SPIRAL out of convexity: Sparsity-regularized algorithms for photon-limited imaging

Zachary T. Harmany[a], Roummel F. Marcia[b], and Rebecca M. Willett[a]

[a]Department of Electrical and Computer Engineering, Duke University, Durham NC, 27708 USA
[b]School of Natural Sciences, University of California, Merced, Merced, CA 95343 USA

ABSTRACT

The observations in many applications consist of counts of discrete events, such as photons hitting a detector, which cannot be effectively modeled using an additive bounded or Gaussian noise model, and instead require a Poisson noise model. As a result, accurate reconstruction of a spatially or temporally distributed phenomenon ($f^\star$) from Poisson data (y) cannot be accomplished by minimizing a conventional ℓ_2-ℓ_1 objective function. The problem addressed in this paper is the estimation of $f^\star$ from y in an inverse problem setting, where (a) the number of unknowns may potentially be larger than the number of observations and (b) $f^\star$ admits a sparse representation. The optimization formulation considered in this paper uses a negative Poisson log-likelihood objective function with nonnegativity constraints (since Poisson intensities are naturally nonnegative). This paper describes computational methods for solving the constrained sparse Poisson inverse problem. In particular, the proposed approach incorporates key ideas of using quadratic separable approximations to the objective function at each iteration and computationally efficient partition-based multiscale estimation methods.

Keywords: Poisson intensity, compressed sensing, optimization, recursive dyadic partitions, wavelets

1. INTRODUCTION

In a variety of applications, ranging from nuclear medicine to night vision and from astronomy to traffic analysis, data are collected by counting a series of discrete events, such as photons hitting a detector or vehicles passing a sensor. The measurements are often inherently noisy due to low count levels, and we wish to reconstruct salient features of the underlying phenomenon from these noisy measurements as accurately as possible. The inhomogeneous Poisson process model[1] has been used effectively in many such contexts.

1.1 Background

Under the Poisson assumption, we can write our observation model as

$$y \sim \text{Poisson}(Af^\star), \tag{1}$$

where $f^\star \in \mathbb{R}_+^m$ is the signal or image of interest, $A \in \mathbb{R}_+^{N\times m}$ linearly projects the scene onto a N-dimensional set of observations, and $y \in \mathbb{Z}_+^N$ is a length-N vector of observed Poisson counts. Specifically, under the model in (1), the likelihood of observing a particular vector of counts y is given by

$$p\left(y\,|Af^\star\right) \;=\; \prod_{j=1}^{N} \frac{(Af^\star)_j^{y_j}}{y_j!} e^{-(Af^\star)_j},$$

where $(Af^\star)_j$ is the j^{th} component of $Af^\star$.

The problem addressed in this paper is the estimation of $f^\star$ from y when (a) the number of unknowns, m, may be larger than the number of observations, N, and (b) $f^\star$ is sparse or compressible in some basis W (i.e., $f^\star = W\theta^\star$ and $\theta^\star$ is sparse). This challenging problem has clear connections to "compressed sensing" (CS),[2–5] but arises in a number

Further author information: (Send correspondence to Z. T. Harmany)
Z. T. Harmany: E-mail: zth@duke.edu Telephone: 1-919-619-9123
R. F. Marcia.: E-mail: rmarcia@ucmerced.edu, Telephone: 1-209-228-4874
R. M. Willett: E-mail: willett@duke.edu, Telephone: 1-919-660-5544

of other settings as well, such as tomographic reconstruction in nuclear medicine, superresolution image reconstruction in astronomy, and deblurring in confocal microscopy.

In recent work,[6,7] we explored some of the theoretical challenges associated with CS in a Poisson noise setting, and in particular highlighted two key differences between the conventional CS problem and the Poisson CS problem:

1. unlike many sensing matrices in the CS literature, the matrix A must contain all nonnegative elements, and
2. the intensity $f^\star$, and any estimate of $f^\star$, must be nonnegative.

Not only are these differences significant from a theoretical standpoint, but they pose computational challenges as well. In this paper, we consider algorithms for computing an estimate $\widehat{f}$ of $f^\star$ from observations y according to

$$\begin{aligned} \widehat{f} &\triangleq \arg\min_f \quad -\log\, p(y\,|Af\,) + \mathrm{pen}(f\,) \qquad \text{subject to } f \geq 0 \\ &= \arg\min_f \quad \sum_{i=1}^{N} \{(Af\,)_i - y_i \log\,[(Af\,)_i]\} + \mathrm{pen}(f\,) \qquad \text{subject to } f \geq 0, \end{aligned} \tag{2}$$

where $\mathrm{pen}(f\,)$ is a penalty term which measures the sparsity or noncomplexity of f in some basis or alternative representation. This paper explores computational methods for solving (2). The nonnegativity of f and A results in challenging optimization problems. In particular, the restriction that f is nonnegative introduces a set of inequality constraints into the minimization setup; these constraints are simple to satisfy when f is sparse in the canonical basis,[8] but they introduce significant challenges when enforcing sparsity in an arbitrary basis. We will consider several variants of the penalty term, including $\|f\|_1$, $\|W^T f\|_1$ for some arbitrary orthonormal basis W, and a complexity regularization penalty based upon recursive dyadic partitions.

1.2 Current methods

Recent advances in the area of compressed sensing (CS) have spurred widespread interest in sparse reconstruction. The majority of the CS literature assumes that there exists a "sparsifying" reference basis W, so that $\theta^\star \triangleq W^T f^\star$ is sparse or lies in a weak-ℓ_p space. When the matrix product AW obeys the so-called *restricted isometry property* (RIP)[2] or some related criterion,[9] and when the noise is bounded or Gaussian, then $\theta^\star$ can be accurately estimated from y by solving the following ℓ_2-ℓ_1 optimization problem (or some variant thereof):

$$\widehat{\theta} \;= \arg\min_\theta \|y - AW\theta\|_2^2 + \tau\|\theta\|_1, \tag{3}$$

where $\tau > 0$ is a regularization parameter.[3,9,10] However, the ℓ_2 data-fitting term, $\|y - AW\theta\|_2^2$, is problematic in the presence of Poisson noise. Because under the Poisson model the variance of the noisy observations is proportional to the signal intensity, ℓ_2 data-fitting terms can lead to significant overfitting in high-intensity regions and oversmoothing in low-intensity regions. Furthermore, photon-limited imaging systems impose hard constraints on the nature of the measurements that can be collected, such as nonnegativity, which are not considered in much of the existing CS literature.

