A Concentration-of-Measure Inequality for Multiple-Measurement Models

Liming Wang[†], Jiaji Huang[†], Xin Yuan[†], Volkan Cevher[‡], Miguel Rodrigues[⋆],
Robert Calderbank[†] and Lawrence Carin[†]
[†]Dept. of Electrical & Computer Engineering, Duke University, Durham, NC 27708, USA
[‡] Laboratory for Information and Inference Systems, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland
[⋆]Dept. of Electronic & Electrical Engineering, University College London, London, U.K.

***Abstract*—Classical compressive sensing typically assumes a single measurement, and theoretical analysis often relies on corresponding concentration-of-measure results. There are many real-world applications involving multiple compressive measurements, from which the underlying signals may be estimated. In this paper, we establish a new concentration-of-measure inequality for a block-diagonal structured random compressive sensing matrix with Rademacher-ensembles. We discuss applications of this newly-derived inequality to two appealing compressive multiple-measurement models: for Gaussian and Poisson systems. In particular, Johnson-Lindenstrauss-type results and a compressed-domain classification result are derived for a Gaussian multiple-measurement model. We also propose, as another contribution, theoretical performance guarantees for signal recovery for multi-measurement Poisson systems, via the inequality.**

I. Introduction

Recently there has been considerable interest in compressive sensing (CS), with the goal of compressively measuring a high-dimensional signal with minimal loss of information [1], [2], [3], [4]. Linear compressive measurements have attracted particular attention, given their simplicity, with compressive measurement realized by projecting the high-dimensional signal $\mathbf{f} \in \mathbb{R}^n$ using a sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ with $m \ll n$.

In particular, it has been revealed that certain classes of randomly constituted sensing matrices facilitate perfect reconstruction of the original high-dimensional signal with overwhelming probability, via tractable l_1 or iterative methods [5], [6], [7]. The theoretical analysis often necessitates the notion that the norm of the signal $\mathbf{f}$ is approximately preserved under $\mathbf{\Phi}$ with high probability, and such a notion is formalized via the *concentration-of-measure* phenomenon [8]. It has been found that required concentration-of-measure results can be manifested by posing proper sub-Gaussian ensembles on $\mathbf{\Phi}$ [9].

In many practical applications, the input data may be naturally presented in discrete blocks, with each block sequentially measured by the system. Alternatively, it may be impractical to consider all data as one input signal $\mathbf{f}$, because of the resources and computational constraints of the acquisition system. Typical examples include distributed sensing systems in which each sensing entity only has access to a subset of the input data, and sensing systems for streaming signals, such as video, that require sequential sensing operations on each frame. For all these situations, multiple measurements $\{\mathbf{y}_k\}$ are sequentially collected for the input signal sequence $\{\mathbf{f}_k\}$ via the identical sensing matrix $\mathbf{\Phi}$.

Almost all previous work is directed towards the case of a *single* compressive measurement, and corresponding concentration-of-measure results for the single-measurement model are formulated [7], [9]. To the best of our knowledge, there are very few papers on deriving concentration-of-measure results for a multiple-measurement model, except for [10], [11] in which a concentration-of-measure inequality for the multiple-measurement model with identical sensing matrix is established via Gaussian-ensembles on $\mathbf{\Phi}$.

In this paper we derive a new concentration-of-measure inequality for a block-diagonal sensing matrix with Rademacher-ensembles, which aims to establish theoretical properties for multiple-measurement models akin to single measurement counterparts.

Our second goal is to adapt the newly-derived concentration-of-measure inequality to the Poisson multiple-measurement model, in view of numerous applications of the Poisson sensing model in X-ray [12], chemical imaging [13], [14] and document classification [14], [15]. In contrast to results in [10], [11] where Gaussian emsembles are manifested, the non-negativity constraint on the Poisson sensing matrix can be readily satisfied for our proposed Rademacher-ensembled $\mathbf{\Phi}$, by adding a constant offset. We then propose maximum-likelihood estimators for signal recovery in the Poisson multiple-measurement model, and theoretical performance guarantees are derived based on the new concentration-of-measure inequality.

