FAST, NEAR-OPTIMAL, MULTIRESOLUTION ESTIMATION OF POISSON SIGNALS AND IMAGES

Rebecca Willett and Robert Nowak

Department of Electrical and Computer Engineering, Rice University
MS 366, Houston, TX 77251-1892, USA
Department of Electrical and Computer Engineering, University of Wisconsin
1415 Engineering Drive, Madison, WI 53706

ABSTRACT

The nonparametric multiscale methods presented here are powerful tools for Poisson signal and image denoising and reconstruction. These methods offer near minimax convergence rates for broad classes of functions. However, the computational burden these methods impose makes them impractical for many applications which involve iterative algorithms, such as deblurring and tomographic reconstruction. The techniques described in this paper allow multiscale Poisson signal and image reconstruction methods to be implemented with significantly less computational complexity than previously possible. Fast translation-invariant Haar denoising for Poisson data is accomplished by deriving the relationship between maximum penalized likelihood tree pruning decisions and the undecimated wavelet transform coefficients. Fast wedgelet and platelet methods are accomplished with a coarse-to-fine technique which detects possible boundary locations before performing wedgelet or platelet fits.

1. MULTIRESOLUTION ANALYSIS OF POISSON DATA

The weighted sum of two Poisson random variables is not Poissonion. This simple little fact can lead to large complications when traditional multiscale methods, such as wavelet analysis, are used to estimate Poisson signals and images. These complications can be circumvented by using Gaussian approximations to the observed data, but these approximations can be especially inaccurate when the number of events is small. Unfortunately, a number of important applications, including density estimation, network traffic analysis, gamma ray burst (GRB) intensity estimation, astronomical imaging, and medical imaging are all characterized by limited observations. Frequently diagnostics or analyses of physical processes depend on the quality of these estimators, making improved Poisson signal and image estimation methods essential [1].

In previous work we introduced new multiscale analysis methods for Poisson and multinomial data based on piecewise constants in one dimension and constants, wedgelets, or platelets in two dimensions [2, 3]. 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. At the heart of these methods lies penalized likelihood estimation on recursive dyadic partitions, which results in multiscale methods that provide spatial adaptivity similar to wavelet-based techniques [4].

These methods have two primary advantages over traditional wavelet denoising. First, the proposed methods are near minimax optimal reconstruction techniques for Poisson data. Second, these methods can mitigate the effect of the dyadic partition underlying classical multiscale techniques and reduce the associated artifacts. This can be accomplished when using the piecewise constant estimator (in one or two dimensions) by averaging over all possible shifts of the underlying partition. In two dimensions, wedgelets and platelets are not restricted to traditional dyadic partitions because the wedgelet splits form a non-rectangular partition of the image. While traditional wavelet-based methods are difficult to analyze for Poisson data and (in two dimensions) less adept at representing image edges, they are very fast: only $O(N)$ operations are required for wavelet denoising and only $O(N \log N)$ operations are required for translation invariant (TI) wavelet denoising, where N is the number of observations. The proposed methods for Poisson data analysis, in contrast, involve redundant transforms which can require up to $O\left(N^2 \log N\right)$ operations. The added complexity can be particularly devastating in the context of deblurring or tomographic reconstruction. Expectation-Maximization reconstruction methods involve computing wedgelet, platelet or TI constant estimates for Poisson data at each step in an iterative algorithm, making fast estimation techniques essential.

In the following sections, we review multiscale Poisson signal and image estimation methods and their associated theoretical performance bounds, and then detail the implementation techniques used to dramatically reduce the computational complexity of these methods. Finally, we demonstrate the effectiveness of these methods on the Shepp-Logan phantom.

