POISSON COMPRESSED SENSING

Rebecca Willett and Maxim Raginsky

Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA

ABSTRACT

Compressed sensing has profound implications for the design of new imaging and network systems, particularly when physical and economic limitations require that these systems be as small and inexpensive as possible. However, several aspects of compressed sensing theory are inapplicable to real-world systems in which noise is signal-dependent and unbounded. In this work we discuss some of the key theoretical challenges associated with the application of compressed sensing to practical hardware systems and develop performance bounds for compressed sensing in the presence of Poisson noise. We develop two novel sensing paradigms, based on either pseudo-random dense sensing matrices or expander graphs, which satisfy physical feasibility constraints. In these settings, as the overall intensity of the underlying signal increases, an upper bound on the reconstruction error decays at an appropriate rate (depending on the compressibility of the signal), but for a fixed signal intensity, the error bound actually grows with the number of measurements or sensors. This surprising fact is both proved theoretically and justified based on physical intuition.

1. INTRODUCTION

The goal of *compressive sampling* or *compressed sensing* (CS) [1,2] is to replace conventional sampling by a more efficient data acquisition framework, which generally requires fewer sensing resources. This paradigm is particularly enticing whenever the measurement process is costly or constrained in some sense. For example, in the context of photon-limited applications (such as low-light imaging), the photomultiplier tubes used within sensor arrays are physically large and expensive. Similarly, when measuring network traffic flows, the high-speed memory used in packet counters is cost-prohibitive. These problems appear ripe for the application of CS.

However, photon-limited measurements [3, 4] and arrivals/departures of packets at a router [5] are commonly modeled with a Poisson probability distribution, posing significant theoretical and practical challenges in the context of CS. One of the key challenges is the fact that the measurement error variance scales with the true intensity of each measurement, so that we cannot assume constant noise variance across the collection of measurements. Futhermore, the measurements, the underlying true intensities, and the system models are all subject to certain physical constraints, which play a significant role in performance. Recent work has *empirically* explored CS in the context of photon limited measurements [6–9], but theoretical performance bounds similar to those widely cited in the conventional CS context previously remained elusive. This is in part because the standard assumption of signal-independent and/or bounded noise (cf. [10,11]) is violated under the Poisson noise models.

This work was supported by NSF CAREER Award No. CCF-06-43947, DARPA Grant No. HR0011-07-1-003, and NSF Grant DMS-08-11062.

The Poisson observation model considered in this paper is

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

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 $y \in \mathbb{Z}_+^N$ is a length-N vector of observed Poisson counts, stipulates that the likelihood of observing a particular vector of counts y is given by

$$\mathbb{P}_{Af^*}(y) = \prod_{j=1}^{N} \mathbb{P}_{(Af^*)_j}(y_j) \tag{1}$$

where $(Af^*)_j$ is the j^{th} component of Af^* and

$$\mathbb{P}_\lambda(z) \triangleq \begin{cases} \frac{\lambda^z}{z!}e^{-\lambda}, & \text{if } \lambda > 0 \\ 1_{\{z=0\}} & \text{otherwise} \end{cases}.$$

Moreover, in order to correspond to a physically realizable linear optical system, the measurement matrix A must be:

- **Positivity** — every element of the sensing matrix A must be nonnegative. This restriction reflects the physical limitations of many sensing systems of interest (e.g., packet routers and counters or linear optical systems) since the associated hardware systems can aggregate events (e.g., photons or packets) but not measure differences. An important byproduct of this restriction is that for any nonnegative input signal f, the projected signal Af must also be nonnegative. Using the standard notation $f \succeq 0$ to denote the nonnegativity of f, we can write this condition as

$$f \succeq 0 \qquad \Longrightarrow \qquad Af \succeq 0.$$

- **Flux-preserving** — for any input signal $f \succeq 0$, the mean total intensity of the observed signal Af must not exceed the total intensity incident upon the system:

$$\sum_{i=1}^{N}(Af)_i \le \sum_{i=1}^{m} f_i.$$

In this paper, we consider two different classes of compressive sensing paradigms: A is a pseudo-random dense matrix, or A corresponds to the normalized adjacency matrix of an expander graph. While the technical details of the ideas presented here have been established in our previous work [12, 13], this paper focuses on the tradeoffs between these two potential sensing modes. We will see that the pseudo-random model must be altered from common CS models to ensure positivity and flux-preservation; this alteration plays a significant role in performance bounds. The expander model is a more natural fit to the Poisson CS problem, in that positivity and flux-preservation are immediately guaranteed, but performance hinges on f^* being sparse (as opposed to sparse in some arbitrary basis). In both cases, the performance bounds are less favorable than the performance predicted by CS theory which does not account for physical constraints associated with Poisson noise.

The quality of a candidate estimator f is measured in terms of the *risk*

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

The main theoretical results presented in this paper shows that, for N sufficiently large and for an α-compressible signal of total intensity I[1],

$$R\left(\widehat{f}, f^*\right) = O\left[N \left(\frac{\log m}{I} \right)^{\frac{2\alpha}{2\alpha+1}} + \frac{\log(m/N)}{N} \right]$$

for pseudo-random dense sensing matrices and

$$R\left(\widehat{f}, f^*\right) = O\left[N \left(\frac{\log m}{NI} \right)^{\frac{2\alpha}{2\alpha+1}} + \left(\frac{\log(m/N)}{N} \right)^{2\alpha} \right]$$

for expander sensing matrices. The expander bounds are somewhat better, but this result is confined to the case where f^* itself is compressible, while the pseudo-random dense sensing matrix result holds when f^* is compressible in some orthonormal basis.

While the rate of error decay as a function of the total intensity, I, coincides with earlier results in denoising contexts, the proportionality of the intensity-dependent term in the error to N may seem surprising at first glance. However, one can intuitively understand this result from the following perspective. If we increase the number of measurements (N) while keeping the expected number of observed photons (I) constant[2], the number of photons per sensor will decrease, so the signal-to-noise ratio (SNR) at each sensor will likewise decrease, thereby degrading performance. Having the number of sensors exceed the number of observed photons is not necessarily detrimental in a *denoising* or *direct measurement* setting (i.e., where A is the identity matrix) because multiscale algorithms can adaptively bin the noisy measurements together in homogeneous regions to achieve higher SNR overall [14, 15]. However, in the CS setting the signal is first altered by the compressive projections in the sensing matrix A, and the raw measurements cannot themselves be binned to improve SNR. In particular, there is no natural way to aggregate measurements across multiple sensors because the aggregation effectively changes the sensing matrix in a way that does not preserve critical properties of A.

One might also be surprised by this main result because in the case where the number of observed photons is very large (so that SNR is quite high and not a limiting factor), our bounds do not converge to the standard performance bounds in CS. This is because our bounds pertain to a sensing matrix A which, unlike conventional CS matrices based on i.i.d. realizations of a zero-mean random variable, is designed to correspond to a feasible physical system. In particular, every element of A must be nonnegative and appropriately scaled, so that the observed photon intensity is no greater than the photon intensity incident on the system (i.e., we cannot measure more light than is available). This rescaling dramatically impacts important elements of any performance bounds, including the form of the restricted isometry property [16, 17], even in the case of Gaussian or bounded noise. (Additional details and interpretation are provided in Section 3.2 after we introduce necessary concepts and notation.)

As a result, incorporating these real-world constraints into our measurement model has a *significant and adverse impact on the expected performance of a Poisson CS system*.

[1] More precisely, I refers to the total intensity integrated over the exposure time, so that increasing I can be associated with more source intensity, longer exposure time per measurement, or both.

[2] In some systems, such as a single-detector system, more measurements might seem to suggest more observed photons. However, holding I fixed while increasing N implies that each measurement is collected over a shorter exposure. Thus increasing N does *not* correspond to an increase in the number of observed events/photons.

