

Sparsity Regularization for Image Reconstruction with Poisson Data

Daniel J. Lingenfelter^a, Jeffrey A. Fessler^a, and Zhong He^b

^aElectrical Engineering and Computer Science, University of Michigan, Ann Arbor;

^bNuclear Engineering, University of Michigan, Ann Arbor

ABSTRACT

This work investigates three penalized-likelihood expectation maximization (EM) algorithms for image reconstruction with Poisson data where the images are known *a priori* to be sparse in the space domain. The penalty functions considered are the ℓ_1 norm, the ℓ_0 “norm,” and a penalty function based on the sum of logarithms of pixel values, $R(\mathbf{x}) = \sum_{j=1}^{n_p} \log(\frac{x_j}{\delta} + 1)$. Our results show that the ℓ_1 penalized algorithm reconstructs scaled versions of the maximum-likelihood (ML) solution, which does not improve the sparsity over the traditional ML estimate. Due to the singularity of the Poisson log-likelihood at zero, the ℓ_0 penalized EM algorithm is equivalent to the maximum-likelihood EM algorithm. We demonstrate that the penalty based on the sum of logarithms produces sparser images than the ML solution. We evaluated these algorithms using experimental data from a position-sensitive Compton-imaging detector, where the spatial distribution of photon-emitters is known to be sparse.

Keywords: Poisson, Compton Imaging, Sparsity, Regularization, Expectation Maximization

1. INTRODUCTION

A challenging problem facing society today is the detection and identification of nuclear materials at ports of entry and in densely populated areas. If law enforcement had access to a device able to produce images of high-energy photon emissions, it would increase their ability to thwart an attack. The radiation measurement group at the University of Michigan is developing a 4π position-sensitive Compton imaging system¹ made from a block of semiconductor material where charge is produced from high-energy photon interactions. A pixelated capacitor is applied to the semiconductor material to measure changes in voltage over time, which gives the position, intensity, and energy of each interacting photon. Nuclear devices are typically small, which means that they should be concentrated in the reconstructed image. Many ordinary materials, such as concrete, contain radioactive isotopes that contribute to the background radiation recorded by the camera. These materials, as well as cosmic radiation, are usually more broadly distributed in space. We would like an imaging system that reveals only small concentrated point-sources of photon-emitting materials while suppressing the coarse background emission distribution. This leads us to the *a priori* knowledge that an image reconstructed using this system should be sparse in the space domain.

The measured photon interactions recorded by such a system can be modeled as a Poisson random vector. The model is

$$y_i \sim \text{Poisson}([\mathbf{A}\mathbf{x}]_i + \bar{r}_i), \quad i = 1 \dots n_d, \quad (1)$$

where y_i is the number of recorded photons in the i^{th} discretized detector bin, n_d is the number of detector bins, \mathbf{A} is the system matrix, $\bar{\mathbf{r}} = [\bar{r}_1, \bar{r}_2, \dots, \bar{r}_{n_d}]^T$ is the mean number of background counts, and $\mathbf{x} = [x_1, x_2, \dots, x_{n_p}]^T$ is the image of radiation emitters around the sphere discretized with n_p pixels. Since the number of possible interactions within the detector bulk is very large, such systems typically use the list-mode approach,² where only the rows of \mathbf{A} corresponding to received photons are computed. A commonly used reconstruction method¹ for Compton imaging employs the Maximum Likelihood Expectation Maximization (MLEM) algorithm³ with a Poisson model similar to those used in PET and SPECT.⁴ Penalized-likelihood algorithms have been developed

This work was supported in part by DND0/NSF grant CBET 0736091.

for image reconstruction problems to incorporate *a priori* knowledge about the reconstructed images.⁵ These algorithms find the solution to the following optimization problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} -L(\mathbf{x}) + \beta R(\mathbf{x}), \quad (2)$$

where $L(\mathbf{x})$ is the log-likelihood of the image given the observed data and $R(\mathbf{x})$ is typically a convex function known as a penalty or regularizer. The log-likelihood $L(\mathbf{x})$ is based on the statistical model (1) and is well understood. In contrast, there are numerous possible choices for the regularizer $R(\mathbf{x})$. In this work, we focus on choices for $R(\mathbf{x})$ that encourage sparsity.

