts" or "cycle-spinning" [41, 42]. 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. We employ this technique in all subsequent experiments. Both the Haar- and platelet-based image estimates improve (both in an MSE sense and in the reduction of visible artifacts) as a result of this technique.

Figure 9(a) depicts an image of a heart obtained from a nuclear medicine study. The image was 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 9(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 [43].

The hereditary Haar and platelet based MPLE results are shown below below. The first set of estimates (Figures 9 and 10 (b) and (c)) were generated by one pass through the denoising algorithm. In both the heart and the spine images, the Haar-based MPLE results effectively capture the image intensity, yet exhibit considerable blocking artifacts. In contrast, the platelet-based MPLE results provide much smoother intensity estimates. Furthermore, the artifacts do not blur structure edges to the same extent as the Haar blocking artifacts. For the second set (Figures 9 and 10 (d) and (e)), we applied averaging over shifts again, which resulted in a significant reduction in both blocking and platelet artifacts. The platelet-based MPLE images are smooth without the blurred appearance of the Haar-based MPLE images. This is particularly evident in the spine, where the platelets enhance the appearance of individual vertebrae.

5 Photon-Limited Image Restoration and Reconstruction

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 $\boldsymbol{\lambda}$. 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

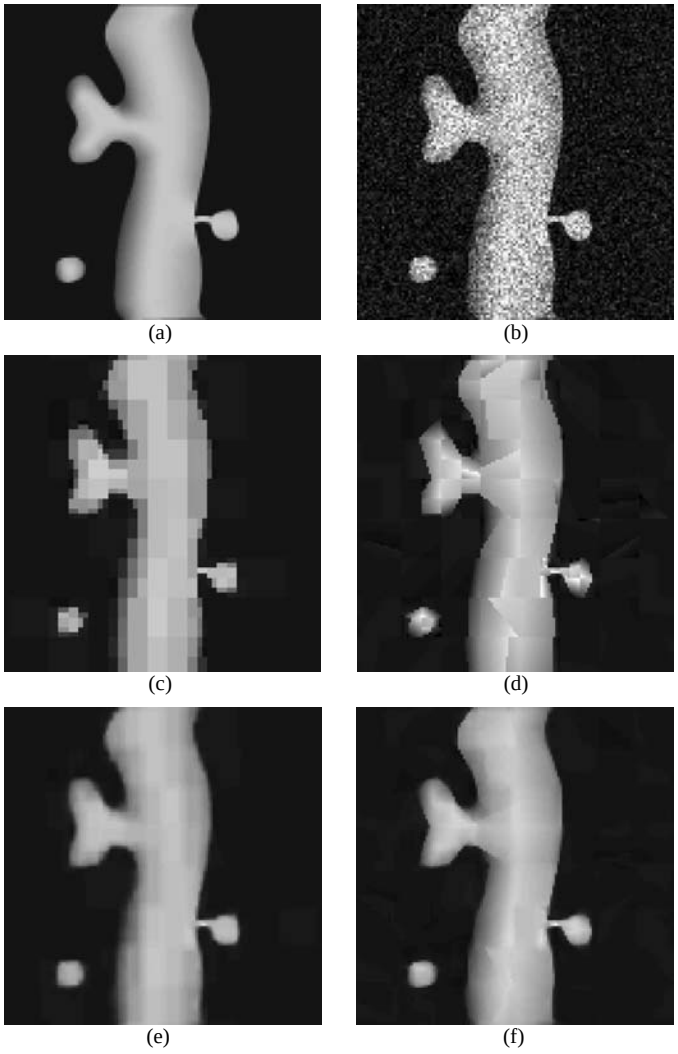

Figure 8: Denoising of confocal microscopy phantom. (a) Phantom image. (b) Noisy image. (c) Haar-based MPLE estimate. MSE=0.0173. (d) Platelet-based MPLE estimate; MSE = 0.0089. (e) Haar-based MPLE estimate averaged over 25 shifts; MSE = 0.0057. (f) Platelet-based MPLE estimate averaged over 25 shifts; MSE = 0.0034.

distributed according to

$$x_{m,n} \sim \text{Poisson}\left(\sum_{k,l} p_{k,l,m,n}\,\lambda_{k,l}\right). \tag{8}$$

Notice that the intensity of the observation is $\sum_{k,l} p_{k,l,m,n}\lambda_{k,l}$, rather than $\lambda_{m,n}$ as in the direct case. The transition probabilities $p_{k,l,m,n}$ represent the blurring or projection process. The recovery $\boldsymbol{\lambda}$ 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 $\boldsymbol{\lambda}$ is denoted $p(\boldsymbol{x}|\boldsymbol{\lambda})$. The log-likelihood is

$$\log\, p(\boldsymbol{x}|\boldsymbol{\lambda}) = \sum_{m=0}^{M_1}\sum_{n=0}^{M_2}\left[-\sum_{k,l=0}^{N-1} p_{k,l,m,n}\lambda_{k,l} + x_{m,n}\log\left(\sum_{k,l=0}^{N-1} p_{k,l,m,n}\lambda_{k,l}\right) - \log x_{m,n}!\right], \tag{9}$$

