

SPARSITY-REGULARIZED PHOTON-LIMITED IMAGING

Zachary T. Harmany¹, Roummel F. Marcia², and Rebecca M. Willett¹

¹Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA

²School of Natural Sciences, University of California, Merced, Merced, CA 95348 USA

ABSTRACT

In many medical imaging applications (*e.g.*, SPECT, PET), the data are a count of the number of photons incident on a detector array. When the number of photons is small, the measurement process is best modeled with a Poisson distribution. The problem addressed in this paper is the estimation of an underlying intensity from photon-limited projections where the intensity admits a sparse or low-complexity representation. This approach is based on recent inroads in sparse reconstruction methods inspired by compressed sensing. However, unlike most recent advances in this area, the optimization formulation we explore uses a penalized negative Poisson log-likelihood objective function with nonnegativity constraints (since Poisson intensities are naturally nonnegative). This paper describes computational methods for solving the nonnegatively constrained sparse Poisson inverse problem. In particular, the proposed approach incorporates sequential separable quadratic approximations to the log-likelihood and computationally efficient partition-based multi-scale estimation methods.

Index Terms— Photon-limited imaging, Poisson noise, sparse approximation, wavelets, tomography

1. INTRODUCTION

Photon-limited imaging arises in several important imaging modalities, including Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), and Confocal Microscopy [1]. Frequently, diagnostics depend on the quality of these images. In these settings, the statistics governing the data follow a Poisson process. If the photon counts are sufficiently high, the data can be modeled accurately using a Gaussian distribution. However, a Poisson statistical model should be employed when the number of photons collected by the detector elements is very small relative to the number of pixels, voxels, or other quantities to be estimated. Doing so will avoid over-fitting in high intensity regions and over-smoothing in low intensity regions, demonstrated by the numerical results herein. While the Poisson distribution accurately reflects the physics of these systems, rigorously accounting for Poisson statistics within a reconstruction algorithm is a challenging task.

Recent advances in image reconstruction theory and algorithms have been based on the assumption that the underlying image is sparse or nearly sparse within some basis or other representation. In particular, compressed sensing theory [2, 3, 4] has resulted in compelling arguments in favor of incorporating prior information about image sparsity into the reconstruction processes, leading to a plethora of new sparsity-promoting reconstruction algorithms (see <http://dsp.rice.edu/cs> for a list of currently available CS methods).

However, these and other algorithms typically measure the reconstruction's fit to the data using a squared error term which corresponds to a negative Gaussian log likelihood. While effective in some settings, in the Poisson noise context it can result in significant over- and under-fitting artifacts. In this paper we present a new approach to sparsity-promoting image reconstruction from Poisson observations in an inverse problem setting.

2. PROBLEM FORMULATION

A variety of photon-limited imaging modalities are well-modeled by an inhomogeneous Poisson process model. In this model, the true scene intensity f^* is passed through an intensity-preserving passive imaging system described by the matrix A , yielding a detector photon intensity of Af^* . We then make Poisson observations, y , of the detector photon intensity, Af^* ; that is

$$y \sim \text{Poisson}(Af^*). \quad (1)$$

where $f^* \in \mathbb{R}^m$ is the signal or image of interest, $A \in \mathbb{R}^{n \times m}$ linearly projects the scene onto an n -dimensional space of observations, and $y \in \{0, 1, 2, \dots\}^n$ 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

$$p(y|Af^*) = \prod_{j=1}^n \frac{(Af^*)_j^{y_j}}{y_j!} e^{-(Af^*)_j},$$

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

In this paper, we propose estimating f^* from y using a penalized Poisson log-likelihood objective function in the context of inverse problems in medical imaging. Our class of algorithms reconstructs an estimate \hat{f} of the true intensity f^* according to the following objective:

$$\hat{f} = \underset{f}{\text{arg min}} \quad \phi(f) + \tau \text{pen}(f) \quad (2)$$

subject to $f \succeq 0$,

where

$$\phi(f) \triangleq \sum_{j=1}^n [(Af)_j - y_j \log(Af)_j],$$

$\text{pen}(\cdot)$ is a sparsity-based penalty function that will be detailed later, and the standard notation $f \succeq 0$ means that the components of f are nonnegative. The constraints reflect the nonnegativity of both the observed intensity and the underlying image. We keep the penalty term $\text{pen}(\cdot)$ general since our class of reconstruction algorithms may utilize one of many modern complexity-penalization schemes.

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.

2.1. Related work

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 $\theta^* \triangleq W^T f^*$ 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 [4], and when the noise is bounded or Gaussian, then θ^* can be accurately estimated from y by solving the following ℓ_2 - ℓ_1 optimization problem (or some variant thereof):

$$\hat{\theta} = \arg \min_{\theta} \|y - AW\theta\|_2^2 + \tau \|\theta\|_1, \quad (3)$$

where $\tau > 0$ is a regularization parameter [3, 4, 5]. This method is highly accurate for sparse images, however the performance degrades as the sparsity decreases [6, 7].

