NONCONVEX RELAXATION FOR POISSON INTENSITY RECONSTRUCTION

Lasith Adhikari, and Roummel F. Marcia

Department of Applied Mathematics, University of California, Merced, Merced, CA 95343 USA

ABSTRACT

Critical to accurate reconstruction of sparse signals from low-dimensional Poisson observations is the solution of nonlinear optimization problems that promote sparse solutions. Theoretically, non-convex ℓ_p-norm minimization $(0 \leq p < 1)$ would lead to more accurate reconstruction than the convex ℓ_1-norm relaxation commonly used in sparse signal recovery. In this paper, we propose an extension to the existing SPIRAL-ℓ_1 algorithm based on the Generalized Soft-Thersholding (GST) function to better recover signals with mostly nonzero entries from Poisson observations. This approach is based on iteratively minimizing a sequence of separable subproblems of the nonnegatively constrained, ℓ_p-penalized negative Poisson log-likelihood objective function using the GST function. We demonstrate the effectiveness of the proposed method, called SPIRAL-ℓ_p, through numerical experiments.

***Index Terms*—** Nonconvex optimization, low-photon imaging, Poisson noise, generalized soft-thersholding, ℓ_p-norm

1. INTRODUCTION

The Poisson process model [1] has been widely used in a variety of imaging applications, including atmospherically degraded and low-light imaging [2] and low photon medical imaging such as Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), and Confocal Microscopy [3]. When the arrival of photons is modeled by the Poisson process model, the observed data $\mathbf{y}$ is said to have a Poisson distribution with mean detector photon intensity $\mathbf{A}\mathbf{f}^*$:

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{f}^*),$$

where $\mathbf{y} \in \mathbb{Z}_+^m$ is a vector of observed photon counts, $\mathbf{f}^* \in \mathbb{R}_+^n$ is the vector of true signal intensity, and $\mathbf{A} \in \mathbb{R}_+^{m\times n}$ is the system matrix that linearly projects the true signal to the detector photon intensity. Since the Poisson parameter is not known, the *maximum likelihood principle* is used to determine $\mathbf{A}\mathbf{f}^*$ such that the probability of observing the vector of photon counts $\mathbf{y}$ is maximized.

In SPIRAL [4], the Poisson intensity reconstruction was achieved by minimizing a sequence of convex subproblems regularized by a variety of penalty terms. One penalty term used, in particular, is the ℓ_1-norm, which has been shown to be a very good approximation to the ℓ_0-norm [5]. While this SPIRAL-ℓ_1 approach yielded reasonably good results, its reconstruction contained some spurious artifacts. In [6], these reconstruction errors can be corrected using a nonconvex ℓ_p-norm penalty term, where $p < 1$. This paper uses this nonconvex ℓ_p penalty within the SPIRAL framework to eliminate the spurious artifacts in the Poisson intensity reconstruction while keeping reconstruction error low. The resulting optimization problem will be nonsmooth and non-convex. The proposed approach is based on the recent work of Zuo et al. [7], who proposed a simple and efficient iterative algorithm for ℓ_p-norm non-convex sparse coding which was an extension to the popular soft-thresholding operator [8].

This work was supported by National Science Foundation Grant CMMI-1333326.

2. PROBLEM FORMULATION

In this section, we first formulate the Poisson intensity reconstruction problem as a constrained optimization problem with an ℓ_p penalty term. Then, based on [4], we describe how it can be solved using a sequence of separable nonconvex subproblems. Finally, we discuss how these subproblems can be solved using generalized soft-thresholding.

2.1. Sparse Poisson Intensity Reconstruction using ℓ_p- norm

The Poisson reconstruction problem has the following constrained optimization form (see e.g., [9]):

$$\begin{aligned} &\underset{\mathbf{f}\in\mathbb{R}^n}{\text{minimize}} && \Phi(\mathbf{f}) \equiv F(\mathbf{f}) + \tau\ \text{pen}(\mathbf{f}) \\ &\text{subject to} && \mathbf{f} \succeq 0, \end{aligned} \qquad (1)$$

where $F(\mathbf{f})$ is the negative Poisson log-likelihood function

$$F(\mathbf{f}) = \mathbf{1}^T A\mathbf{f} - \sum_{i=1}^{m} y_i\ \log(\mathbf{e}_i^T A\mathbf{f} + \beta),$$