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 arbitrarily determined by the detection system, and 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 [16] 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). The complete data are Poisson distributed according to

$$z_{k,l,m,n} \sim \text{Poisson}(\lambda_{k,l}\, p_{k,l,m,n}) \ . \tag{10}$$

Hence the observed data $\boldsymbol{x}$ in (8) 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 $y_{k,l} \sim \text{Poisson}(\lambda_{k,l})$. 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 [44], we developed an EM algorithm that can be adapted to the MPLE considered here. The approach developed in our 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 [44] showed that its performance was competitive with (but could be slightly inferior to) stopped EM-MLE reconstruction methods. Here we demonstrate that the Haar and wedgelet based MPLEs perform *better* than even the best possible stopped EM-MLE procedure both perceptually and in terms of the global MSE.

5.1 Maximum Likelihood Estimation

It is well-known that the maximizer of (9) 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 [45], 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 [16]. 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 (9) [5].

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., [17]). 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.

5.2 EM-MPLE using Platelet Approximations

The maximum penalized likelihood function employed here is

$$L_\gamma(\boldsymbol{\lambda}) \ \equiv \ \log p(\boldsymbol{x} \,|\, \boldsymbol{\lambda}) \ - \ \gamma \, \{\#\boldsymbol{\theta}\}, \tag{11}$$

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., $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta}, \mathcal{P})$. 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 $\boldsymbol{\lambda}$ (without explicit indication of the dependence on $\mathcal{P}$ and $\boldsymbol{\theta}$).

The EM algorithm is based on consideration of the following conditional expectation:

$$L_\gamma^c(\boldsymbol{\lambda}) \;\equiv\; \log p(\boldsymbol{z} \,|\, \boldsymbol{\lambda}) \;-\; \gamma\,\{\#\boldsymbol{\theta}\}, \tag{12}$$

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 $L_\gamma^c(\boldsymbol{\lambda})$ 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 the surrogate function

$$Q(\boldsymbol{\lambda}^{(i)}, \boldsymbol{\lambda}) \;\equiv\; E_{\boldsymbol{\lambda}^{(i)}}\left[L_\gamma^c(\boldsymbol{\lambda})|\boldsymbol{x}\right], \tag{13}$$

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

$$E_{\boldsymbol{\lambda}^{(i)}}\left[L_\gamma^c(\boldsymbol{\lambda})|\boldsymbol{x}\right] = E_{\boldsymbol{\lambda}^{(i)}}\left[\log p(\boldsymbol{z} \,|\, \boldsymbol{\lambda})|\boldsymbol{x}\right] \;-\; \gamma\,\{\#\boldsymbol{\theta}\}. \tag{14}$$

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 $E_{\boldsymbol{\lambda}^{(i)}}\left[\log p(\boldsymbol{z} \,|\, \boldsymbol{\lambda})|\boldsymbol{x}\right]$. The complete data log likelihood $\log p(\boldsymbol{z} \,|\, \boldsymbol{\lambda})$ happens to be a linear function of $\boldsymbol{z}$ and so this calculation simplifies to computing the $\boldsymbol{z}^{(i)} \equiv E_{\boldsymbol{\lambda}^{(i)}}[\boldsymbol{z}|\boldsymbol{x}]$. A closed-form expression for this calculation can be found in [5]. 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 $Q(\boldsymbol{\lambda}^{(i)}, \boldsymbol{\lambda})$ over $\boldsymbol{\lambda}$. The penalty, which is a function of $\boldsymbol{\lambda}$, plays a key role in this step. Based on the derivation of the EM algorithm in [44], the surrogate function can be written as

$$Q(\boldsymbol{\lambda}^{(i)}, \boldsymbol{\lambda}) \;=\; \log p(\boldsymbol{y}^{(i)} \,|\, \boldsymbol{\lambda}) \;-\; \gamma\,\{\#\boldsymbol{\theta}\} \;+\; C(\boldsymbol{x}), \tag{15}$$

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 $\boldsymbol{\lambda}$. Thus, the M-Step is equivalent to maximizing $\log p(\boldsymbol{y}^{(i)} \,|\, \boldsymbol{\lambda}) - \gamma\,\{\#\boldsymbol{\theta}\}$. 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 $y_{k,l}^{(i)} = \sum_{m,n} z_{k,l,m,n}^{(i)}$, 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 $\boldsymbol{\lambda}$) that maximizes $Q(\boldsymbol{\lambda}^{(i)}, \boldsymbol{\lambda})$. This results in the next iterate of the algorithm, $\boldsymbol{\lambda}^{(i+1)}$. Each iteration of this algorithm has the computational complexity stated in Theorem 2.

