PERFORMANCE BOUNDS FOR TRACTABLE POISSON DENOISERS WITH PRINCIPLED PARAMETER TUNING

Chinmay Talegaonkar

Dept. of EE, IIT Bombay,
chinmay0301@iitb.ac.in

Ajit Rajwade[*]

Dept. of CSE, IIT Bombay,
ajitvr@cse.iitb.ac.in

ABSTRACT

We present an algorithm for image denoising under Poisson noise using the theory of variance stabilization transforms. We derive worst-case performance bounds for our algorithm. Our proposed estimator allows for easy and very principled parameter tuning unlike existing approaches which require specification of signal dependent parameters. Moreover our estimator is computationally tractable. We also demonstrate numerical results on image denoising under Poisson noise to support the theoretical results.

***Index Terms*—** Poisson noise, variance stabilization transform, performance bounds in denoising

1. INTRODUCTION

It is well known that the noise affecting images acquired in an optical or X-Ray based system is dominantly Poisson distributed [1, 2]. That is, the observed image intensity y_i at pixel i is given as $y_i \sim \text{Poisson}(x_i)$ where x_i is the intensity of the underlying 'true' image at pixel i. There is the implicit assumption that y_i is a non-negative integer and $x_i \geq 0$. The Poisson noise model is signal dependent. For images affected by Poisson noise, the restoration algorithms should ideally account for this noise model. A plethora of algorithms have been developed for Poisson image denoising [3, 4, 5, 6, 7, 8]. The work in [3, 4] exploits variance stabilizing transforms (VSTs) in different ways, in conjunction with image sparsity priors. Innovative sparsity priors have been developed in [5]. On the other hand, the work in [6, 7, 8] is based on different forms of dictionary learning such as PCA adapted to Poisson noise, variants of the KSVD dictionary learning algorithm [9], or non-negative sparse coding [10].

However, relatively less work has been done on developing theoretical performance bounds for the task of estimating $\{x_i\}_{i=1}^n$ from $\{y_i\}_{i=1}^n$. A notable exception is the work in [11], which estimates the underlying signal using a dictionary-based matrix factorization of the form $\boldsymbol{X} = \boldsymbol{D}\boldsymbol{\Theta}$ where $\boldsymbol{X} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ is the unknown signal matrix, $\boldsymbol{D} \in \mathbb{R}^{n_1 \times K}$ is a dictionary with K columns and $\boldsymbol{\Theta} \in \mathbb{R}^{K \times n_2}$ is a matrix of signal coefficients in $\boldsymbol{D}$, on which sparsity is imposed. The specific estimator used is of the form $\text{argmin}_{\boldsymbol{D},\boldsymbol{\Theta}} - \log p(\boldsymbol{Y}|\boldsymbol{D}\boldsymbol{\Theta}) + \lambda\|\boldsymbol{\Theta}\|_0$. Here the first term is the negative log-likelihood of the Poisson distribution and the second term is a coding-theoretic regularizer. The aforementioned estimator is not computationally tractable due to the $\|\boldsymbol{\Theta}\|_0$ term. While in practice, this term can be softened to $\|\boldsymbol{\Theta}\|_1$ to yield a tractable estimator as in [12], there are no guiding principles on a signal-independent choice of the regularizer parameter λ.

Contributions: Our work in this paper presents the following contributions:

1. We present a constrained but tractable optimization algorithm for Poisson denoising based on VSTs.
2. We present theoretical performance bounds for this algorithm. Our algorithm requires no knowledge of the underlying signal $\boldsymbol{x}$, except for $\|\boldsymbol{x}\|_\infty$ (i.e. maximum rate of the underlying Poisson process) which is needed solely for the theoretical analysis.

Paper Organization: This paper is organized as follows. The estimator is presented in Section 2, and its performance bounds are presented in Section 3, followed by empirical results in Section 5 and a conclusion in Section 6.

