LOGARITHMIC TOTAL VARIATION REGULARIZATION FOR CROSS-VALIDATION IN PHOTON-LIMITED IMAGING

Albert K. Oh[†] *Zachary T. Harmany*[⋆] *Rebecca M. Willett*[†]

[†] Duke University, Department of Electrical and Computer Engineering, Durham, NC
[⋆] University of Wisconsin–Madison, Department of Electrical and Computer Engineering, Madison, WI

ABSTRACT

In fields such as astronomy and medicine, many imaging modalities operate in the photon-limited realm because of the low photon counts available over a reasonable exposure time. Photon-limited observations are often modeled as the composite of a linear operator, such as a blur or tomographic projection, applied to a scene of interest, followed by Poisson noise draws for each pixel. One method to reconstruct the underlying scene intensity is to minimize a regularized Poisson negative log-likelihood, but choosing a good scaling parameter for the regularizer is notoriously difficult. This paper presents a new model that solves for and regularizes the *logarithm* of the true scene, and focuses on the special case of total variation regularization. This method yields considerable gains when used in conjunction with cross-validation for regularization parameter selection, where weighting of the regularization term is automatically determined using observed data.

Index Terms— Convex optimization, image denoising, image reconstruction, total variation, cross-validation

1. INTRODUCTION

In photon-limited imaging, which often arises in astronomy, medical imaging, microscopy, and night vision, an image can be represented using a Poisson model of a discrete number of photons hitting a detector for every pixel. Observations from this model [1] can be expressed as

$$y \sim \mathrm{Poisson}(Af^\star),$$

where $\mathrm{Poisson}(\cdot)$ draws from a Poisson distribution element-wise, $y \in \mathbb{Z}_+^m$ is an m-dimensional vector of observed photon counts, $f^\star \in \mathbb{R}_+^n$ is an n-dimensional vector of the ground truth image of interest (assumed to be piecewise smooth), and $A \in \mathbb{R}_+^{m\times n}$ is a linear projection of the truth to the observation set, such as an operation with an imaging system's point-spread function. From observations y, we are interested in recovering an estimate of $f^\star$.

One approach to this problem is to solve the following constrained optimization problem:

$$\begin{aligned} \hat{f}_\tau = &\arg\min_f \quad -\log\ p(y|Af) + \tau\|f\|_{\mathrm{TV}} \\ &\text{subject to} \quad f \geq 0, \end{aligned} \tag{1}$$

where $p(y|Af)$ is the likelihood of observing y if the underlying intensity were Af and $\|\cdot\|_{\mathrm{TV}}$ is the total variation (TV) semi-norm with tuning parameter $\tau \geq 0$ [2].

The total variation semi-norm is an operator that computes the magnitude of its argument's discrete gradient. It is used in image recovery because natural images tend to have sparse gradients, as they are often piecewise smooth with sparse edge pixels. The TV norm has been shown to yield good signal reconstruction in practice, but theoretical guarantees about its robustness have only recently been found [3]. There are different types of TV, a popular one being isotropic TV for images [4], which will be used in this paper. For an image $f \in \mathbb{R}_+^{n_1\times n_2}$ (note that this is the unvectorized version, where $n_1 n_2 = n$), it is defined as

$$\begin{aligned} \|f\|_{\mathrm{TV}} = &\sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1}\sqrt{(f_{i,j}-f_{i+1,j})^2+(f_{i,j}-f_{i,j+1})^2} \\ &+\sum_{i=1}^{n_1-1}|f_{i,n_2}-f_{i+1,n_2}|+\sum_{j=1}^{n_2-1}|f_{n_1,j}-f_{n_1,j+1}|. \end{aligned}$$

The best value of τ (in terms of either mean squared error or visual quality) differs depending on the data set, and it can be found empirically through trial-and-error parameter tuning, often involving manual choice of the perceived optimal value. One way to automate this process is through cross-validation (CV) via hold out, where the data is split into a training and validation set [5], [6]. For each candidate value of τ, an optimization problem is solved on the training set, then the resulting $\hat{f}_\tau$ is fit onto the validation set using the negative Poisson log-likelihood, ultimately choosing the value of τ which makes this measure smallest.

