PERFORMANCE BOUNDS FOR POISSON COMPRESSED SENSING USING VARIANCE STABILIZATION TRANSFORMS

Deepak Garg

Department of Computer Science and Engg.,
IIT Bombay
thedeepak@cse.iitb.ac.in

Ajit Rajwade *

Department of Computer Science and Engg.,
IIT Bombay
ajitvr@cse.iitb.ac.in

ABSTRACT

The analysis of reconstruction errors for compressed sensing under Poisson noise is challenging due to the signal dependent nature of the noise, and also because the Poisson negative log-likelihood is not a metric. In this paper, we present error bounds for reconstruction of signals which are sparse or compressible under any given orthonormal basis, given compressed measurements corrupted by Poisson noise and acquired in a realistic physical system. The concerned optimization problem is framed based on the well-known Variance Stabilization Transforms which transform the noise to (approximately) Gaussian with a fixed variance. This problem also turns out to be convex. We demonstrate promising numerical results on signals with different sparsity, intensity levels and given different numbers of compressed measurements.

Index Terms— Poisson Noise, nonlinear Compressed Sensing, Variance Stabilization Transform, Reconstruction Error Bounds

1. INTRODUCTION

Compressed Sensing (CS) theory predicts the ability to successfully reconstruct a signal that is sparse or compressible (in some orthonormal basis) from a sub-Nyquist number of measurements. The forward model for compressive acquisition can be expressed as $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{\eta} = \boldsymbol{\Phi \Psi \theta} + \boldsymbol{\eta} = \boldsymbol{A\theta} + \boldsymbol{\eta}$ where $\boldsymbol{y}_{N \times 1}$ is the acquired measurement vector, $\boldsymbol{A}_{N \times m}$ ($N \ll m$) is the product of the known measurement matrix $\boldsymbol{\Phi}_{N \times m}$ and the known orthonormal signal representation matrix $\boldsymbol{\Psi}_{m \times m}$, $\boldsymbol{\theta}_{m \times 1}$ is the *unknown* sparse/compressible vector of coefficients yielding the signal $\boldsymbol{x}_{m \times 1} = \boldsymbol{\Psi}_{m \times m} \boldsymbol{\theta}_{m \times 1}$, and $\boldsymbol{\eta}$ is the additive noise in the measurements. Under Gaussian or bounded uniform noise, the signal $\boldsymbol{x}$ can be obtained near-accurately within proven error limits [1, 2], by solving the optimization problem:

$$\text{(P1)} : \min \|\boldsymbol{\theta}\|_1 \text{such that } \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2 \leq \epsilon. \quad (1)$$

However, the noise in practical imaging systems is known to follow the Poisson distribution [3, 4, 5], *i.e.*, $\boldsymbol{y} \sim \text{Poisson}(\boldsymbol{\Phi x})$ for non-negative signal $\boldsymbol{x}$, for which the problem (P1) is not suitable. The likelihood of $\boldsymbol{y}$ given $\boldsymbol{x}$ and $\boldsymbol{\Phi}$ assuming independence of the individual measurements is given by $p(\boldsymbol{y}|\boldsymbol{\Phi x}) = \prod_{i=0}^{N-1} \frac{[(\Phi x)_i]^{y_i} e^{-(\Phi x)_i}}{y_i!}$ where the subscript i represents the ith component of $\boldsymbol{y}$. The corresponding optimization problem involving the negative log-likelihood turns out to be nonlinear and is given as follows:

$$\text{(P2)} : \min \lambda \|\boldsymbol{\theta}\|_q + \sum_{i=0}^{N-1} [(\Phi x)_i - y_i \log(\Phi x)_i], \quad (2)$$

where q decides the sparsity-promoting penalty and λ is a regularization parameter.

Furthermore, practical imaging systems involve photon-counting which imposes the constraint that $\boldsymbol{y}$ and $\boldsymbol{x}$ must both be non-negative. Such systems are also photon-limited, which imposes an additional constraint that the measurement matrix $\boldsymbol{\Phi}$ must be flux-preserving [6]. Flux-preservation implies that the total photon count of the measurement vector be less than or equal to the incident photon count, *i.e.*, the flux of the original signal. Thus for such a matrix, we must have $\sum_{i=0}^{N-1} (\Phi x)_i \leq \sum_{i=0}^{m-1} x_i$. The sensing matrix created under these two constraints does not obey the restricted isometry property (RIP), a sufficient condition for successful CS recovery. However using concepts from [6], starting from $\boldsymbol{\Phi}$, another sensing matrix $\widetilde{\boldsymbol{\Phi}}$ which obeys RIP can be reconstructed.