2. PROPOSED ESTIMATOR

Consider noisy intensity $y_i \sim \text{Poisson}(x_i)$ at pixel location i. It is well known that $\sqrt{y_i + c}$ is approximately Gaussian distributed with mean $\sqrt{x_i + c}$ and variance $1/4$ where c is a non-negative constant [13]. Moreover the quality of the approximation improves as $x_i \to \infty$. This result is called the VST, specifically the Anscombe transform when $c \triangleq 3/8$ [14]. Let $\boldsymbol{y}$ and $\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\theta}$ respectively be the noisy and original images in vectorized form, where $\boldsymbol{\Psi}$ is a sparsifying orthonormal basis and $\boldsymbol{\theta}$ is a vector of sparse/compressible coefficients. Inspired by the VST, we propose the following estimator:

$$\text{Q1 :min } \|\boldsymbol{\theta}\|_1 \text{ s. t. } \|\sqrt{\boldsymbol{y} + c} - \sqrt{\boldsymbol{\Psi}\boldsymbol{\theta} + c}\|_2 \leq \varepsilon, \boldsymbol{\Psi}\boldsymbol{\theta} \succeq \boldsymbol{0}, \quad \|\boldsymbol{\Psi}\boldsymbol{\theta}\|_\infty \leq L. \tag{1}$$

[*]AR thanks IIT-B Seed Grant 14IRCCSG012

Here L is an upper bound on the maximum pixel intensity value of the unknown image, and ε is a statistically motivated bound on the magnitude of the residual vector $r(\boldsymbol{y}, \boldsymbol{x}) \triangleq \sqrt{\boldsymbol{y}+c} - \sqrt{\boldsymbol{\Psi\theta}+c}$. The choice of ε is based on the following theorem, proved in our unpublished work [15] (see Theorem 1).

Theorem 1: If $\boldsymbol{y}$ is a vector of n Poisson corrupted values with underlying intensity in $\boldsymbol{x}$ such that $\forall i, y_i \sim \text{Poisson}(x_i)$, then we have the following:

1. $\mathbb{E}[\|r(\boldsymbol{y}, \boldsymbol{x})\|_2] \leq \sqrt{n/2}$.
2. Define $v \triangleq \text{Var}[\|r(\boldsymbol{y}, \boldsymbol{x})\|_2]$. If $\forall i, x_i \geq 1$ and $n \geq 29$, then we have $v \leq \frac{3n/4}{n(2c+1)/(8(1+c)^2)} \approx 6.48$ and $\mathbb{P}\Big(\|r(\boldsymbol{y}, \boldsymbol{x})\|_2 \leq \sqrt{n}(\frac{1}{\sqrt{2}} + 2.55)\Big) \geq 1 - 1/n$. ■

The complete proof is omitted here owing to lack of space. The main advantage of this theorem is that it derives a bound on the residual magnitude which is (a) signal-independent, and (b) does not require the residual vector to be Gaussian distributed. A bound based on the Gaussian approximation would not be rigorous enough, because the distribution is only *asymptotically* Gaussian when the elements of $\boldsymbol{x}$ have infinitely large values. Instead, our tail bound is based on Chebyshev's inequality.

Poisson denoising based on VSTs is not new, however existing techniques such as [8, 3] denoise the Anscombe transform of the noisy image using Gaussian denoising techniques, followed by inversion of the Anscombe transform in one of many different ways to remove statistical bias [3]. From a rigorous theoretical standpoint such an approach is not fully accurate, since the elements of the Anscombe transformed image are only *approximately* Gaussian distributed. Moreover proving performance bounds for this method is quite challenging, and has not been attempted so far. As against this, our approach *directly* solves for the underlying image (without any further inversion step) as is evident from the estimator in Eqn. 2. Our approach is partly similar to that in [4], where a regularized estimator of the following form is used:

$$\text{Q2 :min } \lambda\|\boldsymbol{\theta}\|_1 + \|\sqrt{\boldsymbol{y}+c} - \sqrt{\boldsymbol{\Psi\theta}+c}\|_2^2, \text{ s. t. } \boldsymbol{\Psi\theta} \succeq \boldsymbol{0}. \tag{2}$$