In the context of Poisson reconstruction problems, however, cross-validation has a significant pitfall. In particular, say there is a pixel that has zero counts in the training set and a nonzero number of counts in the validation set, which is a common scenario in low-light settings. Then for moderate values of τ, the estimate $\hat{f}_\tau$ may be zero in this location, and the value of the corresponding validation measure (the negative log-likelihood) will be infinite, even if $\hat{f}_\tau$ is an excellent estimate for every other pixel in the image. This value of τ will be discarded in favor of a larger τ which keeps all estimated pixel values away from zero by over-smoothing the reconstruction. The practical effect of this is a CV-selected τ which is far larger than what would minimize the root mean squared error (RMSE). We want to address this practical problem because it would allow for fully automated code to be provided to experimentalists, who do not have access to the underlying ground truth.

In this paper, we present an alternative to the optimization formulation described in (1), which (a) yields accurate and visually appealing results, and (b) facilitates cross-validation for selecting tuning parameters to address the pitfall discussed above. Essentially, instead of estimating $f^\star$ directly, its element-wise *logarithm* $x^\star \equiv \log(f^\star)$ is estimated and regularized via TV. This formulation yields a simple convex optimization problem in the denoising setting (i.e. where A is the identity matrix), and yields a more complex, nonconvex optimization problem for more general inverse problems. We describe approaches to solving both optimization problems.

2. EXPONENTIAL FORMULATION

Using the element-wise relationship $x = \log(f)$ so that $x \in \mathbb{R}^n$, the observation model from (1) can be rewritten as

$$y \sim \mathrm{Poisson}(A\exp(x^\star)),$$

where the exponentiation operator is applied element-wise. Our model contrasts from previous work where the observation model involves $\exp(Ax^\star)$ instead [7]. Note that if A were the identity matrix, both models would correspond to a generalized linear model [8].

In this Poisson model, the likelihood of observing y, assuming its elements are conditionally independent given x, is

$$p(y|A\exp(x)) = \prod_{i=1}^{m} \frac{(e_i^T A\exp(x))^{y_i}}{y_i!} \exp(-e_i^T A\exp(x)),$$

where e_i is the i^{th} unit vector in the standard basis, so the negative Poisson log-likelihood is proportional to

$$F(x) = \sum_{i=1}^{m} e_i^T A\exp(x) - y_i \ \log(e_i^T A\exp(x)),$$

where $\log(y_i!)$ is neglected because it is constant with respect to x. Then $x^\star$ can be estimated using the negative log-likelihood function and TV penalty

$$x_\tau = \arg\min_x \quad \Phi(x) \equiv F(x) + \tau\|x\|_{\mathrm{TV}},$$

which enforces TV regularization on the *logarithm* of the image of interest. This is related to previous work in regularized density estimation, with TV penalization on the logarithm of a density [9].

This negative log-likelihood function, however, is non-convex for general A. To handle this, we adopt the following approach: using the current iterate, compute an approximation of $F(x)$ which is used to formulate a convex optimization problem that can be solved to yield the next iterate. Known convex optimization techniques can then be used to find the minimizer of this convex surrogate.

The key to our approximation is to use a first-order Taylor series expansion on $h_i(x) \equiv \log(e_i^T A\exp(x))$ in the negative log-likelihood expression. The approximation of $h_i(x)$ centered about the j^{th} iterate x^j, denoted $h_i^j(x)$, is

$$\begin{aligned} h_i^j(x) &\equiv h_i(x^j) + \nabla h_i(x^j)^T(x - x^j) \qquad (2)\\ &= \left[h_i(x^j) - \nabla h_i(x^j)^T x^j\right] + \nabla h_i(x^j)^T x \\ &= u_i^j + e_i^T B^j x, \end{aligned}$$

where we define

$$B^j \equiv \mathrm{diag}(A\exp(x^j))^{-1} A\,\mathrm{diag}(\exp(x^j)) \qquad (3)$$
$$u^j \equiv \log(A\exp(x^j)) - B^j x^j, \qquad (4)$$

and where $\mathrm{diag}(\cdot)$ is a function that diagonalizes its vector input and $\log(\cdot)$ is an element-wise logarithm operator.