Instead of considering the sparsity of $f^\star$ in a particular basis, many successful image reconstruction approaches rely on some sense of smoothness in $f^\star$. In the case of emission tomography, Ahn and Fessler[11] consider a penalized maximum likelihood scheme similar to that in (2) where the penalty corresponds to a measure of the image roughness. These penalties are of the general form

$$\mathrm{pen}(f) = \sum_{i=1}^{m} \sum_{k \in \mathcal{N}_i} w_{ik}\psi(f_i - f_k),$$

where $\mathcal{N}_i$ is a neighborhood about the ith pixel, $w_{ik} > 0$ are weighting factors, and ψ is a continuously differentiable, convex potential function, symmetric about the origin with $\psi(0) = 0$. Commonly employed potential functions include the quadratic potential, $\psi(f_i) = f_i^2$, and the hybrid ℓ_2-ℓ_1 Huber potential

$$\psi(f_i) = \begin{cases} \frac{1}{2} f_i^2 & \text{if } |f_i| \leq \delta \\ \delta|f_i| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases} \tag{4}$$

If we consider a particular neighborhood $\mathcal{N}_i$, we see that the penalty would be small if $f_i \approx f_k$ for all $k \in \mathcal{N}_i$, hence this type of penalty favors smooth candidate estimates. Considerable flexibility is achieved via the selection of different weighting schemes, neighborhoods, and potential functions. The objective in these algorithms are often minimized using a variant of the expectation-maximization (EM) algorithm.[12,13] In some respects, penalty functions such as (4) result in relatively simple optimization problems because physical constraints such as the nonnegativity of f can be imposed on the canonical basis. As we describe below, penalty functions which, for example, measure sparsity in some alternative basis result in a more complex set of constraints and optimality conditions.

Alternative complexity penalizations have been considered in other schemes,[14–20] but many of these approaches are difficult to use in inverse problem settings, yield non-convex optimization problems for which it is difficult to find the global minimum, or are associated with reconstruction algorithms with relatively slow convergence. Closely related to the approach described in this paper is the work of Fessler and Erdoğan,[21] in which they approximate a penalized Poisson log-likelihood with a quadratic function at each iteration. The quadratic approximation used in their work, however, yields a more complicated quadratic optimization subproblem which must be solved numerically, while the approach described in the sequel yields a simple subproblem equivalent to nonnegative denoising. As a result, the number of iterations needed by our algorithm will typically be larger, but the computation at each iteration can be very fast.

1.3 Organization of the paper

The paper is organized as follows. We first describe the negative Poisson log-likelihood in Sec. 2. Next, in Sec. 3, we consider three methods for solving the constrained maximum penalized Poisson likelihood estimation problem with different penalty terms. The first two approaches use an ℓ_1 penalty while the third uses partition-based penalty functions. Algorithmic details are further described in Sec. 4, and experimental results comparing the proposed approaches with expectation-maximization (EM) algorithms are presented in Sec. 5. We conclude with some remarks in Sec. 6.

2. NEGATIVE POISSON LOG-LIKELIHOOD

The negative Poisson log-likelihood is given by

$$F(f) \;=\; \sum_{i=1}^{N} \{(Af)_i - y_i \log\left[(Af)_i\right]\} \;=\; \mathbb{1}^T Af - \sum_{i=1}^{m} y_i \log(e_i^T Af), \tag{5}$$

where $\mathbb{1}$ is the vector of ones in $\mathbb{R}^N$ and e_i is the i^{th} column of the $N \times N$ identity matrix. To avoid the singularity at $f = 0$, it is advantageous to introduce a small parameter $\varepsilon > 0$:

$$F(f) = \mathbb{1}^T Af - \sum_{i=1}^{m} y_i \log(e_i^T Af + \varepsilon). \tag{6}$$

This small parameter also appears in the gradient,

$$\nabla F(f) = A^T \mathbb{1} - \sum_{i=1}^{m} \frac{y_i}{e_i^T Af + \varepsilon} A^T e_i, \tag{7}$$

and the Hessian,

$$\nabla^2 F(f) = A^T \left\{ \sum_{i=1}^{m} \frac{y_i}{(e_i^T Af + \varepsilon)^2} e_i e_i^T \right\} A. \tag{8}$$

We state and prove a result that will be used in the convergence analysis of our proposed method.

THEOREM 2.1. *The gradient of the negative log Poisson likelihood with parameter $\varepsilon > 0$ in (7) is Lipschitz continuous over the feasible set $\{f : f \geq 0\}$, i.e., for all $f, \tilde{f} \geq 0$,*

$$\|\nabla F(f) - \nabla F(\tilde{f})\| \leq L_F \|f - \tilde{f}\|$$

with Lipschitz constant

$$L_F \leq \frac{y_{\max}}{\varepsilon^2} \max(A^T \mathbb{1}) \max(A\mathbb{1}),$$

where $y_{\max} = \max_{i=1,\ldots,N} y_i$.

PROOF. A sufficient condition for $\nabla F(f)$ to be Lipschitz continuous is that its derivative, the Hessian matrix $\nabla^2 F(f)$, is uniformly bounded for all $f \geq 0$; and in particular, a bound on the largest eigenvalue of $\nabla^2 F(f)$ is the same as a bound on the Lipschitz constant of the gradient. Now, the Hessian $\nabla^2 F(f)$ in (8) is positive semidefinite since the term in the brackets is a diagonal matrix with nonnegative entries. Therefore, all that is required is to bound

$$\lambda_{\max} = \sup_{f \geq 0} \sup_{\|z\|_2 \leq 1} z^T \nabla^2 F(f) z.$$

Since A is nonnegative, the supremum over $f \geq 0$ is attained at $f = 0$. Therefore we simply need to bound the largest eigenvalue of

$$\nabla^2 F(0) = \frac{1}{\varepsilon^2} A^T \mathrm{Diag}(y) A,$$

using properties of matrix norms we have

$$\lambda_{\max} \;=\; \frac{1}{\varepsilon^2}\|A^T \mathrm{Diag}(y) A\|_2 \;\leq\; \frac{1}{\varepsilon^2}\|A\|_2^2 \|\mathrm{Diag}(y)\|_2 \;\leq\; \frac{y_{\max}}{\varepsilon^2}\|A\|_1 \|A\|_\infty \;=\; \frac{y_{\max}}{\varepsilon^2} \max(A^T \mathbb{1}) \max(A \mathbb{1}).$$