2. PROBLEM FORMULATION

We have a signal or image $f^* \succeq 0$ of size m that we wish to estimate using a detector array of size $N \ll m$. We assume that the total intensity of f^*, given by $I \triangleq \|f^*\|_1 = \sum_{i=1}^m f_i^*$, is known *a priori*. We make Poisson observations of Af^*, $y \sim \text{Poisson}(Af^*)$, where $A \in \mathbb{R}^{N \times m}$ is a positivity- and flux-preserving sensing matrix. Our goal is to estimate $f^* \in \mathbb{R}_+^m$ from $y \in \mathbb{Z}_+^N$.

To recover f^*, we will use a penalized Maximum Likelihood Estimation (pMLE) approach. Let us choose a finite or countable set Γ_I of candidate estimators $f \in \mathbb{R}_+^m$ with $\|f\|_1 = I$, and a *penalty* $\text{pen} : \Gamma_I \to \mathbb{R}_+$ satisfying the *Kraft inequality*[3]

$$\sum_{f \in \Gamma_I} e^{-\text{pen}(f)} \leq 1. \qquad (3)$$

For instance, we can impose less penalty on sparser signals or construct a penalty based on any other prior knowledge about the underlying signal. With these definitions, we consider the following *penalized maximum likelihood estimator (pMLE)*:

$$\widehat{f} \triangleq \underset{f \in \Gamma_I}{\arg\min} \left[-\log \mathbb{P}_{Af}(y) + 2\tau \,\text{pen}(f) \right] \qquad (4)$$

(Our theoretical results use $\tau = 1$. In practice, however, one often prefers to use a value of τ different from what is supported in theory because of slack in the bounds.) One way to think about the procedure in (4) is as a Maximum *a posteriori* Probability (MAP) algorithm over the set of estimates Γ_I, where the likelihood is computed according to the Poisson model (1) and the penalty function corresponds to a negative log prior on the candidate estimators in Γ_I.

The penalty term may be chosen, for instance, to be smaller for sparser solutions $\theta = \Phi^T f$, where Φ is an orthogonal matrix that represents f in its "sparsifying" basis. In fact, a variety of penalization techniques can be used in this framework; see [14, 18] for examples and discussions relating Kraft-compliant penalties to prefix codes for estimators. Many penalization or regularization methods in the literature, if scaled appropriately, can be considered prefix codelengths.

We bound the accuracy with which we can estimate f^*/I according to (2); in other words, we focus on accurately estimating the *distribution* of intensity in f^* independent of any scaling factor proportional to the total intensity of the scene, which is typically of primary importance to practitioners. Since the total number of observed events, n, obeys a Poisson distribution with mean I, estimating I by n is the strategy employed by most methods. However, the variance of this estimate is I, which means that, as I increases,

[3] Many penalization functions can be modified slightly (e.g., scaled appropriately) to satisfy the Kraft inequality. All that is required is a finite collection of estimators (i.e., Γ_I) and an associated prefix code for each candidate estimate in Γ_I. For instance, this would certainly be possible for a total variation penalty, though the details are beyond the scope of this paper.

our ability to estimate the distribution improves, while accurately estimating the unnormalized intensity is more challenging. We chose to assume I is known to discount this effect.

2.1. Summary of notation

Before proceeding to state and prove risk bounds for the proposed estimator, we summarize for the reader's convenience the principal notation used in the sequel:

- m: dimension of the original signal
- $N(\ll m)$: number of measurements (detectors)
- $f^* \in \mathbb{R}_+^m$: unknown nonnegative-valued signal
- $I = \sum_i f_i^*$: total intensity of f^*, assumed known
- $\Gamma_I \subset \mathbb{R}_+^m$: finite or countably infinite set of candidate estimators with a penalty function $\mathrm{pen} : \Gamma_I \to \mathbb{R}_+$ satisfying the Kraft inequality (3)
- $R(f, f^*) = \mathbb{E}\left[\|f^* - f\|_2/I\right]^2$: the risk of a candidate estimator f
- $\widehat{f}$: the penalized maximum-likelihood estimator taking values in Γ_I, given by the solution to (4)