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 (11). Two applications of the MPLE and EM algorithm are examined next.

5.3 Confocal Microscopy

Confocal microscopy is used to obtain volume images of small fluorescent objects with high spatial resolution [3]. 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 [6]. 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, we have created a scale phantom of an image of a dendrite segment with thin, stubby, and mushroom spines. See [46] 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 11(a) contains the 128×128 phantom developed for this experiment, and Figure 11(b) contains the blurred and noisy data. The best MLE estimate in Figure 12(a) 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 (in the L_2 sense) image obtainable with the commonly used stopped-MLE technique [17]. Figure 12(b) is a closeup of a small region of Figure 12(a). The averaged-over-shifts Haar-based MPLE appears in Figure 12(c), and its closeup in Figure 12(d). Likewise, the averaged-over-shifts platelet-based MPLE appears in Figure 12(e) and its closeup in Figure 12(f). These images reveal the nature of the artifacts caused by each of the algorithms. The EM-MLE image exhibits blurred edges and a mottled surface. The Haar-based MPLE image also exhibits blurred edges but a smoother surface with some blocking artifacts. The platelet-based MPLE exhibits sharp edges and some blocking artifacts on the surface. Since the platelets approximate image gradients better than the Haar basis, the platelet blocking artifacts are at a larger scale than the Haar-based blocking artifacts. The blocking artifacts are a result of the underlying recursive dyadic partitioning scheme and common to many multiresolution techniques. Translation invariant methods, which should reduce these effects, are a focus of ongoing work. Finally, Figure 13 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 14 demonstrates the capabilities of platelet analysis on a real confocal microscopy image of a dendritic spine. [1] 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.

5.4 Emission Computed Tomography

Here we consider the application of our 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

[1] The authors would like to thank Tycho Hoogland, Division of Neuroscience, Baylor College of Medicine, for this data and valuable discussions.

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 $\boldsymbol{\lambda}$). The probability transition matrix $\boldsymbol{p}$ is derived from the physics and geometry of the detection device and data collection process [5].

In Figure 15 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 $\boldsymbol{\lambda}$ is a 64×64 square image shown in Figure 15(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 [47]. $\boldsymbol{p}$ was applied to $\boldsymbol{\lambda}$ to obtain $\boldsymbol{\mu}$, and we used a standard Poisson random number generator to synthesize the projection data $\boldsymbol{y}$. The Shepp-Logan model is piecewise constant, and so we employ a wedgelet-based MPLE instead of using platelets.

For comparison, in Figure 15(b) we show the very best likelihood-based reconstruction obtained by stopping the likelihood-based EM algorithm at the reconstruction having the smallest global squared error, which is impossible to determine in practice since the true intensity is unknown. Figures 15 (c) and (d) display the Haar- and wedgelet-based MPLE algorithm results, respectively. A penalty of $\gamma = \frac{1}{50}\log(\#\text{counts})$ was found to give reasonable results in this problem. It was not selected to minimize any error criterion, but because it produced an image comparable in smoothness and overall quality to a state-of-the-art reconstruction described below. Comparison of the Haar- and wedgelet-based MPLE images reveals the effectiveness of the wedgelets in defining boundaries of structures within the image. While the wedgelet-based estimate exhibits an slightly higher overall MSE, the visual quality of the image is superior.

For further comparison, Figure 15(e) shows the filtered back-projection reconstruction and Figure 15(f) shows the best reconstruction obtained using penalized weighted least squares with a partitioned separable paraboloidal surrogate (SPS) coordinate ascent reconstruction algorithm [48], calculated using Fessler's Tomography Toolbox [49]. This algorithm is representative of modern tomographic reconstruction capabilities. In this case, the penalization parameters β and δ were selected after conducting an exhaustive search for the values which minimize the global MSE; such an operation would not be possible in practice. Both in terms of global MSE and visual quality, the Haar- and wedgelet-based MPLE estimates are comparable to this SPS estimate. This implies that the Haar- and wedgelet-based MPLEs may out-perform the SPS algorithm if the penalties were selected in a more sophisticated manner (e.g., cross-validation methods, user-interaction). These favorable results encourage further study of the effect of different initialization and penalization schemes on reconstruction quality. For example, level-dependent penalization [50] is a common practice in inverse problems such as tomography, and its application to Haar- and wedgelet-based MPLEs could further improve reconstructed image quality.

6 Conclusions and Ongoing Work

This paper introduced the platelet representation for the analysis, denoising, and reconstruction of photon-limited medical images. Platelets can outperform conventional wavelet representations because of their ability to 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 confocal microscopy and nuclear medicine demonstrate that platelet-based methods can outperform (by achieving lower global L_2 errors) the popular and widely used stopped EM-MLE method.