2. MULTISCALE PENALIZED LIKELIHOOD ESTIMATION

The piecewise constant, wedgelet, and platelet methods mentioned above calculate estimates by determining the ideal partition of the domain of observations, Ω (assumed to be $[0,1]$ for signals and $[0,1]^2$ for images) and using maximum likelihood estimation to fit a model to each region in the optimal partition, where the model may be a constant, wedgelet, or platelet fit [2]. (This section describes the piecewise constant estimator underlying the shifting and averaging procedure which generates TI estimates; the extension to translation invariance will be discussed later in the paper. Also, note that while piecewise constant estimation is applicable to both signals and images, wedgelet and platelet estimates are only applicable to image reconstruction.) The space of possible partitions is a nested hierarchy defined through a recursive dyadic partition (RDP) of Ω, and the optimal partition is selected by optimally pruning a dyadic tree representation of the observed data. (The unpruned dyadic tree is referred to as a complete RDP.) This gives our estimators the capability of spatially varying the resolution to automatically increase the smoothing in very regular regions of the data and to preserve detailed structure in less regular regions. Pruning decisions are made using a penalized likelihood criterion.

We assume that we observe N noisy samples of an underlying true function, $f : \Omega \rightarrow [0, \infty)$ (assumed to be smooth or piecewise smooth). The N samples, $\mathbf{x}$, are contaminated by Poisson noise; thus the likelihood of observing $\mathbf{x}$, given the corresponding N samples of f, $\mathbf{f}$, is

$$p(\mathbf{x}|\mathbf{f}) = \prod_{i=0}^{N} \frac{e^{-\mathbf{f}_i} \mathbf{f}_i^{\mathbf{x}_i}}{\mathbf{x}_i!}.$$

In general, the RDP framework leads to a model selection problem that can be solved by a tree pruning process. Each of the termi-

Corresponding author: R. Willett (willett@rice.edu). R. Willett was partially supported by the National Science Foundation Graduate Student Fellowship. R. Nowak (nowak@engr.wisc.edu) was partially supported by the National Science Foundation, grants CCR-0310889 and ANI-0099148, and the Office of Naval Research, grant N00014-00-1-0390.

nal regions in the pruned RDP could correspond to a region of homogeneous or smoothly varying intensity. Such a partition can be obtained by merging neighboring regions of (*i.e.* pruning) a complete RDP to form a data-adaptive RDP $\mathcal{P}$ and fitting models to the intensity on the terminal regions of $\mathcal{P}$. Thus the intensity estimate, $\widehat{\mathbf{f}}$, is completely described by $\mathcal{P}$. Model coefficients for each region are chosen to maximize $p(\mathbf{x}|\widehat{\mathbf{f}})$.

We tackle this problem using a minimum description length/coding theoretic approach to regularization [5]. This provides for a very simple framework for penalized likelihood estimation, wherein the penalization is based on the complexity of the underlying partition. The goal here is to find the partition which minimizes the penalized likelihood function:

$$\begin{aligned} \widehat{\mathcal{P}} &\equiv \arg\min_{\mathcal{P}} \left[-\log p(\mathbf{x} \mid \mathbf{f}(\widehat{\mathcal{P}})) + \text{pen}_m(\widehat{\mathcal{P}}) \right] \\ \widehat{\mathbf{f}} &\equiv \mathbf{f}(\widehat{\mathcal{P}}) \end{aligned} \quad (1)$$

where $p(\mathbf{x} \mid \mathbf{f}(\widehat{\mathcal{P}}))$ denotes the likelihood of observing counts $\mathbf{x}$ given the estimate $\mathbf{f}(\widehat{\mathcal{P}})$ and where $\text{pen}_m(\widehat{\mathcal{P}})$ is the penalty associated with the estimate $\mathbf{f}(\widehat{\mathcal{P}})$ when using model m, where m is $constant$, $wedgelet$, or $platelet$. Given the partition $\widehat{\mathcal{P}}$, $\widehat{\mathbf{f}}$ can be calculated by finding the maximum likelihood model-m fit to the observations over each region in $\widehat{\mathcal{P}}$. The resulting estimator $\widehat{\mathbf{f}}$ is referred to as the penalized likelihood estimator (PLE).

