Extreme Compressed Sensing of Poisson Rates From Multiple Measurements

Pavan K. Kota ⓘ, Daniel LeJeune ⓘ, *Member, IEEE*, Rebekah A. Drezek, and
Richard G. Baraniuk ⓘ, *Fellow, IEEE*

Abstract—Compressed sensing (CS) is a signal processing technique that enables the efficient recovery of a sparse high-dimensional signal from low-dimensional measurements. In the multiple measurement vector (MMV) framework, a set of signals with the same support must be recovered from their corresponding measurements. Here, we present the first exploration of the MMV problem where signals are independently drawn from a sparse, multivariate Poisson distribution. We are primarily motivated by a suite of biosensing applications of microfluidics where analytes (such as whole cells or biomarkers) are captured in small volume partitions according to a Poisson distribution. We recover the sparse parameter vector of Poisson rates through maximum likelihood estimation with our novel Sparse Poisson Recovery (SPoRe) algorithm. SPoRe uses batch stochastic gradient ascent enabled by Monte Carlo approximations of otherwise intractable gradients. By uniquely leveraging the Poisson structure, SPoRe substantially outperforms a comprehensive set of existing and custom baseline CS algorithms. Notably, SPoRe can exhibit high performance even with one-dimensional measurements and high noise levels. This resource efficiency is not only unprecedented in the field of CS but is also particularly potent for applications in microfluidics in which the number of resolvable measurements per partition is often severely limited. We prove the identifiability property of the Poisson model under such lax conditions, analytically develop insights into system performance , and confirm these insights in simulated experiments. Our findings encourage a new approach to biosensing and are generalizable to other applications featuring spatial and temporal Poisson signals.

Index Terms—Compressed sensing, sparse recovery, Poisson, maximum likelihood, Monte Carlo methods, microfluidics.

I. INTRODUCTION

AS DATA increasingly informs critical decision-making, efficient signal acquisition frameworks must keep pace. Modern signals of interest are often high-dimensional but can be efficiently recovered by exploiting their underlying structure through signal processing. The field of compressed sensing (CS), reviewed in [1], [2], focuses on the recovery of sparse signals from fewer measurements than the signal dimension. Concretely, an N-dimensional signal $\mathbf{x}^*$ with at most k nonzero entries (in which case $\mathbf{x}^*$ is said to be k-sparse) can be recovered from a measurement vector $\mathbf{y}$ acquired by $M < N$ sensors. The sensors' linear responses to entries of $\mathbf{x}^*$ define a sensing matrix $\mathbf{\Phi}$ such that, compactly, $\mathbf{y} = \mathbf{\Phi}\mathbf{x}^*$. Recovering $\mathbf{x}^*$ from $\mathbf{y}$ is known as the single measurement vector (SMV) problem [3]–[6]. In the *multiple measurement vector* (MMV) problem [7]–[9], D measurements are captured in an $M \times D$ matrix $\mathbf{Y}$ in order to recover $\mathbf{X}^*$, an $N \times D$ signal matrix. $\mathbf{X}^*$ is jointly sparse such that only k rows contain at least some nonzero elements. CS has been applied extensively in imaging [10]–[12] and communications [13]–[15] and only recently in biosensing [16]–[19].

Emerging microfluidics technologies in the field of biosensing motivate a new MMV framework. With microfluidics, a single sample can be split into D small-volume partitions such as droplets or nanowells with D on the order of 10^3 to 10^7 [20]. Microfluidic partitioning captures individual analytes (e.g. cells [21], [22]; genes [23]; proteins [24], [25]; etc.) in partitions, and analyte quantities across partitions are known to follow a Poisson distribution [26], [27]. The common method to detect a library of analytes with large N is to either dilute samples or split samples into more partitions such that the Poisson distributions reduce to either empty or single-analyte capture, i.e., that columns of $\mathbf{X}^*$ satisfy $\|\mathbf{x}_d^*\| \in \{0, 1\}$ [22], [28]. This assumption motivates a straightforward N-class classification problem for each non-empty partition, but it necessitates clear separation between classes even under noise, some prior knowledge of sample concentration, and the generation of many wasteful, empty partitions. We hypothesize that CS could generalize the signal recovery strategy when samples are sparse, a common characteristic of biological samples. For example, samples may contain only a few microbes or mutations of interest among many possibilities [29], [30].

We propose the following generally applicable framework for the MMV problem with Poisson signals (MMVP). Let
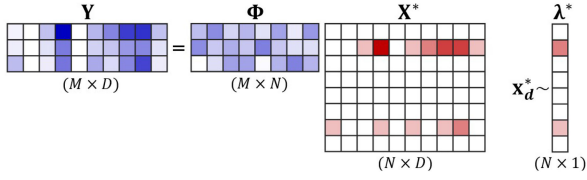
Fig. 1. The *multiple measurement vector* (MMV) problem with Poisson signals (MMVP) with one sensor group and noiseless measurements. White squares are zeroes and darker colors represent larger values. Each column $\mathbf{x}_d^*$ of $\mathbf{X}^*$ is drawn from a Poisson distribution governed by the 2-sparse $\boldsymbol{\lambda}^*$ (i.e., $\mathbf{x}_{d*} \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\boldsymbol{\lambda}^*)$).

each signal $\mathbf{x}_d^*$ be drawn independently from a multivariate Poisson distribution parametrized by the N-dimensional, k-sparse vector $\boldsymbol{\lambda}^*$. That is, $x_{n,d}^* \sim \text{Poisson}(\lambda_n^*)$ are independent. This framework should not be confused with the well-studied "Poisson compressed sensing" problem in imaging where the *measurement noise*, rather than the signal, follows a Poisson distribution [31], [32]. In contrast to typical MMV problems, our primary goal is to find an estimate $\widehat{\boldsymbol{\lambda}} \approx \boldsymbol{\lambda}^*$ from D observations rather than to estimate $\widehat{\mathbf{X}}$. Each signal and measurement pair $(\mathbf{x}_d^*, \mathbf{y}_d)$ is in one of G different *sensor groups*, each with its own sensing matrix $\boldsymbol{\Phi}^{(g)}$ such that $\mathbf{y}_d = \boldsymbol{\Phi}^{(g)} \mathbf{x}_d^*$. The group g associated with each index d is known and deterministic. This measurement group multiplexing is similar to that found in the single pixel camera (SPC) [10]; however, in the SPC the $\mathbf{x}_d^*$ are assumed to be equal, whereas in MMVP they are independently sampled. In microfluidics, several sensor groups can be feasibly achieved by forking an input microfluidic channel into G reaction zones each containing its own set of M sensors. Note that $\mathbf{x}_d^* \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\boldsymbol{\lambda}^*)$ regardless of which group it is in. The statement $\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X}^*$ is the special case without noise where $G = 1$ and is illustrated in Fig. 1. For multiple groups with $\mathbf{X}^{(g)*}$ denoting the submatrix of $\mathbf{X}^*$ in group g, $\mathbf{Y}$ is the following concatenation:

$$\mathbf{Y} = \begin{bmatrix} \boldsymbol{\Phi}^{(1)}\mathbf{X}^{(1)*} & \dots & \boldsymbol{\Phi}^{(G)}\mathbf{X}^{(G)*} \end{bmatrix}. \tag{1}$$

A. Contributions and Findings

We present the first exploration of the MMVP problem and develop a novel recovery algorithm and initial theoretical results. We take a maximum likelihood estimation (MLE) approach, treating $\mathbf{Y}$ as a set of D observations from which to infer $\widehat{\boldsymbol{\lambda}}$. Our core contributions are **1:** the *Sparse Poisson Recovery* (SPoRe) algorithm that tractably estimates $\widehat{\boldsymbol{\lambda}}$ (Section II); **2:** theoretical results on the identifiability of our MMVP model, proving that MLE can asymptotically recover $\boldsymbol{\lambda}^*$ even when $M = G = 1$ for any k, and insights into MLE performance (Section III); and **3:** simulations demonstrating SPoRe's superior performance over existing and custom baseline algorithms (Section IV). The achievable measurement rate in the MMVP model is unprecedented in CS, and although we are unable to provide theoretical guarantees for our algorithm, we analytically derive insights into the influence of various system parameters and confirm these insights in our simulated experiments. We find that system

designers should first maximize M and then increase G as necessary depending on the expected real-world conditions. We found that using $G > 1$ was helpful for cases with high noise or high $\sum_n \lambda_n^*$.

SPoRe's strong performance even with extremely low $M \in \{1, 2, 3\}$ under very high measurement noise uniquely enables sensor-constrained applications in biosensing. Although microfluidics devices can rapidly generate a large number of partitions D at a tunable rate, most optical and electrochemical sensing modalities that can keep pace are limited in M [20], [33]. Commonly, fluorescently tagged sensors reveal droplets' contents rapidly as they flow by a detector, but spectral overlap generally limits M to be five or less without highly specialized, system-specific approaches [34]. High M alternatives such as various spectroscopic techniques limit throughput, necessitate additional instrumentation, or complicate workflows [33], [35]. We speculate that these severe restrictions in M may have forestalled research into CS's potential role in microfluidics.

B. Previous Work

To the best of our knowledge, the MMVP problem has not yet been explored, likely owing to the ongoing maturation of microfluidics and only recent application of CS to biosensing. The Poisson signal model constrains elements of $\mathbf{X}$ to be nonnegative integers under a set of defined probability mass functions. Some aspects of this structure have been studied tangentially, but not the MMVP structure directly.