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, we are 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 roughness penalties with those based on platelets. Quadratic roughness penalties can be interpreted as a penalty weighting applied to a Fourier expansion of the intensity. We have demonstrated theoretically that platelet approximations to piecewise smooth intensity functions can significantly outperform Fourier approximations. Thus, platelet-based MPLE may produce better numerical results than those of quadratically roughness penalized MPLEs.

In contrast, it should be possible to quantify the theoretical error performance of the platelet-based MPLE very precisely. In [29] we show that the Haar-based MPLE is near minimax optimal when the underlying Poisson intensity belongs to Bounded Variation or Besov function spaces [34]. These spaces are characterized by mostly smooth images with isolated point singularities. Our interest in this work is in images with singularities along smooth curves (edges and boundaries) rather than at points, and we expect that improved minimax bounds can be obtained for our platelet-based MPLE in such cases. We are 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 [50, 51]. We are 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 platelet-based MPLE, extensive experimental comparisons between it and competing methods are essential. This includes more extensive experiments with real and synthetic data, receiver-operating-characteristic (ROC) studies [52], and clinical evaluations. Based on our theoretical analysis and initial experimentation presented here, we believe that platelet-based MPLEs may be quite competitive with state-of-the-art methods [48]. This, coupled with the tractable theoretical analysis of the platelet paradigm discussed above, may make it a very attractive tool for medical imaging applications.

Finally, the spatial adaptivity of platelet-based MPLE reconstructions has potential implications on the design of photon-limited imaging systems. Most existing imaging systems bin photon events, placing a lower limit on the achievable spatial resolution. The message of multisresolution approaches, such as platelets, is that the spatial resolution of the reconstruction should not necessarily be uniform across the image; multiresolution estimation algorithms "zoom-in" when the data indicate fine detail structures like edges and "zoom-out" in regions that appear (statistically) smooth to reduce noise. In principle, the multiresolution reconstruction algorithm, not the imaging system, should determine the proper balance between resolution and SNR *in a spatially adaptive manner*. This suggests that an ideal photon-limited imaging system would measure and record the location of each photon event as accurately as possible, and then the reconstruction algorithm could adaptively determine the best resolutions in different regions of the image. Platelet-based MPLE methods could be an effective tool in assessing the associated trade-off between resolution and SNR.

A Proof of Approximation Theorem

Proof of Theorem 1 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 sidelength greater than or equal to 2^{-J}) wedgelet partition. From [32] 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 sidelength $1/m$. To sketch the idea, imagine tiling $[0,1]^2$ with m^2 dyadic squares of sidelength $1/m$. Because the boundary function is Hölder$^\alpha(C_\alpha)$, with $\alpha \in (1,2]$, it must also be Lipschitz (i.e., Hölder$^1(C_1)$, 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 [32], where C_1 is the Lipschitz constant above. This result holds for $2 \le m \le 2^J$.

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$^\alpha(C_\alpha)$ 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 we generate a new partition by subdividing all squares of sidelength 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 sidelength $1/\sqrt{m}$ or less. This partition also has $O(m)$ regions. To see this, note that tiling $[0,1]^2$ with squares of sidelength $1/\sqrt{m}$ 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). Recall that $p_{(x_0,y_0)}(x,y)$ is the Taylor polynomial of order k for $f(x,y)$ at point $(x_0,y_0) \in [0,1]^2$. Because each square has sidelength of at most $1/\sqrt{m}$, the mean value theorem allows us to bound above the approximation error at any point (x,y) in the square:

$$\begin{aligned} |f(x,y) - p_{(x_0,y_0)}(x,y)| &\le C_\beta \left(\sqrt{\frac{2}{m}}\right)^\beta \\ &= O(m^{-\beta/2}) \end{aligned}$$

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 sidelength of $1/m$. According to the wedgelet analysis above, the area between the true boundary and the wedge boundary in each square is $O(m^{-(\alpha+1)} + \delta m^{-1})$. (Since the image surface is in $\Sigma_2(\beta, C_\beta)$, $\beta \in (1,2]$, 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 sidelength 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 sidelength $1/\sqrt{m}$ or less in the smooth parts of the image with the platelet approximation of the boundary on dyadic squares of sidelength $1/m$ 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 [32] to be less than or equal to δ, rather than the loose bound of $O(\delta)$ we used above. Thus, the total squared error is bounded above by $K'_{\alpha,\beta}\,m^{-\min(\alpha,\beta)} + \delta$. ■

B Proof of Computational Complexity Theorem

Proof of Lemma 1 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 [53], 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 the multinomial log likelihood shows that

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

is concave θ. ■

Proof of Theorem 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 sidelength in one coordinate of one boundary endpoint. This is depicted in Figure 16. 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 sidelength of the image is 1 and therefore the sidelength 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.