where $\mathbf{1}$ is the m-vector of ones, $\mathbf{e}_i$ is the i-th column of the $m \times m$ identity matrix, $\beta > 0$ (typically $\beta \ll 1$), and pen : $\mathbb{R}^n \longrightarrow \mathbb{R}$ is a penalty functional. In this paper, we will consider $\text{pen}(\mathbf{f}) = \|\mathbf{f}\|_p^p$ $(0 \leq p < 1)$ as the penalty function in (1). Then the generalized constrained optimization problem can be written as

$$\begin{aligned} \hat{\mathbf{f}} \quad = \quad &\underset{f\in\mathbb{R}^n}{\arg\min} && \Phi(\mathbf{f}) \equiv F(\mathbf{f}) + \tau\ \|\mathbf{f}\|_p^p \\ &\text{subject to} && \mathbf{f} \succeq 0, \end{aligned} \qquad (2)$$

where $\tau > 0$. As described in [4], the solution of the problem (2) can be found by minimizing a sequence of quadratic models to the function $F(\mathbf{f})$ approximated by second-order Taylor series expansion where the Hessian replaced by a scaled identity matrix $\alpha_k\mathbf{I}$ with $\alpha_k > 0$ [10]. Simplifying the second-order approximation yields a sequence of subproblems of the form

$$\begin{aligned} \mathbf{f}^{k+1} \quad = \quad &\underset{\mathbf{f}\in\mathbb{R}^n}{\arg\min} && \frac{1}{2}\ \|\ \mathbf{f} - \mathbf{s}^k\ \|_2^2\ + \frac{\tau}{\alpha_k}\ \ \|\ \mathbf{f}\ \|_p^p \\ &\text{subject to} && \mathbf{f} \succeq 0, \end{aligned} \qquad (3)$$

where

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

Note that the subproblem (3) can be separated into scalar minimization problems of the form

$$f^* = \arg\min_{f \in \mathbb{R}} \quad \Omega(f) = \frac{1}{2}(f-s)^2 + \lambda|f|^p, \quad \text{subject to} \quad f \geq 0. \tag{4}$$

where f and s denote elements of the vectors $\mathbf{f}$ and $\mathbf{s}^k$ respectively and $\lambda = \tau/\alpha_k$. We next describe how to solve (4) in two steps: first without the constraints, then with the constraints.

2.2. Generalized Soft-Thresholding (GST) Function

As shown in [7], for a given regularization parameter $\lambda > 0$ and p-norm for $\Omega(f)$ in (4), there exists a threshold value $\gamma_p(\lambda)$ (that explicitly depends on p and λ) such that if $s \leq \gamma_p(\lambda)$, the global minimum of (4) is $f^* = 0$; otherwise, the global minimum will be a non-zero value (see Fig. 1). We now show how to compute the threshold $\gamma_p(\lambda)$ so that we can compute f^*.

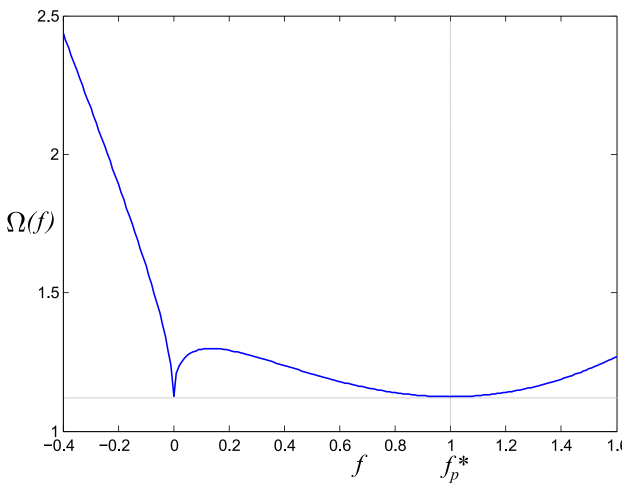

Fig. 1. The plot of the scalar quadratic function $\Omega(f)$ with p-norm penalty term in (4), where $p = 0.5$ and $\lambda = 1.0$. When s is less than or equal to the specific threshold value $\gamma_p(\lambda)$, then $f^* = 0$ is the global minimum. For the graph above $s = \gamma_p(\lambda)$ and there are global minima at $f^* = 0$ and f_p^*. If $s > \gamma_p(\lambda)$, then the global minimum is uniquely at $f_p^* > 0$.