After the linear approximation, we compute the next iterate x_τ^{j+1} by solving the optimization problem

$$x_\tau^{j+1} = \arg\min_x \quad \Phi'(x) \equiv F_j'(x) + \tau\|x\|_{\mathrm{TV}}, \qquad (5)$$

where

$$F_j'(x) = \sum_{i=1}^{m} \exp(u_i^j + e_i^T B^j x) - y_i(u_i^j + e_i^T B^j x) \qquad (6)$$

is convex and has gradient

$$\nabla F_j'(x) = (B^j)^T \left[\exp(u + B^j x) - y\right].$$

Note that (6) is convex because the composition of a non-decreasing convex function (i.e. $\exp(\cdot)$) with any convex function (i.e. its linear argument) is convex [10]. Because the TV semi-norm is convex as well [4], this new objective function is therefore convex.

To solve the subproblem in (5), we use the FISTA algorithm [11]. In our setting, this algorithm iteratively solves for x_τ^{j+1} using a gradient descent step followed by a TV denoising step. With k as an iteration counter for this inner loop and $\tilde{x}^k$ as the subproblem's iterates, we can write

$$\begin{aligned} s^k &= w^k - \frac{1}{L}\nabla F_j'(w^k) \\ \tilde{x}^k &= \arg\min_x \frac{1}{2}\left\|x - s^k\right\|^2 + \frac{\tau}{L}\|x\|_{\mathrm{TV}}, \qquad (7) \end{aligned}$$

where w^k is an update variable introduced in FISTA and is specified in Algorithm 1, and L is an upper bound on the Lipschitz constant of $\nabla F'$ (and is also a step-size parameter in practice). The subproblem in (7) can be solved using the TV-based denoising and deblurring framework by Beck and Teboulle [4].

We refer to our approach as "log-TV reconstruction", which is summarized in Algorithm 1 below.

Algorithm 1 log-TV reconstruction

Initialize f^0. Set τ, A, L, and tolerance.
Set $j = 0$, $x^0 = \log(f^0)$
repeat
  Compute B^j and u^j via (3) and (4)
  $j \leftarrow j + 1$
  Set $k = 0$, $\tilde{x}^0 = x^j$, $t^0 = 1$, $w^1 = x^j$
  repeat
    $k \leftarrow k + 1$
    if using FISTA **then**
      $\tilde{x}^k \leftarrow$ solution of (7)
      $t^k = \frac{1+\sqrt{1+(4t^{k-1})^2}}{2}$
      $w^{k+1} = \tilde{x}^k + (\frac{t^{k-1}-1}{t^k})(\tilde{x}^k - \tilde{x}^{k-1})$
    else if using monotonic FISTA **then**
      $z^k \leftarrow$ solution of (7)
      $t^k = \frac{1+\sqrt{1+(4t^{k-1})^2}}{2}$
      $\tilde{x}^k = \arg\min\{\Phi'(q) : q = z^k, \tilde{x}^{k-1}\}$
      $w^{k+1} = \tilde{x}^k + (\frac{t^{k-1}}{t^k})(z^k - \tilde{x}^k) + (\frac{t^{k-1}-1}{t^k})(\tilde{x}^k - \tilde{x}^{k-1})$
    end if
  until $\|\tilde{x}^k - \tilde{x}^{k-1}\|_2 / \|\tilde{x}^k\|_2 \leq$ tolerance
  $x^j \leftarrow \tilde{x}^k$
until $\|x^j - x^{j-1}\|_2 / \|x^j\|_2 \leq$ tolerance
$\hat{f} = \exp(x^j)$

2.1. Stationary points

Although we are trying to solve a non-convex objective, the convex Taylor series approximation ensures stationarity of gradient-step iterations if at a local minimum. This is due to the fact that the gradient of our approximation matches the gradient of the original objective at the current iterate. To see this, note that using the notation of (2), we have that $h_i^j(x^j) = h_i(x^j)$ and $\nabla h_i^j(x^j) = \nabla h_i(x^j)$. Examining the gradients of the objective and our approximation, we respectively have

$$\nabla F(x) = \sum_{i=1}^{m} \left[\exp(h_i(x)) - y_i\right] \nabla h_i(x) \tag{8}$$

$$\nabla F_j'(x) = \sum_{i=1}^{m} \left[\exp(h_i^j(x)) - y_i\right] \nabla h_i^j(x), \tag{9}$$

and therefore our approximation guarantees $\nabla F(x^j) = \nabla F_j'(x^j)$. If we are at a stationary point for the original problem, we will have $0 \in \nabla F(x^j) + \tau\partial\|x^j\|_{\text{TV}}$, and because (8) and the approximation in (9) are equivalent at $x = x^j$, we have that $0 \in \nabla F_j'(x^j) + \tau\partial\|x^j\|_{\text{TV}}$. Hence, the point x^j also optimizes the subproblem with the approximate objective, thus guaranteeing stationarity of our method.