There has been considerable interest in the case where the measurements are Gaussian distributed. A well-known measure of sparsity, the ℓ_1 norm, is commonly used in sparse approximation and compressed sensing. Using $R(\mathbf{x}) = \|\mathbf{x}\|_1$ for Gaussian noise, (2) can be written as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \beta \|\mathbf{x}\|_1, \quad (3)$$

which can be solved by quadratic programming.⁶ Algorithms that solve (3) yield sparse solutions because of their soft thresholding behavior. In this work, we explore the properties of penalized-likelihood algorithms for Poisson statistics using penalty functions that encourage sparsity.

2. SPARSITY AND REGULARIZATION DESIGN

This section investigates which penalty functions $R(\mathbf{x})$ will encourage sparsity in the reconstructed image. The penalty functions that we consider are:

$$\begin{aligned} R_1(\mathbf{x}) &= \|\mathbf{x}\|_1 \\ R_0(\mathbf{x}) &= \|\mathbf{x}\|_0 \\ R_l(\mathbf{x}) &= \sum_{j=1}^{n_p} \log\left(\frac{x_j}{\delta} + 1\right). \end{aligned}$$

2.1 Unregularized MLEM Algorithm

The well-known MLEM algorithm iteratively finds the maximum-likelihood image based on the observed measurements,¹ or equivalently, finds the solution to the optimization problem

$$\arg \min_{\mathbf{x}} (-L(\mathbf{x})), \quad (4)$$

where the negative log-likelihood corresponding to the model (1) is

$$-L(\mathbf{x}) = \sum_{i=1}^{n_p} \bar{y}_i(\mathbf{x}) - y_i \log \bar{y}_i(\mathbf{x}), \quad (5)$$

and

$$\bar{y}_i(\mathbf{x}) \triangleq [\mathbf{A}\mathbf{x}]_i + \bar{r}_i = \sum_{j=1}^{n_p} a_{ij}x_j + \bar{r}_i. \quad (6)$$

We use the EM algorithm for Poisson data described in⁷ that is equivalent to an optimization transfer approach with the following surrogate function:

$$Q(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{j=1}^{n_p} Q_j(x_j; \mathbf{x}^{(n)}), \quad (7)$$

where

$$Q_j(x_j; \mathbf{x}^{(n)}) \triangleq \sum_{i=1}^{n_d} \left(\frac{a_{ij}(x_j^{(n)} + \gamma_j)}{\bar{y}_i^{(n)}} \right) \left[\frac{x_j + \gamma_j}{x_j^{(n)} + \gamma_j} \bar{y}_i^{(n)} - y_i \log \left(\frac{x_j + \gamma_j}{x_j^{(n)} + \gamma_j} \bar{y}_i^{(n)} \right) \right], \quad (8)$$

and γ_j , for $j = 1 \dots n_p$, are nonnegative constants satisfying

$$\sum_{j=1}^{n_p} a_{ij} \gamma_j \leq \bar{r}_i \quad i = 1, \dots, n_d. \quad (9)$$

Typically, $\gamma_j = 0$, but using $\gamma_j > 0$ can accelerate convergence.⁷ This MLEM algorithm has the following “expectation step” or E-step

$$e_j(\mathbf{x}) = \sum_{i=1}^{n_d} a_{ij} \frac{y_i}{\bar{y}_i(\mathbf{x})}, \quad (10)$$

and a “maximization” step or M-step

$$x_j^{(n+1)} = \left[\frac{x_j^{(n)} + \gamma_j}{a_j} e_j(\mathbf{x}^{(n)}) - \gamma_j \right]_+. \quad (11)$$

The operator $[\cdot]_+$, used in (11), is defined by

$$[u]_+ = \max(u, 0). \quad (12)$$

The a_j term represents the system sensitivity, which is the probability that a photon originating from pixel j will be recorded by the system. Mathematically, we can write

$$a_j = \sum_{i=1}^{n_d} a_{ij}. \quad (13)$$

This quantity is sometimes found by simulation in the list-mode case⁸ because it can be impractical to compute all rows in \mathbf{A} .

In the case where $\mathbf{A} = \text{diag}\{a_i\}$, the cost function is separable and the problem becomes one of denoising rather than reconstruction. In this case, the ML estimate for the source intensity x_i from the measurement y_i is

$$\hat{x}_i = \frac{1}{a_i} [y_i - \bar{r}_i]_+. \quad (14)$$

This ML estimator will be the basis of comparison with the regularized methods that follow.