We note that $\Omega(f)$ is symmetric in s. Thus, without loss of generality, we consider the case $s > 0$. When $s = \gamma_p(\lambda)$, there exists f_p^* such that

$$\Omega(f_p^*) = \Omega(0) \quad \text{and} \tag{5}$$

$$\Omega'(f_p^*) = 0 \tag{6}$$

(see Fig. 1 as an illustration). By solving (5) and (6) simultaneously, we can explicitly find the threshold value $\gamma_p(\lambda)$ for given p and λ values. Specifically, $\gamma_p(\lambda)$ is given by

$$\gamma_p(\lambda) = (2\lambda(1-p))^{\frac{1}{2-p}} + \lambda p(2\lambda(1-p))^{\frac{p-1}{2-p}}. \tag{7}$$

For any $s > \gamma_p(\lambda)$, the unique minimum $f^* = S_p(|s|, \lambda)$ of $\Omega(f)$ is greater than 0 and is obtained by setting Ω' to 0 :

$$\Omega'(S_p(|s|,\lambda)) = S_p(|s|,\lambda) - s + \lambda p(S_p(|s|,\lambda))^{p-1} = 0. \tag{8}$$

The root of Ω' can be computed using fixed-point iteration. More generally, the solution f^* to (4) is given by the Generalized Soft-Thersholding (GST) function

$$T_p(s,\lambda) = \begin{cases} 0, & \text{if } |s| \leq \gamma_p(\lambda) \\ \text{sgn}(s) S_p(|s|,\lambda), & \text{if } |s| > \gamma_p(\lambda) \end{cases} \tag{9}$$

(see [7] for details). We now consider the special cases $p = 0$ and $p = 1$ for $T_p(s, \lambda)$.

When $p = 0$, the GST function $T_0(s, \lambda)$ is called the *hard-thresholding function*, and it solves

$$\underset{f}{\text{minimize}} \quad \frac{1}{2}(f-s)^2 + \lambda|f|^0,$$

where

$$|f|^0 = \begin{cases} 0, & \text{if } f = 0 \\ 1, & \text{if } f \neq 0. \end{cases}$$

In this case, the GST function is given by

$$T_0(s,\lambda) = \begin{cases} 0, & \text{if } |s| \leq \gamma_0(\lambda) \\ s, & \text{if } |s| > \gamma_0(\lambda), \end{cases} \tag{10}$$

where the thresholding value is obtained by evaluating (7) at $p = 0$, i.e., $\gamma_0(\lambda) = (2\lambda)^{1/2}$.

When $p = 1$, the GST function becomes the *soft-thresholding function* (see e.g., [8]), where $\gamma_1(\lambda) = \lambda$, and

$$T_1(s,\lambda) = \begin{cases} 0, & \text{if } |s| \leq \gamma_1(\lambda) \\ \text{sgn}(s)(|s| - \lambda), & \text{if } |s| > \gamma_1(\lambda), \end{cases}$$

In both cases, we do not compute $S_p(|s|, \lambda)$ iteratively in (8), but rather we compute it explicitly.

2.3. Nonnegativity Constraint

Since the subproblems in (3) are nonnegatively constrained, the solution of the ℓ_p-minimization problem (4) also needs to be nonnegative. Therefore the theresholding operator is employed to obtain the next iterate:

$$f^{k+1} = \max(0, T_p(s, \lambda)).$$

We call this nonconvex approach based on SPIRAL [4] and GST [7] the SPIRAL-ℓ_p method.

3. NUMERICAL RESULTS

In this section, we evaluate the effectiveness of the proposed SPIRAL-ℓ_p method by comparing it to the SPIRAL-ℓ_1 method. We implemented the SPIRAL-ℓ_p method in MATLAB (on a PC with Intel Corei7 2.7GHz Processor, 2 cores, 8GB RAM) by modifying the existing MATLAB code of the SPIRAL method [11]. In the experiment, the true signal $\mathbf{f}$ is of length 100,000 with 1,500 nonzero entries (1.5% of sparsity), and the observed vector $\mathbf{y}$ is of length 40,000. We generate Poisson intensity reconstructions for 23 different p-values ranging from 0.99 to 0. For that, we use the parameters in SPIRAL-ℓ_1 as our default parameters in SPIRAL-ℓ_p. More specifically, SPIRAL-ℓ_p is initialized using rescaled $\mathbf{A}^T(\mathbf{y})$ and terminates if consecutive iterates do not significantly change. The regularization parameter τ in (1) is optimized to get the minimum Root Mean Square (RMS) error $\|\mathbf{f}^* - \hat{\mathbf{f}}\|_2/\|\mathbf{f}^*\|_2$ for each p-value.