Note that this quantity can either be analytically determined by exploiting the structure of A, or numerically computed with little effort. This bound on the largest eigenvalue of $\nabla^2 F(f)$ completes the proof. □

3. ALGORITHMS

Our Poisson reconstruction algorithms solves the following constrained optimization problem:

$$\begin{aligned} \underset{f \in \mathbb{R}^m}{\text{minimize}} \quad & \Phi(f) \triangleq F(f) + \tau \mathrm{pen}(f) \\ \text{subject to} \quad & f \geq 0, \end{aligned} \tag{9}$$

where

- $F : \mathbb{R}^m \to \mathbb{R}$ is the negative Poisson log-likelihood in (6)
- $\mathrm{pen} : \mathbb{R}^m \to \mathbb{R}$ is a finite, usually nonsmooth and potentially nonconvex penalty functional.

In previous work,[22] we proposed an approach that is similar to Majorization-Minimization methods:[23] at each iteration, we approximate or majorize $F(f)$ with an appropriate quadratic function, and then minimize this quadratic plus $\mathrm{pen}(f)$ subject to the constraints. The benefit of this approach is that the penalized quadratic problem can often be solved very efficiently and accurately, leading to fast algorithms with good convergence properties. Similar to the SpaRSA framework,[24] we solve (9) via a sequence of subproblems of the form

$$\begin{aligned} f^{k+1} \;=\; \underset{f \in \mathbb{R}^m}{\arg\min} \quad & \phi^k(f) \triangleq \frac{1}{2}\|f - s^k\|_2^2 + \frac{\tau}{\alpha_k}\mathrm{pen}(f) \\ \text{subject to} \quad & f \geq 0, \end{aligned} \tag{10}$$

where

$$s^k \triangleq f^k - \frac{1}{\alpha_k}\nabla F(f^k).$$

This subproblem can be interpreted as a nonnegative denoising method applied to the gradient descent result.

The parameter $\alpha_k > 0$ is chosen via a sequence of two repeated steps. The Barzilai-Borwein method[25,26] is used to choose the initial value of α_k. With $d^k = f^k - f^{k-1}$ and $r^k = \nabla F(f^k) - \nabla F(f^{k-1})$, then

$$\alpha_k = \frac{(d^k)^T r^k}{(d^k)^T d^k}$$

clipped to be within the range $[\alpha_{\min}, \alpha_{\max}]$. This initial choice of α_k is used if the resulting solution of (10) satisfies the acceptance criteria

$$\Phi(f^{k+1}) \leq \max_{i=[k-M]_+,\ldots,k} \Phi(f^i) - \frac{\sigma \alpha_k}{2}\|f^{k+1} - f^k\|_2^2, \tag{11}$$

where $\sigma \in (0,1)$ is a small constant and the operator $[\,\cdot\,]_+ = \max\{0,\cdot\}$. If acceptance criteria (11) is not satisfied, α_k is repeatedly increased by a factor $\eta > 1$ until the solution to (10) satisfies (11). This gentle criteria allows the nonmonotonic objective behavior characteristic of the Barzilai-Borwein methods, yet enforces that the next iterate have a slightly smaller objective than the largest value over the past M iterations. We call our proposed approach Sparse Poisson-Intensity Reconstruction Algorithm (SPIRAL), and for clarity, we describe our general procedure in Algorithm 1.

Algorithm 1 Sparse Poisson-Intensity Reconstruction Algorithm (SPIRAL)

Initialize Choose $\eta > 1$, $0 \leq \alpha_{\min} \leq \alpha_{\max}$, and initial solution f^0. Start iteration counter $k \leftarrow 0$.
repeat
  choose $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$
  repeat
    $f^{k+1} \leftarrow$ solution of (10)
    $\alpha_k \leftarrow \eta\alpha_k$
  until f^{k+1} satisfies acceptance criteria (11)
  $k \leftarrow k+1$
until stopping criterion is satisfied

3.1 Sparsity in canonical basis

Let $\mathrm{pen}(f) = \tau\|f\|_1$, where $\tau > 0$ is a regularization parameter. The minimization problem (9) with this penalty term and that is subject to nonnegativity constraints on f has the following analytic solution:

$$f^{k+1} = \left[s^k - \frac{\tau}{\alpha_k}\mathbb{1}\right]_+$$

where the operation $[\,\cdot\,]_+ = \max\{0,\,\cdot\,\}$ is to be understood component-wise. Thus solving (2) subject to nonnegativity constraints with an ℓ_1 penalty function measuring sparsity in the canonical basis is straightforward to solve. An alternative algorithm for solving this Poisson inverse problem with sparsity in the canonical basis was also explored in the recent literature.[8]

3.2 Sparsity in non-canonical basis

Now suppose that the signal of interest is sparse in some other basis. Then the ℓ_1 penalty term is given by

$$\mathrm{pen}(f) \;\triangleq\; \tau\|W^T f\|_1 = \tau\|\theta\|_1,$$

where $\theta \triangleq Wf$ for some orthonormal basis W, and $\tau > 0$ is some scalar. Then (9) becomes

$$\begin{array}{ll}\underset{\theta\in\mathbb{R}^m}{\text{minimize}} & \Phi(\theta) = F(\theta) + \tau\|\theta\|_1 \\ \text{subject to} & W\theta \geq 0,\end{array} \tag{12}$$

where $F(f) = F(W\theta)$ is now simply rewritten as $F(\theta)$. When the reconstruction $\hat{f} = W\hat{\theta}$ must be nonnegative (i.e., $W\hat{\theta} \geq 0$), the minimization problem