2.2 Penalized-Likelihood Algorithm for $R(\mathbf{x}) = \|\mathbf{x}\|_1$

The ℓ_1 norm has been used previously as a measure of sparsity⁹. It is a convenient penalty function because it is convex and differentiable almost everywhere. This section shows that the influence of $\|\cdot\|_1$ for Poisson noise is quite different from the Gaussian case. The ℓ_1 regularized estimator solves the following optimization problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \Psi_1(\mathbf{x}), \quad \Psi_1(\mathbf{x}) = -L(\mathbf{x}) + \beta \|\mathbf{x}\|_1. \quad (15)$$

We approach this problem using the optimization transfer approach,¹⁰ where one finds a separable surrogate function that satisfies

$$Q(\mathbf{x}; \mathbf{x}^{(n)}) \geq \Psi(\mathbf{x}) \quad (16)$$

$$Q(\mathbf{x}^{(n)}; \mathbf{x}^{(n)}) = \Psi(\mathbf{x}^{(n)}) \quad (17)$$

$$Q(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{i=1}^{n_p} Q_j(x_j; \mathbf{x}^{(n)}). \quad (18)$$

Extending (8), a valid surrogate for the cost function (15) is

$$Q_j(x_j; \mathbf{x}^{(n)}) = (x_j + \gamma_j) a_j - e_j(x_j^{(n)}) (x_j^{(n)} + \gamma_j) \log(x_j + \gamma_j) + \beta x_j, \quad (19)$$

which, minding the nonnegativity constraint on x_j , has the minimizer

$$x_j^{(n+1)} = \left[\frac{x_j^{(n)} + \gamma_j}{a_j + \beta} e_j(\mathbf{x}^{(n)}) - \gamma_j \right]_+, \quad (20)$$

where $e_j(\mathbf{x})$ was defined by the E-step in (10). Note that (20) and (10) represent the M-step and E-step, respectively, of this penalized-likelihood algorithm. Also, (20) is identical to (11) except for a scaling by $a_j + \beta$ instead of a_j . This causes scaling of the iterates, but not thresholding behavior that is desirable in sparse reconstruction. We can gain insight by examining the behavior of ℓ_1 regularization when $\mathbf{A} = \text{diag}\{a_i\}$, where the estimate reduces to

$$\hat{x}_j = \frac{[y_i - \bar{r}_i(1 + \beta/a_i)]_+}{a_i + \beta}. \quad (21)$$

When $\bar{r}_i = 0$, the estimate is just a scaled version of the observation, which is just shrinkage compared to the unregularized estimator in (14). Also, when $\bar{r}_i > 0$, the estimator introduces thresholding where the threshold becomes larger as β increases. However, this thresholding does not occur when one does not have knowledge of the background and $\bar{r}_i = 0$. This behavior differs from the case of Gaussian statistics where

$$\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{r}, 1). \quad (22)$$

The solution to the denoising problem in this case is

$$\hat{x}_i = \begin{cases} \frac{1}{a_i} \left(y_i - \bar{r}_i - \frac{\beta}{a_i} \right) & y_i > \bar{r}_i + \frac{\beta}{a_i} \\ 0 & \bar{r}_i - \frac{\beta}{a_i} \leq y_i \leq \bar{r}_i + \frac{\beta}{a_i} \\ \frac{1}{a_i} \left(y_i - \bar{r}_i + \frac{\beta}{a_i} \right) & y_i < \bar{r}_i - \frac{\beta}{a_i} \end{cases}, \quad (23)$$

which is a soft thresholding estimator.

2.2.1 Non-Uniqueness

The following counterexample shows that the ℓ_1 regularized solution is not unique for Poisson data. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \bar{r}_1 = 0.$$

In this case, $y_1 \sim \text{Poisson}(x_1 + x_2)$. One can show that any pair of nonnegative values \hat{x}_1, \hat{x}_2 that satisfy

$$\hat{x}_1 + \hat{x}_2 = \frac{y_1}{1 + \beta}$$

is a minimizer of the penalized-likelihood cost function Ψ_1 in (15). This non-uniqueness of the ℓ_1 -regularized estimate is consistent with the Gaussian case.

3. ℓ_0 REGULARIZATION FOR POISSON DATA

The ℓ_0 “norm” measures sparsity in its purest sense by counting the number of nonzero elements. Unlike the ℓ_1 norm, it is nonconvex, not continuous, and uniform almost everywhere except zero. This section shows that ℓ_0 regularization can behave quite differently for Poisson data compared to the Gaussian case. The ℓ_0 regularized estimate is the solution to the following optimization problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \Psi_0(\mathbf{x}), \quad \Psi_0(\mathbf{x}) = -\mathbf{L}(\mathbf{x}) + \beta \|\mathbf{x}\|_0, \quad (24)$$