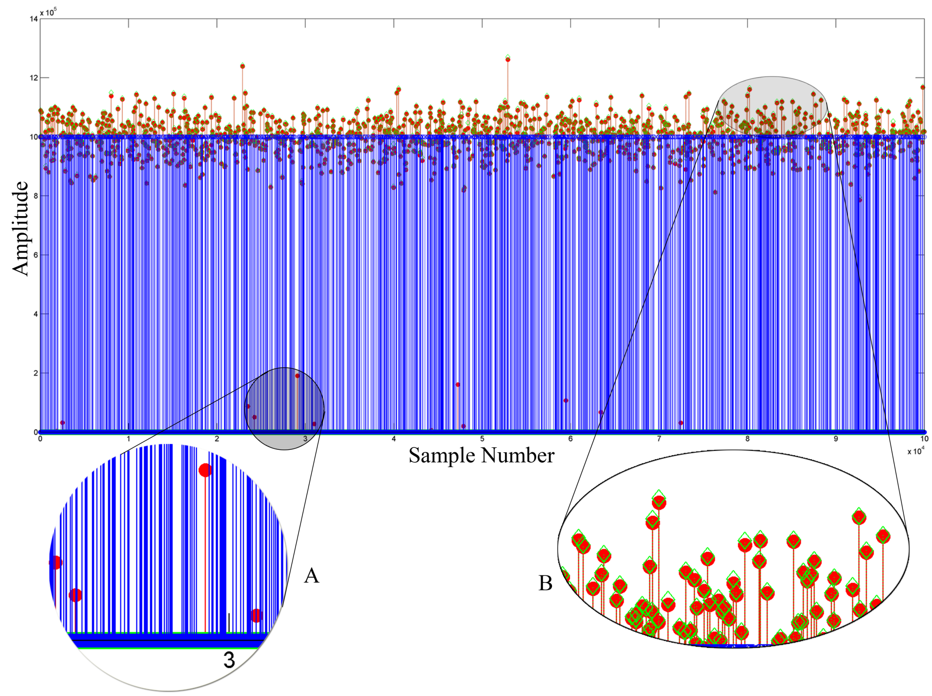

Fig. 3. SPIRAL-ℓ_p intensity reconstruction with $p = 0.05$ (green diamond stems) compared with the SPIRAL-ℓ_1 intensity reconstruction (red filled circle stems). The blue stems depict the true signal. There are 12 spurious solutions in the SPIRAL-ℓ_1 reconstruction. (A) SPIRAL-$\ell_{0.05}$ reconstruction eliminates spurious solutions in the SPIRAL-ℓ_1 reconstruction. (B) SPIRAL-$\ell_{0.05}$ solution generally matches the SPIRAL-ℓ_1 solution.

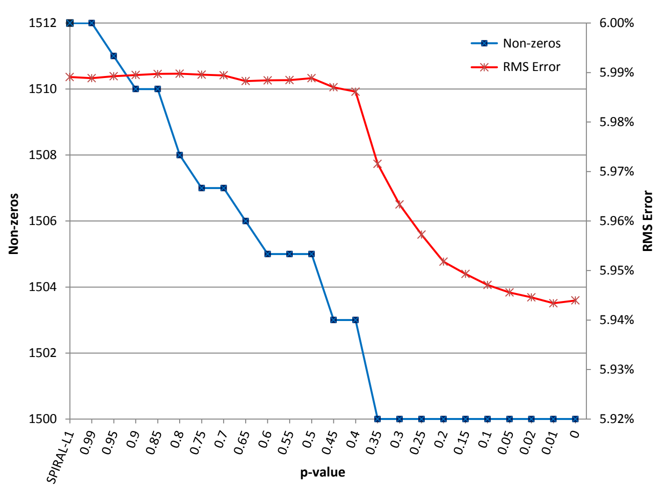

Fig. 2. The plot of the RMS error and the number of nonzero entries in the reconstruction over the p-values ranging from 0.99 to 0. The left most data points in both curves correspond to the error and number of non-zeros of SPIRAL-ℓ_1. There is a steep decrease in the RMS error after $p = 0.4$ while, non-zeros attain their exact value 1500 at $p = 0.35$. RMSE (%) = $100 \cdot \|\mathbf{f}^* - \hat{\mathbf{f}}\|_2 / \|\mathbf{f}^*\|_2$.