The core MMV problem only imposes joint sparsity. Early greedy algorithms for this generalized scenario extend the classic Orthogonal Matching Pursuit (OMP) algorithm [36] into OMPMMV [8], simultaneously developed as Simultaenous OMP (S-OMP) [9]. Generally, OMP-based algorithms iteratively build a support set of an estimated sparse solution $\widehat{\mathbf{x}}$ (or $\widehat{\mathbf{X}}$) by testing for the correlation between columns of $\boldsymbol{\Phi}$ and the residuals between the measurements and previous estimates. A suite of greedy algorithms was recently developed that impose nonnegative constraints to a number of MMV approaches including OMP's analogues, and the nonnegative extensions outperformed their generalized counterparts [37].

The application of integer constraints to the SMV problem has proven challenging in the literature. Some theory involving sensing matrix design includes [38], [39], but practical algorithms have required additional constraints on the possible integers, e.g., $\mathbf{x} \in \{0, 1\}^N$ or other finite-alphabet scenarios [14], [40]–[43]. A recent study verified that these problems, as well as those with unbounded integer signals, are NP-hard [44]. Algorithms for the unconstrained integer SMV problem thus apply greedy heuristics such as OMP-based approaches [45], [46].

Additional structural constraints can also make these problems tractable. The communications problem of multi-user detection (MUD), reviewed in [47], bears some similarity to MMVP. Here, the activity of N users is the signal of interest and generally follows a Bernoulli model where each user is active with the same prior probability p_a [14]. An alternative prior with $\sum_{n=1}^N x_{n,d} \sim \text{Poisson}(\lambda)$ models the mean number of total active users in any given signal [48] although the authors

solely explored an overdetermined system. Applying an MMV framework to MUD enables underdetermined ($M < N$) applications but has generally assumed that any active user is active for the entire frame of observation (a row of $\mathbf{X}$ is entirely zero or nonzero) [49], [50]. Recently, the potency of sensor groups with a $G = 2$ system was demonstrated in the MUD context [51]. Despite some similarities to MMVP with an MMV framework and discrete signals, MUD most fundamentally differs from MMVP in its utilization of the probabilistic structure of $\mathbf{X}$. In MUD, the model parameters governing user activities are assumed and leveraged in recovery of $\mathbf{X}$, whereas in MMVP, the model parameters in $\boldsymbol{\lambda}^*$ themselves are the target of recovery.

Regardless of the particular structural constraints imposed in the variants of MMV above, $M > k$ is necessary to recover $\mathbf{X}^*$ [52]. However, in MMV models where the rows of $\mathbf{X}^*$ are statistically independent, the simpler task of support recovery has been proven to be possible below the $M > k$ regime, as low as $M = \Omega(\sqrt{k})$ [53]–[55]. In particular, multiple measurement sparse Bayesian learning (M-SBL) [53] takes a MLE approach in a setting very similar to MMVP, but where $\mathbf{x}_d^*$ have a Gaussian distribution rather than Poisson, and there $M = \Omega(\sqrt{k})$ is limit for sparse support recovery even for $D \to \infty$. In contrast, we prove that in the MMVP problem, this is improved to $M \geq 1$ (i.e., no minimum number of measurements) regardless of k when $D \to \infty$, even when $G = 1$. Furthermore, the MLE in this setting recovers $\boldsymbol{\lambda}^*$ exactly. We demonstrate empirically that we can recover $\boldsymbol{\lambda}^*$ well for finite D even for extremely low M such as $M \in \{1, 2, 3\}$.

II. SPARSE POISSON RECOVERY (SPoRe) ALGORITHM

A. Notation

We denote by $P(\cdot)$ a probability mass function and by $p(\cdot)$ a probability density function. We use $\mathbb{R}^N$ and $\mathbb{Z}^N$ to represent the N-dimensional Euclidean space and integer lattice, respectively. We denote by $\mathbb{R}_+^N$ and $\mathbb{Z}_+^N$ the non-negative restrictions on these spaces. We use script letters ($\mathcal{A}, \mathcal{B}, ...$) for sets unless otherwise described. We use lowercase and uppercase bold-face letters for vectors and matrices, respectively. We represent their dimensions with uppercase letters (e.g., $\mathbf{X} \in \mathbb{Z}_+^{N \times D}$) that are indexed by their lowercase counterparts. For example, $x_{n,d}$ is the element of $\mathbf{X}$ in the nth row and dth column, and we use the shorthands $\mathbf{x}_n$ and $\mathbf{x}_d$ to represent the entire nth row and dth column vectors, respectively. Other lower case letters (a, b, ϵ, etc.) may represent variable or constant scalars depending on context. We use $\boldsymbol{\lambda}^*$ and $\mathbf{X}^*$ to refer to the true signal values, and we denote estimates $\widehat{\boldsymbol{\lambda}}$ and $\widehat{\mathbf{X}}$ with the source of the estimate (e.g., MLE, SPoRe, baseline algorithm) being implicit from the context. We use $\| \cdot \|$ to denote a norm, $\| \cdot \|_0$ for the number of nonzero elements of a vector, $\| \cdot \|_1$ for the ℓ_1 vector norm, $\| \cdot \|_2$ for the Euclidean norm, and $\| \cdot \|_F$ for the Frobenius norm. We also use $\|\mathbf{X}\|_{Rx} \triangleq \sum_n \max_d |x_{n,d}|$, a relaxation of the row ℓ_0 quasi-norm defined in [56]. We denote the null space of matrix $\mathbf{A}$ by $\mathcal{N}(\mathbf{A})$. As one abuse of notation, for densities of the form $p(\mathbf{y}|\mathbf{x})$, we let the corresponding $\boldsymbol{\Phi}^{(g)}$ applied to $\mathbf{x}$ and the

relevant noise model be implicit. Also, we let the division of two vectors represent element-wise division.

B. Algorithm

If the index d is in sensor group g, we say that linear measurements are corrupted by an additive random noise vector $\mathbf{b}_d$:

$$\mathbf{y}_d = \boldsymbol{\Phi}^{(g)}\mathbf{x}_d + \mathbf{b}_d. \tag{2}$$

We let $\mathbf{b}_d$ be entirely independent (e.g., additive white Gaussian noise (AWGN), as used in our simulations) or dependent on $\mathbf{x}$. With $\mathbf{x}_d \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\boldsymbol{\lambda}^*)$, $\mathbf{y}_d$ are independent across d as well. The MLE estimate maximizes the average log-likelihood of the measurements:

$$\widehat{\boldsymbol{\lambda}}_{MLE} = \arg \max_{\boldsymbol{\lambda}} \prod_{d=1}^{D} p(\mathbf{y}_d|\boldsymbol{\lambda}) \tag{3}$$

$$= \arg \max_{\boldsymbol{\lambda}} \frac{1}{D} \sum_{d=1}^{D} \log \sum_{\mathbf{x} \in \mathbb{Z}_+^N} p(\mathbf{y}_d|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda}). \tag{4}$$

Because the infinite sum over $\mathbf{x}$ yields an intractable posterior distribution, we cannot apply the popular expectation-maximization (EM) algorithm [57] to solve this MLE problem. Instead, our Sparse Poisson Recovery (SPoRe) algorithm (Algorithm 1) optimizes this function with batch stochastic gradient ascent, drawing B elements uniformly with replacement from $\{1, ..., D\}$ to populate a batch set $\mathcal{B}$. First, note that

$$\nabla_{\boldsymbol{\lambda}} P(\mathbf{x}|\boldsymbol{\lambda}) = P(\mathbf{x}|\boldsymbol{\lambda})\left(\frac{\mathbf{x}}{\boldsymbol{\lambda}} - 1\right). \tag{5}$$

Denoting the objective function from the right-hand side of (4) as ℓ, the gradient is

$$\nabla_{\boldsymbol{\lambda}}\ell = \frac{1}{B} \sum_{d \in \mathcal{B}} \frac{\sum_{\mathbf{x} \in \mathbb{Z}_+^N} p(\mathbf{y}_d|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda})\mathbf{x}}{\boldsymbol{\lambda} \sum_{\mathbf{x} \in \mathbb{Z}_+^N} p(\mathbf{y}_d|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda})} - 1. \tag{6}$$

With gradient ascent, each iteration updates $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + \alpha \nabla_{\boldsymbol{\lambda}}\ell$ with learning rate α. However, the summations over all of $\mathbb{Z}_+^N$ are clearly intractable. SPoRe approximates these quantities with a Monte Carlo (MC) integration over S samples of $\mathbf{x}$, newly drawn for each batch gradient step from sampling distribution $Q : \mathbb{Z}_+^N \to \mathbb{R}_+$, such that

$$\sum_{\mathbf{x} \in \mathbb{Z}_+^N} p(\mathbf{y}|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda}) \approx \frac{1}{S} \sum_{s=1}^{S} \frac{p(\mathbf{y}|\mathbf{x}_s)P(\mathbf{x}_s|\boldsymbol{\lambda})}{Q(\mathbf{x}_s)}. \tag{7}$$

The optimal choice of $Q(\mathbf{x}_s)$ is beyond the scope of this work, but we found that $Q(\mathbf{x}_s) = P(\mathbf{x}_s|\boldsymbol{\lambda})$ simplifies the expression, is effective in practice, and draws inspiration from the EM algorithm. In other words, the sampling function is updated at each iteration based on the current estimate of $\boldsymbol{\lambda}$. The gradient thus simplifies to

$$\nabla_{\boldsymbol{\lambda}}\ell = \frac{1}{B} \sum_{d \in \mathcal{B}} \frac{\sum_{s=1}^{S} p(\mathbf{y}_d|\mathbf{x}_s)\mathbf{x}_s}{\boldsymbol{\lambda} \sum_{s=1}^{S} p(\mathbf{y}_d|\mathbf{x}_s)} - 1. \tag{8}$$