where the log-likelihood term $L(\mathbf{x})$ is defined in (5). The solution to this problem is obtained by minimizing the cost function $\Psi(\mathbf{x}) = -L(\mathbf{x}) + \beta \|\mathbf{x}\|_0$. Extending (8) again, we consider the surrogate

$$Q_j(x_j; \mathbf{x}^{(n)}) \triangleq (x_j + \gamma_j) a_j - e_j(x_j^{(n)}) (x_j^{(n)} + \gamma_j) \log(x_j + \gamma_j) + \beta I_{(0, \infty)}(x_j), \quad (25)$$

where $I_{(0, \infty)}(x_j)$ is the indicator function on the interval $(0, \infty)$. The M-step is given by

$$x_j^{(n+1)} = \arg \min_{x_j \geq 0} Q_j(x_j; \mathbf{x}^{(n)}). \quad (26)$$

If the minimizer $\tilde{x}_j^{(n+1)}$ is positive, it is found by setting the derivative equal to zero:

$$\frac{\partial}{\partial x_j} Q_j(x_j; \mathbf{x}^{(n)}) = e_j(\mathbf{x}^{(n)}) \frac{x_j^{(n)} + \gamma_j}{x_j + \gamma_j} - a_j = 0.$$

Solving and minding the non-negativity constraint, we obtain a candidate solution:

$$\tilde{x}_j^{(n+1)} = \left[\frac{x_j^{(n)} + \gamma_j}{a_j} e_j(\mathbf{x}^{(n)}) - \gamma_j \right]_+. \quad (27)$$

However, if $\gamma_j > 0$, the minimizer could either be at the location $\tilde{x}_j^{(n+1)}$ found on the interval $(0, \infty)$, or it could be at $x_j = 0$ because of the non-convexity of the surrogate. If $\tilde{x}_j^{(n+1)} > 0$, one should choose the minimum $\hat{x}_j^{(n+1)} = 0$ if

$$Q_j(0; \mathbf{x}^{(n)}) < Q_j(\tilde{x}_j^{(n+1)}; \mathbf{x}^{(n)})$$

$$e_j(\mathbf{x}^{(n)}) (x_j^{(n)} + \gamma_j) \log\left(\frac{\tilde{x}_j^{(n+1)} + \gamma_j}{\gamma_j}\right) - a_j \tilde{x}_j^{(n+1)} < \beta, \quad (28)$$

otherwise if $\tilde{x}_j^{(n+1)} = 0$, $\hat{x}_j^{(n+1)} = 0$. Combining (27) and (28), we obtain an EM algorithm with E-step given by (10) and M-step:

$$x_j^{(n+1)} = \begin{cases} \tilde{x}_j^{(n+1)} & \tilde{x}_j^{(n+1)} > 0 \text{ AND } e_j(\mathbf{x}^{(n)}) (x_j^{(n)} + \gamma_j) \log\left(\frac{\tilde{x}_j^{(n+1)} + \gamma_j}{\gamma_j}\right) - a_j \tilde{x}_j^{(n+1)} > \beta \\ 0 & \text{else.} \end{cases} \quad (29)$$

This algorithm reduces to the MLEM algorithm when $\gamma_j = 0$ for $j = 1, \dots, n_p$. The behavior of the algorithm depends on these tuning parameters because they shift the cost function so that the discontinuity in the ℓ_0 “norm” and the singularities of the log-likelihood function no longer overlap at $x_j = 0$ for $j = 1, \dots, n_p$. By (9), if $\bar{r}_i = 0$ for $i = 1, \dots, n_d$, γ_j must be zero for all $j = 1 \dots n_p$, which means that the algorithm is equivalent to MLEM in that case. If $\mathbf{A} = \text{diag}\{a_i\}$, the ℓ_0 regularized estimate is

$$\hat{x}_j = \begin{cases} \frac{1}{a_i} [y_i - \bar{r}_i]_+ & \kappa(y_i, \bar{r}_i) > \beta \\ 0 & \text{else,} \end{cases} \quad (30)$$

where $\kappa(u, v)$ denotes the Kullback-Leibler divergence¹¹ given by

$$\kappa(u, v) = \begin{cases} v & u = 0, v \geq 0 \\ u \log\left(\frac{u}{v}\right) - u + v & u > 0, v > 0 \\ \infty & u > 0, v = 0. \end{cases}$$