$$\theta^{k+1} \;\triangleq\; \begin{array}{ll}\underset{\theta\in\mathbb{R}^m}{\arg\min} & \phi^k(\theta) \triangleq \dfrac{1}{2}\|\theta - s^k\|_2^2 + \dfrac{\tau}{\alpha_k}\|\theta\|_1, \\ \text{subject to} & W\theta \geq 0\end{array} \tag{13}$$

no longer has an analytic solution necessarily, as in Sec. 3.1. We can solve this minimization problem by solving its dual. First, we reformulate (13) so that its objective function $\phi^k(\theta)$ is differentiable by defining $u, v \geq 0$ such that $\theta = u - v$. The minimization problem (13) becomes

$$(u^{k+1}, v^{k+1}) \;\triangleq\; \begin{array}{ll}\underset{u,v\in\mathbb{R}^m}{\arg\min} & \dfrac{1}{2}\|u - v - s^k\|_2^2 + \dfrac{\tau}{\alpha_k}\mathbb{1}^T(u+v) \\ \text{subject to} & u, v \geq 0,\;\; W(u-v) \geq 0,\end{array} \tag{14}$$

which has twice as many parameters and has additional nonnegativity constraints on the new parameters, but now has a differentiable objective function. The Lagrangian function corresponding to (14) is given by

$$\mathscr{L}(u, v, \lambda_1, \lambda_2, \lambda_3) = \frac{1}{2}\|u - v - s^k\|_2^2 + \frac{\tau}{\alpha_k}\mathbb{1}^T(u + v) - \lambda_1^T u - \lambda_2^T v - \lambda_3^T W(u - v),$$

where $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}^m$ are the Lagrange multipliers corresponding to the constraints in (14). Let $\beta_k = \tau/\alpha_k$. Setting the derivative of $\mathscr{L}$ with respect to u and v to zero, we obtain

$$u - v = s^k + \lambda_1 - \beta_k \mathbb{1} + W^T\lambda_3 \qquad \text{and} \qquad \lambda_2 = 2\beta_k\mathbb{1} - \lambda_1,$$

which leads to the Lagrangian dual function

$$g(\lambda_1, \lambda_3) = -\frac{1}{2}\|s^k + \lambda_1 - \beta_k\mathbb{1} + W^T\lambda_3\|_2^2 + \frac{1}{2}\|s^k\|_2^2.$$

We define $\gamma \triangleq \lambda_1 - \beta_k\mathbb{1}$. For the Lagrange dual problem corresponding to (14), the Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3 \geq 0$. Since $\lambda_2 = 2\beta_k\mathbb{1} - \lambda_1$, then $-\beta_k\mathbb{1} \leq \gamma \leq \beta_k\mathbb{1}$. Also, let $\lambda = \lambda_3$. The Lagrange dual problem associated with this problem is given by

$$\begin{array}{ll} \underset{\lambda,\gamma\in\mathbb{R}^m}{\text{minimize}} & h(\gamma, \lambda) \triangleq \frac{1}{2}\|s^k + \gamma + W^T\lambda\|_2^2 - \frac{1}{2}\|s^k\|_2^2 \\ \text{subject to} & -\frac{\tau}{\alpha_k}\mathbb{1} \leq \gamma \leq \frac{\tau}{\alpha_k}\mathbb{1} \quad \text{and} \quad \lambda \geq 0 \end{array} \tag{15}$$

and at the optimal values $\gamma^\star$ and $\lambda^\star$, the primal iterate θ^{k+1} is given by

$$\theta^{k+1} \triangleq u^{k+1} - v^{k+1} = s^k + \gamma^\star + W^T\lambda^\star.$$

We note that the minimizers of the primal problem (14) and its dual (15) satisfy $\phi^k(\theta^{k+1}) = -h(\gamma^\star, \lambda^\star)$ since (14) satisfies (a weakened) Slater's condition.[27] In addition, the function $-h(\gamma, \lambda)$ is a lower bound on $\phi^k(\theta)$ at any dual feasible point.

The objective function of (15) can be minimized by alternately solving for λ and γ, which is accomplished by taking the partial derivatives of $h(\gamma, \lambda)$ and setting them to zero. Each component is then constrained to satisfy the bounds in (15). At the j^{th} iteration, the variables can, thus, be defined as follows:

$$\begin{aligned} \gamma^{(j)} &= \text{mid}\left\{-\frac{\tau}{\alpha_k}\mathbb{1}, -s^k - W^T\lambda^{(j-1)}, \frac{\tau}{\alpha_k}\mathbb{1}\right\} \\ \lambda^{(j)} &= \left[-W\left(s^k + \gamma^{(j)}\right)\right]_+, \end{aligned}$$

where the operator mid(a, b, c) chooses the middle value of the three arguments component-wise. Note that at the end of each iteration j, the approximate solution to (13), given by

$$\theta^{(j)} \triangleq s^k + \gamma^{(j)} + W^T\lambda^{(j)},$$

is feasible with respect to the constraint $W\theta \geq 0$:

$$W\theta^{(j)} = Ws^k + W\gamma^{(j)} + \lambda^{(j)} = W(s^k + \gamma^{(j)}) + \left[-W(s^k + \gamma^{(j)})\right]_+ = \left[W\left(s^k + \gamma^{(j)}\right)\right]_+ \geq 0.$$

Note also that the complimentarity conditions $(\lambda^{(j)})^T W\theta^{(j)} = 0$ holds for all j since

$$(\lambda^{(j)})^T W\theta^{(j)} = \left[-W\left(s^k + \gamma^{(j)}\right)\right]_+^T \left[W\left(s^k + \gamma^{(j)}\right)\right]_+ = 0. \tag{16}$$

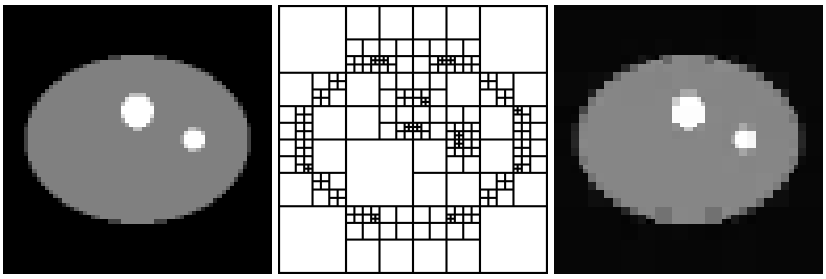
We shall refer to this approach henceforth as SPIRAL-ℓ_1.

