Performance Bounds on Compressed Sensing with Poisson Noise

Rebecca M. Willett
Electrical and Computer Engineering
Duke University
Durham, NC 27708, USA
Email: willett@duke.edu

Maxim Raginsky
Electrical and Computer Engineering
Duke University
Durham, NC 27708, USA
Email: m.raginsky@duke.edu

Abstract— This paper describes performance bounds for compressed sensing in the presence of Poisson noise when the underlying signal, a vector of Poisson intensities, is sparse or compressible (admits a sparse approximation). The signal-independent and bounded noise models used in the literature to analyze the performance of compressed sensing do not accurately model the effects of Poisson noise. However, Poisson noise is an appropriate noise model for a variety of applications, including low-light imaging, where sensing hardware is large or expensive, and limiting the number of measurements collected is important. In this paper, we describe how a feasible positivity-preserving sensing matrix can be constructed, and then analyze the performance of a compressed sensing reconstruction approach for Poisson data that minimizes an objective function consisting of a negative Poisson log likelihood term and a penalty term which could be used as a measure of signal sparsity.

I. Introduction

The basic idea of compressed sensing is that, when the signal of interest is very sparse (i.e., zero-valued at most locations) or highly compressible in some basis, relatively few "incoherent" observations are sufficient to reconstruct the most significant non-zero signal components [1], [2]. Despite the promise of this theory for many applications, very little is known about its applicability to photon-limited imaging systems, where high-quality photomultiplier tubes (PMTs) are expensive and physically large, limiting the number of observations that can reasonably be collected by an imaging system. Limited photon counts arise in a wide variety of applications, including infrared imaging, nuclear medicine, astronomy and night vision, where the number of photons collected by the detector elements is very small relative to the number of pixels, voxels, or other quantities to be estimated. Robust reconstruction methods can potentially lead to many novel imaging systems designed to make the best possible use of the small number of photons collected while reducing the size and cost of the detector array.

However, the signal-independent and bounded noise models which have been considered in the literature (cf. [3], [4]) are not easily adapted to the Poisson noise models used in photon-limited imaging. The Poisson model is often used to model images acquired by photon-counting devices [5]. Under the Poisson assumption, we can write our observation model as

$$y \sim \text{Poisson}(Af^*), \qquad (1)$$

where $f^* \in \mathbb{R}^m$ is the signal or image of interest, $A \in \mathbb{R}^{N \times m}$ linearly projects the scene onto an N-dimensional space of observations, and $y \in \{0, 1, 2, \ldots\}^N$ is a length-N vector of observed Poisson counts. Specifically, under the model in (1), the likelihood of observing a particular vector of counts y is given by

$$p(y|Af^*) = \prod_{j=1}^{N} \frac{(Af^*)_j^{y_j}}{y_j!} e^{-(Af^*)_j},$$

where $(Af^*)_j$ is the jth component of Af^*.

The majority of the compressed sensing literature assumes that there exists a "sparsifying" reference basis W, so that $\theta^* \triangleq W^T f^*$ is sparse or lies in a weak-ℓ_p space. When the matrix product AW obeys the so-called *restricted isometry property* (RIP) [6], [7] or some related criterion, and when the noise is bounded or Gaussian, then θ^* can be accurately estimated from y by solving the following $\ell_2 - \ell_1$ optimization problem (or some variant):

$$\widehat{\theta} = \arg\min_{\theta} \|y - AW\theta\|_2^2 + \tau \|\theta\|_1, \qquad (2)$$

where $\tau > 0$ is a regularization parameter [2], [8], [7].

However, the ℓ_2 data-fitting term, $\|y - AW\theta\|_2^2$, is problematic in the presence of Poisson noise. Because the variance of the noisy observations is proportional to the signal intensity, ℓ_2 data-fitting terms can lead to significant overfitting in high-intensity regions and oversmoothing in low-intensity regions. Furthermore, photon-limited imaging systems implicitly place hard constraints on the nature of the measurements that can be collected, such as non-negativity, which are not considered in much of the existing compressed sensing literature (recent papers of Dai and Milenkovic [9] and of Khajehnejad et al. [10] are notable exceptions).