Note that if only one $\widehat{\mathbf{x}}_d \in \mathbb{Z}_+^N$ satisfied $p(\mathbf{y}_{cd}|\widehat{\mathbf{x}}_d) > 0$ for every $\mathbf{y}_d$, the objective ℓ would be concave with $\widehat{\boldsymbol{\lambda}} = \frac{1}{D}\sum_{d=1}^{D} \widehat{\mathbf{x}}_d$, i.e., the MLE solution if $\mathbf{X}^*$ were directly observed. Of course, with compressed measurements and noise, multiple signals may vie
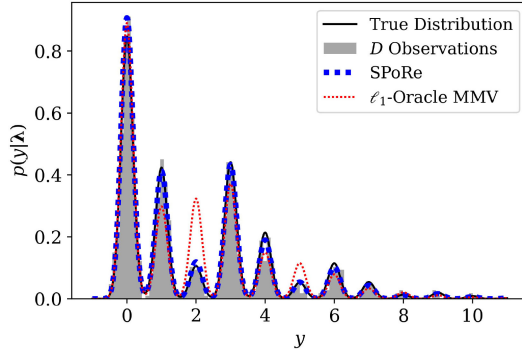
Fig. 2. Example of MMVP and Sparse Poisson Recovery (SPoRe) with $M < k < N$: $\boldsymbol{\Phi} = [1, 2, 3]$, $\boldsymbol{\lambda}^* = [0.5, 0, 0.5]$, and $D = 1000$ measurements under additive white Gaussian noise (AWGN) $\mathbf{b} \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 0.02$. SPoRe attempts to fit the distribution of measurements directly and finds $\widehat{\boldsymbol{\lambda}} \approx [0.45, 0.03, 0.44]$. For comparison, the ℓ_1-Oracle (see Section IV) minimizes the measurement error $\widehat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{X}\|_F$ with $\mathbf{X} \geq 0$ and $\sum_{n,d} x_{n,d} = \sum_{n,d} x_{n,d}^*$ as affine constraints. The estimate $\widehat{\boldsymbol{\lambda}}$ for the ℓ_1-Oracle is then set to the average of the columns of $\widehat{\mathbf{X}}$, and in this example, $\widehat{\boldsymbol{\lambda}} \approx [0.33, 0.31, 0.31]$. The distributions $p(\mathbf{y}|\widehat{\boldsymbol{\lambda}})$ for each estimation method are compared against the true distribution $p(\mathbf{y}|\boldsymbol{\lambda}^*)$ and the empirical histogram of the D observations.

to "explain" any single measurement, but SPoRe's key strength is that it jointly considers independent measurements to directly estimate $\widehat{\boldsymbol{\lambda}}$.

We note that for finite samples, since the MC integration occurs inside a logarithm, the stochastic gradient is biased. However, since it converges in probability to the true gradient, we can expect results comparable to SGD with an unbiased gradient for sufficiently large S [58].

Fig. 2 illustrates key concepts of SPoRe and MMVP with a small example where $M = G = 1$ and $\boldsymbol{\lambda}^* = [0.5, 0, 0.5]$ for which we can numerically compute $p(\mathbf{y}|\boldsymbol{\lambda})$ for various $\boldsymbol{\lambda}$. The measurements $\mathbf{y}_d$ are effectively drawn from an underlying mixture distribution depending on the noise; e.g., under AWGN, $\mathbf{y}_d$ follows a Gaussian mixture. The weights on each mixture component are controlled by $\boldsymbol{\lambda}$. In simulated recovery, SPoRe assigns weights to the mixture via $\widehat{\boldsymbol{\lambda}}$ according to the distribution of measurements, coming close to the true underlying distribution. In contrast, an ℓ_1-Oracle (Section IV) which represents best-case performance for a standard, convex sparse recovery process fails because $M < k$ as shown by its error in $\boldsymbol{\lambda}$ and illustrated by the difference in the distributions. Moreover, by using $\boldsymbol{\Phi} = [1, 2, 3]$, many $\mathbf{x}$ will map to the same $\mathbf{y}$. While CS theory generally focuses on conditions for unique or well-spaced projections of k-sparse signals (e.g., the restricted isometry property, RIP [5]), we demonstrate that such restrictions are unnecessary in MMVP. By accounting for the latent Poisson distribution in the signals, SPoRe succeeds even when $M < k$.

Algorithm 1 summarizes the implementation details of SPoRe. Even though $\boldsymbol{\lambda} \in \mathbb{R}_+^N$, we enforce $\boldsymbol{\lambda} \geq \epsilon$ by clipping ($\epsilon = 10^{-3}$ in our simulations) to maintain exploration of the parameter space. Note that in (8), gradients can become very large with finite sampling as some elements of $\boldsymbol{\lambda}$ approach zero. We found that rescaling gradients to maximum norm γ

Algorithm 1: Sparse Poisson Recovery (SPoRe).

Input: $\boldsymbol{\lambda}^{(0)}, B, S, \gamma, \alpha, \epsilon$
1:   $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda}^{(0)}$
2:   $i = 0$
3:   **repeat**
4:      Draw B columns of $\mathbf{Y}$ uniformly with replacement
5:      Draw S new samples from $Q(\mathbf{x}_s)$
6:      $\boldsymbol{\delta} \leftarrow \alpha\nabla_{\boldsymbol{\lambda}}\ell(\boldsymbol{\lambda})$ $\triangleright$ (8)
7:      **if** $\|\boldsymbol{\delta}_\Gamma\|_2 > \gamma$ **then**
8:         $\boldsymbol{\delta} \leftarrow \frac{\gamma}{\|\boldsymbol{\delta}_\Gamma\|_2}\boldsymbol{\delta}$ $\triangleright$ Rescale gradient step
9:      **end if**
10:     $\lambda_n \leftarrow \max(\lambda_n + \delta_n, \epsilon)$
11: **until** stopping criterion met
12: **return** $\boldsymbol{\lambda}$

---

helps stabilize convergence. For rescaling, we consider only the subvector $\boldsymbol{\delta}_\Gamma$ of the α-scaled gradient $\boldsymbol{\delta}$, defining indices $n \in \Gamma \subseteq \{1, \ldots, N\}$ if $\lambda_n + \delta_n > \epsilon$. This restriction ensures that rescaling is solely based on the indices still being optimized, excluding those clipping to ϵ.

For our stopping criterion, we evaluate a moving average of $\widehat{\boldsymbol{\lambda}}$ for convergence. We also track the median of $p(\mathbf{y}_d|\boldsymbol{\lambda})$ for $d \in \mathcal{B}$, accounting for stochasticity in likelihood approximations and batch selections such that we may reduce α if no improvements in the median have been seen within a patience window and terminate if α is reduced five times. We conducted all experiments on commodity personal computing hardware. Ultimately, recovery of $\widehat{\boldsymbol{\lambda}}$ takes a few minutes on a single core for the problem sizes we consider ($M \leq 15$, $N \leq 50$), and SPoRe can be easily parallelized in the future for faster performance.

C. Practical Considerations

Within an iteration, we found that using the same $S = 1000$ samples for all $d \in \mathcal{B}$ helped to vectorize our implementation to dramatically improved speed over sampling S times for *each* drawn $\mathbf{y}_d$. This simplification had no noticeable influence on performance. While we found random initializations with a small offset $\boldsymbol{\lambda}^{(0)} \sim \text{Uniform}(0, 1) + \nu$ (with $\nu = 0.1$) to be effective in general, we encountered a numerical issue when under low-variance AWGN. Even though AWGN results in nonzero probabilities everywhere, $p(\mathbf{y}_d|\mathbf{x}_s)$ may numerically evaluate to zero for all drawn samples in low-noise settings. These zeros across all samples result in undefined terms in the summation over $d \in \mathcal{B}$ in (8). SPoRe simply ignores such undefined terms, but when this numerical issue occurs for *all* of $\mathcal{B}$, SPoRe takes no gradient step. With very low noise and large N dampening the effectiveness of random sampling, SPoRe may stop prematurely as it appears to have converged. This problem did not arise with larger noise variances where even inexact samples pushed $\widehat{\boldsymbol{\lambda}}$ in the generally appropriate direction until better samples could be drawn (recall that $Q(\mathbf{x}) = P(\mathbf{x}|\widehat{\boldsymbol{\lambda}})$ at each iteration). Nonetheless, we decided to set $\boldsymbol{\lambda}^{(0)} = \nu$ for consistency across all simulated experiments. We speculate that setting $\boldsymbol{\lambda}^{(0)}$ to a small value helped encourage sampling sparse

x's in early iterations to help find $\mathbf{x}_s$ with nonzero $p(\mathbf{y}_d|\mathbf{x}_s)$, bypassing the numerical issue altogether.

III. THEORY AND ANALYSIS

The summation over $\mathbf{x}$ inside the logarithm of the objective function complicates the precise analysis of our SPoRe algorithm. However, we can consider the asymptotic MMVP problem as $D \to \infty$ and its MLE solution to gain insight into the superior recovery performance of SPoRe. In this setting, we obtain the powerful result that $\boldsymbol{\lambda}^*$ is exactly recoverable even when $M = G = 1$ under a simple null space condition on $\boldsymbol{\Phi}$. We then characterize the loss in Fisher Information for MMVP and show how losses accrue with the increase of signals that map to the same measurements, an effect that increases with k. Lastly, we derive insights into the influence of sensor groups through a small-scale analysis. From a system design standpoint, we find that designers should first increase M as much as feasible and then increase G as needed. All proofs can be found in the Appendix.

A. Identifiability of MMVP Models

Our primary theoretical result is that the asymptotic MLE of the MMVP problem is exactly equal to $\boldsymbol{\lambda}^*$ as long as a simple null space condition on $\boldsymbol{\Phi}$ is satisfied. This result improves upon the $M = \Omega(\sqrt{k})$ limitations of correlation-aware support recovery methods [53], [55], placing no restrictions on k and M. The reason for this improvement is that $\boldsymbol{\Phi}$ transforms an integer lattice rather than a linear subspace such that the resulting model is still *identifiable*.