The RMS error curve in the Fig. 2 shows that there is no considerable change in the error for the p-values ranging from 0.99 to 0.4. But when $p < 0.4$, the RMS error decreases drastically and

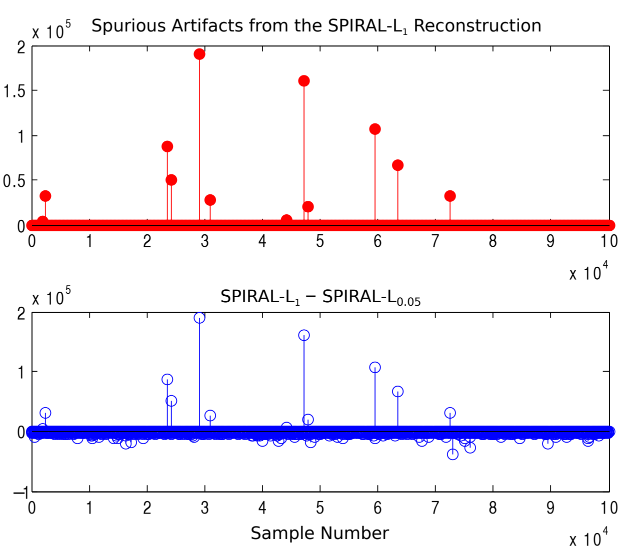

Fig. 4. The accuracy and the strength of the SPIRAL-$\ell_{0.05}$ reconstruction after thresholding at amplitude 7×10^5. Top: Absolute difference between the true signal and SPIRAL-ℓ_1 solutions. Note that there are 12 spurious solutions in the SPIRAL-ℓ_1 reconstruction. Bottom: Difference between SPIRAL-ℓ_1 and SPIRAL-$\ell_{0.05}$ solutions. The height of the positive stems reveals that there are no spurious solutions in the SPIRAL-$\ell_{0.05}$ reconstruction.

is less than the SPIRAL-ℓ_1 RMS error. Meantime, the number of nonzero entries of the reconstruction also converge to the exact sparsity as p value decreases. These results reveal that the SPIRAL-ℓ_p with p value ranging from 0.35 to 0 can generate better reconstruction than SPIRAL-ℓ_1 method. For instance, when $p = 0.05$, Figs. 3 and 4 depict the high accurate SPIRAL-ℓ_p reconstruction without the spurious solutions appear in the SPIRAL-ℓ_1 reconstruction. Furthermore, the SPIRAL-$\ell_{0.05}$ intensity reconstruction exaclty matches the sparsity of the true signal. In additional, we note that the amplitude of the SPIRAL-$\ell_{0.05}$ reconstruction is greater than the SPIRAL-ℓ_1 reconstruction (see Fig. 4).

While SPIRAL-ℓ_p generates high accurate, high strength reconstruction for small p-values, it requires more computational time than the SPIRAL-ℓ_1 method (see Fig. 5). More precisely, SPIRAL-ℓ_1 takes less than 1 second to obtain the reconstruction, while SPIRAL-ℓ_p takes on average, 66 seconds. The SPIRAL-ℓ_0 method requires low computational time (21 seconds) compared to other SPIRAL-ℓ_p methods because when $p = 0$, the GST function reduces to the hard-thresholding function (10), which has a closed-form solution.

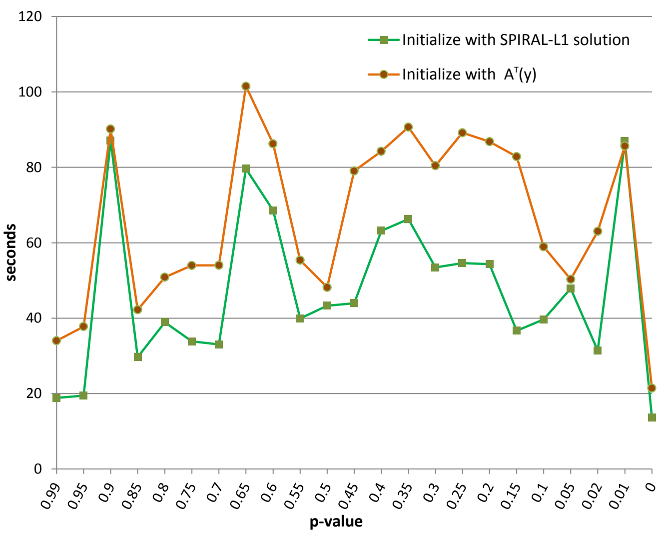

Fig. 5. Computation times of the SPIRAL-ℓ_p over p-values ranging from 0.99 to 0. The plot corresponding to SPIRAL-ℓ_p initialized using the SPIRAL-ℓ_1 solution lies below the plot corresponding to initializing with $\mathbf{A}^T(\mathbf{y})$ almost everywhere. Note in particular, that there is a significant computational time reduction from $p = 0.45$ to $p = 0.15$.