In this paper, we propose estimating f^* from y using a regularized Poisson log-likelihood objective function as an alternative to (2), and we present risk bounds for recovery of a compressible signal from Poisson observations. Specifically, in the Poisson noise setting we maximize the log-likelihood while minimizing a penalty function that, for instance, could

measure the sparsity of $\theta = W^T f$:

$$\widehat{f} = \underset{f}{\arg\min} \quad \sum_{j=1}^{N} (-y_j \log(Af)_j) + \tau \operatorname{pen}(f) \tag{3}$$
$$\text{subject to} \quad Af \succeq 0, \ f \succeq 0, \ \sum_{i=1}^{m} f_i = I$$

where $\operatorname{pen}(\cdot)$ is a penalty function that will be detailed later, I is the total intensity of the unknown f^* (assumed known), and the standard notation $v \succeq 0$ means that the components of v are nonnegative. The constraints reflect the nonnegativity of both the observed intensity and the underlying image and the known total intensity of the underlying image.

II. PROBLEM FORMULATION

We have a signal or image f^* of length m that we wish to estimate using a detector array of length $N \ll m$. We assume that $f^* \succeq 0$. We will bound the accuracy with which we can estimate f^*/I, where $I \triangleq \sum_{i=1}^{m} f_i^*$; in other words, we focus on accurately estimating the *shape* of f^* independent of any scaling factor proportional to the total intensity of the scene. We assume that the total intensity I is known, and our candidate estimators will also be constrained to have total intensity I. The quality of a candidate estimator f will be measured in terms of the *risk*

$$R(f^*, f) \triangleq \left\| \frac{f^*}{I} - \frac{f}{I} \right\|_2^2.$$

We construct our sensing matrix A as follows. Let $Z \in \{-1, +1\}^{N \times m}$ be a matrix whose entries $Z_{i,j}$ are independent Rademacher random variables, i.e., $\mathbb{P}[Z_{i,j} = -1] = \mathbb{P}[Z_{i,j} = +1] = 1/2$ independently of all other $Z_{i',j'}$. Let $\widetilde{A} = (1/N)Z$. Most compressed sensing approaches would proceed by assuming that we make (potentially noisy) observations of the product $\widetilde{A}f^*$, but elements of $\widetilde{A}f^*$ could be negative and thus not physically realizable in photon-counting systems. However, we can use $\widetilde{A}$ to generate a positivity-preserving sensing matrix A as follows. Let $\mathbf{1}_{r \times s}$ denote the $r \times s$ matrix all of whose entries are equal to 1. Then we let

$$A \triangleq \widetilde{A} + (1/N)\mathbf{1}_{N \times m}.$$

Note that $A \in \{0, 2/N\}^{N \times m}$ and, as a consequence, A indeed preserves positivity: for any $f \in \mathbb{R}_+^m$, $Af \succeq 0$.

We make Poisson observations of Af^*, $y \sim \operatorname{Poisson}(Af^*)$, and our goal is to estimate $f^* \in \mathbb{R}_+^m$ from $y \in \{0, 1, 2, \dots\}^N$. To this end, we propose solving the following optimization problem:

$$\widehat{f} \triangleq \underset{f \in \Gamma}{\arg\min} \left[-\log p(y|Af) + 2\operatorname{pen}(f) \right], \tag{4}$$

where $\operatorname{pen}(f)$ is a penalty term. We assume that $\Gamma \equiv \Gamma(m, I)$ is a countable set of feasible estimators $f \in \mathbb{R}_+^m$ satisfying $\sum_{i=1}^{m} f_i = I$, and that the penalty function satisfies the Kraft inequality:

$$\sum_{f \in \Gamma} e^{-\operatorname{pen}(f)} \leq 1. \tag{5}$$

Note that, by construction of A, $f \in \Gamma$ implies that $Af \succeq 0$. Furthermore, while the penalty term may be chosen to be

smaller for sparser solutions $\theta = W^T f$, where W is an orthogonal matrix that represents f in its "sparsifying" basis, our main result only assumes that (5) is satisfied. We can think of (4) as a discretized-feasibility version of (3), where we optimize over a countable set of feasible vectors that grows in a controlled way with signal length m.