We penalize the estimates according to a codelength required to uniquely describe each model with a prefix code. The codelengths are proportional to the size of the partition associated with each model, and thus penalization leads to estimates that favor smaller partitions. This is appropriate given our assumption of piecewise smooth densities. In particular,

$$\begin{aligned} \text{pen}_{constant,1D} &= |\mathcal{P}| \left((1/2) \log_e n + 2 \log_e 2 \right) \\ \text{pen}_{constant,2D} &= |\mathcal{P}| \left((1/2) \log_e n + (4/3) \log_e 2 \right) \\ \text{pen}_{wedgelet} &= |\mathcal{P}| \left((2/3) \log_e n + (2/3) \log_e 2 \right) \\ \text{pen}_{platelet} &= |\mathcal{P}| \left((4/3) \log_e n + (2/3) \log_e 2 \right) \end{aligned}$$

where n is the total number of events observed and $|\mathcal{P}|$ is the number of terminal regions in the partition $\mathcal{P}$. These codelengths are based on a breadth-first tree traversal argument and the quantization of each coefficient, where the quantization level is chosen to ensure the appropriate rate of convergence, as discussed below. The penalties can also be interpreted as negative log-prior densities on the space of models. The penalties are designed to give good guaranteed performance by balancing between fidelity to the data (likelihood) and the estimates complexity (penalty), which effectively controls the bias-variance trade-off. The next section demonstrates that the proposed form of penalization results in near minimax optimal density estimates.

3. UPPER BOUNDS ON ESTIMATION ERROR

In this section, we establish statistical risk bounds associated with the above procedures in the context of density estimation. Note that density estimation is almost equivalent to Poisson intensity estimation, except that the final estimate must be scaled appropriately to ensure that the probability mass function (pmf) sums to one.

The risk is defined to be proportional to the expected squared Hellinger distance between the true and estimated densities as in [6]; that is,

$$R(\widehat{\mathbf{f}}, \mathbf{f}) \equiv \frac{1}{N} \mathbb{E}_{\mathbf{f}} \left[\int \left[\sqrt{p(\mathbf{x}|\widehat{\mathbf{f}})} - \sqrt{p(\mathbf{x}|\mathbf{f})} \right]^2 \nu(\mathbf{x}) \right] \quad (2)$$

where $\mathbb{E}_{\mathbf{f}}$ denotes expectation with respect to the distribution $\mathbf{f}$ and ν is the dominating measure. The squared Hellinger distance is an appropriate error metric here for several reasons. First, it is a general nonparametric measure appropriate for any density. In addition, the Hellinger distance provides an upper and lower bound on the L_1 error, as detailed in our technical report [3]. Finally, using the squared Hellinger distance allows us to take advantage of a key information-theoretic inequality derived by Li and Barron [6] and extended by Kolaczyk and Nowak [5] to bound the risk associated with multiscale piecewise constant estimators. This result holds for one dimensional data, as seen in Gamma Ray Bursts or density estimation. Using the penalty stated above, we have

$$R(\widehat{\mathbf{f}}, \mathbf{f}) \leq C \left(\frac{\log^2 n}{n} \right)^{\frac{2}{3}} \quad (3)$$

for some constant C when the true density/intensity is in a one dimensional Hölder-1 smoothness class; this upper bound is within a logarithmic factor of the minimax lower bound on the risk for all estimators of functions in the class.

In two dimensions, photon-limited (Poisson) images found in astronomical and medical imaging applications can be analyzed with piecewise constant, wedgelet, or platelet estimators. The above risk bounds were extended to bound the risk associated with fitting higher-order polynomial models to RDPs in [3], which makes it possible to bound the risk associated with platelet analysis. The power of platelets is realized in connection with m-term approximations (*e.g.* approximate images composed of the m best basis functions). Specifically, consider images which are Hölder-2 smooth apart from a Hölder-2 smooth boundary over $[0,1]^2$:

Lemma 1 *Consider the class of images*

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

$\forall (x,y) \in [0,1]^2$ *where* $f_i \in$ *Hölder-2,* $i = 1, 2$*, and* $H \in$ *Hölder-2. Suppose that* $2 \leq m \leq 2^J$*, with* $J > 1$*. The squared* L_2 *error of* m*-term,* J*-scale platelet approximation to images in this class is less than or equal to* $Cm^{-2} + \delta$*, for some constant* C*, where* δ *defines the resolution of the platelet dictionary.*