For Q2, no performance bounds are derived in [4], and there is no theoretical treatment for choice of λ, which is instead picked by cross-validation. It should be noted that Q1 (just like Q2) is a convex estimator as the objective function as well as constraints are all convex. The constraint on $\|r(\boldsymbol{y}, \boldsymbol{x})\|_2$ is convex as it represents the sub-level set of a convex set (see Section 3.1.6 of [16]). Moreover it has bounded derivatives even at low intensity levels due to the factor $c \neq 0$ (unlike the negative log-likelihood of the Poisson distribution). Hence Q1 is a tractable estimator.

3. PERFORMANCE BOUNDS FOR OUR ESTIMATOR

In this section, we put forth our key theorem regarding performance of the estimator Q1.

Theorem 2: Consider $\boldsymbol{y} \sim \text{Poisson}(\boldsymbol{x})$ where $\boldsymbol{x} = \boldsymbol{\Psi\theta} \in \mathbb{R}^n_{\geq 0}$ as defined in Theorem 1, and $\boldsymbol{y} \in \mathbb{Z}^n$. Let $\|\boldsymbol{\Psi\theta}\|_\infty \leq L$, $\|\boldsymbol{\Psi\theta}\|_1 = I$. Let $\boldsymbol{\theta^\star}$ be the result of the optimization problem in Q1 with ε set to $\sqrt{n}(\frac{1}{\sqrt{2}} + 2.55/\kappa)$ for some $\kappa > 0$. Then, if $n \geq 29$ and $\forall i, x_i \geq 1$, we have the following inequality with probability exceeding $1 - 1/n$ for any integer $s \leq \mathcal{O}(n/\log n)$:

$$\frac{\|\boldsymbol{\theta} - \boldsymbol{\theta^\star}\|_2}{I} \leq 16(2.55/\kappa + 1/\sqrt{2})\frac{\sqrt{n(L+c)}}{I} + \frac{2s^{-\frac{1}{2}}\|\boldsymbol{\theta} - \boldsymbol{\theta_s}\|_1}{I},$$

where $\boldsymbol{\theta_s}$ is the best s-sparse approximation to $\boldsymbol{\theta}$. ■

Comments on Theorem 2: Even though it is required for the theoretical proof, the constraint $\|\boldsymbol{\Psi\theta}\|_\infty \leq L$ is not required in practice in the numerical experiments with Q1 (see Section 5). The bounds presented in this theorem scale as $\mathcal{O}(\sqrt{n})$. Those in [11] scale as $\mathcal{O}(\sqrt{(n+s)\log n})$, where the (i) $\mathcal{O}(\sqrt{n \log n})$ factor is due to the fact that the dictionary matrix $\boldsymbol{\Psi}$ is also inferred from the noisy data, and (ii) the $\mathcal{O}(\sqrt{s \log n})$ factor is because of the employment of a LASSO-like estimator [17] (see Theorem 11.1 and Eqn. 11.15 and comments on it) in constrast to our estimator that is similar to basis pursuit denoising [18].

4. PROOF OF THEOREM 2

The proof presented here for denoising error bounds using the residual magnitude estimator is inspired from the proof presented in [15] for the task of Poisson compressive reconstruction with a sensing matrix $\boldsymbol{\Phi}$. The proof is adapted here meticulously for the case of Poisson denoising where $\boldsymbol{\Phi}$ is the $n \times n$ identity matrix.