1.1. Main Contribution

In this paper, we derive relative reconstruction error (RRE) bounds for Poisson CS for a practical imaging system, using the well-known variance stabilization transforms (VST) of the form $\sqrt{y_i + c} = \sqrt{(\Phi x)_i + c} + \eta_i$ given constant c, where it is known that η_i is approximately Gaussian distributed with mean 0 and variance $\frac{1}{4}$. The corresponding optimization problem has been previously criticized in [6, 7] as being unsuitable for Poisson inverse problems due to its

nonlinear nature. But in this paper, we present error bounds for this problem, which are easy to derive and competitive with the state of the art. To the best of our knowledge, *this is the first attempt to derive error bounds for Poisson CS using the VST method.* Our paper can also be viewed as a contribution to the less explored sub-field of *nonlinear* compressed sensing [8]. Moreover the proposed optimization problem is also convex and hence affords efficient implementations, for which we present extensive numerical results.

1.2. Organization of Paper

In section 2, we briefly discuss the variance stabilization methods for Poisson noise. The proof for the RRE bounds with our method is provided in Section 3. Section 4 presents our experimental results. Finally, we present comparisons to existing literature on error bounds in Poisson CS, and conclude in Section 5.

2. VARIANCE STABILIZATION TRANSFORM

A VST essentially involves applying a function f to the underlying data x so that the transformed data $f(x)$ have a variance that is (nearly) constant and independent of x. For Poisson data where the variance is equal to the mean, the function f turns out to be the square root. The transformed data, *i.e.* $\sqrt{x}$, have a Gaussian distribution with mean equal to $\sqrt{x}$ and variance approximately $\frac{1}{4}$ when x is large. In [9] (equations 2.8 to 2.12), it was proved that changing $f(x)$ to $\sqrt{x + \frac{3}{8}}$ produced a variance that is closest to being a constant value of $\frac{1}{4}$ for x as small as 4. This is called the Anscombe transform. Thus given Poisson-corrupted compressed measurements of the form $y_i \sim \text{Poisson}((\Phi x)_i)$, we apply the Anscombe transform to yield data of the form

$$\forall i \in \{0, ..., N-1\}, \sqrt{y_i + \frac{3}{8}} \sim \mathcal{N}\left(\sqrt{(\Phi x)_i + \frac{3}{8}}, \frac{1}{4}\right). \tag{3}$$

In this paper, we derive error bounds for the case when $c = 0$, though the core of the analysis is applicable (with trivial modifications) to the case $c = \frac{3}{8}$ as well as to other VSTs for Poisson noise such as the Freeman-Tukey transform [10] $f(x) = \sqrt{x} + \sqrt{x+1}$.

3. DERIVATION

3.1. Sensing Matrix

Sensing matrices drawn randomly from the Gaussian or ± 1 Bernoulli distributions do not satisfy the non-negativity and flux preservation constraints for realistic Poisson imaging systems. Therefore, by referring to the approach in [6], we construct a sensing matrix Φ with entries either ones (scaled)

or zeros. Let $\mathbf{Z}_{N \times m}$ be a matrix with iid random variables taking only zeros and scaled ones as entries, defined as

$$Z_{i,j} = \begin{cases} -\sqrt{\frac{1-p}{p}} & \text{with probability } p, \\ \sqrt{\frac{p}{1-p}} & \text{with probability } 1-p. \end{cases}$$

A new matrix $\widetilde{\Phi} \triangleq \frac{\mathbf{Z}}{\sqrt{N}}$ is defined. When $p = \frac{1}{2}$, it is found that $\widetilde{\Phi}$ (and hence $\widetilde{\Phi}\Psi$) obey RIP of order 2s with high probability [11]. Note that $\widetilde{\Phi}$ contains negative values with high probability which is in contrast to our required constraints. Thus, to obtain a flux-preserving and non-negative matrix Φ from $\widetilde{\Phi}$, the following construction is used:

$$\Phi = \sqrt{\frac{p(1-p)}{N}}\widetilde{\Phi} + \frac{1-p}{N}\mathbf{1}_{N \times m} = \frac{1}{2\sqrt{N}}\widetilde{\Phi} + \frac{1}{2N}\mathbf{1}_{N \times m}. \tag{4}$$

The matrix Φ has entries either 0 or $\frac{1}{N}$ and its non-negativity and flux-preservation can be easily verified.