III. PROPERTIES OF THE SENSING MATRIX A

Our main result, stated and proved in the next section, makes use of the several properties of the sensing matrix A (and $\widetilde{A}$). The most important of these properties is that, with high probability, $\widetilde{A}$ acts near-isometrically on certain subsets of $\mathbb{R}^m$. The usual formulation of this phenomenon is known in the compressed sensing literature as the *restricted isometry property* (RIP) [6], [7], where the subset of interest consists of all vectors with a given sparsity. In fact, the RIP is a special case of a much broader circle of results concerning the behavior of random matrices whose entries are drawn from a subgaussian isotropic ensemble [11]. The Rademacher ensemble is an instance of this, and the following two theorems can be extracted from the results of [11]:

Theorem 1 *There exist absolute constants $c_1, c_2 > 0$, such that, with probability at least $1 - e^{-c_1 N}$,*

$$\|u - v\|_2 \leq \sqrt{2}N\|\widetilde{A}(u - v)\|_2 + c_2\sqrt{\frac{\log(c_2 m/N)}{N}}$$

for all $u, v \in \mathbb{R}^m$ such that $\|u\|_1 = \|v\|_1 = 1$.

Theorem 2 *There exist absolute constants $c_3, c_4 > 0$, such that the following holds. Let $\mathcal{S}$ be a finite subset of the unit sphere in $\mathbb{R}^m$. Then, with probability at least $1 - e^{-c_3 N}$,*

$$1/2 \leq N\|\widetilde{A}s\|_2^2 \leq 3/2, \qquad \forall s \in \mathcal{S}$$

provided $N \geq c_4 \log_2 |\mathcal{S}|$.

We will also rely on the following properties of A and $\widetilde{A}$:

- With probability at least $1 - N2^{-m}$, every row of Z has at least one positive entry. Let $f \in \mathbb{R}^m$ be an arbitrary vector of intensities satisfying $f \succeq (cI)\mathbf{1}_{m \times 1}$ for some $c > 0$. Then

$$Af \succeq (2cI/N)\mathbf{1}_{N \times 1}. \tag{6}$$

- With probability at least $1 - 2me^{-N/8}$,

$$\left| \sum_{i=1}^{N} \widetilde{A}_{i,j} \right| \leq 1/4, \qquad \forall j \in \{1, \dots, m\} \tag{7}$$

(this is a simple consequence of the Chernoff bound and the union bound).

- If the event (7) holds, then

$$(3/4)I \leq \sum_{i=1}^{N} \sum_{j=1}^{m} A_{i,j} f_j \leq (5/4)I, \qquad \forall f \in \mathbb{R}_+^m \tag{8}$$

IV. AN ORACLE INEQUALITY FOR THE EXPECTED RISK

We now state and prove our main result, which gives an upper bound on the expected risk $\mathbb{E}R(f^*, \widehat{f})$ that holds for any target signal $f^* \succeq 0$ satisfying the normalization constraint $\sum_{i=1}^m f_i^* = I$, without assuming anything about the sparsity properties of f^*. Conceptually, our bound is an *oracle inequality*, which states that the expected risk of our estimator is within a constant factor of the best regularized risk attainable by estimators in Γ with full knowledge of the underlying signal f^*. More precisely, for each $f \in \Gamma$ define

$$R^*(f^*, f) \triangleq \left\| \frac{f^*}{I} - \frac{f}{I} \right\|_2^2 + \frac{2\,\mathrm{pen}(f)}{I},$$

and for every $\Gamma' \subseteq \Gamma$ let $R^*(f^*, \Gamma') \triangleq \min_{f \in \Gamma'} R^*(f^*, f)$. Note that $R^*(f^*, \Gamma')$ is the best penalized risk that can be attained over Γ' by an oracle that has full knowledge of f^*. We then have the following:

Theorem 3 *Suppose that the feasible set Γ also satisfies the condition*

$$f \succeq (cI)\mathbf{1}_{m \times 1}, \qquad \forall f \in \Gamma \tag{9}$$

for some $0 < c < 1$. Let $\mathcal{G}_N$ be the collection of all subsets $\Gamma' \subseteq \Gamma$, such that $|\Gamma'| \leq 2^{N/c_4}$. Then the following holds with probability at least $1 - me^{-KN}$ for some positive $K = K(c_1, c_3)$ (with respect to the realization of $\widetilde{A}$):

$$\mathbb{E}R(f^*, \widehat{f}) \leq C_N \min_{\Gamma' \in \mathcal{G}_N} R^*(f^*, \Gamma') + \frac{2c_2^2 \log(c_2 m/N)}{N}, \tag{10}$$

where $C_N = \max(20, 15/c)N$, and the expectation is taken with respect to $y \sim \mathrm{Poisson}(Af^*)$.

Remark 1. A positivity condition similar to (9) is natural in the context of estimating vectors with nonnegative entries from count data. In particular, it excludes the possibility of assigning zero intensity to an input of a detector when at least one photon has been counted [12]. However, as will be clear from the proof below, condition (9) can be replaced with a more general (weaker) condition

$$Af \succeq (c'I/N)\mathbf{1}_{N \times 1}, \qquad \forall f \in \Gamma$$

for some $c' > 0$, which is more appropriate when the signal f^* is sparse in the canonical basis, because then it is in fact desirable to allow candidate estimators with zero components.

Proof: With high probability, the following chain of estimates holds:

$$\frac{1}{I^2} \|f^* - \widehat{f}\|_2^2$$
$$\leq \frac{4N}{I^2} \|\widetilde{A}(f^* - \widehat{f})\|_2^2 + \frac{2c_2^2 \log(c_2 m/N)}{N}$$
$$= \frac{4N}{I^2} \|A(f^* - \widehat{f})\|_2^2 + \frac{2c_2^2 \log(c_2 m/N)}{N}$$
$$\leq \frac{4N}{I^2} \|A(f^* - \widehat{f})\|_1^2 + \frac{2c_2^2 \log(c_2 m/N)}{N},$$

where the first inequality is a consequence of Theorem 1, and the remaining steps follow from definitions and from standard inequalities for ℓ_p norms. Moreover, with high probability,

$$\|A(f^* - \widehat{f})\|_1^2$$
$$= \left(\sum_{i=1}^N \left| (Af^*)_i^{1/2} - (A\widehat{f})_i^{1/2} \right| \cdot \left| (Af^*)_i^{1/2} + (A\widehat{f})_i^{1/2} \right| \right)^2$$
$$\leq \sum_{i,j=1}^N \left| (Af^*)_i^{1/2} - (A\widehat{f})_i^{1/2} \right|^2 \left| (Af^*)_j^{1/2} + (A\widehat{f})_j^{1/2} \right|^2$$
$$\leq 5I \sum_{i=1}^N \left| (Af^*)_i^{1/2} - (A\widehat{f})_i^{1/2} \right|^2,$$

where the first inequality is due to Cauchy–Schwarz, and the second inequality is a consequence of (8) and the inequality between the arithmetic mean and the geometric mean. It is a matter of straightforward algebra to show that