1. Define a vector $\boldsymbol{h} \triangleq \boldsymbol{\theta} - \boldsymbol{\theta^\star}$. The vector denoted by $\boldsymbol{h}_T$ is equal to $\boldsymbol{h}$ only for index set T and zero for other indices. Define T_0 as the set having s largest absolute value indices of $\boldsymbol{\theta}$, similarly T_1 is the set containing s largest absolute value indices of $h_{T_0^c}$ and so on, where T^c is the complement of the set T. Hence $\boldsymbol{h}$ can be broken down as the sum of $\boldsymbol{h}_{T0}$,$\boldsymbol{h}_{T_1}$,$\boldsymbol{h}_{T_2}$,...
2. Following simple algebraic manipulation we have

$$\begin{aligned}\|\boldsymbol{\Psi h}\|_2^2 &= \|\boldsymbol{\Psi}(\boldsymbol{\theta} - \boldsymbol{\theta^\star})\|_2^2 \\ &= \Sigma_{i=1}^N((\sqrt{(\boldsymbol{\Psi\theta})_i + c} - \sqrt{(\boldsymbol{\Psi\theta^\star})_i + c})^2 \\ &\quad (\sqrt{(\boldsymbol{\Psi\theta})_i + c} + \sqrt{(\boldsymbol{\Psi\theta^\star})_i + c})^2).\end{aligned} \tag{3}$$

   (a) Consider an upper bound of ε on $\|\sqrt{\boldsymbol{y}+c} - \sqrt{\boldsymbol{\Psi\theta}+c}\|_2$. At the end of the proof, we shall assign a statistical meaning to ε based on Theorem 1. Based on the triangle inequality and the constraint in Q1,

we get $\|\sqrt{\boldsymbol{\Psi\theta}+c}-\sqrt{\boldsymbol{\Psi\theta^\star}+c}\|_2 \leq \|\sqrt{\boldsymbol{y}+c}-\sqrt{\boldsymbol{\Psi\theta}+c}\|_2 + \|\sqrt{\boldsymbol{y}+c}-\sqrt{\boldsymbol{\Psi\theta^\star}+c}\|_2 \leq 2\varepsilon$.

(b) For scalars $v_1 \geq 0, v_2 \geq 0$, we have $(\sqrt{v_1}+\sqrt{v_2})^2 \leq 4\text{max}(v_1, v_2)$. We also have $(\boldsymbol{\Psi\theta})_i = x_i \leq L$. Likewise we also have $(\boldsymbol{\Psi\theta^\star})_i \leq \|\boldsymbol{x^\star}\|_\infty = L$ based on the constraints defined in Q1. Hence we obtain $(\sqrt{(\boldsymbol{\Psi\theta})_i+c}+\sqrt{(\boldsymbol{\Psi\theta^\star})_i+c})^2 \leq 4(L+c)$.

(c) Using the earlier two results with Step 2(a), we have $\|\boldsymbol{\Psi h}\|_2 \leq 4\varepsilon\sqrt{L+c}$.

3. As shown in [15] to prove the bound on $\|\mathbf{h}_{(T_0\cup T_1)^c}\|_2$, on following steps similar to [18] we get

$$\|\mathbf{h}_{(T_0\cup T_1)^c}\|_2 \leq \|\mathbf{h}_{(T_0)}\|_2 + 2s^{-1/2}\|\boldsymbol{\theta}-\boldsymbol{\theta_s}\|_1. \quad (4)$$

4. The following construct was used to prove error bounds on $\|\mathbf{h}_{(T_0\cup T_1)}\|_2$ based on steps in [18]. Consider the restricted isometry property (RIP) of the identity matrix with a restricted isometry constant (RIC) $\delta_{2s} = 0$ for order s such that $n \geq \mathcal{O}(s\log n)$. Using this and the Cauchy-Schwartz inequality, we can prove that

$$\|\boldsymbol{h}_{T_0\cup T_1}\|_2 \leq C'\varepsilon + C''s^{-1/2}\|\boldsymbol{\theta}-\boldsymbol{\theta^\star}\|_1 \quad (5)$$