3. DENSE PSEUDO-RANDOM SENSING MATRICES

3.1. Construction and properties of the sensing matrix

We construct our sensing matrix A as follows. Let Z be an $N \times m$ matrix whose entries are i.i.d. according to

$$Z_{i,j} = \begin{cases} -1 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases}.$$

We observe that

$$\mathbb{E}Z_{i,j} = 0 \qquad \text{and} \qquad \mathbb{E}Z_{i,j}Z_{k,\ell} = \delta_{ik}\delta_{j\ell}$$

for all $1 \leq i, k \leq N$ and $1 \leq j, \ell \leq m$. Most compressed sensing approaches would proceed by assuming that we make (potentially noisy) observations of $\widetilde{A}f^*$, where $\widetilde{A} \triangleq Z/\sqrt{N}$. However, $\widetilde{A}$ will, with high probability, have at least one negative entry, which will render this observation model physically unrealizable in physical systems of interest. Therefore, we use $\widetilde{A}$ to generate a feasible 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 take

$$A \triangleq \frac{1}{2\sqrt{N}}\widetilde{A} + \frac{1}{2N}\mathbf{1}_{N \times m}.$$

We can immediately deduce the following properties of A:

- It is positivity-preserving because each of its entries is either 0 or $1/N$.
- It is flux-preserving, i.e., for any $f \in \mathbb{R}_+^m$ we have
$$\|Af\|_1 \leq \|f\|_1.$$
- With probability at least $1 - N2^{-m}$ (w.r.t. the realization of $\{Z_{i,j}\}$), every row of A has at least one nonzero entry. Assume that this event holds. 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
$$Af \succeq \frac{cI}{N}\mathbf{1}_{N \times 1}.$$
- Furthermore, and most importantly, 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) [16, 17], where the subset of interest consists of all vectors with a given sparsity.

3.2. DC offset and noise

The intensity underlying our Poisson observations can be expressed as

$$Af^* = \frac{1}{2\sqrt{N}}\widetilde{A}f^* + \frac{I}{2N}\mathbf{1}_{N \times 1}.$$

The *idealized* sensing matrix $\widetilde{A}$ has a RIP-like property which can lead to certain performance guarantees if we could measure $\widetilde{A}f^*$ directly; in this sense, $\widetilde{A}f^*$ is the *informative* component of each measurement. However, a constant DC offset proportional to I is added to each element of $\widetilde{A}f^*$ before Poisson measurements are collected, and elements of $\widetilde{A}f^*$ will be very small relative to I. Thus the intensity and variance of each measurement will be proportional to I, overwhelming the informative elements of $\widetilde{A}f^*$. (To see this, note that y_i can be approximated as $(Af^*)_i + \sqrt{(Af^*)_i}\xi_i$, where ξ_i is a Gaussian random variable with variance one.)

As we will show in this paper, the Poisson noise variance associated with the DC offset, necessary to model feasible measurement systems, leads to very different performance guarantees than are typically reported in the CS literature. The necessity of a DC offset is certainly not unique to our choice of a Rademacher sensing matrix; it has been used in practice for a wide variety of linear optical CS architectures (cf. [19–22]). A notable exception to the need for DC offsets is the expander-graph approach to generating non-negative sensing matrices, as discussed in Section 4.1; however, theoretical results there are limited to signals which are sparse in the canonical (i.e. Dirac delta or pixel) basis.