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 $O(N^3)$. ■

References

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

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

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

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

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

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

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

[14] 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.

[15] 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.

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

[17] 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.

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

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

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

[21] 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.

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

[23] 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.

[24] 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.

[25] 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.

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

[27] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *International Conference on Image Processing*, 23-26 Oct. 1995, Washington, DC, USA, 1995, vol. 3, pp. 444–447.

[28] 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.

[29] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," submitted to Annals of Stat. August, 2001. Available at `http://cmc.rice.edu/`.

[30] 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.

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

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

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

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

[35] A. P. Korostelev and A. B. Tsybakov, *Minimax theory of image reconstruction*, Springer-Verlag, New York, 1993.

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

[37] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imaging*, vol. 19, no. 6, pp. 601–15, 2000.

[38] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and Their Applications*, J. S. Byrnes, J. L. Byrnes, K. A. Hargreaves, and K. Berry, Eds. Kluwer Academic, Boston, 1994.

[39] 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.

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

[41] 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.

[42] 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.

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

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

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

[46] 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.

[47] 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.

[48] Saowapak Sotthivirat and J. A. Fessler, "Partitioned separable paraboloidal surrogate coordinate ascent algorithm for image restoration," in *Proc. IEEE Intl. Conf. on Image Processing*, 2000, vol. 1, pp. 109–12.

[49] J. A. Fessler, "Matlab m-files for image reconstruction (tomography toolbox)," `http://www.eecs.umich.edu/~fessler/code/`.

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

[51] J. Lakifa and S. Mallat, "Thresholding estimators for linear inverse problems and deconvoluations"," to appear in Annals of Statistics, available at `http://www.cmap.polytechnique.fr/~kalifa/`, 2002.

[52] H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. iii. roc metrics, ideal observers, and likelihood-generating functions," *J. Opt. Soc. Am. A*, vol. 15, no. 6, 1998.

[53] T. R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1972.

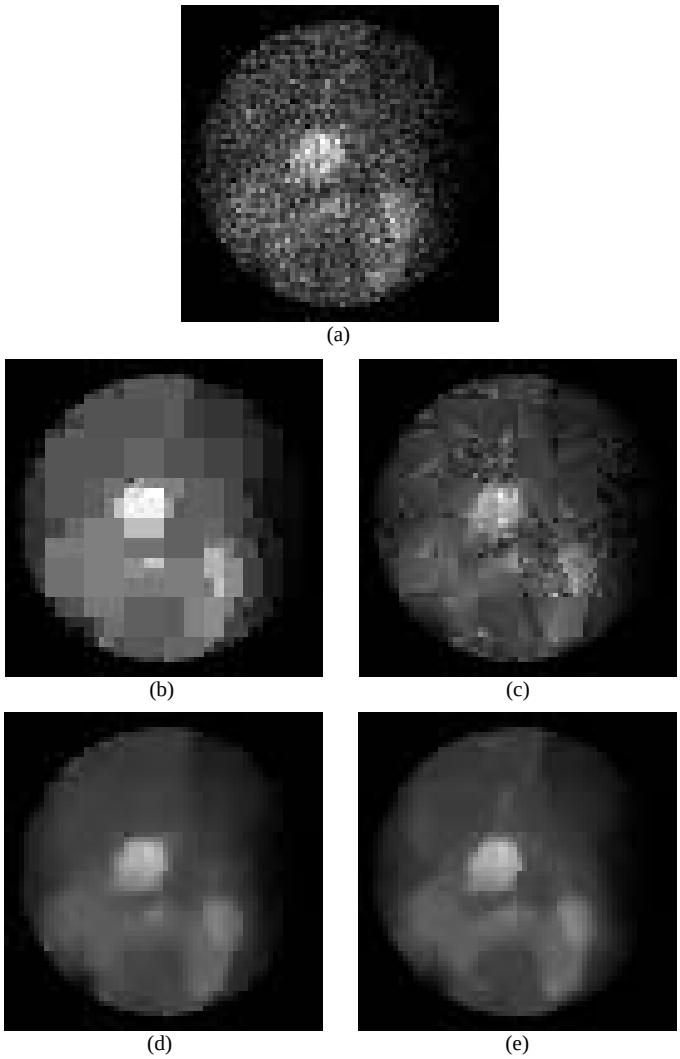

Figure 9: Denoising in nuclear medicine, cardiac example. (a) "Raw" nuclear medicine cardiac image (64×64 pixels). (b) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (c) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (d) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts. (e) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts.