Although this method works well in the Gaussian noise case, 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 over-fitting in high-intensity regions and over-smoothing in low-intensity regions (as evidenced in the numerical simulations below). Furthermore, photon-limited imaging systems impose hard constraints on the nature of the measurements that can be collected, such as non-negativity, which are not considered in much of the existing CS literature.

Instead of considering the sparsity of f^* in a particular basis, many successful image reconstruction approaches rely on some sense of smoothness in f^* . In the case of emission tomography, the approaches in [8] 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

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

where \mathcal{N}_i is a neighborhood about the i th 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, $\psi(x) = x^2$, and the hybrid ℓ_2 - ℓ_1 Huber potential

$$\psi(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq \delta \\ \delta|x| - \frac{1}{2}\delta^2 & \text{otherwise,} \end{cases}$$

where δ is the Huber threshold. 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.

The penalties or regularization terms considered in this paper can be considered negative log prior probabilities on candidate reconstructions, and hence our approach can also be formulated as a Bayesian maximum a posteriori (MAP) estimate. While a wide variety of other Bayesian priors have been proposed for image reconstruction tasks, the multiscale penalties used in this paper effectively capture prior image information and have important theoretical performance guarantees [9].

3. RECONSTRUCTION ALGORITHM

In previous work [10], we described an optimization method called *SPIRAL* (*Sparse Poisson Intensity Reconstruction Algorithms*) for solving (2). Here we describe the key ideas and features of this optimization method.

Sequential quadratic problems. The SPIRAL approach sequentially approximates the objective function in (2) by separable quadratic problems that are easier to minimize. In particular, at the k^{th} iteration we use the second-order Taylor expansion of ϕ around f^k and approximate the Hessian by a positive scalar (η_k) multiple of the identity matrix, resulting in the following minimization problem:

$$f^{k+1} = \arg \min_f \left\| s^k - f \right\|_2^2 + \frac{2\tau}{\eta_k} \text{pen}(f) \quad (4)$$

subject to $f \succeq 0$,

where $s^k = f^k - \frac{1}{\eta_k} \nabla \phi(f^k)$. The optimization problem (4) can be viewed as a denoising subproblem applied to the gradient descent, and as such, it gives us considerable flexibility in designing effective penalty functions.

ℓ_1 -penalty. Much of existing CS optimization literature focuses on ℓ_1 -penalty functions, *i.e.*, $\text{pen}(f) \propto \|W^T f\|_1$, where W is some orthonormal basis, such as a wavelet basis, and $\theta \triangleq W^T f$ are the expansion coefficients in that basis. However, because the ℓ_1 norm is not differentiable, we cannot simply solve (4) with an ℓ_1 -penalty using gradient-based methods. Instead, we introduce additional variables and constraints so that the derivative of the objective function always exists. The resulting optimization problem is more complicated, but the solution can be obtained by solving the corresponding Lagrangian dual problem, which, in this case, is much easier to solve (see [10] for details). We refer to this ℓ_1 -based reconstruction algorithm as SPIRAL- ℓ_1 .

Model-based penalty. We can build on the framework of *recursive dyadic partitions* (RDP), which are described in detail in [9]. 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 constant fit to s^k in (4) 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)). Furthermore,

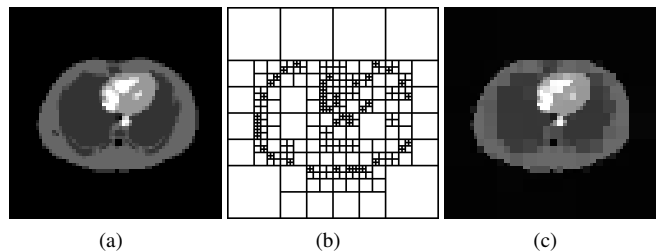


Fig. 1. Example partition-based estimate: (a) true emission image, (b) partition associated with (c) the partition-based approximation of the true emission image.

a penalty can be assigned to the resulting estimator which is proportional to $|P|$, the number of cells in P . Thus we set

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

A search over \mathcal{P} can be computed quickly using a dynamic program. 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 can be implemented with high computational efficiency [9], resulting in a much smoother estimator. We refer to our RDP-based methods as SPIRAL-RDP, and SPIRAL-RDP-TI when the translation-invariant method is employed.

It can be shown that partition-based methods 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 [9]. This constraint is akin to wavelet-tree ideas which exploit persistence of significant wavelet coefficients across scales and have been shown highly useful in compressed sensing settings [11].

4. NUMERICAL RESULTS

Although our algorithms 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 [8].