where $C' \triangleq \frac{8\sqrt{(1+\delta_{2s})(L+c)}}{1-\delta_{2s}(\sqrt{2}+1)}$ and $C'' \triangleq \frac{2\sqrt{2}\delta_{2s}}{1-\delta_{2s}(\sqrt{2}+1)}$. Since $\delta_{2s} = 0$ for an identity matrix, $C' = 8\sqrt{L+c}$ and $C'' = 0$. Putting these values in the above equation we get

$$\|\boldsymbol{h}_{T_0\cup T_1}\|_2 \leq 8\varepsilon\sqrt{L+c}. \quad (6)$$

5. Combining the bounds on $\|\boldsymbol{h}_{T_0\cup T_1}\|_2$ and $\|\boldsymbol{h}_{T_0\cup T_1{}^c}\|_2$, we have

$$\|\boldsymbol{h}\|_2 \leq 16\varepsilon\sqrt{L+c} + C_2\sqrt{2}\|\boldsymbol{\theta}-\boldsymbol{\theta_s}\|_1 \quad (7)$$

where $C_2 \triangleq 2 + 2C'' = 2$ (as $C'' = 0$ for this case).

Finally, we divide by I to obtain upper bounds on the relative reconstruction error:

$$\frac{\|\boldsymbol{\theta}-\boldsymbol{\theta^\star}\|_2}{I} \leq 16\varepsilon\frac{\sqrt{L+c}}{I} + \frac{2s^{-\frac{1}{2}}\|\boldsymbol{\theta}-\boldsymbol{\theta_s}\|_1}{I}. \quad (8)$$

Using Theorem 1, we see that $\varepsilon \leq \sqrt{n}(2.55/\kappa + 1/\sqrt{2})$ with a probability of $1-\kappa^2/n$ for any $\kappa > 0$. This proves Theorem 2. ■

5. EXPERIMENTAL RESULTS

5.1. Experiments on Sparse 1D Signals

The denoising experiments were first performed on non-negative 1D Poisson corrupted signals with 250 elements each. Each signal was generated from a sparse random linear combination of DCT basis vectors, with the DC component adjusted to yield non-negative signals. Different signals had different supports in the DCT basis, and the coefficient values were independently drawn from Unif$[0, 1]$. Corresponding to each Total Intensity ($I \triangleq \|\boldsymbol{x}\|_1$) and sparsity level s, we generated 20 signals. Post denoising by one of different competing methods (see later in this section), the relative mean squared reconstruction error (RMSE = $\|\boldsymbol{x_{true}} - \boldsymbol{x_{est}}\|_2/\|\boldsymbol{x_{true}}\|_2$) was recorded for each of these signals, and a median error was recorded. The signals were scaled suitably so that they had a desired Total Intensity I. In Fig. 1, we show plots of median RMSE vs Total Intensity at a fixed sparsity level $s = 50\%$, and likewise for RMSE vs sparsity at a fixed Total Intensity $I = 10000$. The box-plots for these are provided in the supplemental material [19].

Methods Compared:

1. Q1 using CVX package with $c = \frac{3}{8}$, $\varepsilon = 0.2\sqrt{n}$ and without the $\|\boldsymbol{\Psi\theta}\|_\infty \leq L$ constraint (as its inclusion made very little difference to the results).
2. Penalized negative log-likelihood (PNLL) using the method 'SPIRAL-TAP' from [12] with an omniscient choice of λ for every I value, i.e. the optimal value of λ (in the RMSE sense, assuming full knowledge of the underlying image) for every I was selected by brute force from a set $\mathcal{L}$ of 100 evenly spaced candidate values of λ in the range $[0.01, 1]$.
3. PNLL with λ obtained using cross-validation, i.e. the optimal value of λ was obtained by brute force from the set $\mathcal{L}$ defined earlier, for an image with a fixed I and s value, and then the same value of λ was used for images with other values of I.