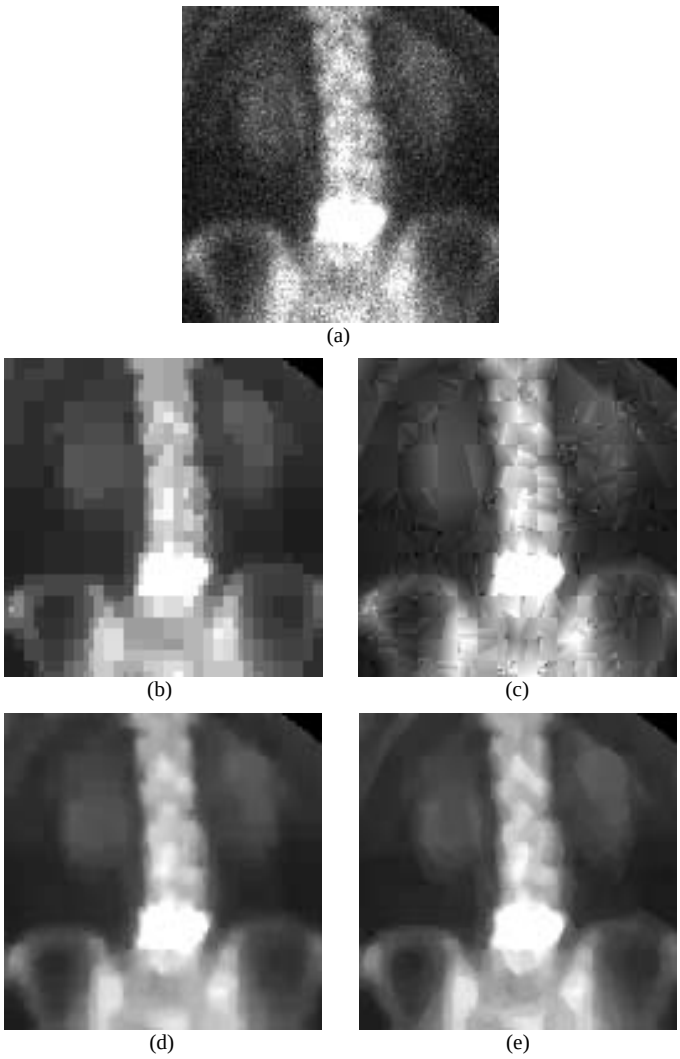

Figure 10: Denoising in nuclear medicine, spine example. (a) "Raw" nuclear medicine spine image (256×256 pixels). (b) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (c) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$. (d) Haar-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts. (e) Platelet-based MPLE, $\gamma = \frac{1}{5}\log(\#\text{counts})$, averaged over 25 shifts.

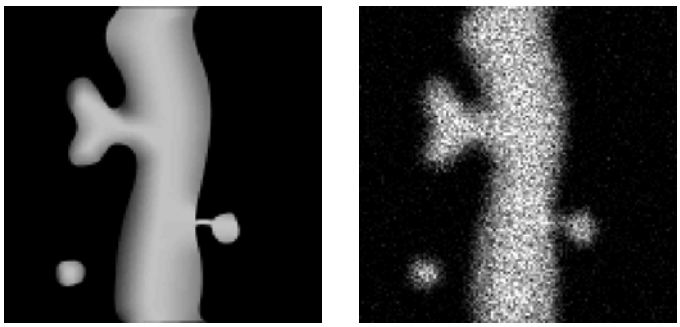

Figure 11: Confocal microscopy simulation. (a) Phantom (128×128 pixels). (b) Blurred and noisy phantom.