$$\sum_{i=1}^N \left| (Af^*)_i^{1/2} - (A\widehat{f})_i^{1/2} \right|^2$$
$$= -2\log \sum_{i=1}^N \exp\left(-\frac{1}{2} \left[(Af^*)_i^{1/2} - (A\widehat{f})_i^{1/2} \right] \right)^2$$
$$= 2\log \left(\int \sqrt{p(y|Af^*)p(y|A\widehat{f})}\,d\nu(y) \right)^{-1},$$

where ν is the counting measure on $\{0, 1, 2, \ldots\}^N$. Now, the same techniques as in Li and Barron [13] (see also the proof of Theorem 7 in [14]) can be used to show that

$$2\mathbb{E}\log \left(\int \sqrt{p(y|Af^*)p(y|A\widehat{f})}\,d\nu(y) \right)^{-1}$$
$$\leq \min_{f \in \Gamma} \left[\mathrm{KL}\left(p(\cdot|Af^*) \big\| p(\cdot|Af) \right) + 2\,\mathrm{pen}(f) \right], \tag{11}$$

where $\mathrm{KL}(\cdot \| \cdot)$ is the *Kullback–Leibler (KL) divergence*, which for the Poisson likelihoods has the form

$$\mathrm{KL}\left(p(\cdot|Af^*) \big\| p(\cdot|Af) \right)$$
$$= \sum_{i=1}^N \left[(Af^*)_i \log \frac{(Af^*)_i}{(Af)_i} - (Af^*)_i + (Af)_i \right].$$

Using the inequality $\log t \leq t - 1$ together with (9) and (6), we can bound the KL divergence as

$$\sum_{i=1}^N \left[(Af^*)_i \log \frac{(Af^*)_i}{(Af)_i} - (Af^*)_i + (Af)_i \right]$$
$$\leq \sum_{i=1}^N \left[(Af^*)_i \left(\frac{(Af^*)_i}{(Af)_i} - 1 \right) - (Af^*)_i + (Af)_i \right]$$
$$= \sum_{i=1}^N \frac{1}{(Af)_i} \left[(Af)_i^2 - 2(Af)_i(Af^*)_i + (Af^*)_i^2 \right]$$
$$\leq \frac{N}{2cI} \|A(f^* - f)\|_2^2$$
$$= \frac{N}{2cI} \|\widetilde{A}(f^* - f)\|_2^2.$$

Now, choose any $\Gamma^* \in \mathcal{G}_N$, such that

$$R^*(f^*, \Gamma^*) = \min_{\Gamma' \in \mathcal{G}_N} R^*(f^*, \Gamma').$$

Then, applying Theorem 2 to the set $\left\{\frac{f^* - f}{\|f^* - f\|_2} : f \in \Gamma^*\right\}$, we have, with high probability, that

$$N\|\widetilde{A}(f^* - f)\|_2^2 \le (3/2)\|f^* - f\|_2^2, \qquad \forall f \in \Gamma^*.$$

Combining everything, we get the bound $\mathbb{E}R(f^*, \widehat{f}) \le$

$$\max\left(20, \frac{15}{c}\right) N \min_{f \in \Gamma^*} \left[\left\|\frac{f^*}{I} - \frac{f}{I}\right\|_2^2 + \frac{2\operatorname{pen}(f)}{I}\right]$$

$$+ \frac{2c_2^2 \log(c_2 m/N)}{N}$$

which holds with high probability w.r.t. the realization of $\widetilde{A}$. Let $C_N = \max(20, 15/c)N$. The theorem is proved. ∎

V. RISK BOUNDS FOR COMPRESSIBLE SIGNALS

We now show how the bound in Theorem 3 can be used to analyze how the performance of the proposed estimator when the target signal f^* is compressible (i.e., admits a sparse approximation) in some reference orthonormal basis.