3.2. Key Theorem and Proof

Theorem 1 : Consider a non-negative signal x with total intensity $I = \|x\|_1$ expressed using the orthornormal basis Ψ in the form $x = \Psi\theta$. Consider Poisson corrupted CS measurements of the form $y \sim \text{Poisson}(\Phi x)$ where Φ is constructed as per Eqn. 4. Let $\theta^\star$ be the result of the following optimization problem:

$$(P3) : \min\|\theta\|_1 \text{ such that } \|\sqrt{y} - \sqrt{A\theta}\|_2 \le \epsilon, \|\Psi\theta\|_1 = I, \Psi\theta \succeq \mathbf{0}. \tag{5}$$

If $\widetilde{\Phi}$ obeys RIP of order 2s with RIC $\delta_{2s} < \sqrt{2} - 1$, and θ_s denotes a vector containing the s largest magnitude elements of θ with the rest being 0, then we have:

$$\frac{\|\theta - \theta^\star\|_2}{I} \le \frac{C_1\epsilon}{\sqrt{I}} + \frac{C_2 s^{-\frac{1}{2}}\|\theta - \theta_s\|_1}{I}, \text{ and} \tag{6}$$

$$E\left(\frac{\|\theta - \theta^\star\|_2}{I}\right) \le \frac{C_1\sqrt{N}}{\sqrt[4]{I}} + \frac{C_2 s^{-\frac{1}{2}}\|\theta - \theta_s\|_1}{I}. \tag{7}$$

Proof: We provide a sketch of the proof below, inspired from [12], but modified to suit our problem.

1. Define a vector $h \triangleq \theta - \theta^\star$. Denote vector h_T to be equal to h only for index set T and zero for other indices. Let T_0 be the set containing s largest absolute value indices of θ, T_1 be 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. Thus, vector h can be decomposed as the sum of $h_{T0}, h_{T_1}, h_{T_2}, ...$

2. Define $A \triangleq \Phi\Psi$. We have

$$\|Ah\|_2^2 = \|A(\theta - \theta^\star)\|_2^2$$
$$= \Sigma_i(\sqrt{(A\theta)_i} - \sqrt{(A\theta^\star)_i})^2(\sqrt{(A\theta)_i} + \sqrt{(A\theta^\star)_i})^2. \tag{8}$$

(a) By triangle inequality and the nature of the constraint in (P3), we have

$$\|\sqrt{A\theta} - \sqrt{A\theta^{\star}}\|_2 \leq \|\sqrt{y} - \sqrt{A\theta}\|_2 + \quad (9)$$
$$\|\sqrt{y} - \sqrt{A\theta^{\star}}\|_2 \leq 2\epsilon.$$

(b) For scalars $v_1 \geq 0, v_2 \geq 0$, we have $(\sqrt{v_1} + \sqrt{v_2})^2 \leq 4\max(v_1, v_2)$. We also have $(A\theta)_i = (\Phi x)_i = \Sigma_j \Phi_{ij} x_j \leq \frac{\|x\|_1}{N} = \frac{I}{N}$. Likewise $(A\theta^{\star})_i \leq \frac{I}{N}$ as well, since $\|x^{\star}\|_1 = I$. Hence $(\sqrt{(A\theta)_i} + \sqrt{(A\theta^{\star})_i})^2 \leq \frac{4I}{N}$.

(c) Combining the earlier two results with Eqn. 8, we have $\|Ah\|_2 \leq 4\epsilon\sqrt{\frac{I}{N}}$.