The remainder of the paper is organized as follows. Section II introduces two multiple-measurement models of interest, *i.e.,* Gaussian and Poisson models. Section III describes a sensing matrix constructed using Rademacher-ensembles, and derives a new concentration of measure inequality. Section IV discusses several applications of the concentration-of-measure inequality to stable embedding and compressive classification for Gaussian multiple-measurement systems. Section V presents an application to the signal-recovery problem for multiple Poisson measurements and establishes theoretical guarantees for the proposed estimators. Section VI concludes the paper.

978-1-4673-7704-1/15/$31.00 ©2015 IEEE

II. Multiple-Measurement Models

Let $\mathbf{f}_k \in \mathbb{R}^n$ denote the k-th input signal of the sensing system. For a Gaussian multiple-measurement model, the k-th measurement $\mathbf{y}_k \in \mathbb{R}^m$ is obtained via the following random transformation:

$$\mathbf{y}_k = \mathbf{\Phi f}_k + \mathbf{w}_k, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ is the sensing matrix with $m \ll n$ and $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ is Gaussian noise. Assuming K such measurements, (1) can be equivalently rewritten as

$$\mathbf{y} = \mathbf{Af} + \mathbf{w}, \tag{2}$$

where

$$\mathbf{y} \stackrel{\text{def}}{=} \left[\mathbf{y}_1^\top, \ldots, \mathbf{y}_K^\top\right]^\top, \tag{3}$$

$$\mathbf{f} \stackrel{\text{def}}{=} \left[\mathbf{f}_1^\top, \ldots, \mathbf{f}_K^\top\right]^\top, \tag{4}$$

$$\mathbf{w} \stackrel{\text{def}}{=} \left[\mathbf{w}_1^\top, \ldots, \mathbf{w}_K^\top\right]^\top. \tag{5}$$

$\mathbf{A} \in \mathbb{R}^{Km\times Kn}$ is constituted via block-diagonalization $\mathbf{A} \stackrel{\text{def}}{=} \mathbf{I}_K \otimes \mathbf{\Phi}$, $\mathbf{I}_K$ denotes the $K \times K$ identity matrix, and $\otimes$ represents the Kronecker (tensor) product.

The Poisson model serves as another important example, and its k-th measurement $\mathbf{y}_k \in \mathbb{Z}_+^m$ is a Poisson-distributed count vector:

$$\mathbf{y}_k \sim \operatorname{Pois}(\mathbf{\Phi f}_k) \stackrel{\text{def}}{=} \prod_{i=1}^m \operatorname{Pois}((\mathbf{\Phi f}_k)_i), \quad \forall k = 1, \ldots, K, \tag{6}$$

where $\mathbf{f}_k \succeq 0$, for all $k = 1, \ldots, K$ and $\succeq$ denotes the entry-wise inequality. $\mathbf{\Phi} \in \mathbb{R}_+^{m\times n}$ is the sensing matrix, with $m \ll n$, and $(\cdot)_i$ denotes the i-th entry of the argument vector. The first $\operatorname{Pois}(\cdot)$ in (6) has a vector argument for the rate, and corresponds to a Poisson distribution implemented independently on each component of the rate vector; the second $\operatorname{Pois}(\cdot)$ in (6) denotes the common scalar Poisson distribution with the argument rate. Akin to its Gaussian counterpart, the Poisson multiple-measurement model may be expressed concisely as

$$\mathbf{y} \sim \operatorname{Pois}(\mathbf{Af}). \tag{7}$$

Despite the resemblance of (2) and (7) to their single-measurement counterparts, the major difference is that $\mathbf{A}$ is required to be of block-diagonal structure for multiple-measurement models.

III. Concentration-of-Measure Inequality

A concentration-of-measure is a phenomenon describing the tendency of certain functions of a high-dimensional random process to concentrate sharply around their means [8]. It is common to randomly constitute the sensing matrix, from which a concentration-of-measure inequality can be derived [9]. We propose to constitute the matrix $\mathbf{\Phi}$ via Rademacher-ensembles as follows. First generate $\mathbf{Z} \in \{1, -1\}^{m\times n}$, with each entry drawn i.i.d. from the Rademacher distribution (*i.e.,* random variables take values 1 or -1 with equal probability). Let $\mathbf{\Phi} \stackrel{\text{def}}{=} \frac{\mathbf{Z}}{\sqrt{m}}$ and the block-diagonal sensing matrix is $\mathbf{A} = \mathbf{I}_K \otimes \mathbf{\Phi}$. We derive a concentration-of-measure inequality for the block-diagonal Rademacher-distributed matrix $\mathbf{A}$.