2.2. Special case: denoising

When the linear projection A is the identity matrix, our problem reduces to a Poisson denoising problem. In this special case, $u = 0$ and $B = I_n$, meaning the outer Taylor series expansion loop is unnecessary because $F(x) = F'(x)$. This denoising problem is convex, so only the FISTA inner loop is needed to solve for x.

3. CROSS-VALIDATION (CV)

Performance depends heavily on the choice of τ, and selecting this value accurately is important. τ can be found by manually tuning parameters and adjusting based on trial-and-error, but it is more practical to select this automatically using cross-validation via hold out [5], [6]. In this setup, we assume that each photon count is independent of all others, so a new model can be written where a certain fraction p of the observations are used as a training set y^{train}, and the remaining fraction $1 - p$ are used as a validation set y^{val}:

$$y^{\text{train}} \sim \text{Poisson}(pAf)$$
$$y^{\text{val}} \sim \text{Poisson}((1-p)Af).$$

To separate photons into these two groups, we take the original set of observations and draw training data from an element-wise Binomial distribution so that $y^{\text{train}} \sim \text{Binomial}(y, p)$, where the validation set is the complement $y^{\text{val}} = y - y^{\text{train}}$. For a collection of test values T for τ, we can use the training set to compute

$$\hat{f}_\tau = \arg\min_f -\log\ p(y^{\text{train}}|pAf) + \tau\|f\|_{\text{TV}}.$$

For each $\tau \in T$, the negative Poisson log-likelihood of the validation set is calculated, and the τ corresponding to the minimum is determined to be the best tuning parameter

$$\hat{\tau} = \arg\min_{\tau\in T} -\log\ p(y^{\text{val}}|(1-p)A\hat{f}_\tau).$$

The final estimate is then $\hat{f}^{\text{TV}} := \hat{f}_{\hat{\tau}}$. Similarly, we can perform cross-validation using log-TV regularization via:

$$\hat{x}_\tau = \arg\min_x -\log\ p(y^{\text{train}}|pA\exp(x)) + \tau\|x\|_{\text{TV}}$$
$$\hat{\tau} = \arg\min_{\tau\in T} -\log\ p(y^{\text{val}}|(1-p)A\exp(\hat{x}_\tau))$$
$$\hat{f}^{\log\text{TV}} := \exp(\hat{x}_{\hat{\tau}}).$$

4. EXPERIMENTAL RESULTS

To demonstrate the performance of our algorithm, we used the Shepp-Logan head phantom available in Matlab's Image Processing Toolbox; the grayscale image used had size 128×128, scaled to have a mean 4.432 photons per pixel. Performance was evaluated visually and by the root mean squared error (RMSE) percentage given by $100 \times \|\hat{f} - f^\star\|_2/\|f^\star\|_2$; note that the truth $f^\star$ was known in these simulations for the sake of evaluation, but would be unknown in a real setting.

We compare our log-TV method with TV reconstruction described in [2] that solves (1), where both methods were run until the relative difference between consecutive iterates converged to $\|x^j - x^{j-1}\|_2/\|x^j\|_2 \leq 10^{-5}$ and $\|f^j - f^{j-1}\|_2/\|f^j\|_2 \leq 10^{-5}$, respectively. The methods were initialized with $x^0 = \log(A^T y)$ and $f^0 = A^T y$, respectively.