Using this result (detailed in [2]) combined with the estimation error bounding technique mentioned above and detailed in our technical report [3], it is straightforward to bound platelet risk as follows:

Theorem 1 *Let* $\mathbf{f}$ *be an array consisting of* n *samples of an image,* f*, defined as in Lemma 1. Further assume* $0 < C_\ell \leq f(\cdot,\cdot) \leq C_u$*. Let* $\widehat{\mathbf{f}}$ *be the platelet estimator satisfying (1) using the penalty* $pen = |\mathcal{P}| \left((4/3) \log_e n + (2/3) \log_e 2 \right)$*. Then*

$$R(\widehat{\mathbf{f}}, \mathbf{f}) \leq C \left(\frac{\log^2 n}{n} \right)^{\frac{2}{3}} \quad (4)$$

for some constant C*.*

The penalization structure employed here minimizes the upper bound on the risk. Furthermore, this upper bound is within a logarithmic factor of the lower bound on the minimax risk ($n^{-2/3}$), demonstrating the near-optimality of the proposed method [7]. Furthermore, the proposed estimator can adapt to spatially varying degrees of smoothness in the data, resulting in a partition with non-uniform resolution.

4. COMPUTATIONAL CONSIDERATIONS

The estimation in (1) can be performed optimally but not always efficiently. In previous work, we proposed an algorithm which begins at the leaf nodes in a dyadic tree representation of the initial RDP and traverses upwards, performing a tree-pruning operation at each stage. For each node (*i.e.,* dyadic region) at a particular scale, we determine the maximum likelihood model-m coefficient vector

and then calculate the penalized log likelihoods for (a) the model fit and (b) the optimal fit calculated at the previous, finer scale of the dyadic tree. We then select the model which maximizes the penalized likelihood for each region and proceed to the next coarser level. The partition underlying the PLE is pruned to a coarser scale (lower resolution) in areas with low SNR and where the data suggest that the intensity is fairly smooth. Thus the PLE provides higher resolution and detail in areas of the intensity where there are dominant edges or singularities with higher count levels (higher SNR).

The computational complexity of the above techniques motivates our work in this section. In the case of constant model fits, the optimization can be computed very quickly, but the resulting intensity estimate may contain artifacts caused by the dependence of the estimator on the dyadic partition. This can be alleviated through a process called "averaging over shifts" or "cycle-spinning", which entails circularly shifting the raw data, denoising, and then shifting the estimate back to its original position, and repeating this procedure for each possible shift [8]. It has been shown that cycle-spinning can result in estimators with both fewer artifacts and a higher orders of approximation [9]. However, this process increases the computational complexity from $O\left(N \log N\right)$ to $O\left(N^2 \log N\right)$ operations. Our experiments have revealed that wavelet coefficient thresholding can be more robust to noise when the thresholding rule has a hereditary constraint, *i.e.,* when a coefficient can only be thresholded if all its descendants are also thresholded. We demonstrate here that hereditary TI piecewise constant (Haar) denoising of Poisson data can be performed with the same computational complexity as traditional TI wavelet thresholding techniques [8].

Wedgelet- and platelet- based methods for Poisson image reconstruction are also promising, particularly for recovering boundaries in tomographic reconstruction, but these methods pose similar computational hurdles. In the case of wedgelet or platelet model fits, the algorithms must exhaustively apply all possible wedgelet splits at each of the $O(N)$ leafs in the bottom-up pruning process. For example, in the wedgelet case, the number of models that must be exhaustively examined at each leaf may be as large as $O(N^{2/3})$, resulting in an overall computational complexity of $O(N^{5/3})$. In this paper we describe a coarse-to-fine approach to wedgelet and platelet estimation which significantly reduces the computational burden without reducing accuracy.

In Section 4.1, we discuss the relationship between the above cycle-spinning multiscale constant estimator and the translation invariant Haar wavelet coefficients. Using this relationship, we can compute the cycle-spinning multiscale constant estimator with only $O\left(N \log N\right)$ operations. Next, in Section 4.2, we describe a new, coarse-to-fine wedgelet and platelet estimation technique which reduces the complexity of these estimators to $O\left(N\right)$.