Theorem 1. *Let* $\mathbf{A}$ *be generated as described above and let* $\Delta = \{\mathbf{f} | \mathbf{f} \in \mathbb{R}^{Kn}\}$ *be a countable or finite set. Then the matrix* $\mathbf{A}$ *satisfies the following concentration-of-measure inequality*

$$\mathbb{P}(\left|\|\mathbf{Af}\|_2^2 - \|\mathbf{f}\|_2^2\right| \geq \epsilon\|\mathbf{f}\|_2^2) \leq e \cdot \exp\left(-\frac{c_1\epsilon^2\|\mathbf{f}\|_2^4}{mn^2}\right) \quad \forall \mathbf{f} \in \Delta,\ \epsilon \in (0,1), \tag{8}$$

where $c_1 > 0$ *is a constant and* e *denotes the base of the natural logarithm.*

In contrast to many previous concentration-of-measure results for matrices populated with i.i.d. sub-Gaussian entries [16], the decay rate indicated by Theorem 1 depends on the signal being measured. Our concentration-of-measure result provides a new strategy, by constituting $\mathbf{\Phi}$ via the Rademacher distribution. Compared to the result in [11], [10], where the best possible decay rate for the exponential is of $\Theta(K)$, our results may provide a better result, as $\Theta(K^2)$, at a cost of a countability constraint on the signal class. The big-theta Θ notation is used here to denote two-sided boundedness for functions of real numbers. For example, $f(K) \sim \Theta(g(K))$ means $c_1 g(K) \leq f(K) \leq c_2 g(K)$ with constants $c_1, c_2 > 0$ for all K large enough. More importantly, as we elaborate on later, such a Rademacher configuration facilitates an easy construction of a non-negative sensing matrix necessary for Poisson sensing, by simply adding a constant offset. The Gaussian configuration proposed in [11], [10] cannot be easily adapted to guarantee such a non-negative constraint. Nevertheless, in addition to its own theoretical value, our concentration-of-measure results also shed light on CS for multiple Gaussian measurements; in the next section, we present some applications of the concentration-of-measure result for that case.

IV. Gaussian Multiple-Measurement Model

The concentration-of-measure inequality is a powerful characterization for the behavior of a random operator, which possesses a number of implications in various areas [16]. We formulate a modified version of the Johnson-Lindenstrauss (JL) Lemma [17] for block-diagonal matrices. First, recall the definition of the stable embedding [18]

Definition 1. *For* $U, V \subset \mathbb{R}^n$*, a map* $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ *is called an* ϵ*-stable embedding of* (U, V) *if*

$$(1-\epsilon)\|\mathbf{x}-\mathbf{y}\|_2^2 \leq \|\Phi(\mathbf{x}-\mathbf{y})\|_2^2 \leq (1+\epsilon)\|\mathbf{x}-\mathbf{y}\|_2^2, \quad \forall \mathbf{x} \in U,\ \mathbf{y} \in V. \tag{9}$$

In other words, a map is a stable embedding of (U, V) if it almost preserves all pairwise distances between U and V. The classical JL Lemma [17] assures the existence of such an ϵ-stable embedding of (U, U) if $m \sim \Theta(\frac{\log|U|}{\epsilon^2})$.

Via Theorem 1, a modified version of the JL Lemma can be stated as follows.

Theorem 2. *Let $U, V \subset \mathbb{R}^{Kn}$ be two finite sets and $\mathbf{A} \in \mathbb{R}^{Km \times Kn}$ be generated as previously described. For $0 < \rho < 1$ being fixed, $\mathbf{A}$ is an ϵ-stable embedding of (U, V) with*

$$\epsilon = \sqrt{\frac{mn^2(\log \rho + 1 + \log |U| + \log |V|)}{c_1 \min_{\substack{\mathbf{x} \in U, \mathbf{y} \in V \\ \mathbf{x} \neq \mathbf{y}}} \|\mathbf{x} - \mathbf{y}\|_2^4}}, \tag{10}$$

which holds with probability at least $1 - \rho$, where $c_1 > 0$ is a constant.