3.3. Risk bounds for compressible signals

Following [23], we assume that there exists an orthonormal basis of $\mathbb{R}^m$ with matrix representation $\Phi = [\phi_1, \ldots, \phi_m]$ of $\mathbb{R}^m$, such that f^* is α-*compressible* (for some $\alpha > 0$) in Φ in the following sense. Let the vector θ^* denote the basis coefficients of f^*; i.e., $\theta_j^* = \langle f^*, \phi_j \rangle$ or $f^* = \Phi\theta^*$. Let $\theta_{(1)}^*, \ldots, \theta_{(m)}^*$ be the decreasing rearrangement of θ^*: $|\theta_{(1)}^*| \geq |\theta_{(2)}^*| \geq \ldots \geq |\theta_{(m)}^*|$. Given any $1 \leq k \leq m$, let $\theta^{(k)}$ denote the best k-term approximation to θ^*. Then

$$\frac{1}{I^2}\|\theta^* - \theta^{(k)}\|_2^2 = O(k^{-2\alpha}).$$

Theorem 1 *There exist a finite set of candidate estimators Γ_I corresponding to estimators of the form $f = \Phi\theta$ with each element of θ quantized to one of $\sqrt{m}$ levels and with the property*

$$f \succeq (cI)\mathbf{1}_{m \times 1}, \quad \forall f \in \Gamma_I$$

and a penalty function $\mathrm{pen}(f) \propto \|\Phi^T f\|_0 \log(m)$ satisfying Kraft's inequality, such that the bound

$$R\left(\widehat{f}, f^*\right) = O\left(N \min_{1 \leq k \leq k_*}\left[k^{-2\alpha} + \frac{k}{m} + \frac{k\log m}{I}\right] + \frac{\log(m/N)}{N}\right),$$

where $k_* = O(N/\log(m))$ holds with probability at least $1 - me^{-KN}$ for some positive K. The constants obscured by the $O(\cdot)$ are independent of m, n, α, and N.

The quantity k_* is related to the cardinality of Γ_I, which grows like 2^N and must be sufficiently high of our analysis to hold. This can be interpreted as a *threshold effect*, i.e., the existence of a critical number of measurements N^*, below which the expected risk may not monotonically decrease with N or I.

In the low-intensity setting $I \leq m\log(m)$, if $k_* \geq (\alpha I/\log m)^{1/(2\alpha+1)}$ (i.e. if we have sufficiently many measurements), then we can further obtain

$$R\left(\widehat{f}, f^*\right) = O\left[N\left(\frac{\log(m)}{I}\right)^{\frac{2\alpha}{2\alpha+1}} + \frac{\log(m/N)}{N}\right]. \quad (5)$$

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

The factor of N in this expression does not appear in conventional CS performance bounds, and is directly related to the modifications we made to the "ideal" sensing matrix $\widetilde{A}$ to make it positive and flux-preserving. In particular, consider that when $\widetilde{A}$ satisfies the RIP, we have $\|g\|_2^2 \approx \|\widetilde{A}g\|_2^2$ for all sparse vectors g. However, if we neglect the DC offset for the moment and just consider that $\widetilde{A} \propto \sqrt{N}A$, we see that in our setting $\|g\|_2^2 \approx N\|Ag\|_2^2$ for all sparse vectors g. This relationship was used twice in the proof of the above theorem and yielded the unfavorable result.

4. EXPANDER SENSING MATRICES

The above performance analysis suggested that taking a conventional CS sensing matrix and scaling and shifting it to satisfy physical sensing constraints yields bounds which scale unfavorably with N. At the heart of the problem was the fact that the RIP played a central role in our analysis. In an attempt to address this challenge, we turn our attention to another class of sensing matrices based on *expander graphs*. As detailed below, these sensing matrices have a RIP-like property in the ℓ_1 (as opposed to ℓ_2) norm which allows us to sidestep the above normalization error.

We emphasize that the below results are applicable when f^* (not its coefficients in an arbitrary orthonormal basis) are compressible. In the following, we assume

$$\frac{1}{I^2}\|f^* - f^{(k)}\|_1^2 = O(k^{-2\alpha})$$

where $f^{(k)}$ is the best k-term approximation of f^*.