Following [1], we assume that there exists an orthonormal basis $\Phi = \{\phi_1, \ldots, \phi_m\}$ of $\mathbb{R}^m$, such that f^* is *compressible* in Φ in the following sense. Let W be the orthogonal matrix with columns $\phi_1, \ldots, \phi_m$. Then the vector θ^* of the coefficients $\theta_j^* = \langle f^*, \phi_j \rangle$ of f^* in Φ is related to f^* via $f^* = W\theta^*$. Let $\theta_{(1)}^*, \ldots, \theta_{(m)}^*$ be the entries of θ^* arranged in the order of decreasing magnitude: $|\theta_{(1)}^*| \ge |\theta_{(2)}^*| \ge \ldots \ge |\theta_{(m)}^*|$. We assume that there exist some $0 < q < \infty$ and $\rho > 0$, such that for each $1 \le j \le m$

$$|\theta_{(j)}^*| \le \rho I j^{-1/q}. \tag{12}$$

Note that for every $1 \le j \le m$ we have

$$|\theta_{(j)}^*| \le \|\theta^*\|_2 = \|f^*\|_2 \le \|f^*\|_1 = I,$$

so we can take ρ to be a constant independent of I or m. Any θ^* satisfying (12) is said to belong to the *weak-ℓ_q ball of radius ρI*. The weak-ℓ_q condition (12) translates into the following approximation estimate: given any $1 \le k \le m$, let $\theta^{(k)}$ denote the best k-term approximation to θ^*. Then we can show that

$$\left\|\frac{\theta^*}{I} - \frac{\theta^{(k)}}{I}\right\|_2^2 \le C\rho^2 k^{-2\alpha}, \qquad \alpha = 1/q - 1/2 \tag{13}$$

for some constant $C > 0$ that depends only on q. We also assume that f^* satisfies the condition (9) for some $c \in (0, 1)$, a lower bound on which is assumed known.

In order to apply Theorem 3, we will form a suitable finite class of estimators Γ and set a penalty function $\operatorname{pen}(f)$ over this class which (a) is smaller for sparser $\theta = W^T f$ and (b) satisfies (5). The family Γ is constructed as follows.

1) Define the sets

$$\Theta \triangleq \{\theta \in \mathbb{R}^m : \|\theta\|_\infty \le I;$$

each θ_i uniformly quantized to one of $\sqrt{m}$ levels}

and $\mathcal{F} \triangleq \{f \in \mathbb{R}^m : f = W\theta, \theta \in \Theta\}$.

2) For each $f \in \mathcal{F}$, let $\bar{f}$ denote the ℓ_2 projection of f onto the closed convex set

$$\mathcal{C} \triangleq \left\{g \in \mathbb{R}^m : g \succeq (cI)\mathbf{1}_{m \times 1} \text{ and } \sum_{i=1}^m g_i = I\right\},$$

i.e.,

$$\bar{f} \triangleq \arg\min_{g \in \mathcal{C}} \|f - g\|_2.$$

3) Finally, let $\Gamma \triangleq \{\bar{\theta} = W^T \bar{f} : f \in \mathcal{F}\}$.

Note that the projection $\bar{f}$ satisfies the *Pythagorean identity*

$$\|g - f\|_2^2 \ge \|g - \bar{f}\|_2^2 + \|\bar{f} - f\|_2^2, \qquad \forall g \in \mathcal{C}$$

(see, e.g., Theorem 2.4.1 in [15]). In particular, $\|g - f\|_2^2 \ge \|g - \bar{f}\|_2^2$, and, since $f^* \in \mathcal{C}$, we have

$$\|f^* - \bar{f}\|_2^2 \le \|f^* - f\|_2^2, \qquad \forall f \in \mathcal{F}. \tag{14}$$

Consider the penalty

$$\operatorname{pen}(f) = \log_2(m+1) + (3/2)\|\theta\|_0 \log_2(m), \qquad \theta = W^T f.$$

This corresponds to the following prefix code for $\theta \in \Theta$ (that is, we encode the elements of Θ, before they are subjected to the deterministic operation of projecting onto $\mathcal{C}$):

1) First we encode $\|\theta\|_0$, the number of nonzero components of θ, which can be encoded with $\log_2(m+1)$ bits.
2) For each of the $\|\theta\|_0$ nonzero components, we store its location in the θ vector; since there are m possible locations, this takes $\log_2(m)$ bits per component.
3) Next we encode each coefficient value, quantized to one of $\sqrt{m}$ uniformly sized bins.