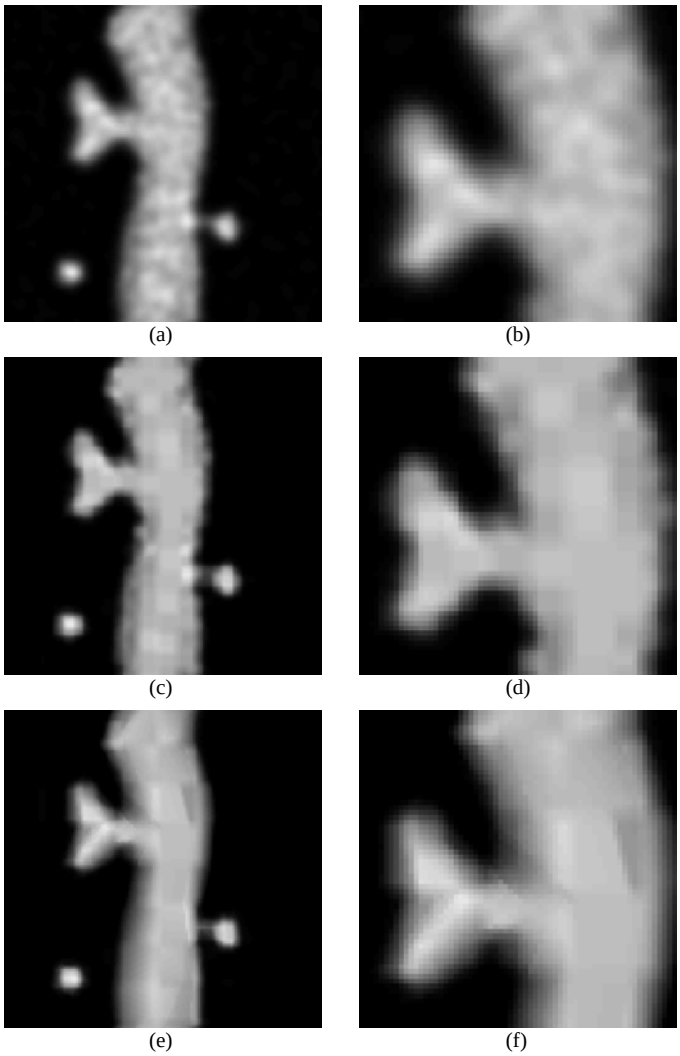

Figure 12: Confocal microscopy simulation. (a) Best EM-MLE restoration. (b) Best EM-MLE (zoomed). (c) Hereditary Haar MPLE (averaged over 5×5 shifts). (d) Hereditary Haar MPLE (zoomed). (e) Platelet MPLE (averaged over 5×5 shifts), (f) Platelet MPLE (zoomed). In all cases, $\gamma = \frac{1}{5}\log(\#\text{counts})$ and convergence was declared when $||\boldsymbol{\lambda}^{(i+1)} - \boldsymbol{\lambda}^{(i)}||_2 / ||\boldsymbol{\lambda}^{(i)}||_2 < 10^{-5}$ (21 iterations in this case).

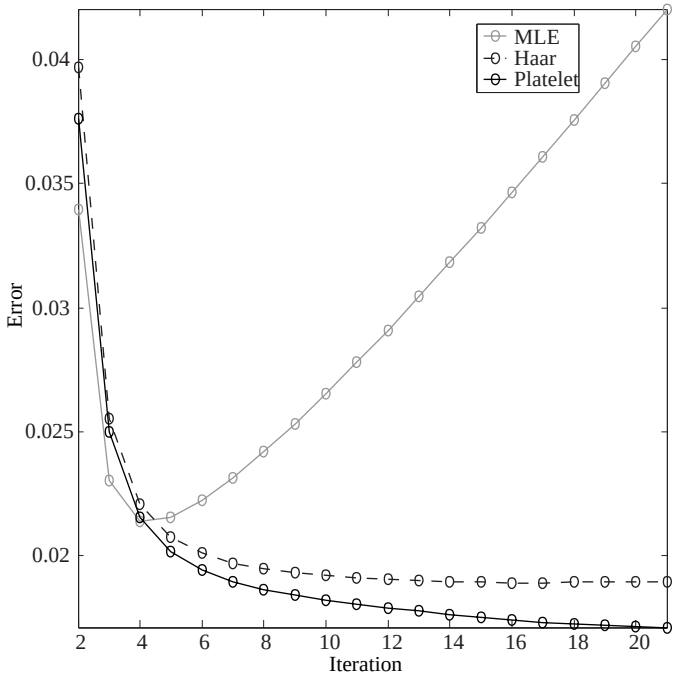

Figure 13: Confocal microscopy simulation – error at each by iteration.