4.1. Construction and properties of the sensing matrix

We start by defining an *unbalanced bipartite vertex-expander graph*.

Definition 1 *We say that a bipartite simple graph $G = (F, Y, E)$ with (regular) left degree[4] d is a (s,ϵ)-*expander *if, for any $S \subset F$ with $|S| \leq s$, the set of neighbors $\mathcal{N}(S)$ of S has size $|\mathcal{N}(S)| > (1-\epsilon)d\,|S|$.*

[4]That is, each node in F has the same number of neighbors in Y.

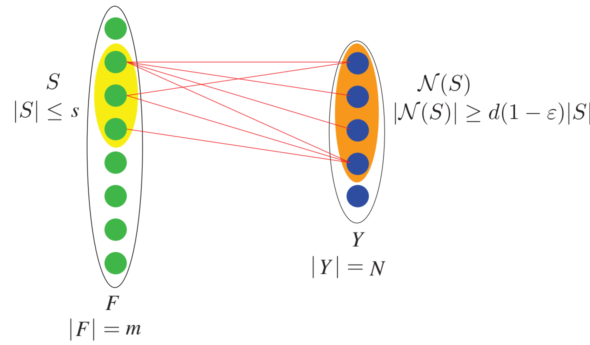

Fig. 1. A (s,ϵ)-expander. In this example, the green nodes correspond to F, the blue nodes correspond to Y, the yellow oval corresponds to the set $S \subset F$, and the orange oval corresponds to the set $\mathcal{N}(S) \subset Y$. There are three colliding edges.

Figure 1 illustrates such a graph. Intuitively a bipartite graph is an expander if any sufficiently small subset of its variable nodes has a sufficiently large neighborhood. In the CS setting, F (resp., Y) corresponds to the components of the original signal (resp., its compressed representation). Hence, for a given $|F|$, a "high-quality" expander should have $|Y|$, d, and ϵ as small as possible, while s should be as close as possible to $|Y|$.

Expanders have been recently proposed as a means of constructing efficient compressed sensing algorithms [24–27]. In particular, it has been shown that any m-dimensional vector that is s-sparse can be fully recovered using $O\left(s\log\left(\frac{m}{s}\right)\right)$ measurements in $O\left(m\log\left(\frac{m}{s}\right)\right)$ time [24, 26]. It has been also shown that, even in the presence of noise in the measurements, if the noise vector has low ℓ_1 norm, expander-based algorithms can approximately recover any s-sparse signal [25,26,28]. One reason why expander graphs are good sensing candidates is that the adjacency matrix of any (s,ϵ)-expander almost preserves the ℓ_1 norm of any s-sparse vector [25]. In other words, if the adjacency matrix of an expander is used for measurement, then the ℓ_1 distance between two sufficiently sparse signals is preserved by measurement. This property is known as the "Restricted Isometry Property for ℓ_1 norms" or the "RIP-1" property. Berinde et al. have shown that this condition is sufficient for sparse recovery using ℓ_1 minimization [25].

The precise statement of the RIP-1 property, whose proof can be found in [24], goes as follows:

Lemma 1 (RIP-1 property of the expander graphs) *Let B be the $N \times m$ adjacency matrix of a (s,ϵ) expander graph G. Then for any s-sparse vector $f \in \mathbb{R}^m$ we have:*

$$(1-2\epsilon)d\|f\|_1 \leq \|Bf\|_1 \leq d\|f\|_1$$

For future convenience, we introduce the following piece of notation. Given m and $1 \leq s \leq m/4$, we denote by $G_{s,m}$ a $(2s, 1/16)$-expander with left set size m whose existence is guaranteed [13]. Then $G_{s,m} = (F, Y, E)$ has

$$|F| = m, \quad |Y| = N = O(s\log(m/s)), \quad d = O(\log(m/s)).$$

4.2. Risk bounds for compressible signals

Let us choose a convenient $1 \leq s \leq m/4$ and take A to be the normalized adjacency matrix of the expander $G_{s,m}$: $A \triangleq B/d$.