Since this corresponds to a uniquely decodable code for $f \in \mathcal{F}$ (or $\theta \in \Theta$), we see that $\operatorname{pen}(f)$ satisfies the Kraft inequality.

Now, given $\theta^* = W^T f^*$, let $\theta^{(k)}$ be its best k-term approximation, $\theta_q^{(k)} \in \Theta$ the quantized version of $\theta^{(k)}$, for which we have

$$\left\|\frac{\theta_q^{(k)}}{I} - \frac{\theta^{(k)}}{I}\right\|_2^2 \le \frac{k}{m},$$

and $\bar{\theta}_q^{(k)}$ the element of Γ obtained by projecting $f_q^{(k)} = W\theta_q^{(k)}$ onto $\mathcal{C}$ and then transforming back into the basis Φ: $\bar{\theta}_q^{(k)} = W^T \bar{f}_q^{(k)}$. Then, using (14) and (13), we get

$$
\begin{aligned}
\|f^* - \bar{f}_q^{(k)}\|_2^2 &\le \|f^* - f_q^{(k)}\|_2^2 \\
&= \|\theta^* - \theta_q^{(k)}\|_2^2 \\
&\le 2\|\theta^* - \theta^{(k)}\|_2^2 + 2\|\theta^{(k)} - \theta_q^{(k)}\|_2^2 \\
&\le I^2\left(2Ck^{-2\alpha} + \frac{2k}{m}\right).
\end{aligned}
$$

Given each $1 \le k \le m$, let $\Gamma_k \subseteq \Gamma$ be the set of all $\bar{\theta} \in \Gamma$, such that the corresponding $\theta \in \Theta$ satisfies $\|\theta\|_0 \le k$. Then $|\Gamma_k| = \binom{m}{k} m^{k/2}$, so that $\log_2 |\Gamma_k| \le 2k \log_2 m$, and therefore $\Gamma_k \in \mathcal{G}_N$ whenever $k \le k_*(N)$, where $k_*(N) \triangleq$

$N/(2c_4 \log_2 m)$. Then the first term on the right-hand side of (10) can be bounded by

$$C_N \min_{1 \le k \le k_*(N)} R^*(f^*, \Gamma_k)$$

$$\le O(N) \min_{1 \le k \le k_*(N)} \left[\left\| \frac{\theta^*}{I} - \frac{\bar{\theta}_q^{(k)}}{I} \right\|_2^2 + \frac{2 \operatorname{pen}(f_q^{(k)})}{I} \right]$$

$$\le O(N) \min_{1 \le k \le k_*(N)} \left[k^{-2\alpha} + \frac{k}{m} + \frac{k \log_2 m}{I} \right],$$

where the constant obscured by the $O(\cdot)$ notation depends only on C and c. We can now consider two cases:

1) $I \le m \log m$, i.e., the penalty term dominates the quantization error, then we get the risk bound $\mathbb{E}R(f^*, \widehat{f}) \le$

$$O(N) \min_{1 \le k \le k_*(N)} \left[k^{-2\alpha} + \frac{2k \log_2 m}{I} \right] + \frac{2c_2^2 \log(c_2 m/N)}{N}.$$

If $k_*(N) \ge (\alpha I / \log_2 m)^{1/(2\alpha+1)}$, then we can further obtain

$$\mathbb{E}R(f^*, \widehat{f}) \le O(N) \left(\frac{I}{\log m} \right)^{-\frac{2\alpha}{2\alpha+1}} + \frac{2c_2^2 \log(c_2 m/N)}{N}.$$

If $k_*(N) < (\alpha I / \log_2 m)^{1/(2\alpha+1)}$, there are not enough measurements, and the estimator saturates, although its risk can be controlled.