Identifiability refers to the uniqueness of the model parameters that can give rise to a distribution of observations. A model $\mathcal{P} = \{p(\cdot|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \mathbb{R}_+^N\}$ is a collection of distribution functions which are indexed by the parameter $\boldsymbol{\lambda}$; in the MMVP problem, each choice of $\boldsymbol{\Phi}$ and noise give rise to a different model $\mathcal{P}$. Through an optimization lens, if our model is identifiable, then $\boldsymbol{\lambda}^*$ is the unique global optimum of the data likelihood as $D \to \infty$. Recall that $p(\mathbf{y}|\boldsymbol{\lambda}) = \sum_{x \in \mathbb{Z}_+^N} p(\mathbf{y}|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda})$, meaning that we can interpret this model as each sensor group consisting of a mixture whose elements' positions are governed by $\boldsymbol{\Phi}^{(g)}\mathbf{x}$, distributions by the noise model, and weights by $P(\mathbf{x}|\boldsymbol{\lambda})$. We focus in this analysis on $G = 1$, since as $D \to \infty$, at least one sensor group contains infinite measurements. If the corresponding $\boldsymbol{\Phi}^{(g)}$ satisfies the conditions we describe here, then the model is identifiable. Formally:

Definition 3.1 (Identifiability): The model $\mathcal{P}$ is identifiable if $p(\mathbf{y}|\boldsymbol{\lambda}) = p(\mathbf{y}|\boldsymbol{\lambda}') \; \forall \mathbf{y} \Rightarrow \boldsymbol{\lambda} = \boldsymbol{\lambda}'$ for all $\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in \mathbb{R}_+^N$.

The identifiability of mixtures is well-studied [59], [60]; if a mixture is identifiable, the mixture weights uniquely parametrize possible distributions. For finite mixtures, a broad set of distributions including multivariate exponential and Gaussian have been proven to be identifiable [61]. A finite case may manifest in realistic MMVP systems where measurements $\mathbf{y}$ must eventually saturate; all sensors have a finite dynamic range of values they can capture. In the most general case, $p(\cdot|\boldsymbol{\lambda})$ is a countably infinite mixture. Although less studied, countably infinite mixtures are identifiable under some classes of distributions [62]. The

AWGN that we use in our simulations is identifiable for both the finite and countably infinite cases. Characterizing the full family of noise models that are identifiable under countably infinite mixtures is beyond the scope of this work. Our contribution is that *given* a noise model that yields identifiable mixtures, equal mixture weights induced by $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ imply $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$. We prove the sufficiency of the following simple conditions on $\boldsymbol{\Phi}$ for identifiability:

Theorem 3.2 (Identifiability of Mixture Weights): Let $\mathbf{b}$ be additive noise drawn from a distribution for which a countably infinite mixture is identifiable. If $\mathcal{N}(\boldsymbol{\Phi}) \cap \mathbb{R}_+^N = \{\mathbf{0}\}$ and $\boldsymbol{\phi}_n \neq \boldsymbol{\phi}_{n'} \; \forall n, n' \in \{1, \ldots, N\}$ with $n \neq n'$, then $\mathcal{P}$ is identifiable.

The null space condition essentially says that any nonzero vector in $\mathcal{N}(\boldsymbol{\Phi})$ must contain both positive and negative elements. Many practical $\boldsymbol{\Phi}$ satisfy this constraint (e.g., any $\boldsymbol{\Phi}$ with at least one strictly negative or positive row). The second condition is trivial: no two columns of $\boldsymbol{\Phi}$ can be identical. We also obtain a separate sufficient condition, that $\boldsymbol{\Phi}$ drawn from any continuous distribution results in identifibiablity.

Corollary 3.3 (Identifiability with Random Continuous $\boldsymbol{\Phi}$): Let $\mathbf{b}$ be additive noise drawn from a distribution for which a countably infinite mixture is identifiable. If the elements of $\boldsymbol{\Phi}$ are independently drawn from any continuous distribution, then $\mathcal{P}$ is identifiable.

We emphasize the general result of Theorem 3.2, since discrete sensing is common in biomedical systems. For example, sensors are often designed to bind to an integer number of known target sites and yield "digital" measurements [26], [63]. Discrete $\boldsymbol{\Phi}$ can give rise to what we call *collisions*. Formally:

Definition 3.4 (Collisions and Collision Sets): Let $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ be a sensing matrix applied to signals $\mathbf{x} \in \mathbb{Z}_+^N$. A *collision* occurs between $\mathbf{x}$ and $\mathbf{x}'$ when $\boldsymbol{\Phi}\mathbf{x} = \boldsymbol{\Phi}\mathbf{x}'$. A *collision set* for an arbitrary $\mathbf{u} \in \mathbb{Z}_+^N$ is the set $\mathcal{C}_{\mathbf{u}} = \{\mathbf{x} : \boldsymbol{\Phi}\mathbf{x} = \boldsymbol{\Phi}\mathbf{u}; \mathbf{x} \in \mathbb{Z}_+^N\}$.

If the distribution from which $\mathbf{b}$ is drawn is fixed (e.g., AWGN) or a function of $\boldsymbol{\Phi}\mathbf{x}$, then the mixture weights are the probability mass of each collision set. Let the set of collision sets be $\mathcal{U}$ with $\mathcal{C}_{\mathbf{u}} \in \mathcal{U}$ being an arbitrary collision set.

$$p(\mathbf{y}|\boldsymbol{\lambda}) = \sum_{\mathcal{C}_{\mathbf{u}} \in \mathcal{U}} p(\mathbf{y}|\mathbf{x} \in \mathcal{C}_{\mathbf{u}})P(\mathcal{C}_{\mathbf{u}}|\boldsymbol{\lambda}) \tag{9}$$

$$P(\mathcal{C}_{\mathbf{u}}|\boldsymbol{\lambda}) = \sum_{\mathbf{x} \in \mathcal{C}_{\mathbf{u}}} P(\mathbf{x}|\boldsymbol{\lambda}). \tag{10}$$

The weights of the mixture elements are governed by $P(\mathcal{C}_{\mathbf{u}}|\boldsymbol{\lambda})$. Given a noise model that yields identifiable mixtures, the same distribution of observations $\mathbf{y}$ implies that the mixture weights are identical, i.e. $P(\mathcal{C}_{\mathbf{u}}|\boldsymbol{\lambda}) = P(\mathcal{C}_{\mathbf{u}}|\boldsymbol{\lambda}') \; \forall \mathbf{u}$. We prove that $P(\mathcal{C}_{\mathbf{u}}|\boldsymbol{\lambda}) = P(\mathcal{C}_{\mathbf{u}}|\boldsymbol{\lambda}') \; \forall \mathbf{u}$ implies $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$, which implies the identifiability of $\mathcal{P}$ under both the conditions of Theorem 3.2 and Corollary 3.3.

Our proofs are based on the existence and implications of single-vector collison sets $\mathcal{C}_{\mathbf{x}} = \{\mathbf{x}\}$. When (9) holds, $\mathbf{u}$ indexes both the mixture elements and the collision sets. In the general case where $\mathbf{b}$ is dependent on $\mathbf{x}$ and not simply $\boldsymbol{\Phi}\mathbf{x}$, signals participating in the same mixture element may have different noise distributions. These differences can only further subdivide

collision sets and leaves single-vector collision sets unaffected. Thus, our results also cover the general noise case.

B. Fisher Information of MMVP Measurements

While identifiability confirms that $\boldsymbol{\lambda}^*$ is a unique global optimum of the MLE problem given infinite observations, Fisher Information helps characterize estimation of $\boldsymbol{\lambda}^*$ as D increases. The Fisher Information matrix $\mathcal{I}$ is the (negative) Hessian of the expected log-likelihood function at the optimum $\boldsymbol{\lambda}^*$, and it is well-known that under a few technical conditions the MLE solution is asymptotically Gaussian with covariance $\mathcal{I}^{-1}/D$. Intuitively, higher Fisher Information implies a "sharper" optimum that needs fewer observations for stable recovery. For direct observations of Poisson signals $\mathbf{x}_d^*$ rather than $\mathbf{y}_d$, $\mathcal{I}$ is diagonal with $\mathcal{I}_{n,n} = 1/\lambda_n^*$. In MMVP with observations of noisy projections ($\mathbf{y}_d$), $\mathcal{I}$ and its inverse are difficult to analyze. We can, however, instead characterize the reduction in $\mathcal{I}_{n,n}$ in MMVP caused by the noisy measurement of $\mathbf{x}_d^*$ and derive an insight that we empirically confirm in Section IV-D. Concretely, elements of $\mathcal{I}$ follow

$$\mathcal{I}_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial \lambda_i^*} \log p(\mathbf{y}|\boldsymbol{\lambda}^*)\right)\left(\frac{\partial}{\partial \lambda_j^*} \log p(\mathbf{y}|\boldsymbol{\lambda}^*)\right)\right]. \quad (11)$$

We denote the shorthand $w_\mathbf{x} \triangleq p(\mathbf{y}|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda})$ and note that $\sum_\mathbf{x} w_\mathbf{x} = p(\mathbf{y}|\boldsymbol{\lambda})$. Following a similar derivation for the partial derivatives in (8), it can be shown that the general expression for diagonal elements $\mathcal{I}_{n,n}$ is

$$\mathcal{I}_{n,n} = \int \left(\frac{\sum_\mathbf{x} w_\mathbf{x} x_n}{\sum_\mathbf{x} w_\mathbf{x} \lambda_n^*} - 1\right)^2 \left(\sum_\mathbf{x} w_\mathbf{x}\right) d\mathbf{y}. \quad (12)$$