In the above theorem, we note that the performance of the stable-embedding depends on the pairwise distance between two classes U and V and the total number of measurement K is not directly revealed there. Theorem 2 is of particular interest when the minimal energy of signal difference between two classes U and V scales with the number of measurements K, *i.e.*, $\min_{\substack{\mathbf{x} \in U, \mathbf{y} \in V \\ \mathbf{x} \neq \mathbf{y}}} \|\mathbf{x} - \mathbf{y}\|_2^2 \sim \Theta(K)$. Whenever such an assumption is valid, it is straightforward to see that embedding performance will keep improving with increasing K.

This assumption can be satisfied for many common classes of signals, akin to the notion of "favorable signal class" proposed in [10]. Specifically, for video signals and frequency-sparse signals satisfying additional assumptions, it has been justified empirically and theoretically in [10] that the energy of the signal difference $\|\mathbf{x}_i - \mathbf{y}_i\|_2^2, i = 1, \ldots, K$ tends to be uniformly distributed. In other words, we have $\|\mathbf{x} - \mathbf{y}\|_2^2 = \sum_{i=1}^K \|\mathbf{x}_i - \mathbf{y}_i\|_2^2 \sim \Theta(K)$. We refer readers to [10] for details and more examples. The following corollary summarizes the previous discussion, which explicitly reveals the effect of K on the embedding performance.

Corollary 1. *Let $U, V \subset \mathbb{R}^{Kn}$ be two finite sets and $\mathbf{A} \in \mathbb{R}^{Km \times Kn}$ be generated as previously described. Assume that $\min_{\substack{\mathbf{x} \in U, \mathbf{y} \in V \\ \mathbf{x} \neq \mathbf{y}}} \|\mathbf{x} - \mathbf{y}\|_2^2 \sim \Theta(K)$. For fixed ρ with $0 < \rho < 1$, $\mathbf{A}$ is an ϵ-stable embedding of (U, V) with*

$$\epsilon = \sqrt{\frac{mn^2(\log \rho + 1 + \log |U| + \log |V|)}{c_2 K^2}}, \tag{11}$$

which holds with probability at least $1 - \rho$, where $c_2 > 0$ is a constant.

Indeed, the classical JL Lemma can be derived from Corollary 1 by considering a finite set U satisfying the assumptions. Setting $m = 1$, Theorem 1 essentially claims the existence of an ϵ-stable embedding for (U, U) which maps U to $\mathbb{R}^K$, provided $K \sim \Theta(\frac{\log |U|}{\epsilon^2})$. Note that this result coincides with the classical JL Lemma and can be regarded as a JL Lemma result for sequential embedding of a sequence of signals. Furthermore, rather than being a pure existence result, as in the classical JL Lemma, this result provides a randomized method to realize such a stable-embedding.

With the derived stable-embedding results, it is possible to apply them to various applications for signal processing in the compressed domain, where the stable-embedding result plays a pivotal role [18]. We now present such an application to compressed domain classification for Gaussian multiple-measurement model, which extends the results in [18] for Gaussian single measurement model.

Often classification of a signal among multiple hypotheses is demanded, where one such signal $\mathbf{f} \in \mathbb{R}^{Kn}$ is provided to the acquisition system in the form of multiple inputs $\{\mathbf{f}_i\}_{i=1}^K$ with $\mathbf{f} = \left[\mathbf{f}_1^\top, \ldots, \mathbf{f}_K^\top\right]^\top$. Assume L classes of signals $\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(L)}$ and the compressed multiple measurements $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T \in \mathbb{R}^{Km}$ is obtained via a Gaussian measurement model $\mathbf{y}_i = \mathbf{\Phi}\mathbf{f}_i + \mathbf{w}_i$ with i.i.d. $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$. Let $\mathbf{A} \stackrel{\text{def}}{=} \mathbf{I}_K \otimes \mathbf{\Phi}$. Given the multiple-measurement $\mathbf{y}$, we wish to classify $\mathbf{y}$ via the following L hypotheses:

$$\mathcal{H}_i : \mathbf{y} = \mathbf{A}\mathbf{f}^{(i)} + \mathbf{w},\ i = 1, \ldots, L, \tag{12}$$

where $\mathbf{w} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_K^T]^T$. The minimal classification error is achieved by minimizing the sufficient statistic $\|\mathbf{y} - \mathbf{A}\mathbf{f}^{(i)}\|_2^2$ when each hypothesis are equally likely [19]. Like results in [18], the following theorem for the performance of the classifier $i = \arg\min_i \|\mathbf{y} - \mathbf{A}\mathbf{f}^{(i)}\|_2^2$ can be established via our previous stable-embedding results.

Theorem 3. *Let $\mathbf{A}$ be generated as previously described and fix $0 < \rho < 1$. Assume that input signal $\mathbf{f}$ is chosen from L classes of signals $\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(L)}$ and the multiple-measurement is obtained as $\mathbf{y} = \mathbf{A}\mathbf{f}^{(i)} + \mathbf{w}$ with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{Km})$, for $i = 1, \ldots, K$. We further assume that each class is equally likely. Then the classifier $i = \arg\min_i \|\mathbf{y} - \mathbf{A}\mathbf{f}^{(i)}\|_2^2$ will produce a correct classification with probability at least*

$$1 - \frac{L-1}{2} \exp\left\{-\frac{\delta^2(1-\epsilon)}{8\sigma^2}\right\} - 2\rho, \tag{13}$$

where $\delta = \min_{i \neq j} \|\mathbf{f}^{(i)} - \mathbf{f}^{(j)}\|_2$ and $\epsilon = \sqrt{\frac{mn^2(\log \rho + 1 + 2\log L)}{c_1 \delta^4}}$ and $c_1 > 0$ is a constant.

V. Poisson Multiple-Measurement Model

Provided with K measurements $\{\mathbf{y}_k\}_{k=1}^K$ via the Poisson multiple-measurement model in (6), we consider the goal of recovering the underlying signals $\{\mathbf{f}_k\}_{k=1}^K$.

When developing the theory, we make the following assumptions:

A1) The intensity of each signal $\mathbf{f}_k$ is known and fixed, *i.e.*, $\|\mathbf{f}_k\|_1 = S$ for $k = 1, \ldots, K$; such an assumption is practical for many real applications and a similar one was made in [20], and is necessary to make $\{\mathbf{f}_k\}$ and $\mathbf{\Phi}$ identifiable.

A2) $\mathbf{A}\mathbf{f} \succeq cS\mathbf{1}_{Km}$, where constant $c > 0$ and $\succeq$ denotes the entry-wise inequality. $\mathbf{1}_{Km}$ denotes a vector of dimension Km with all 1 entries. This is used to exclude the singular case, where some Poisson rates asymptotically approach zero.

We propose to estimate $\mathbf{f}$ via the following maximum-likelihood estimator (MLE):

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f} \in \Gamma} \left\{-\log \mathrm{Pois}[\mathbf{y}; \mathbf{A}\mathbf{f}] + 2\,\mathrm{pen}(\mathbf{f})\right\}, \tag{14}$$

where Γ is a collection of all candidate estimators.

In practice, the signal energy level S may not be known precisely. In order to increase the flexibility of regularizers in (14), one may relax the MLE in (14) to the following form:

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}\in\Gamma} \left\{ -\log \mathrm{Pois}[\mathbf{y}; \mathbf{A}\mathbf{f}] + \tau_1 \,\mathrm{pen}(\mathbf{f}) \right\}, \tag{15}$$

where $\tau_1 > 0$ is a preset constant. We refer to the above estimator as the relaxed-MLE.