Observations and Comments: From the plots in Fig. 1, we see that the RMSE decreases with I for the Q1 estimator (as for other estimators too). For Q1, the plots show relative insensitivity to s (the RMSE varies from 0.012 at low s to 0.024 for higher s), because the intensity term dominates the bound (see Theorem 2) and we are considering $I = 10^4$, which is large in value. At low I, the omniscient PNLL approach outperforms Q1 with $\varepsilon = 0.2\sqrt{n}$ (which is a fixed, statistically motivated choice) although at higher I their performance is almost the same. The cross-validation based PNLL approach suffers in performance at low as well as high I. This experiment therefore highlights the sensitivity of PNLL based methods to choice of the regularization parameter λ, which is in fact dependent on properties of the underlying signal. The cross validation approach ensures good performance only close to the intensity level on which the cross-validation was performed, while for other I the performance is adversely affected. In other words, PNLL outperforms Q1 (with a fixed $\varepsilon = 0.2\sqrt{n}$) only when the λ value for the former is picked in an intensity-dependent fashion. Further analysis of box plots at various intensities and sparsities is provided in the supplemental material[19].

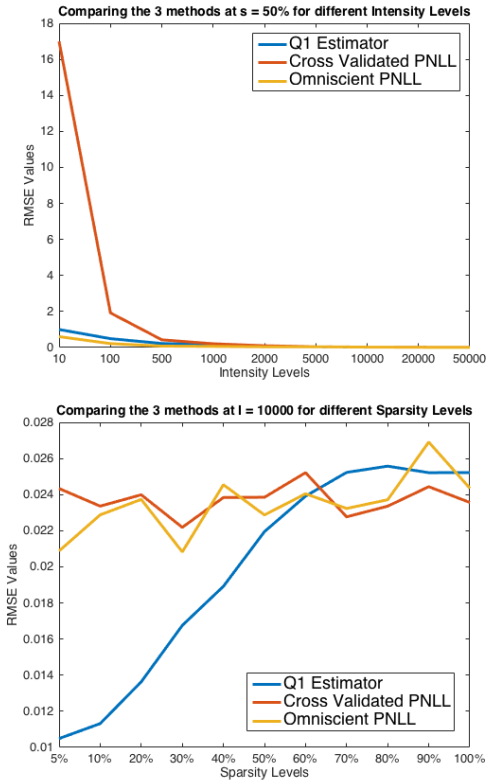

Fig. 1: Median RMSE plots for Q1, PNLL with omniscient λ and PNLL with cross validated λ. Top plot: (a) Across different values of I with $s = 0.5n$, (b) Across different values of s with $I = 10^4$.

5.2. Image Denoising Experiments

We also performed denoising experiments for two different images: Barbara and House. Our estimator Q1 was tested on portions of size 128×128 from the images at 3 different intensity levels. We used an overlapping patch based denoising approach, with sliding window averaging to avoid patch-seam artifacts. We denoised 8×8-sized image patches, and used a window stride of 2 pixels. Reconstruction RMSE for both images, each at different intensity levels given by fI_0 where $0 < f \leq 1$ and $I_0 \triangleq \|\boldsymbol{x}\|_1/n$, are reported in Table 1. I_0 values for the Barbara and House image are 127.89 and 145.837 respectively. Denoising results are shown in Fig. 2. These show the performance of Q1 on actual image. The results improve with increase in I as expected intuitively, and predicted from the term $\sqrt{L}/I$ in Theorem 2. We compared the performance of Q1 with that of PNLL for the same patch-size and window stride. For PNLL, the λ value was picked omnisciently from a global denoising routine using SPIRAL-TAP, and then used for patch-based denoising with the SPIRAL-TAP method. Results for another image - a collage of textures - are shown in [19].

Fig. 2: The rows corresponds to Intensities $0.4I_0$ (row 1 and 3) and $0.1I_0$ (row 2 and 4). I_0 denotes the average value per pixel of the original image which is then Poisson corrupted after scaling to a desired intensity level.