3.3 Partition-based methods

We can build on the framework of *recursive dyadic partitions (RDP)*, which we summarize here and are described in detail in.[28] Let $\mathcal{P}$ be the class of all recursive dyadic partitions of $[0,1]^2$ where each cell in the partition has a sidelength at least $1/\sqrt{m}$, and let $P \in \mathcal{P}$ be a candidate partition. The intensity on P, denoted $f(P)$, is calculated using a nonnegative least-squares method to fit a model (such as a constant) to s^k in (13) in each cell in the RDP. As an example, consider Fig. 1. Here we approximate the true emission image (Fig. 1(a)) on the recursive dyadic partition defined in Fig. 1(b)). The result is a piecewise constant approximation to the emission image (Fig. 1(c)). We see that the partition model is able to accurately capture the image in clear multiresolution fashion: large homogeneous regions are well-modeled by large cells, whereas edges are approximated via the deeper recursive partitioning. Furthermore, a penalty can be assigned to the resulting estimator which is proportional to $|P|$, the number of cells in P. Thus we set

$$\begin{aligned} \widehat{P}^{k+1} &= \arg\min_{P\in\mathcal{P}} \frac{1}{2}\|s^k - f(P)\|_2^2 + \frac{\tau}{\alpha_k}|P|, \\ \widehat{f}^{k+1} &= f(\widehat{P}^{k+1}). \end{aligned} \tag{17}$$

A search over $\mathcal{P}$ can be computed quickly using a dynamic program. When using constant partition cell models, the nonnegative least-squares fits can be computed non-iteratively in each cell by simply using the maximum of the average of s^k in that cell and zero. Because of this, enforcing the constraints is trivial and can be accomplished very quickly. The disadvantage of using constant model fits is that it yields piecewise constant estimates. However, a cycle-spun translation-invariant (TI) version of this approach[28] can be implemented with high computational efficiency and be used for solving this nonnegative regularized least-squares subproblem that results in a much smoother estimator. We refer to these approaches as SPIRAL-RDP and SPIRAL-RDP-TI.



(a) (b) (c)

Figure 1. Example of a partition-based estimate: (a) true emission image, (b) partition associated with (c) the partition-based approximation of the true emission image.

It can be shown that partition-based denoising methods such as this are closely related to Haar wavelet denoising with an important hereditary constraint placed on the thresholded coefficients—if a parent coefficient is thresholded, then its children coefficients must also be thresholded.[28] This constraint is akin to wavelet-tree ideas which exploit persistence of significant wavelet coefficients across scales and have recently been shown highly useful in compressed sensing settings.[29]

4. ALGORITHMIC DETAILS

4.1 Termination criteria

In this section, we list criteria by which SPIRAL decides whether the iterates in the subproblem (10) and for the main problem (2) are acceptable approximations to the true minimizers to terminate the algorithm. Here, we only provide criteria for the SPIRAL-ℓ_1 subproblem since the global solution to the partition-based SPIRAL subproblem (17) can be easily and exactly obtained using a non-iterative tree-pruning algorithm,[28] even though its objective function is nonconvex (due to the nonconvex penalty).

4.1.1 SPIRAL-ℓ_1 subproblem

The criterion for termination for the SPIRAL-ℓ_1 subproblem measures the duality gap. Since the objective function $\phi^k(\theta)$ in (13) is convex and all the constraints are affine, then (a weaker) Slater's condition holds[27] and, therefore, the duality gap is zero, i.e.,

$$\phi^k(\theta^{k+1}) = -h(\gamma^\star, \lambda^\star),$$

where $(\gamma^\star, \lambda^\star)$ solves (15). Recall that, at the j^{th} iterate, $\theta^{(j)} = s^k + \gamma^{(j)} + W^T\lambda^{(j)}$ can be viewed as an approximate solution to θ^{k+1}. Thus, the SPIRAL approach for the ℓ_1 subproblem will consider the iterates to be sufficiently close to the optimal solution if

$$\frac{|\phi^k(\theta^{(j)}) + h(\gamma^{(j)}, \lambda^{(j)})|}{|\phi^k(\theta^{(j)})|} \leq \texttt{tol}_{\texttt{SUB}},$$

where $\texttt{tol}_{\texttt{SUB}} > 0$ is some small constant. In our numerical experiments, we often found that it is not necessary to solve this subproblem very accurately, especially at the beginning of the algorithm where the iterate θ^k is still far from the optimal solution.

4.1.2 SPIRAL

Since the global minimizer $\theta^\star$ of (12) is not known *a priori*, criteria to terminate the SPIRAL algorithm must be established to determine whether a computed minimizer $\widehat{\theta}$ is an acceptable solution. We list two such criteria. The first of these criteria is simple: terminate if consecutive iterates or the corresponding function values do not change significantly, i.e.,

$$\frac{\|\theta^{k+1} - \theta^k\|_2}{\|\theta^k\|_2} \leq \texttt{tolP} \qquad \text{or} \qquad \frac{|\Phi(\theta^{k+1}) - \Phi(\theta^k)|}{|\Phi(\theta^k)|} \leq \texttt{tolP},$$

where tolP is a small positive constant. The advantage of these criteria is that they apply to general penalty functions. The disadvantage, however, is that it is possible that the change between two consecutive iterates may be small or that they result in only small improvements in the objective function even though iterates are still far from the true solution. However, we have yet to observe this premature termination in practice.

The next criterion applies only to SPIRAL-ℓ_1, where the penalty is convex and, after a change of variables, differentiable. This criterion is based on the Karush-Kuhn-Tucker (KKT) conditions for optimality: at the k^{th} iteration, given θ^k and the corresponding Lagrange multipliers λ^k computed from (15), determine whether

$$\|\nabla\Phi(\theta^k) - W^T\lambda^k\|_2 \leq \texttt{tolP}, \tag{18}$$

The left hand side corresponds to the gradient of the Lagrangian function, which criterion (18) forces to be sufficiently close to zero. A complimentarity condition could also be required, but by construction, it is always satisfied, i.e., $(\lambda^k)^T W\theta^k = 0$ (see (16)).