When $\bar{r}_i = 0$, (30) simplifies to (14) because $\kappa(y_i, 0) = \infty$ for all y_i . Again, in the case of ℓ_0 regularization, when $\bar{r}_i = 0$, there is no thresholding behavior. The ineffectiveness of ℓ_0 regularization for Poisson noise is different from that for Gaussian noise modeled by (22), where the denoising estimator is

$$\hat{x}_i = \begin{cases} \frac{y_i - \bar{r}_i}{a_i} & |y_i - \bar{r}_i| > \sqrt{2\beta} \\ 0 & \text{else,} \end{cases} \quad (31)$$

which is hard thresholding. The ineffectiveness of Poisson regularization in the ℓ_0 case is due to the singularity in the Poisson log-likelihood $-L(x)$ at $x = 0$. Figure 1 shows the Gaussian and Poisson likelihoods for a denoising problem where the observation $y = 3$. Note that the Gaussian likelihood is well-behaved near $x = 0$, but the Poisson likelihood approaches infinity. This singularity causes the cost function to be arbitrarily large near $\mathbf{x} = 0$, regardless of the regularizer chosen.

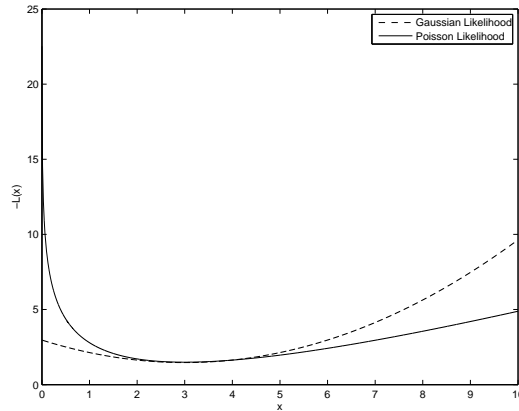


Figure 1. Gaussian and Poisson log-likelihoods vs. x for denoising case with measurement $y = 3$

4. log REGULARIZATION

The penalty function $R(\mathbf{x}) = \sum_{j=1}^{n_p} \log\left(\frac{x_j}{\delta} + 1\right)$ is suggested as an alternative to the ℓ_1 and ℓ_0 norms⁹. It is nonconvex and thus the minimizer is not guaranteed to be unique. However, it is a continuous function that is differentiable almost everywhere. It is also a better measure of sparsity than the ℓ_1 norm in the sense that it better approximates the true measure of sparsity, the ℓ_0 “norm.” The minimization problem is:

$$\arg \min_{\mathbf{x}} \Psi_l(\mathbf{x}), \quad \Psi_l(\mathbf{x}) = -L(\mathbf{x}) + \beta \sum_{j=1}^{n_p} \log\left(\frac{x_j}{\delta} + 1\right). \quad (32)$$

We first examine the case where \mathbf{A} is diagonal, i.e., $\mathbf{A} = \text{diag}(\mathbf{a})$, for which the cost function is separable. This denoising case provides insight into the behavior of the regularizer for a coupled cost function. The minimizer of the separable cost function is given by

$$\hat{x}_i = \arg \min_{x_i \geq 0} \Psi_i(x_i), \quad \Psi_i(x_i) = \sum_{i=1}^n a_i x_i + \bar{r}_i - y_i \log(a_i x_i + \bar{r}_i) + \beta \log\left(\frac{x_i}{\delta} + 1\right), \quad (33)$$

which has nonnegative minimizer

$$\hat{x}_i(y_i) = \frac{1}{2} \left(-\delta - \frac{1}{a_i}(\bar{r}_i + \beta - y_i) + \frac{1}{a_i} \sqrt{\delta^2 a_i^2 + 2\delta a_i(\bar{r}_i + \beta - y_i) + (\bar{r}_i + \beta - y_i)^2 + 4a_i\delta(y_i - \bar{r}_i) - 4\beta\bar{r}_i} \right). \quad (34)$$

Since it is difficult to visualize the behavior of the estimate with respect to δ and β , Figure 4 illustrates the estimator for various values of β and δ with $\bar{r}_i = 0$ and $a_i = 1$ for all i .

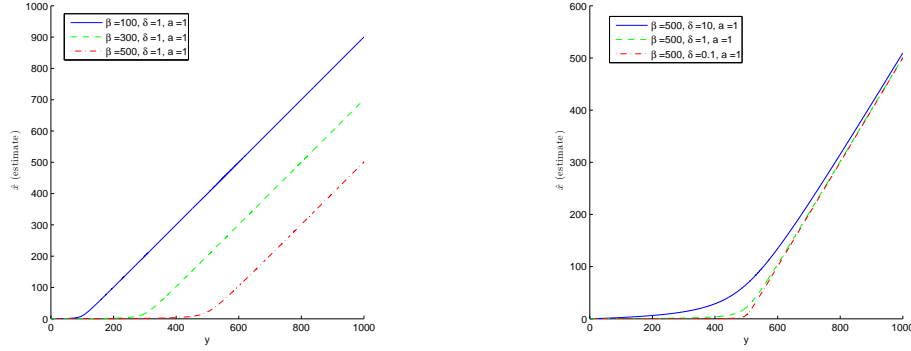


Figure 2. \hat{x}_i vs y_i with $\bar{r}_i = 0$ for three β values and $\delta = 1$ (right), and three δ values and $\beta = 100$ (left)