3. To prove the bound on $\|h_{(T_0 \cup T_1)^c}\|_2$, we follow steps similar to [12] to obtain

$$\|h_{(T_0 \cup T_1)^c}\|_2 \leq \|h_{(T_0)}\|_2 + 2s^{-1/2}\|\theta - \theta_s\|_1. \quad (10)$$

4. To prove error bounds on $\|h_{(T_0 \cup T_1)}\|_2$, we adopt the following steps.

(a) Given the construction for Φ in Eqn. 4, we have

$$\Phi\Psi(\theta - \theta^{\star}) = \frac{1}{2\sqrt{N}}\widetilde{\Phi}\Psi(\theta - \theta^{\star}) +$$
$$(\|\Psi\theta\|_1 - \|\Psi\theta^{\star}\|_1)$$
$$= \frac{1}{2\sqrt{N}}\widetilde{\Phi}\Psi(\theta - \theta^{\star}) \quad (11)$$

since we know that $\|\Psi\theta\|_1 = \|\Psi\theta^{\star}\|_1 = I$. Defining $B \triangleq \widetilde{\Phi}\Psi$, we get

$$\|Bh\|_2 = 2\sqrt{N}\|Ah\|_2 \leq 8\epsilon\sqrt{I}. \quad (12)$$

(b) Following steps in [12] using the RIP and the Cauchy-Schwarz inequality, we can prove that

$$\|h_{T_0 \cup T_1}\|_2 \leq C'\epsilon\sqrt{I} + C''s^{-1/2}\|\theta - \theta^{\star}\|_1 \quad (13)$$

where $C' \triangleq \frac{8\sqrt{1+\delta_{2s}}}{1-\delta_{2s}(\sqrt{2}+1)}$ and $C'' \triangleq \frac{2\sqrt{2}\delta_{2s}}{1-\delta_{2s}(\sqrt{2}+1)}$.

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

$$\|h\|_2 \leq C_1\epsilon\sqrt{I} + C_2 s^{-\frac{1}{2}}\|\theta - \theta_s\|_1 \quad (14)$$

where $C_1 \triangleq 2C'$ and $C_2 \triangleq 2 + 2C''$.

Finally, we divide by I to obtain upper RRE bounds:

$$\frac{\|\theta - \theta^{\star}\|_2}{I} \leq \frac{C_1\epsilon}{\sqrt{I}} + \frac{C_2 s^{-\frac{1}{2}}\|\theta - \theta_s\|_1}{I}. \quad (15)$$

Since ϵ is equal to the magnitude of a vector with elements drawn from $\mathcal{N}(0, \frac{1}{4})$, it follows a chi-distribution with N degrees of freedom. Therefore, the expected RRE (expectation over noise instances for fixed A, x) can be given as:

$$E\left(\frac{\|\theta - \theta^{\star}\|_2}{I}\right) \leq \frac{C_1\sqrt{N}}{4\sqrt{I}} + \frac{C_2 s^{-\frac{1}{2}}\|\theta - \theta_s\|_1}{I}. \quad (16)$$

Remarks on Proof:

1. Our proof architecture is inspired from [12], but the points of departure are steps 2(a), 2(b), 2(c) as well as step 4(a) which gives a relationship between $\|Ah\|_2$ and $\|Bh\|_2$. These steps exploit the non-negativity and flux-preserving property of Φ.

2. Given that we are dealing with a *Poisson* inverse problem, it is more intuitive to analyze the *relative* reconstruction error (RRE) rather than the (absolute) reconstruction error. This is because as the mean of the Poisson distribution increases, so does its variance, causing an increase in the mean squared error but a decrease in the relative mean squared error.

3. Notice that our derived RRE bound is inversely proportional to the signal intensity I. For a fixed I, if N is increased, the incident photon flux I is distributed across the N measurements, causing a decrease in SNR per measurement and possibly degrading performance. This phenomenon differs from CS under Gaussian noise, and has earlier been noted in [6, 13, 14].

4. Prior knowledge of total signal intensity I might look like a strong assumption, but in some compressive camera architectures, such as the Rice Single Pixel Camera [3], I can be easily estimated during acquisition. Moreover our experimental results in the next section show that knowledge of I is not necessary, although we required it for our theoretical analysis.

5. The RRE bounds are also applicable to the Anscombe or Freeman-Tukey transforms with minor changes to the constant C_1.