We assume that Γ is a countable or finite set, and all the candidate estimators in Γ satisfy the constraints

$$\Gamma \subset \left\{ \mathbf{f} \,\middle|\, \begin{array}{l} \mathbf{f} \succeq 0,\ \|\mathbf{f}_k\|_1 = S, \forall k = 1, \ldots, K; \\ \mathbf{A}\mathbf{f} \succeq cS\mathbf{1}_{Km}. \end{array} \right\}. \tag{16}$$

The penalty term $\mathrm{pen}(\mathbf{f})$ is required to satisfy the Kraft inequality $\sum_{\mathbf{f}\in\Gamma} e^{-\mathrm{pen}(\mathbf{f})} \leq 1$. The $\mathrm{pen}(\cdot)$ acts as a logarithmic prior on the signal and can be designated as many popular penalty functions, when a proper scaling is applied in order to satisfy the Kraft inequality. Typical choices for the $\mathrm{pen}(\cdot)$ include the l_1 norm, of interest for sparse signals [3], and the total-variation norm for smooth signals [21]. In fact, this Kraft-compliant penalty is related to the prefix codes for estimators, and more concrete examples of this penalty functions are presented in [22]. Since the performance bounds are built upon Theorem 1, the countability assumption reflected in Γ is therein inherited.

We note that the above MLE and relaxed-MLE can be solved efficiently via an alternating proximal-gradient method for many popular choices of $\mathrm{pen}(\cdot)$, such as l_p norm $\|\cdot\|_p (p \geq 1)$ [23] and total-variation norm [21] $\|\cdot\|_{\mathrm{TV}}$, etc.

In order to comply with non-negativity of $\mathbf{\Phi}$ for the Poisson multiple-measurement model, a constant offset is applied to guarantee such a constraint. Specifically, we constitute $\mathbf{\Phi}$ as follows. Let $\mathbf{\Psi} \stackrel{\text{def}}{=} \frac{\mathbf{Z}}{\sqrt{m}}$ and $\mathbf{\Phi} \stackrel{\text{def}}{=} \mathbf{\Psi} + \frac{1}{\sqrt{m}}\mathbf{1}_{m\times n}$ and the sensing matrix is $\mathbf{A} = \mathbf{I}_K \otimes \mathbf{\Phi}$ and, for use below, we define $\tilde{\mathbf{A}} = \mathbf{I}_K \otimes \mathbf{\Psi}$. Note that $\mathbf{\Phi}$ is a matrix with entries being either 0 or $\frac{2}{\sqrt{m}}$. In other words, the sensing matrix $\mathbf{A}$ consists of a scaled-Rademacher matrix $\mathbf{\Psi}$ and a constant offset $\frac{1}{\sqrt{m}}\mathbf{1}_{m\times n}$ keeping the sensing matrix non-negative.

In particular, for the estimator candidates set Γ defined in (16), we have the following theorem, upon which performance guarantees of MLE and relaxed-MLE are derived.

Theorem 4. *Let $\tilde{\mathbf{A}}$ be generated as previously described. We have*

$$(1-\epsilon)\|\mathbf{f}\|_2^2 \leq \|\tilde{\mathbf{A}}\mathbf{f}\|_2^2 \leq (1+\epsilon)\|\mathbf{f}\|_2^2, \quad \forall \mathbf{f} \in \Gamma,\ \epsilon \in (0,1) \tag{17}$$

with probability at least $1 - e\cdot\exp\left(-\frac{c_1\epsilon^2K^2S^4}{mn^4}\right)$, *where* $c_1 > 0$ *is a constant.*

We consider a performance analysis for the proposed MLE and relaxed-MLE as in (14) and (15), with estimate $\hat{\mathbf{f}}$ for true $\mathbf{f}^*$ evaluated via the risk function

$$R(\hat{\mathbf{f}}, \mathbf{f}^*) = \frac{1}{K}\frac{\|\hat{\mathbf{f}} - \mathbf{f}^*\|_2}{\|\mathbf{f}^*\|_2}. \tag{18}$$

Note that the above risk function calculates the average total-estimation error per measurement, where the error term has been normalized. The adopted risk function measures the average recovery error per measurement. Although small average error does not necessarily lead to small total recovery error, the employed risk function reflects the average recovery performance of the sensing system and serves as a meaningful evaluation criterion. We assume that $\mathbf{f}^*$ is drawn from a distribution whose support satisfies assumptions A1 and A2, and we present a performance guarantee for the MLE which quantifies the *expected* risk bounds with respect to that distribution.