4.2 Convergence proof

In this section we consider the minimization problem

$$\underset{\theta\in\mathbb{R}^m}{\text{minimize}} \;\; \Phi(\theta) \triangleq F(\theta) + \tau\text{pen}(\theta) \qquad \text{subject to } \; B\theta \geq 0, \tag{19}$$

where F is the negative Poisson log-likelihood defined in (5) in our context, possibly after a coordinate transformation as in (12). Note that $\Phi(\theta)$ is convex and has a Lipschitz continuous gradient over the feasible set so that

$$\|\nabla\Phi(\theta) - \nabla\Phi(\theta')\| \leq L_\Phi\|\theta - \theta'\|;$$

this follows from Theorem 2.1 when the penalty function is strictly convex with a Lipschitz continuous gradient. Let θ^0 denote the initial iterate, and let

$$\begin{aligned} \Theta &\triangleq \{\theta : B\theta \geq 0\} \\ \mathcal{S} &\triangleq \{\theta \in \Theta : \Phi(\theta) \leq \Phi(\theta^0)\}. \end{aligned}$$

We note that $\mathcal{S}$ is a convex closed set and assume it is bounded. At iteration k, let $\phi^k(\theta)$ be a function satisfying the following conditions:

CONDITION 1: $\Phi(\theta) - \Phi(\theta^k) \leq \phi^k(\theta) - \phi^k(\theta^k)$ for all $\theta \in \mathcal{S}$;

CONDITION 2: $\phi^k(\theta)$ is strictly convex, twice differentiable, and continuous with a Lipschitz continuous gradient over $\mathcal{S}$ so that

$$\|\nabla\phi^k(\theta) - \nabla\phi^k(\theta')\| \leq L_\phi\|\theta - \theta'\|,$$

where L_ϕ does not depend on k;

CONDITION 3: the gradient of ϕ^k matches the gradient of Φ at θ^k: $\nabla\Phi(\theta^k) = \nabla\phi^k(\theta^k)$;

CONDITION 4: θ^{k+1} solves the subproblem

$$\underset{\theta\in\mathbb{R}^m}{\text{minimize}}\ \phi^k(\theta) \text{subject to } B\theta \geq 0,$$

so that there exists a Lagrange multiplier λ^{k+1} such that the subproblem KKT conditions are satisfied:

$$\begin{aligned} \nabla\phi^k(\theta^{k+1}) - B^T\lambda^{k+1} &= 0 \\ \lambda^{k+1} &\geq 0 \\ B\theta^{k+1} &\geq 0 \\ (\lambda^{k+1})^T B\theta^{k+1} &= 0; \end{aligned}$$

CONDITION 5: and the minimum eigenvalue of the Hessian $\nabla^2\phi^k(\theta)$ exceeds some $C > 0$.

Note that these conditions are satisfied by the approach described in Section 3. In particular, for α_k chosen appropriately, ϕ^k as in (13) (or defined similarly via θ is an alternative basis) satisfies Conditions 1- 5. From these conditions, we have the following:

LEMMA 4.1.[30] *The iterates $\{\theta^k\}$ yield monotonic decreases in $\Phi(\theta^k)$, and hence $\theta^k \in \mathcal{S}$.*

LEMMA 4.2.[30] *Given $\tilde{\epsilon} > 0$, there exists some constant $K_{\tilde{\epsilon}} > 0$ such that for all $k > K_{\tilde{\epsilon}}$, $\|\theta^{k+1} - \theta^k\| < \tilde{\epsilon}$.*

Using these lemmas, we can prove the following weak convergence result.

THEOREM 4.3. *Given $\epsilon > 0$, there exists some $K_\epsilon > 0$ such that for all $k > K_\epsilon$, the primal-dual pair $(\theta^{k+1}, \lambda^{k+1})$ nearly satisfies the KKT conditions associated with (19); that is, for all k sufficiently large, we have*

$$\begin{aligned} \|\nabla\Phi(\theta^{k+1}) - B^T\lambda^{k+1}\| &< \epsilon && (20)\\ \lambda^{k+1} &\geq 0 && (21)\\ B\theta^{k+1} &\geq 0 && (22)\\ (\lambda^{k+1})^T B\theta^{k+1} &= 0. && (23) \end{aligned}$$

PROOF: The final three KKT conditions (i.e. (21), (22), and (23)) are satisfied as a result of Condition 4 above. To see (20), note

$$\begin{aligned} \|\nabla\Phi(\theta^{k+1}) - B^T\lambda^{k+1}\| &= \|\nabla\Phi(\theta^{k+1}) - \nabla\phi^k(\theta^{k+1})\| && (24)\\ &\leq \underbrace{\|\nabla\Phi(\theta^{k+1}) - \nabla\Phi(\theta^k)\|}_{T_1} + \underbrace{\|\nabla\Phi(\theta^k) - \nabla\phi^k(\theta^k)\|}_{T_2} + \underbrace{\|\nabla\phi^k(\theta^k) - \nabla\phi^k(\theta^{k+1})\|}_{T_3}, && (25) \end{aligned}$$

where (24) follows from the first KKT condition of the subproblem described in Condition 4 and (25) follows from the triangle inequality. Given $\epsilon > 0$, showing that there exists K_ϵ such that for all $k > K_\epsilon$, $T_1 + T_2 + T_3 < \epsilon$, will complete the proof of the theorem. First, note that $T_2 \equiv 0$ by Condition 3. Then, for T_1, by Lemma 4.2, given $\tilde{\epsilon}_\Phi \triangleq \epsilon/(2L_\Phi)$, there exists $K_{\tilde{\epsilon}_\Phi}$ such that for all $k > K_{\tilde{\epsilon}_\Phi}$,

$$T_1 = \|\nabla\Phi(\theta^{k+1}) - \nabla\Phi(\theta^k)\| \leq L_\Phi\|\theta^{k+1} - \theta^k\| < L_\Phi\tilde{\epsilon}_\Phi = \frac{\epsilon}{2},$$