4. EXPERIMENTAL RESULTS

We perform experiments on 1D signals that are sparse in the canonical basis, though our analysis and experiments are equally applicable to 1D or higher-D signals sparse or compressible in an arbitrary orthonormal basis Ψ. In particular, we solve the following optimization problem (which is equivalent to P3 except for the $\|x\|_1 = I$ constraint):

$$(P4): \|\sqrt{y} - \sqrt{Ax}\|_2^2 + \lambda\|x\|_1 \text{ s.t. } x \succeq 0 \quad (17)$$

where λ is the regularization parameter. Since P4 is convex, we use the CVX package [15] with the MOSEK solver [16] for all our experiments. The chosen signals had a length of 100 and the sensing matrices were as per Eqn. 4. The signals and sensing matrices were generated randomly subject to required constraints. We generated a box-plot of the relative mean square error (RMSE) values computed as $\text{RMSE}(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\|\boldsymbol{x}-\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}\|_2}$ where $\boldsymbol{x}$ and $\boldsymbol{x}^*$ denote the true and estimated signals respectively. In Fig 1, the RMSE box-plots were generated against differing values of I, signal sparsity s and number of measurements N. For all experiments, the value of λ that gave the best RMSE was selected from a range of values, assuming the original signal was known. Figure 1 clearly shows (a) rapid improvement in performance with increase in I, as is to be expected for Poisson noise, (b) a degradation in performance as s increases, and (c) that as the number of measurements increases, the RMSE decreases but stabilizes to a non-zero value unlike the case of CS with Gaussian noise. We found that the results with problem (P4) were similar to those obtained with the LASSO or with problem (P2). Also, though our theoretical analysis requires the explicit constraint $\|\boldsymbol{x}^*\|_1 = I$ where I is known *a priori*, we obtained good estimation results in practice even without this constraint.

5. DISCUSSION AND RELATION TO PRIOR WORK

Error bounds for the problem (P2) with $q = 0$ were derived in the pioneering work in [6] and further tightened in [17], but the associated estimators are computationally intractable. Moreover the analysis of [17] applies to strictly sparse (not compressible) signals. The bounds were extended to the problem of low-rank matrix estimation under Poisson noise in [14, 13]. Error bounds for other tractable estimators for Poisson CS have been derived: for example, LASSO and weighted-LASSO in [18] (which assume a Gaussian approximation with appropriate variance), and problem (P2) in [19]. However, unlike our work, these papers do not account for the flux-preservation constraint explicitly which is important for realistic imaging systems. Our work stands out as (1) it is based on an elegant extension of the bounds from [12], (2) it produces error bounds for a computationally tractable estimator, and (3) is designed for a flux-preserving sensing matrix. VSTs have been used for Poisson inverse problems such as deconvolution in earlier work [20], but no error bounds were derived, and the case of Poisson CS was not dealt with.