Theorem 5. *Let ϵ be an arbitrary constant in $(0,1)$. With assumptions A1-A2 and the designed sensing matrix $\mathbf{A}$ generated as previously described, the expected risk bound between the true signal $\mathbf{f}^*$ and the estimate $\hat{\mathbf{f}}$ output by the MLE in (14) is bounded by*

$$\begin{aligned} &\mathbb{E}[R(\hat{\mathbf{f}}, \mathbf{f}^*)] \\ &\leq \mathbb{E}\left\{ \sqrt{C_1 \min_{\mathbf{f}\in\Gamma} \left\{ \left(\frac{mK}{cS}\left((1+\epsilon)\|\mathbf{f} - \mathbf{f}^*\|_2^2\right) + 2\,\mathrm{pen}(\mathbf{f}) \right) \right\}} \right\} \end{aligned} \tag{19}$$

with probability at least $1 - \frac{m}{2^n} - e\cdot\exp\left(-\frac{c_1\epsilon^2KS^4}{mn^4}\right)$, *where* $C_1 = \left(\frac{8n\sqrt{mn}}{(1-\epsilon)SK\sqrt{K}}\right)$ *and* $c_1 > 0$ *is a constant. The expectation is taken with respect to an arbitrary joint distribution of* $\mathbf{f}^*$ *whose support satisfies assumptions A1-A2.*

Similarly, we also establish a performance bound for the relaxed-MLE in (15).

Theorem 6. *Let ϵ be an arbitrary constant in $(0,1)$ and fix $\tau_1 \geq 2$. With assumptions A1-A2 and the designed sensing matrix $\mathbf{A}$ generated as previously described, the expected risk bound between the true signal $\mathbf{f}^*$ and the estimate $\hat{\mathbf{f}}$ output by the MLE in (15) is bounded by*

$$\begin{aligned} &\mathbb{E}[R(\hat{\mathbf{f}}, \mathbf{f}^*)] \\ &\leq \mathbb{E}\left\{ \sqrt{C_1 \min_{\mathbf{f}\in\Gamma} \left\{ \left(\frac{mK}{cS}\left((1+\epsilon)\|\mathbf{f} - \mathbf{f}^*\|_2^2\right) + \tau_1\,\mathrm{pen}(\mathbf{f}) \right) \right\}} \right\} \end{aligned} \tag{20}$$

with probability at least $1 - \frac{m}{2^n} - e\cdot\exp\left(-\frac{c_1\epsilon^2KS^4}{mn^4}\right)$, *where* $C_1 = \left(\frac{8n\sqrt{mn}}{(1-\epsilon)SK\sqrt{K}}\right)$ *and* $c_1 > 0$ *is a constant. The expectation is taken with respect to an arbitrary joint distribution of* $\mathbf{f}^*$ *whose support satisfies assumptions A1-A2.*

Theorems 5 and 6 provide quantitative performance characterizations for the MLE algorithm with respect to the number of measurements K. According to the theorem, when the assumptions are valid and m, n, K, ϵ are fixed, the performance of the proposed MLEs are governed by the minimization terms as in (19) and (20), which represent the minimal error one could achieve over all the candidate estimators in Γ. In practice, the Γ set can be simply selected as a quantized version of the continuous search area of interests. In the case of uniform quantization, the minimization terms in (19) and (20) are controlled by the designated quantization level.

The coefficient C_1 is clearly a decreasing functions of K, via which it has been suggested that the average performance of MLE can be potentially improved by increasing the number of measurements K. This result has rigorously justified the intuition that reconstruction quality enhances when more measurements are available; a similar phenomenon under the Gaussian measurement model has been observed and justified in [10].

The case of a Poisson single-measurement model has been considered in [24], [20]. As previously mentioned, a fundamental difference is that the sensing matrix $\mathbf{A}$ for the Poisson multiple-measurement model is limited to a *block-diagonal* structure, rather than being arbitrary, as in the single-measurement case. The block-diagonal measurement matrix poses a more-challenging (and practical) problem. Hence, the proof techniques from [24], [20] cannot be applied to the multiple-measurements case, and the non-negativity constraint on the sensing matrix also invalidates adaptation of the results in [11], [10]. Furthermore, it is undesirable to derive a performance bound by repeatedly applying the single measurement result in [20] for each individual measurement, which claims a performance bound valid with probability p for recovering a single measurement. This simple strategy would yield a bound for recovering multiple measurements valid with probability p^K, and this probability decays to 0 with increasing number of measurements K, thereby eventually invalidating the derived performance bound.