Initialization. For the same p values as in the previous experiment, we initialize the SPIRAL-ℓ_p using the solution of the SPIRAL-ℓ_1. The resulting reconstructions have similar RMS error values and the same number of nonzero entries as in the Fig. 2. However, initializing the SPIRAL-ℓ_p with the SPIRAL-ℓ_1 solution improves computational time. On average, we can obtain 30% improvement in computational time (see Fig. 5) if we solve the SPIRAL-ℓ_1 problem first and leverage its solution to initialize SPIRAL-ℓ_p.

Finally, we ran the proposed SPIRAL-ℓ_p method with $p = 0.05$ for ten different simulated measurement vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{10}$ with Poisson noise. Specifically, the Poisson noise levels in y_i's are around 16%, where noise $(\%) = 100 \cdot \|\mathbf{Af}^* - \mathbf{y}_i\|_2 / \|\mathbf{y}_i\|_2$. The resulting RMS error and the number of nonzeros for each of the final reconstruction are shown in the Table 1. In particular, we were able to recover the exact sparsity of the true signal in all ten different experiments with an average of 5.998% RMS error. Therefore, we conclude that for this experimental setup, the proposed SPIRAL-ℓ_p method is robust with respect to different Poisson noise realizations.

Experiment	RMSE (%)	Non-zeros
1	5.945	1500
2	5.947	1500
3	5.959	1500
4	5.991	1500
5	6.140	1500
6	6.077	1500
7	5.827	1500
8	5.955	1500
9	5.973	1500
10	6.162	1500
Average	5.998	1500

Table 1. RMS error and number of non-zeros in reconstructions using 10 different Poisson measurements. Here, RMSE (%) = $100 \cdot \|\mathbf{f}^* - \hat{\mathbf{f}}\|_2 / \|\mathbf{f}^*\|_2$.

Analysis. In the proposed SPIRAL-ℓ_p method, the p-values range from 0 to 1. But when p gets closer to 0, numerical issues arise. In particular, when $p = 0.02$, the objective function value in Eq. (1) is $O(10^{165})$. If we decrease p further, the objective function values become very large, which affect the steplength α_k in (3). While the SPIRAL-ℓ_p method still converges to a solution (using the difference between iterates as a termination criterion rather than a decrease in the objective function), the monotonic behavior of the objective function in the algorithm can no longer be enforced with arbitrarily small p-values.

4. CONCLUSION

In this paper, we have formulated the nonnegatively constrained sparse Poisson intensity reconstruction algorithm as a ℓ_p nonconvex regularized minimization problem (2). We have showed that this approach can be uncoupled into the separable ℓ_p-minimization problems in the form of (4), with each scalar minimization problem is solved using Generalized Soft-Thresholding (GST) function (9). We have demonstrated that the proposed SPIRAL-ℓ_p reconstruction for small p values eliminates the spurious artifacts found in the SPIRAL-ℓ_1 reconstruction. While the proposed method leads to more accurate and high strength reconstructions, it requires more computational effort because evaluating the GST function requires solving a zero-finding problem (8) iteratively. We have found that computational time can be decreased significantly by using the SPIRAL-ℓ_1 solution to initialize the SPIRAL-ℓ_p method.

5. REFERENCES

[1] D. L. Snyder and M. I. Miller, "Random point processes in space and time," *Springer-Verlag, New York, NY*, 1991.

[2] J. W. Goodman and J. F. Belsher, "Fundamental limitations in linear invariant restoration of atmospherically degraded images," *Proc. SPIE*, vol. 0075, pp. 141–154, 1976.

[3] J. A. Sorenson and M. E. Phelps, *Physics in Nuclear Medicine*, Grune & Stratton, 1980.

[4] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms - theory and practice," *IEEE Trans. Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[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] R. Chartrand, "Nonconvex compressed sensing and error correction," in *Proceedings of 2007 IEEE ICASSP*, Honolulu, Hawaii, April 2007.

[7] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, Dec 2013, pp. 217–224.

[8] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[9] J. A Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized em algorithms," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1417–1429, 1995.

[10] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.

[11] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "The Sparse Poisson Intensity Reconstruction ALgorithms (SPIRAL) Toolbox," `http://drz.ac/code/spiraltap/`.