4.1 Choosing δ

As $\delta \rightarrow 0$, the estimator $\hat{x}_i(y_i)$ in (34) converges pointwise to the following function (minding the non-negativity constraint):

$$\hat{x}_i(y_i)^{ideal} = \lim_{\delta \rightarrow 0} \hat{x}_i(y_i) = \begin{cases} 0 & (\bar{r}_i + \beta - y_i)^2 - 4\beta\bar{r}_i < 0 \\ \frac{1}{2a_i} \left(y_i - \beta - \bar{r}_i + \sqrt{(\bar{r}_i - y_i + \beta)^2 - 4\beta\bar{r}_i} \right) & (\bar{r}_i + \beta - y_i)^2 - 4\beta\bar{r}_i \geq 0, \end{cases} \quad (35)$$

which is an approximation to hard thresholding. Figure 3 shows a graph of this function for a set of particular parameters. When $\bar{r}_i = 0$, which is of interest in 4π imaging, (35) simplifies to:

$$\hat{x}_i(y_i)^{ideal} = \frac{1}{2a_i} (y_i - \beta + |y_i - \beta|) = \frac{1}{a_i} [y_i - \beta]_+. \quad (36)$$

Thus, when $\bar{r}_i = 0$, smaller δ results in better approximation of soft thresholding. However, the sharp curvature of $\hat{x}_i(y_i)$ when δ is small can slow convergence of the EM algorithm in the case where \mathbf{A} is not diagonal. In our experiments, we found that $10 \leq \delta \leq 100$ worked well for the Compton imaging problem.

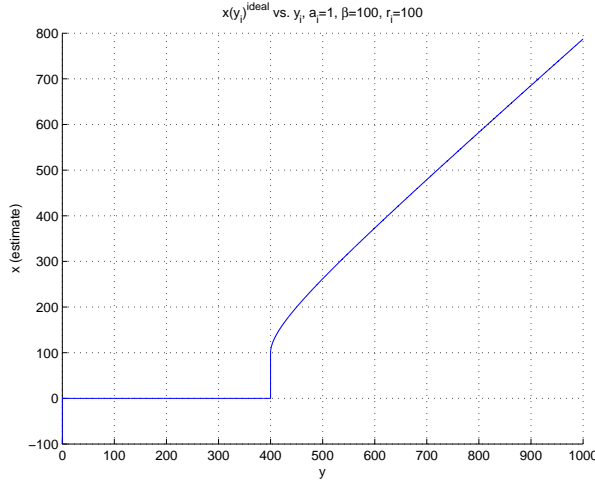


Figure 3. $\hat{x}_i(y_i)^{ideal}$ vs y_i for $\beta = 100, a_i = 1, \bar{r}_i = 100$

4.2 Choosing β

If $\bar{r}_i = 0$ the estimator approximates soft thresholding as shown in (36) and when $\bar{r}_i > 0$, the estimator approximates hard thresholding as shown in Figure 3, where the threshold is a function of β . In both cases, increasing β increases the threshold. The choice of β involves a trade-off between sensitivity to noise and failure to image low-intensity point sources.

4.3 EM Algorithm for log Regularization

To develop an EM algorithm for the cost function (32), we use the following separable surrogate functions to iteratively find the minimizer:

$$Q_j(x_j; \mathbf{x}^{(n)}) = (x_j + \gamma_j) a_j - e_j(x_j^{(n)}) (x_j^{(n)} + \gamma_j) \log(x_j + \gamma_j) + \beta \log\left(\frac{x_j}{\delta} + 1\right). \quad (37)$$

The E-step is given by $e_j(\mathbf{x})$ in (10) and the M-step has the same form as (26). Differentiating (37) and zeroing yields the following quadratic formula which one can solve to compute $\mathbf{x}_j^{(n+1)}$:

$$0 = \frac{a_j}{\delta} x_j^2 + \left(a_j + \delta^{-1} \left(\gamma_j a_j - e_j(\mathbf{x}^{(n)}) (x_j^{(n)} + \gamma_j) + \beta \right) \right) x_j + a_j \gamma_j - e_j(\mathbf{x}^{(n)}) (x_j^{(n)} + \gamma_j) + \frac{\beta}{\delta} \gamma_j. \quad (38)$$

The behavior of this algorithm is difficult to visualize in the reconstruction problem when \mathbf{A} is not diagonal, but the separable analysis of the denoising problem gives some intuition. The algorithm will perform an approximation of soft thresholding when $\bar{r}_i = 0$ for $i = 1, \dots, n_d$, which will lead to sparser reconstructed images than ML estimate.