Our main bound on the performance of the pMLE is as follows:

Theorem 2 *Let A be the normalized adjacency matrix of $G_{s,m}$, let $f^* \in \mathbb{R}_+^m$ be the original signal compressively sampled in the presence of Poisson noise and let $\widehat{f}$ be obtained through* (4). *Then*

$$R\left(\widehat{f}, f^*\right) = O\left(s^{-2\alpha} + \min_{f \in \Gamma_I}\left[\frac{KL(\mathbb{P}_{Af^*} \parallel \mathbb{P}_{Af})}{I} + \frac{2\,\mathrm{pen}(f)}{I}\right]\right),$$

where

$$KL(\mathbb{P}_g \| \mathbb{P}_h) \triangleq \sum_{y \in \mathbb{Z}_+^N} \mathbb{P}_g(y) \log \frac{\mathbb{P}_g(y)}{\mathbb{P}_h(y)}$$

is the Kullback–Leibler divergence (relative entropy) between $\mathbb{P}_g$ and $\mathbb{P}_h$ [29].

4.3. A bound in terms of ℓ_1 error

At first blush, Theorem 2 is encouraging because there is no poor scaling with N; this is a results of using RIP-1 in our analysis. However, the bound of Theorem 2 is not always useful since it bounds the ℓ_1 risk of the pMLE in terms of the relative entropy. A bound purely in terms of ℓ_1 errors would be more desirable, but it is not easy to obtain without imposing extra conditions either on f^* or on the candidate estimators in Γ_I. This follows from the fact that the divergence $\mathrm{KL}(\mathbb{P}_{Af^*} \| \mathbb{P}_{Af})$ may take the value $+\infty$ if there exists some y such that $\mathbb{P}_{Af}(y) = 0$ but $\mathbb{P}_{Af^*}(y) > 0$.

One way to eliminate this problem is to impose an additional requirement on the candidate estimators in Γ_I: there exists some $c > 0$, such that

$$Af \succeq c, \qquad \forall f \in \Gamma_I \tag{6}$$

Because every $f \in \Gamma_I$ satisfies $\|f\|_1 = I$, the constant c cannot be too large. In particular, a necessary condition for (6) to hold is $c \leq I/N$. Under this condition, we now develop a risk bound for the pMLE purely in terms of the ℓ_1 error.

Theorem 3 *Suppose that all the conditions of Theorem 2 are satisfied. In addition, suppose that the set Γ_I satisfies the condition* (6). *Then*

$$R\left(\widehat{f}, f^*\right) = O\left(s^{-2\alpha} + \min_{f \in \Gamma_I}\left[N\frac{\|f - f^*\|_1^2}{I^2} + \frac{\mathrm{pen}(f)}{I}\right]\right). \tag{7}$$

Effectively, this means that, under the positivity condition (6), the ℓ_1 error of $\widehat{f}$ is the sum of the s-term approximation error of f^* plus N times the best penalized squared ℓ_1 approximation error. The first term in (7) is smaller for sparser f^*, and the second term is smaller when there is a $f \in \Gamma_I$ which is simultaneously a good ℓ_1 approximation to f^* and has a low penalty.

If we consider $\mathrm{pen}(f) \propto \|f\|_0 \log(m)$ and note $s \approx \log(m/N)/N$, we have

$$R\left(\widehat{f}, f^*\right) = O\left[N\left(\frac{\log(m)}{NI}\right)^{\frac{2\alpha}{2\alpha+1}} + \left(\frac{\log(m/N)}{N}\right)^{2\alpha}\right]. \tag{8}$$

5. CONCLUSIONS

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. [20]), but was lacking the theoretical support provided by this paper.