In the ideal scenario, we observe $\mathbf{x}_d^*$ directly such that

$$\mathcal{I}_{n,n}^{\text{ideal}} = \sum_\mathbf{x} P(\mathbf{x}|\boldsymbol{\lambda})\left(\frac{x_n}{\lambda_n} - 1\right)^2 \left(\int p(\mathbf{y}|\mathbf{x}) d\mathbf{y}\right). \quad (13)$$

It can be easily shown that (13) reduces to the canonical $1/\lambda_n^*$. The integration of $p(\mathbf{y}|\mathbf{x})$ evaluates to one, but we can manipulate it algebraically to re-express the quantity as

$$\mathcal{I}_{n,n}^{\text{ideal}} = \int \sum_\mathbf{x} \left[w_\mathbf{x}\left(\frac{x_n}{\lambda_n} - 1\right)^2\right] d\mathbf{y}. \quad (14)$$

Let $\mathcal{I}_n^{\text{loss}} \triangleq \mathcal{I}_{n,n}^{\text{ideal}} - \mathcal{I}_{n,n}$ and let $\sum_{(\mathbf{x},'\boldsymbol{\chi})}$ denote the sum over all pairs of signals $\mathbf{x},' \boldsymbol{\chi} \in \mathbb{Z}_+^N$. Expanding Equations (12) and (14) and simplifying yields

$$\mathcal{I}_n^{\text{loss}} = \frac{1}{\lambda_n^{*\,2}} \int \left(\sum_\mathbf{x} w_\mathbf{x} x_n^2 - \frac{(\sum_\mathbf{x} w_\mathbf{x} x_n)^2}{\sum_\mathbf{x} w_\mathbf{x}}\right) d\mathbf{y}$$

$$= \frac{1}{\lambda_n^{*\,2}} \int \frac{1}{\sum_\mathbf{x} w_\mathbf{x}} \left(\sum_{\forall \mathbf{x},'\boldsymbol{\chi}} w_{\mathbf{x}'} w_{\boldsymbol{\chi}} (x_n' - \chi_n)^2\right) d\mathbf{y}. \quad (15)$$

Note that $\mathcal{I}_n^{\text{loss}}$ is non-negative such that $\mathcal{I}_{n,n} \leq \mathcal{I}_{n,n}^{\text{ideal}}$ and that pairs of signals with $x_n' \neq \chi_n$ can contribute to $\mathcal{I}_n^{\text{loss}}$. Also note that, $w_{\mathbf{x}'} w_{\boldsymbol{\chi}} = p(\mathbf{y}|\mathbf{x}')p(\mathbf{y}|\boldsymbol{\chi})P(\mathbf{x}'|\boldsymbol{\lambda}^*)P(\boldsymbol{\chi}|\boldsymbol{\lambda}^*)$ and that $P(\mathbf{x}|\boldsymbol{\lambda}^*) > 0$ only when supp$(\mathbf{x}) \subseteq$ supp$(\boldsymbol{\lambda}^*)$. Thus, the Fisher Information is only reduced over the direct Poisson observation case when there are pairs of signals that are well-explained by the same $\mathbf{y}$ and by the same $\boldsymbol{\lambda}^*$. Clearly, $\boldsymbol{\lambda}^*$ with higher k will result in more of such pairs, and we empirically confirm this

influence of k on Fisher Information in Section IV-D. Although further precise analysis via Fisher Information is challenging, we provide deeper analysis of the special case of the MMVP problem with small $\boldsymbol{\lambda}$ through a different lens in the next section.

C. Small Scale Analysis

With identifiability, we know that $\boldsymbol{\lambda}^*$ uniquely maximizes the expected log-likelihood. However, because SPoRe uses stochastic gradient ascent to optimize the empirical log-likelihood, it will typically achieve a $\widehat{\boldsymbol{\lambda}}$ that is near but not equal to $\boldsymbol{\lambda}^*$. We therefore wish to understand how the neighborhood of $\boldsymbol{\lambda}^*$ changes given the parameters of the problem. The natural way to do this for MLE problems is to consider the Fisher Information matrix as in the previous section, but the presence of a sum inside the logarithm makes analysis difficult. Instead, we consider a particular $\widetilde{\boldsymbol{\lambda}}$ near $\boldsymbol{\lambda}^*$ that solves an optimization related to the original likelihood maximization problem. To further simplify the setting, we consider the "small scale" case where $\sum_n \lambda_n^*$ is small enough that there is almost never a case where $\sum_n x_n^* > 1$. We emphasize that although this setting is simple, the MLE approach can still drastically outperform a trivial solution such as $\widehat{\boldsymbol{\lambda}} = \mathbb{E}[\widehat{\mathbf{x}}]$, where $\widehat{\mathbf{x}} = \arg\max_\mathbf{x} p(\mathbf{y}|\mathbf{x})$, since with sufficient noise, $\widehat{\mathbf{x}} \neq \mathbf{x}^*$ with arbitrary probability (Section IV-B).

At the small scale, the distribution of each x_n^* becomes Bernoulli with parameter λ_n, and the probability that $x_n^* = 1$ and $x_{n'}^* = 1$ for $n \neq n'$ vanishes. Let $n^* \triangleq$ (the first nonzero index of $\mathbf{x}^*$, 0 if none), which has a categorical distribution with parameter $\boldsymbol{\lambda}^*$. We abuse notation so that $\boldsymbol{\phi}_0 = \mathbf{0}$, λ_0^* is the probability that $n^* = 0$, and $\sum_{n=0}^N \lambda_n^* = 1$. Applying Jensen's inequality to the log-likelihood for the conditional expectation given n^*, we obtain

$$\mathbb{E}\left[\log \sum_{n=0}^N p(\mathbf{y}|n)\lambda_n\right] \leq \mathbb{E}_{n^*}\left[\log \sum_{n=0}^N \mathbb{E}_{\mathbf{y}|n^*}\left[p(\mathbf{y}|n)\right]\lambda_n\right]. \quad (16)$$

Call the right-hand side of this inequality the *Jensen bound*. This Jensen bound via the logarithm has the attractive property of having a gradient that is equal to a first-order Taylor approximation of the gradient of the original likelihood.[1] To see this, consider the partial derivatives for a single λ_n:

$$\mathbb{E}\left[\frac{p(\mathbf{y}|n)}{\sum_{n'=0}^N p(\mathbf{y}|n')\lambda_{n'}}\right] \approx \mathbb{E}_{n^*}\left[\frac{\mathbb{E}_{\mathbf{y}|n^*}[p(\mathbf{y}|n)]}{\sum_{n'=0}^N \mathbb{E}_{\mathbf{y}|n^*}[p(\mathbf{y}|n')]\lambda_{n'}}\right]. \quad (17)$$

Thus, we can expect the optimizer of the Jensen bound to be close to $\boldsymbol{\lambda}^*$ (this is particularly true as measurement noise vanishes and the bound becomes tight).

In the case where $G = 1$ under AWGN, we have the following result characterizing the solution of the Jensen bound.

Proposition 3.5: If $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\phi}_{n^*}, \sigma^2 \mathbf{I})$ and

$$\mathbf{K} = \left(\exp\left\{-\frac{1}{4\sigma^2}\|\boldsymbol{\phi}_n - \boldsymbol{\phi}_{n'}\|_2^2\right\}\right)_{n,n'=0}^N \quad (18)$$

---

[1]The Taylor expansion is of $f(u,v) = u/v$, for which a first-order approximation yields $\mathbb{E}[U/V] \approx \mathbb{E}[U]/\mathbb{E}[V]$ for random variables U, V.

is invertible, then the maximizer $\widetilde{\boldsymbol{\lambda}}$ of the Jensen bound satisfies

$$\widetilde{\boldsymbol{\lambda}} \propto \mathbf{K}^{-1}\left(\frac{\boldsymbol{\lambda}^*}{\mathbf{K}^{-1}(\mathbf{s} - \boldsymbol{\mu})}\right). \tag{19}$$

where $\mathbf{s} \in \partial\|\widetilde{\boldsymbol{\lambda}}\|_1$ and for all n, $\mu_n \geq 0$ and $\mu_n \widetilde{\lambda}_n = 0$.

In the case where all entries of $\widetilde{\boldsymbol{\lambda}}$ are positive, $\mathbf{s} - \boldsymbol{\mu} = \mathbf{1}$. $\mathbf{K}$ has values of one along the diagonal and smaller values off the diagonal, so it mimics the identity matrix. Clearly, as $\mathbf{K} \rightarrow \mathbf{I}$, $\widetilde{\boldsymbol{\lambda}} \rightarrow \boldsymbol{\lambda}^*$. However, given nonzero σ^2, $\mathbf{K}$ is bounded away from $\mathbf{I}$. Furthermore, since it is impossible to find a set of more than $M+1$ equidistant points in $\mathbb{R}^M$, the off-diagonal values of $\mathbf{K}$ will differ when $M < N$, introducing distortion in the transformation.

However, even if $M < N$, if $\mathbf{y}$ is a measurement from a *random sensor group*, then the effect of this distortion can be mitigated such that $\widetilde{\boldsymbol{\lambda}}$ is a reliable estimator of $\boldsymbol{\lambda}^*$ from a support recovery perspective:

Theorem 3.6: If $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\phi}_{n^*}^{(g)}, \sigma^2\mathbf{I})$, g is distributed uniformly, and $\boldsymbol{\phi}_n^{(g)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I})$, then if $G \rightarrow \infty$ and all elements of the maximizer $\widetilde{\boldsymbol{\lambda}}$ of the Jensen bound are stricly positive, there exist $c_1 \geq 0$, $c_2 \in \mathbb{R}$ such that $\widetilde{\lambda}_n = c_1 \lambda_n^* + c_2$ for $1 \leq n \leq N$.