using the Lipschitz continuity of $\nabla\Phi(\theta)$. Similarly, for T_3, by Lemma 4.2, given $\tilde{\epsilon}_\phi \triangleq \epsilon/(2L_\phi)$, there exists $K_{\tilde{\epsilon}_\phi}$ such that for all $k > K_{\tilde{\epsilon}_\phi}$,

$$T_3 = \|\nabla\phi^k(\theta^k) - \nabla\phi^k(\theta^{k+1})\| \leq L_\phi\|\theta^{k+1} - \theta^k\| < L_\phi\tilde{\epsilon}_\phi = \frac{\epsilon}{2},$$

using Condition 2. Thus, these inequalities show that by letting $K_\epsilon = \max\left\{K_{\tilde{\epsilon}_\Phi}, K_{\tilde{\epsilon}_\phi}\right\}$,

$$\|\nabla\Phi(\theta^{k+1}) - B^T\lambda^{k+1}\| \leq T_1 + T_3 < \epsilon$$

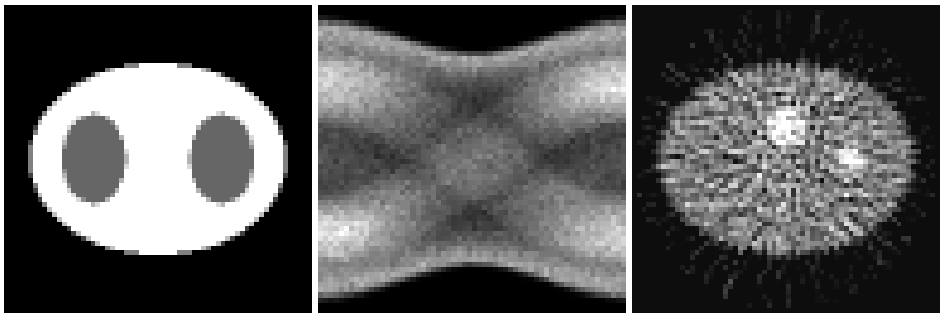
for all $k > K_\epsilon$. □

This theorem says that for any arbitrarily small tolerance level ϵ, we can satisfy the KKT conditions to within that tolerance if the number of iterations, k, is sufficiently large.

5. NUMERICAL EXPERIMENTS

Although the algorithms described heretofore are applicable to a wide range of imaging contexts, we demonstrate the effectiveness of the proposed estimation algorithms on a simulated emission tomography dataset. We compare our algorithm with the currently available state-of-the-art emission tomography reconstruction algorithms.[11]

In this experimental setup, we wish to reconstructe the true axial emission map $f^\star$ (Fig. 1a) as accurately as possible. The photon flux desribed by this true emission map is subject to attenuation effects caused by the various densities of tissue through which the photons must travel to reach the detector array. The simulated attenuation map, μ, which is assumed known during the reconstruction process, is shown in Fig. 2(a). (This and the true axial emission map are standard test images included in Prof. Fessler's Image Reconstruction Toolbox.[31]) Denoting G as the tomographic projection operation, the sensing matrix in the emission tomography setup is given by $A \equiv \mathrm{diag}[\exp(-G\mu)]G$, with the noisy tomographic data simulated according to the inhomogeneous Poisson process (1). The noisy sinogram observations are shown in Fig. 2(b); in this image, we have a total photon count of 4.99×10^5, a mean count over the support of the tomographic projections of 166.35, and a maximum count of 365.



(a) (b) (c)

Figure 2. Simulation Setup: (a) attenuation map, (b) noisy sinogram observations, (c) filtered backprojection solution used as an initialization for all the algorithms considered.

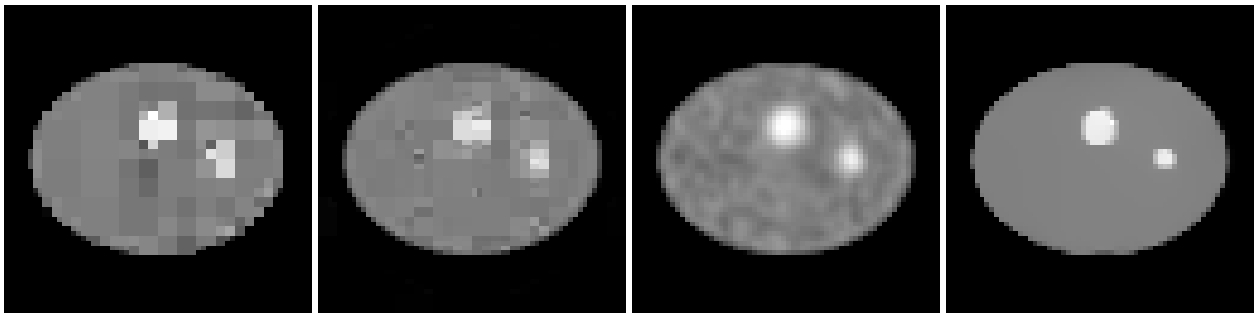
The methods we evaluate include the proposed SPIRAL-ℓ_1 algorithm, using the Daubechies-2 (Haar) wavelet basis for W, and the proposed SPIRAL-RDP and SPIRAL-RDP-TI partition-based algorithms. We compare our proposed approaches with a competing approach, suggested by Prof. Fessler as representative of the current state-of-the-art in emission tomographic reconstruction. This method, denoted E-ML-EM-3, employs an incremental penalized Poisson likelihood EM algorithm utilizing the Huber potential function (see (4)). This method is available using Prof. Fessler's Image Reconstruction Toolbox;[31] specifically, we used the `epl_inc` function from the toolbox. For each method, we independently search over an appropriate range of regularization values to minimize the RMS error with respect to the true emission map $f^\star$. Each algorithm was initialized with the filtered backprojection solution (Fig. 2(c)), thresholded to be nonnegative (i.e., feasible).