4.1. Deblurring

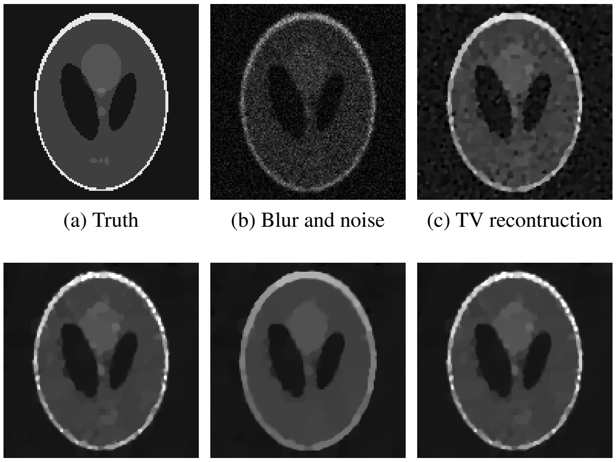
Results for the deblurring case can be seen in Figure 1, where all of the reconstructions are from the same Poisson noise realization and are displayed on the same intensity scale as the truth. The truth was convolved with a Gaussian blur matrix and was subjected to Poisson noise, and these resulting observations had a mean value of 4.426 photons per pixel. The log-TV reconstruction used the monotonic version of FISTA for each convex subproblem with $L = 100$, and the TV reconstruction was performed enforcing near-monotonic iterations [12].

TV and log-TV reconstructions were used to estimate $\hat{f}$ clairvoyantly by manually choosing values of τ to minimize the RMSE. It can be seen that log-TV yields a reconstruction that more accurately matches the true intensities. The TV reconstruction also has more noise artifacts in both the background and foreground, even estimating some pixels within the main body as having small value. Cross-validation (with $p = 0.9$ and the candidate collection of τ values considered spaced 0.02 apart for TV and 0.001 apart for log-TV) shows that TV offers a poorer reconstruction that is very over-smoothed; log-TV with CV, on the other hand, gives a better estimate that is comparable to the clairvoyant log-TV result. The RMSE averaged over 50 different Poisson noise realizations for these comparisons can be seen in Table 1, where it is seen that log-TV has less error than TV.

The huge practical benefits of log-TV reconstruction can be seen in the cross-validation behavior. As discussed in Section 3, CV selects the optimal τ value based on the minimum negative Poisson log-likelihood from various validation sets. A more accurate method would be to select the value that minimizes the RMSE, but this cannot be the criteria in a real setting because the truth is unknown. Figure 2 shows that for TV reconstruction, there is a large difference between the τ chosen via the validation set negative log-likelihood and via the training set reconstruction RMSE known in simulation, while the difference is much smaller for log-TV. This highlights the advantage that log-TV reconstruction with CV has much less over-smoothing because of its smaller τ value choice.

	TV reconstruction	log-TV reconstruction
Tuning τ	31.66%	28.48%
CV τ	43.96%	28.49%

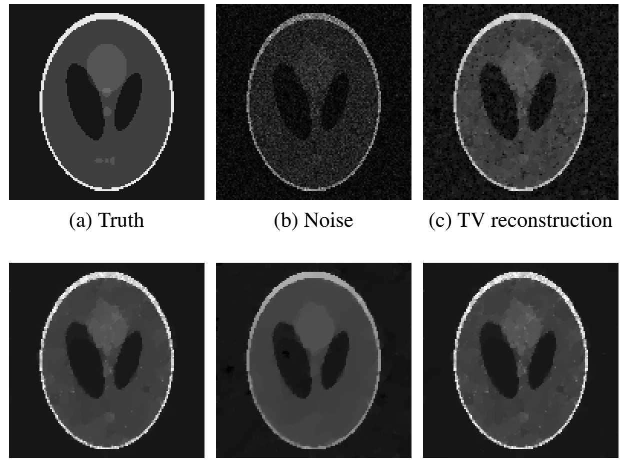
Table 1: RMSE comparison for deblurring, averaged over 50 trials



(a) Truth (b) Blur and noise (c) TV recontruction

(d) log-TV reconstruction (e) TV with CV (f) log-TV with CV

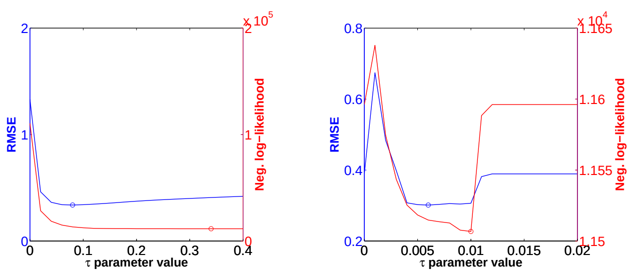
Fig. 1: Deblurring results



(a) Truth (b) Noise (c) TV reconstruction

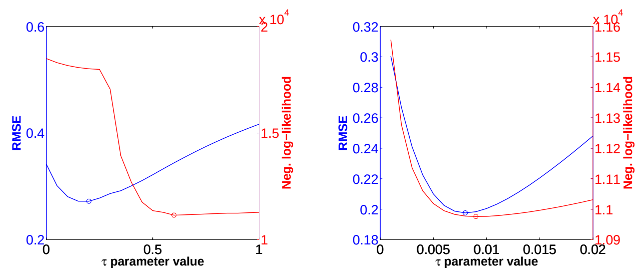
(d) log-TV reconstruction (e) TV with CV (f) log-TV with CV