Currently our upper bounds are directly proportional to $\sqrt{N}$ though Figure 1 (third) shows a decreasing trend w.r.t. N. This is a common feature between our work as well as [6, 13, 14]. Future work would involve tightening these error bounds, extensions to other Poisson regression problems such as low rank matrix recovery, developing criteria for automated selection of λ (related to the unknown ϵ in P3), and refining theoretical analysis to avoid dependence on knowledge of I.
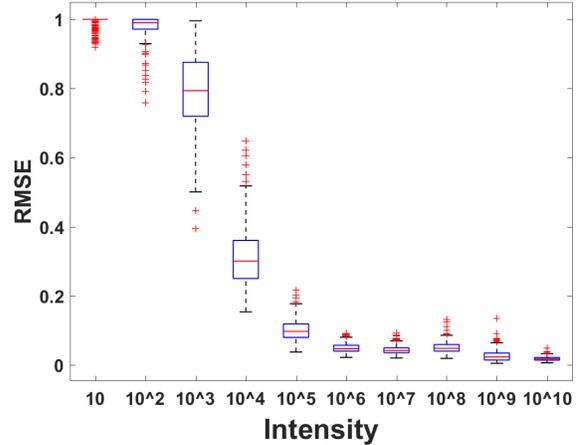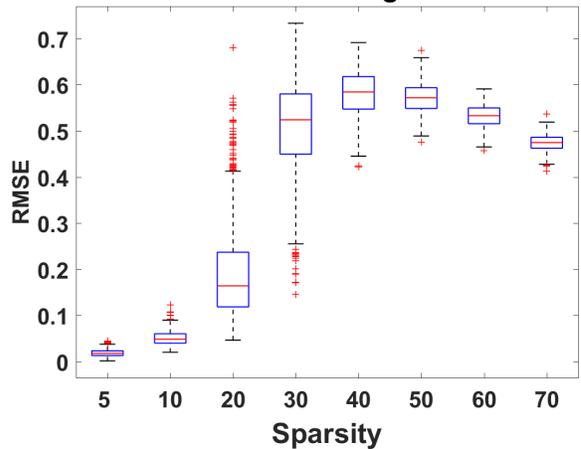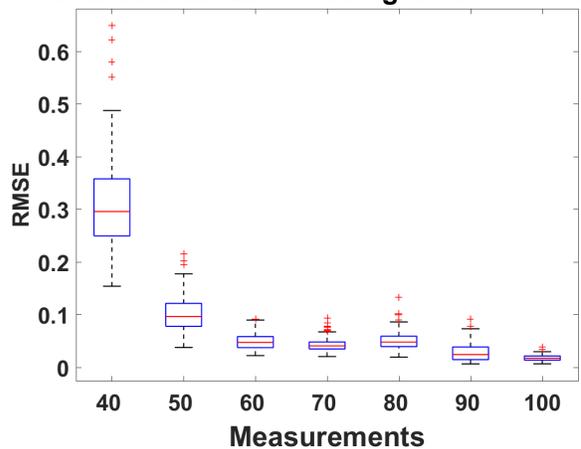
Fig. 1. Top to bottom: $\text{RMSE}(\boldsymbol{x}, \boldsymbol{x}^*)$ v/s Intensity I at $s = 10, N = 50$; $\text{RMSE}(\boldsymbol{x}, \boldsymbol{x}^*)$ v/s Sparsity at $I = 10^8, N = 50$; $\text{RMSE}(\boldsymbol{x}, \boldsymbol{x}^*)$ v/s Measurements at $s = 10, I = 10^8$

6. REFERENCES

[1] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[2] D. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83, 2008.

[4] F. Alter, Y. Matsushita, and X. Tang, "An intensity similarity measure in low-light conditions," in *ECCV*, 2006.

[5] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J-B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 442454, 2010.

[6] M. Raginsky, R. Willett, Z. Harmany, and R. Marcia, "Compressed sensing performance bounds under poisson noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 3990–4002, 2010.

[7] R. Willett, "The dark side of image reconstruction," https://www.youtube.com/watch?v=ea0A1xXjwCI, Jan 2016.

[8] T. Blumensath, "Compressed sensing with nonlinear observations and related nonlinear optimization problems," *IEEE Trans. Information Theory*, vol. 59, no. 6, pp. 3466–3474, 2013.

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

[10] M. Freeman and J. Tukey, "Transformations related to the angular and the square root," *The Annals of Mathematical Statistics*, vol. 21, no. 4, pp. 607611, 1950.

[11] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[12] Emmanuel J Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[13] Y. Xie, Y. Chi, and R. Calderbank, "Low-rank matrix recovery with Poisson noise," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 622–622.

[14] Y. Cao and Y. Xie, "Poisson matrix recovery and completion," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1609–1620, March 2016.

[15] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[16] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*, 2015.

[17] X. Jiang, G. Raskutti, and R. Willett, "Minimax optimal rates for Poisson inverse problems with physical constraints," *IEEE Trans. Information Theory*, vol. 61, no. 8, pp. 4458–4474, 2015.

[18] X. Jiang, P. Reynaud-Bouret, V. Rivoirard, L. Sansonnet, and R. Willett, "A data-dependent weighted lasso under poisson noise," *arXiv preprint arXiv:1509.08892*, 2015.

[19] M.-H. Rohban, V. Saligrama, and D.-M. Vaziri, "Minimax optimal sparse signal recovery with Poisson statistics," *IEEE Trans. Signal Processing*, vol. 64, no. 13, pp. 3495–3508, 2016.

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