VI. Conclusion

A new concentration-of-measure inequality for a block-diagonal Rademacher-ensembled sensing matrix has been derived, which aims to bridge classical single-measurement CS results to multiple-measurement models. A few applications of the newly-derived concentration-of-measure inequality have been presented for the Gaussian multiple-measurement case. In particular, Johnson-Lindenstrauss-type results and a compressed-domain classification result have been derived.

By constituting the random sensing matrix via Rademacher-ensembles, we have been able to adapt the new concentration-of-measure inequality as well to the Poisson multiple-measurement model, and theoretical performance guarantees for signal recovery via proposed MLE algorithms have been established. By revealing the performance guarantees, one may apply the proposed MLE algorithms to address several practical signal reconstruction problems for Poisson multiple-measurement model, known to be relevant to emerging applications such as X-ray scattering imaging [12].

References

[1] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[2] 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.

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

[4] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Sig. Proc. Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[5] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[6] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[7] 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.

[8] M. Ledoux, *The Concentration of Measure Phenomenon*, vol. 89, American Mathematical Society, 2005.

[9] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an l1-minimization decoder," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 275–288, 2009.

[10] J. Y. Park, H. L. Yap, C. J. Rozell, and M. B. Wakin, "Concentration of measure for block diagonal matrices with applications to compressive signal processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5859–5875, 2011.

[11] A. Eftekhari, H. L. Yap, C. J. Rozell, and M. B. Wakin, "The restricted isometry property for random block diagonal matrices," *Applied and Computational Harmonic Analysis*, vol. 35, pp. 1–31, 2015.

[12] J. A. Greenberg, K. Krishnamurthy, and D. J. Brady, "Snapshot molecular imaging using coded energy-sensitive detection," *Optics express*, vol. 21, no. 21, pp. 25480–25491, October 2013.

[13] D. S. Wilcox, G.T. Buzzard, B.J. Lucier, P. Wang, and D. Ben-Amotz, "Photon level chemical classification using digital compressive detection," *Analytica Chimica Acta*, vol. 755, pp. 17–27, 2012.

[14] L. Wang, D. Carlson, M. Rodrigues, D. Wilcox, R. Calderbank, and L. Carin, "Designed measurements for vector count data," in *Advances in Neural Information Processing Systems*, 2013.

[15] L. Wang, D. Carlson, M. Rodrigues, R. Calderbank, and L. Carin, "A Bregman matrix and the gradient of mutual information for vector Poisson and Gaussian channels," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2611–2629, 2014.

[16] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Foundations and Trends in Communications and Information Theory*, vol. 10, no. 1-2, pp. 1–246, 2013.

[17] W. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemporary Mathematics*, vol. 26, pp. 189–206, 1984.

[18] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, 2010.

[19] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, 1994.

[20] 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.

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[22] R. Willett and R. Nowak, "Multiscale Poisson intensity and density estimation," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3171–3187, 2007.

[23] Q. Tran-Dinh, A. Kyrillidis, and V. Cevher, "Composite self-concordant minimization," *ArXiv e-prints arXiv:1308.2867*, Aug. 2013.

[24] M. Raginsky, S. Jafarpour, Z. T. Harmany, R. F. Marcia, R.M. Willett, and R. Calderbank, "Performance bounds for expander-based compressed sensing in Poisson noise," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4139–4153, 2011.

[25] L. A. Kontorovich and K. Ramanan, "Concentration inequalities for dependent random variables via the martingale method," *The Annals of Probability*, vol. 36, no. 6, pp. 2126–2158, 2008.

[26] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[27] J. Li and A. Barron, "Mixture density estimation," in *Advances in Neural Information Processing System 12*, 1999.

[28] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of Cal. Math. Soc.*, vol. 35, no. 1, pp. 99–109, 1943.