Fig. 3: Denoising results



(a) TV reconstruction (b) log-TV reconstruction

Fig. 2: Cross-validation behavior (deblurring)



(a) TV reconstruction (b) log-TV reconstruction

Fig. 4: Cross-validation behavior (denoising)

4.2. Denoising

Denoising simulation results can be seen in Figure 3, where the reconstructions are from the same Poisson noise realization and are shown on the same intensity scales. After the truth was subjected to Poisson noise, the observation mean was 4.450 photons per pixel. The log-TV method used non-monotonic FISTA iterations for each convex subproblem with $L = 200$, and TV reconstruction was performed enforcing near-monotonic iterations [12].

In the manual parameter tuning case, log-TV performed better than TV, yielding a smoother reconstruction that matched true intensities better. For TV, noise artifacts are seen in the background and certain pixels within the body are estimated to be near-zero. For cross-validation (with $p = 0.9$ and the candidate τ values considered spaced 0.05 apart for TV and 0.001 apart for log-TV), TV gives an over-smoothed reconstruction due to a large selected parameter value; its intensities are much lower than that of the truth, especially around the outer edge of the phantom. The log-TV reconstruction is also comparable to its clairvoyantly-chosen counterpart, which cannot be said about TV. Average RMSEs for all cases can be seen in Table 2, where it is seen that log-TV empirically performs better.

Cross-validation again shows the practical benefits of this new formulation. As seen in Figure 4, there is a large difference between the τ value chosen via the validation set negative log-likelihood for TV and that of the training set reconstruction RMSE; log-TV reconstruction chose a more reasonable τ that avoids over-smoothing.

	TV reconstruction	log-TV reconstruction
Tuning τ	21.69%	14.87%
CV τ	29.30%	14.98%

Table 2: RMSE comparison for denoising, averaged over 50 trials

5. CONCLUSION

We have presented a new formulation for photon-limited imaging inverse problems involving logarithmic TV regularization. This method has been shown to perform better than its untransformed counterpart in visual quality and empirical error of reconstructions. We have also shown significant improvements in selecting optimal regularization parameters via cross-validation, where this new formulation prevents over-smoothing of reconstructions. This has beneficial implications in experimental settings, where the truth is unknown and must be accurately estimated. This work has further extensions beyond TV and can be extended to other convex regularizations, such as ℓ_1 and learned dictionaries (*e.g.*, [13]), and to more general frameworks not limited to images.

6. REFERENCES

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

[2] R. M. Willett, Z. T. Harmany, and R. F. Marcia, "Poisson image reconstruction with total variation regularization," in *Proc. IEEE International Conference on Image Processing*, 2010.

[3] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," arXiv:1202.6429, 2012.

[4] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, November 2009.

[5] Z. T. Harmany, *Computational Optical Imaging Systems: Sensing Strategies, Optimization Methods, and Performance Bounds*, Ph.D. thesis, Duke University, Durham, NC, 2012.

[6] P. L. Bartlett, S. Boucheron, and G. Lugosi, "Model selection and error estimation," *Machine Learning*, vol. 48, pp. 85–113, 2002.

[7] H. Erdŏgan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Transactions on Medical Imaging*, vol. 18, no. 9, pp. 801–814, September 1999.

[8] A. J. Dobson, *Introduction to Generalized Linear Models*, Chapman and Hall/CRC, 2nd edition, 1945.

[9] R. Koenker and I. Mizera, "Density estimation by total variation regularization," in *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A Doksum*. 2007, pp. 613–634, World Scientific.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[11] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[12] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms – Theory and Practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, March 2012.

[13] J. Salmon, Z. Harmany, C. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," submitted to *Journal of Mathematical Imaging and Vision*. arXiv:1206:0338, 2012.