2) $I > m \log m$, i.e., the quantization error dominates the penalty term. Then we obtain $\mathbb{E}R^*(f^*, \widehat{f}) \le$

$$O(N) \min_{1 \le k \le k_*(N)} \left[k^{-2\alpha} + \frac{2k}{m} \right] + \frac{2c_2^2 \log(c_2 m/N)}{N}.$$

If $k_*(N) \ge (\alpha m)^{1/(2\alpha+1)}$, then we can further get

$$\mathbb{E}R(f^*, \widehat{f}) \le O(N) m^{-\frac{2\alpha}{2\alpha+1}} + \frac{2c_2^2 \log(c_2 m/N)}{N},$$

Again, if $k_*(N) < (\alpha m)^{1/(2\alpha+1)}$, there are not enough measurements, and the estimator saturates.

Note that, when $I \asymp m$ and $N \asymp m^{1/p}$ for some $p > 1 + 1/2\alpha$, we get (up to log factors) the rates

$$\mathbb{E}R(f^*, \widehat{f}) = O\left(m^{-\beta} \right),$$

where $\beta = \frac{2\alpha - (2\alpha+1)/p}{2\alpha+1} > 0$.

VI. CONCLUSION

We have derived upper bounds on the compressed sensing estimation error under Poisson noise for sparse or compressible signals. We specifically prove error decay rates for the case where the penalty term is proportional to the ℓ_0-norm of the solution; this form of penalty has been used effectively in practice with a computationally efficient Expectation-Maximization algorithm (cf. [16]), but was lacking the theoretical support provided by this paper. Furthermore, the main theoretical result of this paper holds for any penalization scheme satisfying the Kraft inequality, and hence can be used to assess the performance of a variety of potential reconstruction strategies besides sparsity-promoting reconstructions.

One significant aspect of the bounds derived in this paper is that they grow with N, the size of the measurement array, which is a major departure from similar bounds in the Gaussian or bounded-noise settings. It does not appear that this is a simple artifact of our analysis. Rather, this behavior can be intuitively understood to reflect that elements of y will all have similar values at low light levels, making it very difficult to infer the relatively small variations in $\widetilde{A}f^*$. Hence, Poisson compressed sensing using shifted Rademacher sensing matrices is fundamentally difficult when the data are very noisy. It may be possible to address these limitations through alternative constructions of sensing matrices which introduce more variation in the signal Af^*.

REFERENCES

[1] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

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

[3] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] J. Haupt and R. D. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[5] D. L. Snyder, A. M. Hammond, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Amer. A*, vol. 10, pp. 1014–1023, 1993.

[6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 15, no. 12, pp. 4203–4215, 2005.

[7] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[9] W. Dai and O. Milenkovic, "Weighted superimposed codes and constrained integer compressive sensing," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2215–2229, May 2009.

[10] M. A. Khajehnejad, A. G. Dimakis, W. Xu, and B. Hassibi, "Sparse recovery of positive signals with minimal expansion," 2009, submitted.

[11] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geom. Funct. Anal.*, vol. 17, pp. 1248–1282, 2007.

[12] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Annals of Stat.*, vol. 19, no. 4, pp. 2032–2066, 1991.

[13] Q. Li and A. Barron, "Mixture density estimation," in *Advances in Neural Information Processing*, vol. 12. MIT Press, 2000.

[14] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of Stat.*, vol. 32, pp. 500–527, 2004.

[15] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford Univ. Press, 1997.

[16] M. Gehm, R. John, D. Brady, R. Willett, and T. Schultz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14 013–14 027, 2007.