Table 1: Tabulated RMSE values

	Barbara ($0.1I_0$)	Barbara ($0.4I_0$)	House ($0.4I_0$)	House ($0.1I_0$)
Q1	5.72%	1.35%	1.80%	5.06%
PNLL	11.89%	2.40%	4.17%	13.98%
Noisy	25.11%	8.30%	9.72%	21.19%

6. CONCLUSION

We have presented an algorithm for Poisson image denoising inspired from the Anscombe transform. The proposed estimator is computationally tractable and we have also proved performance bounds for it with principled choice for ε. This is unlike earlier denoising estimators that are either not tractable, or for which provable bounds do not exist, or for which principled choices of estimator parameters have not been developed. A notable exception is the work in [20] which uses a weighted LASSO or weighted group LASSO formulation for denoising. However the image model used there is of the form $\boldsymbol{x} = \exp(\boldsymbol{\Psi\theta})$, unlike the linear model $\boldsymbol{x} = \boldsymbol{\Psi\theta}$ considered here. The numerical results of our estimator Q1 are promising and comparable to negative log-likelihood based estimators. There are many directions for future work. First, our framework could be extended to jointly denoise groups of similar patches which could yield superior performance. Second, the prospect of deriving performance bounds in conjunction with tractable inference of $\boldsymbol{\Psi}$ as well as $\boldsymbol{\theta}$ *in situ* from the noisy data is also very interesting.

7. REFERENCES

[1] H. Trussell and R. Zhang, "The dominance of poisson noise in color digital cameras," in *ICIP*, 2012.

[2] J. Fessler, "Lecture notes: X-ray imaging: noise and SNR," https://web.eecs.umich.edu/~fessler/course/516/l/c6-noise.pdf.

[3] M. Makitalo and A. Foi, "Optimal inversion of the generalized anscombe transformation for Poisson-gaussian noise," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 91–103, Jan 2013.

[4] F.-X. Dupé, J. Fadili, and J.-L. Starck, "A proximal iteration for deconvolving poisson noisy images using sparse representations," *IEEE Trans. Image Processing*, vol. 18, no. 2, pp. 310–321, 2009.

[5] D. Lingenfelter, J. Fessler, and Z. He, "Sparsity regularization for image reconstruction with Poisson data," vol. 7246, 2009.

[6] J. Salmon and et al, "Poisson noise reduction with non-local PCA," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279–294, 2014.

[7] R. Giryes and M. Elad, "Sparsity-based poisson denoising with dictionary learning," *IEEE Trans. Image Processing*, vol. 23, no. 12, pp. 5057–5069, 2014.

[8] S. Patil and A. Rajwade, "Poisson noise removal for image demosaicing," in *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*, 2016.

[9] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[10] P. Hoyer, "Non-negative sparse coding," in *Neural Networks for Signal Processing*, 2002, pp. 557–565.

[11] A. Soni and J. Haupt, "Estimation error guarantees for poisson denoising with sparse and structured dictionary models," in *International Symposium on Information Theory*, 2014.

[12] 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.

[13] J. Curtiss, "On transformations used in the analysis of variance," *The Annals of Mathematical Statistics*, vol. 14, no. 2, pp. 107–122, 1943.

[14] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[15] D. Garg, P. Bohra, K. S. Gurumoorthy, and A. Rajwade, "Reconstruction error bounds for compressed sensing under poisson or poisson-gaussian noise using variance stabilization transforms," https://www.cse.iitb.ac.in/~ajitvr/Poisson_PoissonGaussian_VST.pdf, 2018.

[16] S. Boyd, *Convex Optimization*. Cambridge University Press, 2004.

[17] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, 2015.

[18] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 910, pp. 589 – 592, 2008.

[19] C. Talegaonkar and A. Rajwade, "Supplement material," https://www.cse.iitb.ac.in/~ajitvr/Chinmay-supplemental-GlobalSIP.pdf, 2018.

[20] S. Ivanoff, F. Picard, and V. Rivoirard, "Adaptive Lasso and group-Lasso for functional Poisson regression," *JMLR*, vol. 17, no. 1, 2016.