In particular, in studying Poisson compressed sensing we examined two different measurement paradigms: one based on dense pseudorandom matrices generated using a Bernoulli distribution, and one based on the adjacency matrices of expander graphs. We found that the expander graphs are a very natural fit to the physical constraints often associated with Poisson noise and yield somewhat better performance bounds in (8); however, these bounds are only applicable when the signal (as opposed to a basis representation) is sparse or compressible. Pseudorandom dense matrices yield larger bounds which are applicable to a broader class of sparsity and compressibility models in (5).

One significant aspect of the bounds derived in this paper is that their signal-dependent portions grow with N, the size of the measurement array, which is a major departure from similar bounds in the Gaussian or bounded-noise settings. While lower bounds are not yet known, 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 $\widetilde{A}f^*$. These results suggest that compressed sensing may be *fundamentally* difficult when the data are Poisson observations.

6. REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[2] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] 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.

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

[5] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, 1992.

[6] D. Lingenfelter, J. Fessler, and Z. He, "Sparsity regularization for image reconstruction with Poisson data," in *Proc. SPIE Computational Imaging VII*, 2009, vol. 7246.

[7] Z. Harmany, R. Marcia, and R. Willett, "Sparse Poisson intensity reconstruction algorithms," in *IEEE Workshop on Statistical Signal Processing*, 2009, Available at http://arxiv.org/abs/0905.0483.

[8] M. Figueiredo and J. Bioucas-Dias, "Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization," in *Proc. IEEE Workshop on Statistical Signal Processing*, 2009, Available at http://arxiv.org/abs/0904.4868.

[9] J.-L. Starck and J. Bobin, "Astronomical data analysis and sparsity: from wavelets to compressed sensing," *Proc. IEEE: Special Issue on Applications of Sparse Representation and Compressive Sensing*, 2010, In press.

[10] 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.

[11] J. Haupt and R. D. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, September 2006.

[12] 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.

[13] M. Raginsky, S. Jafarpour, Z. Harmany, R. Marcia, R. Willett, and R. Calderbank, "Performance bounds for expander-based compressed sensing in poisson noise," Accepted for publication in *IEEE Transactions on Signal Processing*, 2010.

[14] R. Willett and R. Nowak, "Multiscale Poisson intensity and density estimation," *IEEE Trans. Inform. Theory*, vol. 53, no. 9, pp. 3171–3187, September 2007.

[15] R. Willett and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 332–350, 2003.

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

[17] 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, March 2006.

[18] K. Krishnamurthy, M. Raginsky, and R. Willett, "Multiscale photon-limited hyperspectral image reconstruction," Submitted to SIAM Journal on Imaging Sciences, 2010.

[19] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[20] M. Gehm, R. John, D. Brady, R. Willett, and T. Schultz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[21] R. DeVerse, R. Coifman, A. Coppi, W. Fateley, F. Geshwind, R. Hammaker, S. Valenti, F. Warner, and G. Davis, "Application of spatial light modulators for new modalities in spectrometry and imaging," in *Proc. of SPIE*, 2003, vol. 4959.

[22] M.A. Neifeld and J. Ke, "Optical architectures for compressive imaging," *Appl. Opt.*, vol. 46, pp. 5293–5303, 2007.

[23] 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, December 2006.

[24] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 4299–4308, September 2009.

[25] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: a unified approach to sparse signal recovery.," *46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 798–805, September 2008.

[26] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the $\ell 1$ norm," *Proc. 49th Ann. IEEE Symp. on Foundations of Computer Science (FOCS)*, pp. 199–207, 2008.

[27] V. Guruswami, J. Lee, and A. Razborov, "Almost euclidean subspaces of ℓ_1 via expander codes," *Proceedings of the 19th annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 353–362, January 2008.

[28] R. Berinde, P. Indyk, and M. Ruzic, "Practical near-optimal sparse recovery in the ℓ_1 norm," *46th Annual Allerton Conf. on Comm., Control, and Computing*, 2008.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 2 edition, 2006.