4.1 Translation-Invariant Haar Estimation

Cycle-spinning, as originally proposed, requires $O\left(N \log N\right)$ operations, but was derived in the context of undecimated wavelet coefficient thresholding in the presence of Gaussian noise. The above multiscale tree-pruning method can be modified to produce the same effect by averaging over shifts, but the increase in quality comes at a high computational cost. The method proposed in this paper is based on the key observation that averaging over shifts involves unnecessarily computing the various penalized log likelihoods multiple times. This idea can be used to make TI-Haar Poisson estimation as computationally efficient as TI-Haar wavelet thresholding for Gaussian data.

For example, consider computing a cycle-spinning piecewise constant estimate of the length-4 signal $[x_0, x_1, x_2, x_3]$, as shown in Figure 1. In this figure, internal nodes of the tree are marked as $x_{ij} \equiv x_i + x_j$. Computing the penalized likelihood estimate for Shift 0 involves calculating $p([x_0, x_1]|[\frac{x_{01}}{2}, \frac{x_{01}}{2}])$ and $p([x_0, x_1]|[x_0, x_1])$; note that the first likelihood depends on the scaling coefficients at two different scales, which is why this procedure is less straightforward than traditional coefficient thresholding. Also, these same calculations must be conducted when computing the estimate at Shift 2. In fact, we are left with a total of only $N \log N$ unique pruning decisions. These terms can then be used to perform non-linear estimation on the $N \log N$ TI-Haar wavelet coefficients, and the inverse undecimated wavelet transform can be used to reconstruct the estimated signal in $O\left(N \log N\right)$ operations.

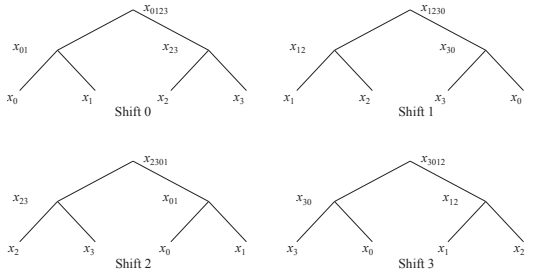

Figure 1: Four different trees used for producing a cycle-spinning estimate of a length-4 signal.

The relationship between the pruning decisions and the wavelet coefficients is delicate and does not correspond to traditional hard or soft thresholding schemes. In the case of multiscale penalized likelihood estimation, wavelet coefficients are scaled depending on their ancestors' pruning decisions. That is, each wavelet coefficient is weighted by the percentage of different shifts in which the corresponding node was not pruned. The exact relationship is detailed in [10]. Unlike traditional wavelet thresholding techniques, this method is near minimax optimal for Poisson noise distributions and is robust to noise due to the hereditary nature of the tree-pruning process.

4.2 Coarse-to-Fine Wedgelet Estimation

Multiscale image estimation techniques such as wedgelets and platelets are promising because of their ability to accurately represent images consisting of smooth regions separated by smooth boundaries and are near minimax optimal for Poisson noise distributions. However, the wedgelet and platelet dictionaries are overcomplete and can be quite computationally demanding to implement. This computational hurdle motivates our work here. We consider a sequential, coarse-to-fine image estimation strategy. The basic idea is to first examine the data on a coarse grid, and then refine the analysis and approximation in regions of interest. By carefully examining the bias and variance trade-offs in each stage, we can show that images can be optimally recovered through a sequential process in linear time.

In the smooth regions of the function f the piecewise constant or linear approximation estimates perform well. The task that drives the rates of convergence of the MSE is the estimation of f in the vicinity of boundaries. The main idea is therefore to perform the global estimation task in a sequential, coarse to fine, fashion: In the first step (denoted *Preview* step), a coarse estimation of the field is performed, using only (translation dependent) piecewise constant approximations, as an attempt to identify the coarse location of any boundaries. In a second step (denoted *Refinement* step) we perform a piecewise linear boundary fit on the areas that were identified as possible boundary regions. If the Preview step is effective then we will perform a search over the dictionary of possible wedgelet splits (which are computationally demanding) only in the regions where they are needed, instead of using them throughout the entire domain.