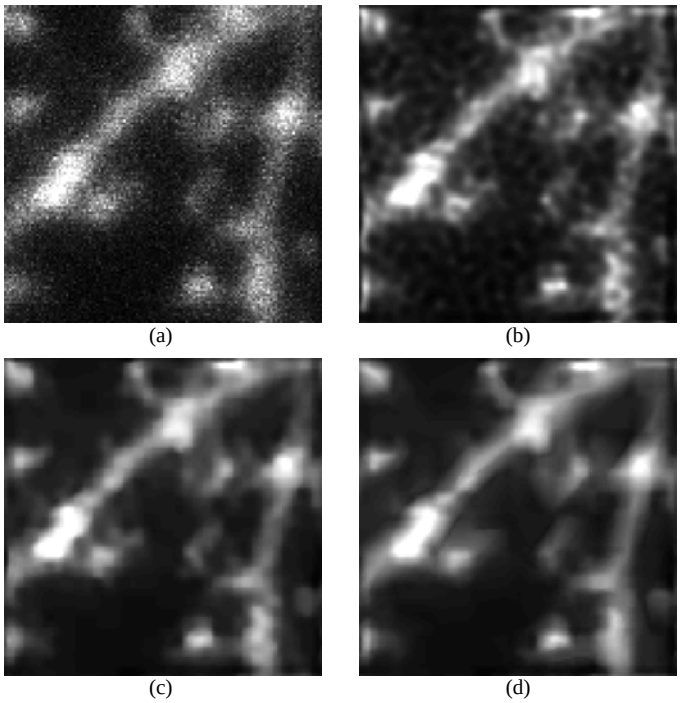

Figure 14: Confocal microscopy application. (a) Raw data (128×128 pixels). (b) EM-MLE restoration after 19 iterations. (c) Haar-based MPLE (averaged over 5×5 shifts). (d) Platelet-based MPLE (averaged over 5×5 shifts). In these cases, $\gamma = \frac{1}{5} \log(\#\text{counts})$ and convergence was declared when $||\boldsymbol{\mu}^{(i+1)} - \boldsymbol{\mu}^{(i)}||_2 / ||\boldsymbol{\mu}^{(i)}||_2 < 10^{-5}$ (19 iterations in this case).

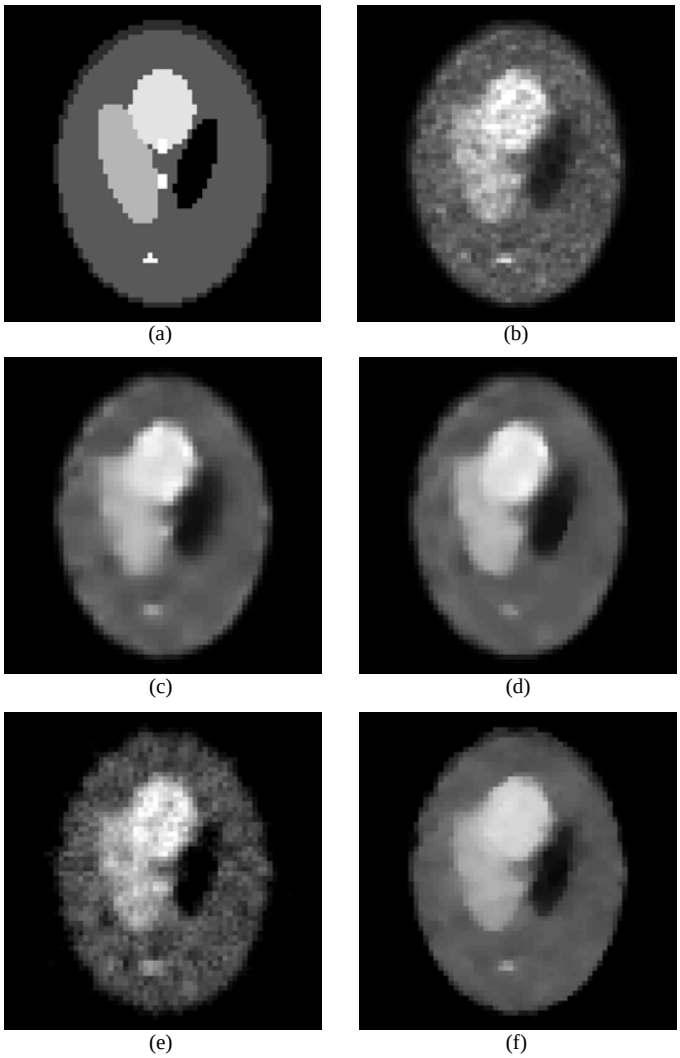

Figure 15: Tomography simulation. (a) Shepp-Logan phantom (64×64). (b) Best EM-MLE reconstruction (stopped at 14 iterations which gave the minimum global squared error reconstruction); MSE = 0.0461. (c) Non-hereditary Haar-based MPLE (averaged over 5×5 shifts); MSE = 0.0374. (d) Wedgelet-based MPLE (averaged over 5×5 shifts); MSE = 0.0378. (e) Filtered back projection reconstruction; MSE = 0.0807. (f) Best PWLS-SPS reconstruction; MSE = 0.0382. In all cases, $\gamma = \frac{1}{50}\log(\#\text{counts})$ and convergence was declared when $||\boldsymbol{\lambda}^{(i+1)} - \boldsymbol{\lambda}^{(i)}||_2 / ||\boldsymbol{\lambda}^{(i)}||_2 < 10^{-5}$ (at most 32 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.

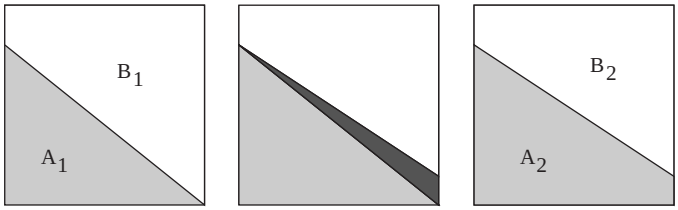

Figure 16: Sequential calculation of wedgelet likelihoods