If $\widehat{\boldsymbol{\lambda}}$ has the same rank ordering as $\boldsymbol{\lambda}^*$, the exact support can be recovered. Therefore, we expect an increase in G to improve performance in tasks such as support recovery. From this result, however, we expect gains due to increasing G to be less immediate than those due to increasing M (and indeed, we see this in our simulations in Section IV-C). To see this, contrast the asymptotic nature of Theorem 3.6 in G with the fact that for a finite choice of M (specifically $M = N$) we can select all $\boldsymbol{\phi}_n$ equidistant (or that for M even smaller we can select $\boldsymbol{\Phi}$ satisfying a RIP with some acceptable distortion) and obtain the same reliability result.

IV. SIMULATIONS

In this section, we present comparisons of SPoRe against existing and custom baseline algorithms and follow with focused experimentation on SPoRe's performance and limitations. For most baseline algorithms that find an estimate $\widehat{\mathbf{X}}$, we set their estimates $\widehat{\boldsymbol{\lambda}} = \frac{1}{D}\sum_{d=1}^{D}\widehat{\mathbf{x}}_d$, i.e., the canonical Poisson MLE if $\mathbf{X}^*$ were observed directly. For a performance metric, we chose cosine similarity between $\widehat{\boldsymbol{\lambda}}$ and $\boldsymbol{\lambda}^*$ as it captures the relative distribution of elements of the solution which we believe is most useful to a user in biosensing applications. Although comparisons of cosine similarity mask differences in magnitude, estimates with high cosine similarity also exhibited low mean-squared error in our experience (results not shown). We plot cosine similarity alone for brevity. In all simulations, we use AWGN and set $\phi_{m,n}^{(g)} \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, 1)$ since many sensors are restricted to nonnegative measurements. For each parameter combination, we evaluate over 50 trials in which we draw new $\boldsymbol{\Phi}^{(g)}$ and $\boldsymbol{\lambda}^*$ for each trial. Due to high performance variability for some baseline algorithms, all error bars are scaled to $\pm\frac{1}{2}$ standard deviation for consistency and readability.

A. Comparison Against Existing Baselines

With no existing algorithm designed for Poisson signals, we compare against a number of algorithms with various relevant structural assumptions. We compare against both greedy and convex optimization approaches. First, we use DCS-SOMP [64], a generalization of the common baseline Simultaneous Orthogonal Matching Pursuit (S-OMP) [9] that assumes no structure and greedily solves MMV problems for any value of G. Next, we use NNS-SP and NNS-CoSaMP [37], two greedy algorithms for nonnegative MMV CS motivated by subspace pursuit (SP) [65] and compressive sampling matching pursuit (CoSaMP) [66] which exhibited the best empirical performance in [37]. For integer-based recovery, we use PROMP [45], an SMV algorithm for unbounded integer sparse recovery, to recover an estimate for each signal $\widehat{\mathbf{x}}_d$. We also use two support recovery algorithms with $M = \Omega(\sqrt{k})$ measurement rates: M-SBL [53] and a convex, ℓ_1-based variance recovery algorithm [54]. Support recovery is generally not quantitative and appears to make cosine similarity an inappropriate metric for comparison. However, these algorithms produce an estimate of the variance of $\mathbf{x}_n^*$, which equals $\boldsymbol{\lambda}_n^*$ for the Poisson distribution, making them coincidentally reasonable baseline approaches for MMVP directly.

For comparison against best-possible performance of the baselines and to avoid hyperparameter search where possible (for regularization weights, stopping criteria, etc.), we arm the baselines with relevant oracle knowledge of $\boldsymbol{\lambda}^*$ or $\mathbf{X}^*$. While NNS-SP and NNS-CoSaMP require k as an input, we also give DCS-SOMP and PROMP, algorithms that iteratively and irreversibly select support elements, knowledge of k and have them stop after k elements have been chosen. Additionally, we created three oracle-enabled convex algorithms that minimize $\sum_{g=1}^{G}\|\mathbf{Y} - \boldsymbol{\Phi}^{(g)}\mathbf{X}^{(g)}\|_F$ under norm constraints with oracle knowledge of the value of the norm. The ℓ_1 norm is commonly used as a penalty for convex solvers to encourage sparsity in sparse recovery. Our ℓ_1-Oracles include SMV and MMV versions, where in the SMV case, $\mathbf{Y}$ is collapsed to a single vector by summing $\sum_d \mathbf{y}_d$, and a vector $\widehat{\mathbf{x}}$ is recovered from which $\widehat{\boldsymbol{\lambda}} = \widehat{\mathbf{x}}/D$. The ℓ_1-Oracle SMV and ℓ_1-Oracle MMV are both given $\sum_{n,d} x_{n,d}^*$. In [56], $\|\mathbf{X}\|_{Rx}$ is suggested as a better alternative for MMV, so that our Rx-Oracle is given $\|\mathbf{X}^*\|_{Rx}$. In the support recovery algorithm from [54], the vector of sample variances $(\text{Var}(\mathbf{x}_n))_{n=1}^{N}$ is recovered under a constraint on $\sum_n \text{Var}(\mathbf{x}_n)$. Because we provide the sum of variances, we call this method the $\sum \text{Var}$-Oracle. We also enforce nonnegativity of the optimization variables in all convex problems, which we solve using the convex optimization package CVX in Matlab [67], [68]. We use no oracle knowledge for M-SBL, but we use the same initialization as for SPoRe and run its fixed point update until convergence.

From Fig. 3(a), we see the crucial result that the $M < k$ regime is only feasible with SPoRe, while conventional CS algorithms, both SMV and MMV, fail. Such a result is expected; generally speaking, CS algorithms seek to minimize measurement error ($\|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{X}\|_F$) while constraining the sparsity of the recovered solution. CS theory focuses on $M > k$ since if
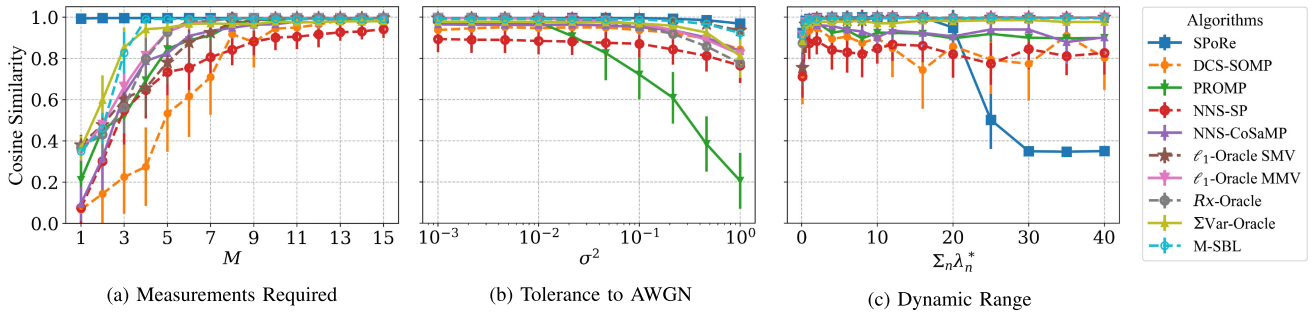
Fig. 3. Performance of SPoRe vs. compressed sensing baseline algorithms over 50 trials. Common settings unless otherwise specified are $M = 10$, $k = 3$, $N = 20$, $D = 100$, $G = 1$, $\sum_n \lambda_n^* = 2$. **(a)** Performance as a function of M, with $\sigma^2 = 10^{-6}$ for comparison in an effectively noiseless setting. **(b)** Performance as a function of AWGN variance, with $M = 10$. **(c)** Performance as a function of $\sum_n \lambda_n^*$, with $\sigma^2 = 10^{-2}$ and $M = 10$.

$M < k$, $M \times k$ submatrices of $\mathbf{\Phi}$ yield underdetermined systems in general. In other words, there simply cannot be unique k-sparse minimizers of measurement error alone with $M < k$, so the conventional CS problem is not well-posed, unlike in the MMVP problem. Support recovery algorithms M-SBL and the $\sum$Var-Oracle exhibit improved performance over other CS baselines as M decreases, but SPoRe remains far superior as $M \rightarrow 1$. Next, in Fig. 3(b), we set $M = 10$, a regime where most baselines performed nearly perfectly according to (Fig. 3(a)), and we increased the AWGN variance. We see that even in the conventional regime of $N > M > k$, SPoRe exhibits the highest noise tolerance which reflects the fact that its leverage of the Poisson assumption minimizes its dependence on accurate measurements. Lastly, however, in Fig. 3(c), SPoRe has the unique disadvantage of struggling to recover cases with high $\sum_n \lambda_n^*$. We observed that as $\sum_n \lambda_n^*$ increases, SPoRe's finite sampling results in few to no gradient steps taken as "good" samples with nonzero (numerically) $p(\mathbf{y}|\mathbf{x}_s)$ were drawn increasingly rarely, and SPoRe mistakenly terminates. Under AWGN, larger λ_n^* raises the signal-to-noise ratio but can paradoxically compromise SPoRe's performance. If $M \gg k$ is a practical design choice, practitioners should consider existing MMV approaches if $\sum_n \lambda_n^*$ may be highly variable.

B. Comparison Against Integer Baselines: $M < k$

In the $M < k$ regime, since with high probability we can bound the elements of $\mathbf{X}^*$, we might expect the discrete nature of the problem to admit at the least a brute-force solution for obtaining $\widehat{\mathbf{X}}$ that we can use to obtain $\widehat{\boldsymbol{\lambda}}$. Indeed, if measurement noise is low, then the integer signal that minimizes measurement error for $\mathbf{y}_d$ is likely to be $\mathbf{x}_d^*$. But a finite search space alone has not enabled integer-constrained CS research to achieve $M < k$ in general.