In this experimental setup, we wish to reconstruct the true axial emission map f^* (Fig. 3(a)) as accurately as possible. The photon flux described 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 (IRT) [12].) The tomographic projection operation G , corresponding to a parallel strip-integral geometry with 66 radial samples and 102 angular samples uniformly spaced over 180 degrees, was also generated by the IRT software [12]. The resulting sensing matrix in the emission tomography setup is given by $A \equiv \text{diag}[\exp(-G\mu)]G$, with the noisy tomographic data y 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 199647, a mean count over the support of the tomographic projections of 45.4, and a maximum count of 95.

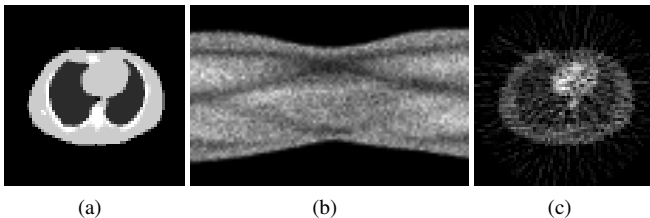


Fig. 2. Simulation Setup: (a) attenuation map, (b) noisy sinogram observations, and (c) filtered backprojection solution used as an initialization for all the algorithms considered.

The methods we evaluate include the proposed SPIRAL- ℓ_1 algorithm, using the Daubechies-8 wavelet basis for W , and the proposed SPIRAL-RDP and SPIRAL-RDP-TI partition-based algorithms. We compare our proposed approaches with two competing methods. The first, denoted SPS-OS, uses a separable paraboloidal surrogate with ordered subsets algorithm utilizing the Huber potential function [8]. The second, denoted E-ML-EM-3, employs an incremental penalized Poisson likelihood EM algorithm utilizing the Huber potential function. This method is available using Prof. Fessler’s Image Reconstruction Toolbox [12]; specifically, we used the `eplinc` function from the toolbox. This was suggested by Prof. Fessler as representative of the current state-of-the-art in emission tomographic reconstruction.

In addition to these two Poisson reconstruction methods, we also compare to the GPSR algorithm [13]. This algorithm solves the ℓ_2 - ℓ_1 problem (3), using the Daubechies-8 wavelet basis for W . As the solution to (3) is not guaranteed to be nonnegative, we threshold the result to obtain a feasible – and therefore more accurate – solution. Including this result allows us to demonstrate the effectiveness of solving the formulation (2) that utilizes the Poisson likelihood.

For each method, we independently search over an appropriate range of regularization values to minimize the RMS error $\|\hat{f} - f^*\|_2 / \|f^*\|_2$ with respect to the true emission map f^* . Tuning the regularization in this manner is convenient in a simulation study. However, in the absence of truth, one typically resorts to a cross-validation procedure to determine an appropriate level of regularization. Each algorithm was initialized with the filtered backprojection solution (Fig. 2(c)), thresholded to be nonnegative. Table 1 shows the results of the reconstruction accuracy averaged over 10 trials of the experiment (different Poisson realizations of the data y).

From the results presented in Table 1 and Fig. 3, we see that out of the proposed SPIRAL approaches, the method based on the TI partitions achieves the lowest 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 recursive dyadic partition structure for which there are no fortuitous alignment of the breakpoints of the partition with strong edges in the image. From Table 1, we also see that the proposed SPIRAL-RDP-TI approach outperforms the competing SPS-OS and E-ML-EM-3 approaches in terms of the average RMS error.

Examining the reconstructed images in Fig. 3, we see that the SPIRAL- ℓ_1 estimate in (b) has significant wavelet artifacts. As expected, the SPIRAL-RDP estimate in (c) exhibits blocking artifacts due to the recursive dyadic partition structure not fortuitously aligning to any strong edges in the image. The SPIRAL-RDP-TI reconstruction in (d) is highly accurate in the homogeneous regions in the

Reconstruction Method	Mean RMSE (%)
GPSR ℓ_2 - ℓ_1 (DB-8)	24.070
SPIRAL-RDP	21.707
SPIRAL- ℓ_1 (DB-8)	20.127
E-ML-EM-3	18.818
SPS-OS	18.493
SPIRAL-RDP-TI	17.769

Table 1. Mean reconstruction accuracy statistics over 10 trials of the experiment (different realizations of the data y). Note: RMSE (%) = $100 \cdot \|\hat{f} - f^*\|_2 / \|f^*\|_2$.

image, although the minimal RMS solution over-smoothes some of the small features such as the smaller bright and dark spots in the true image (circled in red and yellow in Fig. 3(a)). However, this can be avoided by de-emphasizing the regularizer to yield a less-smooth estimate in (e), which recovers this isolated bright spot. Our proposed approaches are comparable to the competing SPS-OS method in (f) and E-ML-EM-3 method in (g). Both of these methods are able to accurately estimate the locations of the brighter points in the image, but perform less well in some of the regions of homogeneous intensity or near low-contrast boundaries. For instance, the E-ML-EM-3 estimate shows many dark spots that could be misinterpreted as regions of low uptake. The GPSR solution in (h) has higher RMS error and is clearly visually inferior to all the reconstruction that utilize a Poisson log-likelihood term in the objective.