5. RESULTS

This section gives experimental results of the algorithms presented in this paper for a list-mode 4π Compton imaging system. The scenes monitored by this system do not have a known background emission distribution, so we set the expected background contribution to each measurement, \bar{r}_i , to be zero. These images were produced from real laboratory measurements of a Cesium source and a CdZnTe 4π position-sensitive detector. The source is located at $\phi = 90^\circ, \theta = 90^\circ$.

5.1 ℓ_0 Regularization

As mentioned in Sec. 3, if $\bar{r}_i = 0$ for all i , then $\gamma_j = 0$ for all j . We can see from (29) that the algorithm in this case simplifies to unregularized MLEM. The unregularized ML estimates are shown in the upper left of Figures 4 and 5.

5.2 ℓ_1 Regularization

Figure 4 shows reconstructed images using the ℓ_1 regularized algorithm. Note that the image corresponding to $\beta = 0$ is the ML estimate. Also note that the images appear identical to the eye except the pixel values are scaled by a constant. This is apparent from the color scales located to the right of each image. We investigated numerous other values of β and observed only pixel scaling. This behavior is consistent with the analysis for the separable system in (21).

5.3 log Regularization

The log regularization algorithm which solves (32) does indeed increase the sparsity of the reconstructed images as β increases. Figure 5 shows four reconstructed images using the log regularization estimate for different values of β . Following the reasoning of Sec. 4.1, these images were reconstructed with the parameter $\delta = 100$. However, the choice of δ could impact the number of iterations needed to achieve convergence.

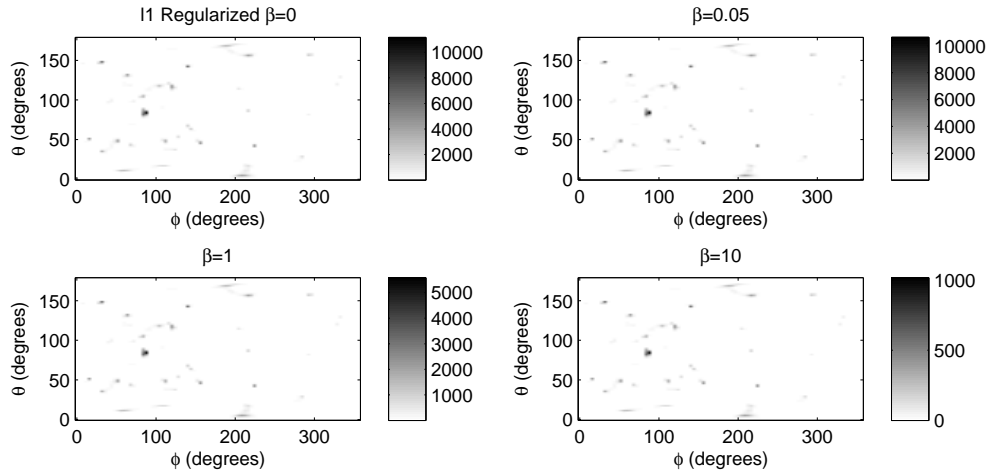


Figure 4. Reconstructed images using ℓ_1 regularization with various values of β . These images were reconstructed from 500 recorded photons in a real CdZnTe detector. The algorithm was run for 200 iterations.