One may wonder whether SPoRe is simply taking advantage of this practically finite search space and, by virtue of MC sampling over thousands of iterations, is effectively finding the right solution by brute force. To address this possibility, we compare against an ℓ_0-Oracle that is given k and the maximum value in $\mathbf{X}^*$ in order to test all $\binom{N}{k}$ combinations of $\mathbf{X}$'s support. For each combination, it enumerates the $(\max(\mathbf{X}^*) + 1)^k$ possibilities for each $\mathbf{x}_d$ and selects $\widehat{\mathbf{x}}_d = \arg\min_{\mathbf{x}} \|\mathbf{y}_d - \mathbf{\Phi}\mathbf{x}\|_2$. Finally,

it returns the k-sparse solution with the lowest minimized measurement error. This algorithm is the only Poisson-free approach in this section.

Comparing SPoRe and other Poisson-enabled baselines against the ℓ_0-Oracle characterizes the effect of incorporating the Poisson assumption on recovery performance. An early solution of ours for tackling MMVP, which we now use as a baseline, was an alternating optimization framework to update estimates of $\widehat{\mathbf{X}} = \arg\max_{\mathbf{X}} p(\mathbf{X}|\mathbf{Y}, \widehat{\boldsymbol{\lambda}})$ and $\widehat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}} p(\boldsymbol{\lambda}|\mathbf{Y}, \widehat{\mathbf{X}})$. Noting that $p(\mathbf{X}|\mathbf{Y}, \widehat{\boldsymbol{\lambda}}) \propto p(\mathbf{Y}|\mathbf{X})p(\mathbf{X}|\widehat{\boldsymbol{\lambda}})$, this MAP framework for solving for $\mathbf{X}$ under AWGN with variance σ^2 is

$$\widehat{\mathbf{X}} = \arg\max_{\mathbf{X}} \frac{1}{D} \sum_{d=1}^{D} \log P(\mathbf{y}_d|\mathbf{x}_d) + \log p(\mathbf{x}_d|\widehat{\boldsymbol{\lambda}}) \quad (20)$$

$$= \arg\max_{\mathbf{X}} \frac{1}{D} \sum_{d=1}^{D} \left[-\frac{1}{2\sigma^2} \|\mathbf{y}_d - \mathbf{\Phi}\mathbf{x}_d\|_2^2 \right.$$

$$\left. + \sum_{n=1}^{N} \left(x_{n,d} \log \hat{\lambda}_n - \log \Gamma(x_{n,d} + 1) \right) \right], \quad (21)$$

where the Gamma function $\Gamma(\cdot)$ is the continuous extension of the factorial ($\Gamma(x_{n,d} + 1) = x_{n,d}!$) and is log-concave in the space of positive reals $\mathbb{R}_{++}^N$. We implemented the classic branch-and-bound (BB) algorithm [69] to find the optimal integer-valued solution $\widehat{\mathbf{X}}$ of the concave objective. Once an estimate $\widehat{\mathbf{X}}$ is available, the update to $\widehat{\boldsymbol{\lambda}}$ is also concave with the closed form solution $\widehat{\boldsymbol{\lambda}} = \frac{1}{D} \sum_d \widehat{\mathbf{x}}_d$. The biconcavity of this objective function in $\mathbf{X}$ and $\boldsymbol{\lambda}$ makes this approach attractive, but it is unclear how to best initialize $\widehat{\boldsymbol{\lambda}}$. We refer to this alternating baseline algorithm with the prefix "Alt" followed by the method of initialization. For example, for Alt-Random, we use random initialization with a small offset ($\hat{\lambda}_n \sim \text{Uniform}(0, 1) + 0.1$) to avoid making any particular λ_n irretrievable from the start.

We also explore a few "guided" initialization processes. The quantity $\sum_n \lambda_n^*$ can hypothetically be estimated from data if $P(\mathbf{x}_d = \mathbf{0}|\boldsymbol{\lambda})$ is significant and easily estimated from $\mathbf{Y}$. In fact, quantification in microfluidics often relies on a clear identification of empty sample partitions (that is, where $\mathbf{x}_d = \mathbf{0}$) [26]. This motivates a strategy of relaxing the problem by optimizing $\mathbf{X}$ with a Poisson assumption on the sum of each column $\sum_n x_{n,d}$ rather than each element of $\mathbf{X}$ individually. The $\sum_n \lambda_n^*$-Oracle
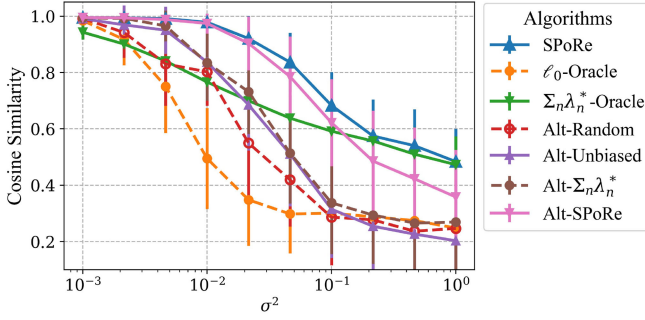
Fig. 4.   AWGN tolerance of integer-restricted algorithms over 50 trials with $M = 2, k = 3, N = 10, D = 100, G = 1, \sum_n \lambda_n^* = 2$.

is given $\sum_n \lambda_n^*$ and optimizes for $P(\sum_n x_{n,d} | \sum_n \lambda_n^*)$ in place of $P(\mathbf{x}_d | \widehat{\boldsymbol{\lambda}})$ in (20). It is straightforward to show that this objective is also concave. Each estimate $\widehat{\mathbf{x}}_d$ is solved via BB from which $\widehat{\boldsymbol{\lambda}} = \frac{1}{D} \sum_d \widehat{\mathbf{x}}_d$. We use the $\sum_n \lambda_n^*$-Oracle as its own baseline and as an initialization to our alternating framework (Alt-$\sum_n \lambda_n^*$). For the next guided initialization, we again use $\sum_n \lambda_n^*$ for an unbiased initialization where the first estimate of $\hat{\lambda}_n = (\sum_{n'} \lambda_{n'}^*)/N$ for all n (Alt-Unbiased). Finally, we used the output of SPoRe as an initial value for $\widehat{\boldsymbol{\lambda}}$ (Alt-SPoRe). Alt-SPoRe can be understood as a way to use SPoRe to estimate $\widehat{\mathbf{X}}$ if needed.

Fig. 4 illustrates that SPoRe has the greatest tolerance to measurement noise whereas the ℓ_0-Oracle has the least. This comparison illustrates the value of incorporating the Poisson assumption in recovery; specifically, the integer and sparsity structures (perfectly captured by the ℓ_0-Oracle) are not sufficient for recovery under measurement noise. The alternating optimization algorithm's behavior was unexpected; initialization (other than with Alt-SPoRe) does not appear to have a major influence on performance. Surprisingly, comparing Alt-$\sum_n \lambda_n^*$ versus $\sum_n \lambda_n^*$-Oracle and Alt-SPoRe versus SPoRe, alternating seems to worsen the performance under high noise. Our interpretation is that in high noise settings, the ability of SPoRe to not "overcommit" to a particular solution $\widehat{\mathbf{x}}_d$ may be especially effective when $\boldsymbol{\lambda}^*$ is the signal of interest rather than $\mathbf{X}^*$. Any given measurement $\mathbf{y}_d$ may make the specific estimate $\widehat{\mathbf{x}}_d$ arbitrarily unreliable. In our alternating framework with $\widehat{\mathbf{x}}_d$ recovered separately for each d, errors on individual estimates accumulate. SPoRe instead makes gradient steps based on batches of observations, helping it maintain awareness of the full distribution of measurements.

C. Sparsity and $\sum_n \lambda_n^*$

We empirically tested the limitations of SPoRe's recovery performance under very challenging conditions of $M = 2$, $3 \leq k \leq 7, N = 50, \sigma^2 = 10^{-2}$. Here we set $D = 1000$ to better reflect the typical capabilities of biomedical systems, whereas $D = 100$ in our baseline comparisons was due to our budget on computational time strained by solving BB for $\widehat{\mathbf{x}}_d$. From our analysis and previous simulations, we expect that both k and the

magnitudes of λ_n^* will influence recovery. Fig. 5 probes when and why SPoRe fails. Fig. 5(a) illustrates SPoRe's performance decreases with increasing k and $\sum_n \lambda_n^*$. To elucidate the cause of poor performance, Fig. 5(b) shows SPoRe's performance under the same conditions when initialized at the optimum. SPoRe's maintenance of high cosine similarity in this case means that in Fig. 5(a), SPoRe is converging to incorrect optima (or terminating before convergence). These two figures depict fundamental limitations of stochastic optimization in a challenging landscape.

Moreover, Fig. 5(c) illustrates that MC gradients decrease in quality with high $\sum_n \lambda_n^*$ and k. In SPoRe, recall that we set a minimum $\hat{\lambda}_n \geq \epsilon = 10^{-3}$ so that nonzero x_n^* have a chance of being sampled for all n. We keep S fixed as we increase k and $\sum_n \lambda_n^*$, and we see that the variance of the gradient increases at coordinates where $\hat{\lambda}_n = \epsilon$ and $\lambda_n^* = 0$. Such an effect accounts for some drift from the optimum observed in Fig. 5(b) that increases with k, and we believe that it helps to explain the inability to converge to the optimum in Fig. 5(a). Future work can explore alternative techniques for stochastic optimization and sampling. Practitioners may benefit significantly from reducing $\sum_n \lambda_n^*$ if faced with limitations in M. In a biosensing context, these results indicate that given a microfluidics system with fixed D, having fewer total analytes in the sample (i.e., smaller $D \sum_n \lambda_n^*$) can counterintuitively improve performance and may be particularly important if both M and G are limited.