From the results presented in Table 1, we see that out of the proposed SPIRAL approaches, the method based on the TI partitions achieves the best RMS error. This is to be expected as the ℓ_1-based approach does not consider any structure

in the estimate beyond that of a sparse representation in the basis W, and the non-cycle-spun partitions are a highly biased estimate due to considering only a single shift of the RDP structure. In addition, from Table 1, we see that the proposed SPIRAL-RDP-TI approach achieves a lower RMS error than the competing E-ML-EM-3 approach. Examining the reconstructed images, we see that the SPIRAL-RDP estimate in Fig. 3(a) exhibits blocking artifacts due to the RDP structure not fortuitously aligning to any strong edges in the image, while the SPIRAL-ℓ_1 estimate in Fig. 3(b) has noticable wavelet artifacts. The competing E-ML-EM-3 method clearly is able to accurately estimate the locations of the brighter portions of the image in Fig. 3(c), however, it performs less well in the regions of homogeneous intensity. For instance, there are many darker spots that could be misinterpreted as regions of low uptake. The SPIRAL-RDP-TI reconstruction in Fig. 3(d) is highly accurate in the homogeneous regions in the image, although it does smooth the strong edges around the white regions. Visually, the proposed SPIRAL-RDP-TI approach seems to outperform the previous three.

Reconstruction Method	RMS Error (%)
SPIRAL-RDP	14.713
SPIRAL-ℓ_1 (DB-2)	11.103
E-ML-EM-3	9.996
SPIRAL-RDP-TI	8.875

Table 1. Reconstruction accuracy for the minimal RMS error solutions. Note: RMS Error (%) $= 100 \cdot \|\widehat{f} - f^\star\|_2 / \|f^\star\|_2$.



(a) SPIRAL-RDP (b) SPIRAL-ℓ_1 (DB-2) (c) E-ML-EM-3 (d) SPIRAL-RDP-TI

Figure 3. Reconstruction emission images of the considered methods. The regularization parameter for each of the methods was chosen to yield the minimal RMS error.

6. CONCLUSION

We have formulated the general goal of reconstructing an image from photon-limited measurements as a penalized maximum Poisson likelihood estimation problem. To obtain a solution to this problem, we have proposed an algorithm that allows for a flexible choice of penalization methods, and focused particularly on sparsity-promoting penalties. In particular, we detail the case where the penalty corresponds to the sparsity-promoting ℓ_1 norm of the expansion coefficients in a sparsifying basis, and also the case where the penalty is related to the complexity of a RDP-based estimate. We establish mild conditions for which this algorithm has desirable convergence properties, although in practice it is beneficial to relax these conditions to attain faster (albeit nonmonotonic) convergence. We demonstrate the effectiveness of our methods through a simulated emission tomography example. When our cycle-spun TI partition method is applied to this problem set, the resulting estimates outperform the current state-of-the-art approaches developed specifically for emission tomography.

ACKNOWLEDGMENTS

This work was supported by NSF CAREER Award No. CCF-06-43947, DARPA Grant No. HR0011-07-1-003, and NSF Grant DMS-08-11062.

REFERENCES

[1] D. Snyder, *Random Point Processes*, Wiley-Interscience, New York, NY, 1975.

[2] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory* **15**, pp. 4203–4215, December 2005.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory* **52**, pp. 1289–1306, April 2006.
[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory* **52**(2), pp. 489 – 509, 2006.
[5] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?," *IEEE Trans. Inform. Theory* **52**, pp. 5406–5425, December 2006.
[6] R. Willett and M. Raginsky, "Performance bounds for compressed sensing with Poisson noise," in *Proc. of IEEE Int. Symp. on Inf. Theory*, 2009.
[7] M. Raginsky, Z. T. Harmany, R. F. Marcia, and R. Willett, "Performance bounds for compressed sensing with Poisson noise," *IEEE Trans. on Signal Proc.* , 2009. Submitted.
[8] D. Lingenfelter, J. Fessler, and Z. He, "Sparsity regularization for image reconstruction with Poisson data," in *Proc. SPIE Computational Imaging VII*, **7246**, 2009.
[9] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inform. Theory* **52**, pp. 1030–1051, March 2006.
[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B* **58**(1), pp. 267–288, 1996.
[11] S. Ahn and J. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Trans. Med. Imaging* **22**, pp. 613–626, May 2003.
[12] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Royal statistical Society B* **39**, pp. 1–38, 1977.
[13] Y. Vardi, L. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *Journal of the American Statistical Association* **80, No.389**, pp. 8–20, 1985.
[14] J.-L. Starck and J. Bobin, "Astronomical data analysis and sparsity: from wavelets to compressed sensing," *Proc. IEEE: Special Issue on Applications of Sparse Representation and Compressive Sensing* , 2010. In press.
[15] E. Kolaczyk, "Wavelet shrinkage estimation of certain poisson intensity signals using corrected thresholds," *Statistica Sinica* **9**, pp. 119–135, 1999.
[16] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Ann. Statist.* **32**, pp. 500–527, 2004.
[17] R. Nowak and E. Kolaczyk, "A multiscale statistical framework for Poisson inverse problems," *IEEE Trans. Info. Theory* **46**, pp. 1811–1825, 2000.
[18] K. Timmermann and R. Nowak, "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging," *IEEE Transactions on Information Theory* **45**(3), pp. 846–862, 1999.
[19] R. Willett and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Transactions on Medical Imaging* **22**(3), pp. 332–350, 2003.
[20] J. Liu and P. Moulin, "Complexity-regularized denoising of Poisson-corrupted data," in *Proc. Intl. Conf. on Image Processing*, **3**, pp. 254–257, 2000.
[21] J. A. Fessler and H. Erdoğan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1998.
[22] Z. Harmany, R. Marcia, and R. Willett, "Sparse Poisson intensity reconstruction algorithms," in *IEEE Workshop on Statistical Signal Processing*, 2009. Available at http://arxiv.org/abs/0905.0483.
[23] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician* **58**, pp. 30–37, 2004.
[24] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.* , 2009 (to appear).
[25] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J Numer Anal* **8**(1), pp. 141–148, 1988.
[26] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Trans. Sel. Top. Signal Process.* **1**(4), pp. 586–597, 2007.
[27] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
[28] R. Willett and R. Nowak, "Multiscale Poisson intensity and density estimation," *IEEE Trans. Inform. Theory* **53**(9), pp. 3171–3187, 2007.
[29] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. on Inform. Theory* , 2008. Submitted.
[30] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Im. Proc.* **4**(10), pp. 1417–1429, 1995.
[31] J. A. Fessler *et al.*, "Image reconstruction toolbox." `http://www.eecs.umich.edu/~fessler/code`.