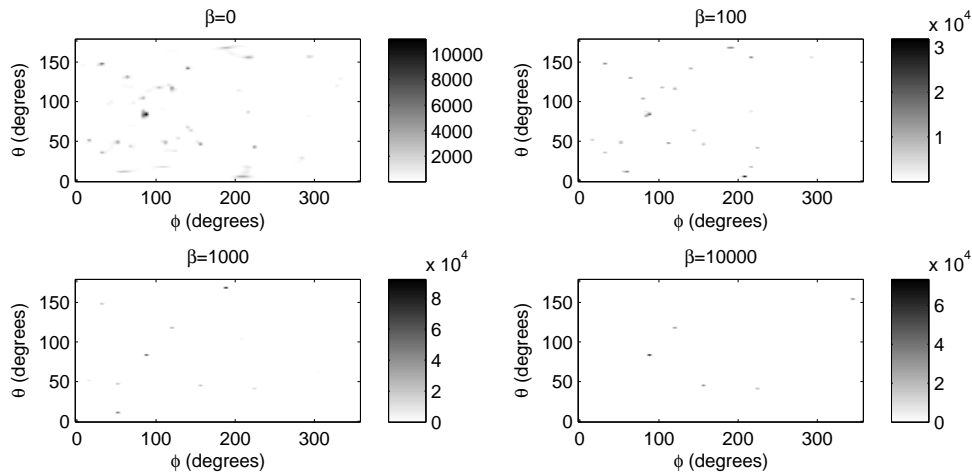


Figure 5. Reconstructed images using log regularization with various values of β . These images were reconstructed from 500 recorded photons in a real CdZnTe detector. The algorithm was run for 200 iterations and the parameter $\delta = 100$.

6. CONCLUSION

We derived three penalized-likelihood algorithms for image reconstruction of Poisson data when the images are known to be sparse in the space domain. We found that, unlike the case of Gaussian measurements, the ℓ_1 and ℓ_0 norms do not yield sparser reconstructed images than unregularized algorithms when the background is assumed to be zero. The penalty based on the logarithm of the pixel values given in Sec. 4 was found to enhance the sparsity of the reconstructed images (even for zero background), which is an advantage over the ℓ_1 and ℓ_0 penalties. In the case where there is a known number of mean background counts contributing to each measurement, the ℓ_1 and ℓ_0 reconstruction algorithms perform soft and hard thresholding, respectively. Future work would investigate the benefit of ℓ_1 and ℓ_0 regularization when the background is known. Mathematically, the ℓ_1 and ℓ_0 regularized algorithms do not behave in the Poisson case as they do in the Gaussian case because of the singularity of the Poisson log-likelihood at zero and the nonnegativity constraint. The singularity removes the effect of the discontinuity of the ℓ_0 “norm” at zero, and the nonnegativity constraint causes the ℓ_1 norm to be a linear function of the pixel intensities, which causes uniform shrinkage of the pixel values.

ACKNOWLEDGMENTS

The authors would like to thank Christopher Wahl of the University of Michigan for supplying the data used to produce the reconstructed images shown in this paper. The authors also thank Professor Clayton Scott of the University of Michigan for discussions about ℓ_1 regularization.

REFERENCES

- [1] Xu, D., He, Z., Lehner, C. E., and Zhang, F., “4-pi Compton imaging with single 3D position-sensitive CdZnTe detector,” in [*Proc. SPIE 5540, Hard X-Ray and Gamma-Ray Detector Physics VI*], 144–55 (2004).
- [2] Barrett, H. H., White, T., and Parra, L. C., “List-mode likelihood,” *J. Opt. Soc. Am. A* **14**, 2914–23 (Nov. 1997).
- [3] Dempster, A. P., Laird, N. M., and Rubin, D. B., “Maximum likelihood from incomplete data via the EM algorithm,” *J. Royal Stat. Soc. Ser. B* **39**(1), 1–38 (1977).
- [4] Shepp, L. A. and Vardi, Y., “Maximum likelihood reconstruction for emission tomography,” *IEEE Trans. Med. Imag.* **1**, 113–22 (Oct. 1982).
- [5] De Pierro, A. R., “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography,” *IEEE Trans. Med. Imag.* **14**, 132–7 (Mar. 1995).
- [6] Figueiredo, M., Nowak, R., and Wright, S. J., “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE J. Sel. Top. Sig. Proc.* **1**, 586–97 (Dec. 2007).
- [7] Fessler, J. A. and Hero, A. O., “Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms,” *IEEE Trans. Im. Proc.* **4**, 1417–29 (Oct. 1995).
- [8] Qi, J. and Huesman, R. H., “Propagation of errors from the sensitivity image in list mode reconstruction,” *IEEE Trans. Med. Imag.* **23**, 1094–9 (Sept. 2004).
- [9] Candès, E. J., Wakin, M. B., and Boyd, S., “Enhancing sparsity by reweighted l1 minimization,” (2007).
- [10] Lange, K., Hunter, D. R., and Yang, I., “Optimization transfer using surrogate objective functions,” *J. Computational and Graphical Stat.* **9**, 1–20 (Mar. 2000).
- [11] Byrne, C., [*Applied iterative methods*], A K Peters, Wellesley, MA (2008).