A detailed description of this technique, along with a complete error analysis, is available in [11].

5. SIMULATIONS

We demonstrate the effectiveness of the proposed methods using the Shepp-Logan phantom brain image, commonly used in medical imaging simulations. The 512×512 grid of noisy measurements

(a) (b) (c)

Figure 3: Shepp-Logan simulation results. (a) Standard Haar estimate, MSE $= 1.23e-2$. (b) Shepp-Logan phantom estimate formed by fitting one wedgelet or constant to each of the unpruned squares from the preview stage; MSE $= 3.72e-3$. (c) Shepp-Logan phantom estimate formed using hereditary TI-Haar denoising; MSE $= 5.87e-3$.

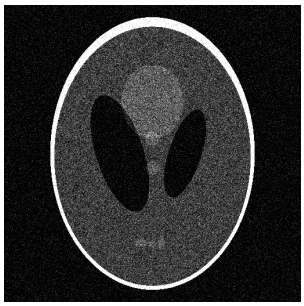

Figure 2: Shepp-Logan phantom. 512×512 noisy measurements with an average of 8 photons per pixel, MSE $= 8.58e-2$.

is displayed in Figure 2; in this example, there is an average of 8 photons counted per pixel. For these simulations, the penalty used for constant fits was $1/3 \log N$ per leaf node and the penalty used for wedgelet fits was $2/3 \log N$ per leaf node. In Figure 3(a) is the standard Haar estimate. The rectangular artifacts mentioned above are clearly visible in this image.

The coarse-to-fine wedgelet estimate after the refinement stage is displayed in Figure 3(b). Note the sharp boundaries in this estimate, particularly in regions of high contrast, as compared to the standard Haar estimate. In this case, the preview stage is initialized with an RDP of 4096 8×8 coarse resolution squares. After the preview stage pruning, the initial coarse resolution RDP has been pruned back to a nonuniform RDP with only 955 8×8 squares remaining. Recall that this procedure requires $O(N)$ operations, as opposed to the $O\left(N^{5/3}\right)$ required for standard wedgelet estimation.

The TI-Haar estimate is displayed in Figure 3(c). While this estimate has a higher mean square error than the coarse-to-fine wedgelet estimate, it should be noted that this estimate exhibits fewer rectangular artifacts, particularly in regions of low SNR. Recall that this procedure requires $O(N \log N)$ operations, the same complexity as standard wavelet TI denoising.

All the experiments describe here can be reproduced using our MATLAB toolbox for Poisson estimation available online at http://www.ece.rice.edu/∼willett/Research/software.html.

6. CONCLUSIONS

The multiscale methods discussed in this paper are near minimax optimal for intensities in certain smoothness classes [2, 3], and with the accelerated implementations presented here, the techniques become practical and effective for a variety of applications with Poisson noise. The denoising techniques detailed here are easily extended to Expectation-Maximization algorithms for deblurring and tomographic reconstruction.

REFERENCES

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

[2] R. Willett and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, 2003.

[3] R. Willett and R. Nowak, "Multiscale density estimation," Tech. Rep. TREE0303, Rice University, 2003, Available at http://www.ece.rice.edu/∼willett/papers/WillettTREE0303.pdf.

[4] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?," *Journal of the Royal Statistical Society*, vol. 57, pp. 301–337, 1995.

[5] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," submitted to Annals of Stat. August, 2001. Available at http://www.ece.wisc.edu/ nowak/pubs.html.

[6] Q. Li and A. Barron, *Advances in Neural Information Processing Systems 12*, chapter Mixture Density Estimation, MIT Press, 2000.

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

[8] 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.

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

[10] R. Willett and R. Nowak, "Fast multiresolution photon-limited image reconstruction," 2004, to appear in the Proceedings of the IEEE International Symposium on Biomedical Imaging.

[11] R. Castro, R. Willett, and R. Nowak, "Coarse-to-fine manifold learning," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP*, 2004.