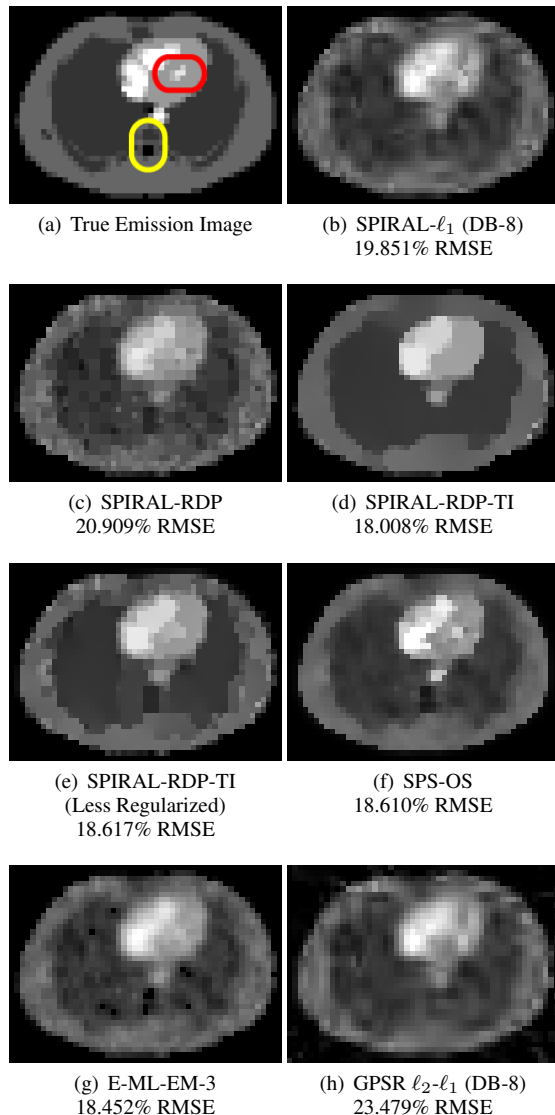


Fig. 3. Reconstructed emission images for one of 10 experiments performed: (a) is the true emission image, (b)–(d) and (f)–(h) are the minimal RMS error solutions. We include a less-regularized SPIRAL-RDP-TI estimate (e) that recovers more features at the expense of a slight increase in RMS error.

5. CONCLUSIONS

We have formulated the general goal of reconstructing an image from photon-limited measurements as a penalized maximum Poisson likelihood estimation problem. To solve this problem we proposed an algorithm that allows for a flexible choice of penalization methods, and focused on penalization schemes that discourage overly-complex or non-sparse estimates. We consider the case where the penalty corresponds to the sparsity-promoting ℓ_1 norm of the coefficients in a sparsifying basis, and the case where the penalty is related to the complexity of a partition-based estimate. When a cycle-spun translationally-invariant partition method is applied to the particular example of an emission tomography problem, the resulting estimates are comparable to the current state-of-the-art approaches developed specifically for emission tomography.

6. REFERENCES

- [1] D. Snyder and M. Miller, *Random Point Processes in Time and Space*, New York: Springer-Verlag, 1991.
- [2] E. J. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inform. Theory*, vol. 15, no. 12, pp. 4203–4215, 2005.
- [3] D. Donoho, “Compressed sensing,” *IEEE Trans. on Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [4] J. A. Tropp, “Just relax: convex programming methods for identifying sparse signals in noise,” *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [5] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [6] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.
- [7] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.
- [8] S. Ahn and J. Fessler, “Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms,” *IEEE Trans. Med. Imaging*, vol. 22, no. 5, pp. 613–626, May 2003.
- [9] R. Willett and R. Nowak, “Multiscale Poisson intensity and density estimation,” *IEEE Trans. Inform. Theory*, vol. 53, no. 9, pp. 3171–3187, 2007.
- [10] Z. T. Harmany, R. F. Marcia, and R. Willett, “Sparse Poisson intensity reconstruction algorithms,” in *IEEE Workshop on Statistical Signal Processing - SSP2009*, 2009.
- [11] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Trans. on Inform. Theory*, 2008, Submitted.
- [12] J. Fessler *et al.*, “Image reconstruction toolbox,” <http://www.eecs.umich.edu/~fessler/code>.
- [13] 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.*, vol. 1, no. 4, pp. 586–597, 2007.