However, note in (8) that SPoRe's gradients are defined by an average of $\mathbf{x}_s$ weighted by $p(\mathbf{y}_d | \mathbf{x}_s)$. The previous result from Fig. 3(c) in which SPoRe performed well with $\sum_n \lambda_n^* \leq 20$ when $M = 10$ illustrates that limitations of MC sampling may be offset by improving the ability of $p(\mathbf{y} | \mathbf{x}_s)$ to guide gradients. In Fig. 6, we explore this notion further for M-constrained systems by increasing G. One may wonder how increasing G compares to a CS problem with GM measurements (i.e., $\bar{\boldsymbol{\Phi}} \in \mathbb{R}^{GM \times N}$). Although $\boldsymbol{\lambda}^*$ is fixed across groups, the $\mathbf{x}_d$ are random such that there is no reasonable method to directly stack individual measurements $\mathbf{y}_d$ from multiple groups. Instead, we created a new baseline ℓ_1-Oracle GM SMV. Denote the average of measurements and signals in each group g as $\bar{\mathbf{y}}^{(g)}$ and $\bar{\mathbf{x}}^{(g)}$, respectively. Our new baseline stacks all $\bar{\mathbf{y}}^{(g)}$ into one measurement vector $\bar{\mathbf{y}} \in \mathbb{R}^{GM}$ and minimizes $\|\bar{\mathbf{y}} - \bar{\boldsymbol{\Phi}}\boldsymbol{\lambda}\|_2$ with respect to $\boldsymbol{\lambda}$ given $\sum_{n,d} x_{n,d}^*$ and $\boldsymbol{\lambda} \geq 0$. Stacking measurements and sensing matrices implicitly assumes that for each group, $\bar{\mathbf{y}}^{(g)} \approx \boldsymbol{\Phi}^{(g)}\boldsymbol{\lambda}^*$, or that $\bar{\mathbf{x}}^{(g)} \approx \boldsymbol{\lambda}^* \; \forall g$. It can be easily shown that the relative errors in these approximations reduce with increasing D or $\boldsymbol{\lambda}^*$.

Although increasing D is feasible in microfluidics, it generally corresponds with a reduction in $\boldsymbol{\lambda}^*$ since a sample's total analyte content is fixed. Therefore, in Fig. 6, we focus on the influence of the magnitude of $\boldsymbol{\lambda}^*$. In Fig. 6(a), we used the most challenging settings from Fig. 5 with $k = 7$ and $\sum_n \lambda_n^* = 10$. As expected from our analysis in Section III-C, larger choices of M make SPoRe much more effective per sensor group, but near perfect recovery is achievable even with $M = 1$. However, note that the new oracle baseline performs almost identically to SPoRe, with SPoRe exhibiting a modest advantage only when GM is comparable to or less than k. When we reduce $\sum_n \lambda_n^*$ to
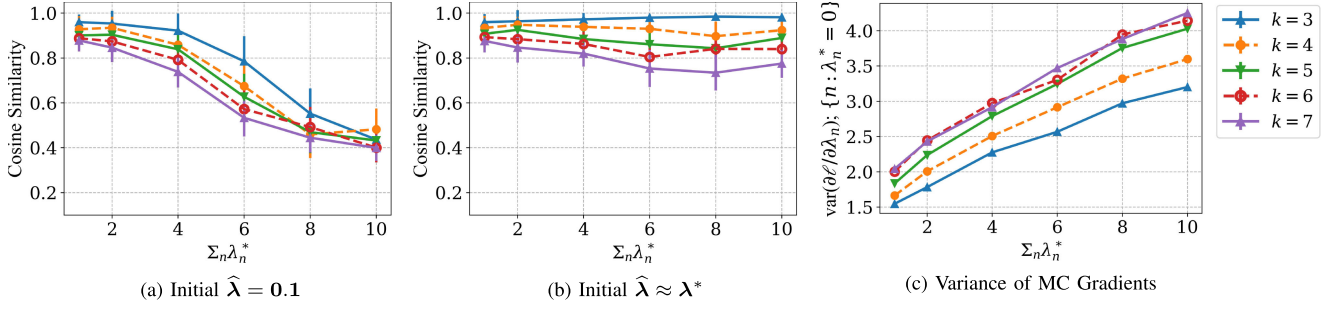
Fig. 5. SPoRe's performance and behavior as a function of k and $\sum_n \lambda_n^*$ over 50 trials. Common settings unless otherwise specified are $M = 2$, $N = 50$, $D = 1000$, $G = 1$, $\sigma^2 = 10^{-2}$. **(a)** Performance when initialized with standard $\widehat{\boldsymbol{\lambda}} = \mathbf{0.1}$. **(b)** Performance when initialized with $\widehat{\boldsymbol{\lambda}} \approx \boldsymbol{\lambda}^*$, specifically $\hat{\lambda}_n = \max\{\epsilon, \lambda_n^*\}$. **(c)** Average variance of partial derivatives for indices $n \notin \mathrm{supp}(\boldsymbol{\lambda}^*)$ evaluated at $\widehat{\boldsymbol{\lambda}} \approx \boldsymbol{\lambda}^*$.
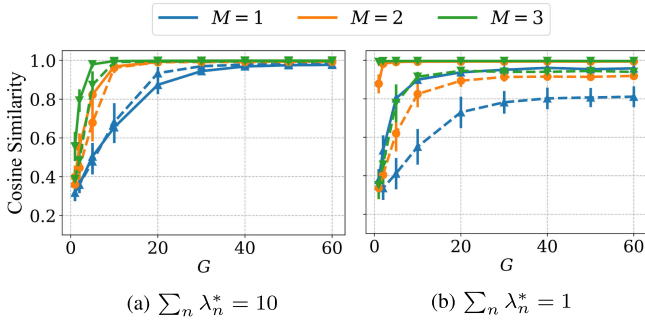

Fig. 6. Performance of SPoRe (solid) vs. ℓ_1-Oracle GM SMV (dashed) as a function of G. Common settings are $k = 7$, $N = 50$, $D = 1000$, $\sigma^2 = 10^{-2}$. **(a)** Comparison with $\sum_n \lambda_n^* = 10$, motivated by Fig. 5(a). **(b)** Comparison with $\sum_n \lambda_n^* = 1$.
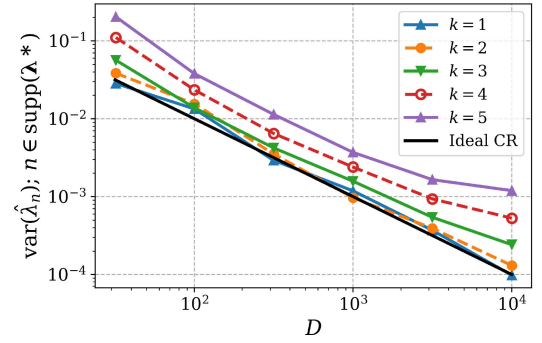

Fig. 7. Comparison of average variance of $\hat{\lambda}_n$ from SPoRe versus the ideal Cramér–Rao (CR) bound as a function of D over 50 trials with $n \in \mathrm{supp}(\boldsymbol{\lambda}^*)$, $\lambda_n^* = 1$, $M = 2$, $N = 50$, $G = 20$, $\sigma^2 = 10^{-2}$.

1 in Fig. 6(b), the assumption that $\bar{\mathbf{x}}^{(g)} \approx \boldsymbol{\lambda}^*$ becomes far less valid. As a result, the performance improvement with SPoRe is dramatic. For instance, what SPoRe achieves with $M = 1$ is only matched by the oracle baseline when $M = 3$. For applications in which $\bar{\mathbf{x}}^{(g)} \approx \boldsymbol{\lambda}^*$ and $GM > k$, practitioners could consider reformulating the recovery problem as a standard CS problem. However, SPoRe is uniquely suited for systems with $GM < k$ and is the best generalized approach for applications with smaller $\boldsymbol{\lambda}^*$ or D.

D. Efficiency

For system design, it is helpful to know how many observations D are necessary and sufficient for stable estimation of $\widehat{\boldsymbol{\lambda}}$. Such insight is often derived from the analysis of Fisher Information $\mathcal{I}$. Recall that for direct observations of Poisson signals $\mathbf{x}_d^*$, the ideal case, the MLE solution $\hat{\lambda}_n = \sum_d x_{n,d}^* / D$ is an *efficient estimator* and achieves the *Cramér–Rao* bound such that $\mathrm{var}(\hat{\lambda}_n) = \lambda_n^*/D$.

In Section III-B, we derived the reduction in $\mathcal{I}_{n,n}$ from MMVP measurements and found reason to expect that the reduction increases with k. Here, we empirically characterize this effect (Fig. 7). The matrix $\mathcal{I}$ is evaluated at $\boldsymbol{\lambda}^*$, so we only consider $n \in \mathrm{supp}(\boldsymbol{\lambda}^*)$. In MMVP, the variance of $\hat{\lambda}_n$ will depend on $\boldsymbol{\Phi}$ and n, but by redrawing random $\boldsymbol{\Phi}$ and $\boldsymbol{\lambda}^*$ over

50 trials, we hope to smooth out these dependencies and capture the broader effect of low dimensional projections. For a concise comparison considering $n \in \mathrm{supp}(\boldsymbol{\lambda}^*)$, we set all $\lambda_n^* = 1$, pool all of $\hat{\lambda}_n$ across all trials, and compute a single average variance for each k and D. Because the Fisher Information describes the optimization landscape near the optimum, we chose parameter settings ($M = 2$, $G = 20$) based on our results in Fig. 6(a) to be confident that SPoRe is arriving near the optimum and the estimation variance is not confounded by poor estimates.

Fig. 7 shows a noticeable increase in the estimation variance and verifies that this deviation from the ideal bound is exponentially worsened in k. However, we argue that the variance quickly becomes negligible at reasonable D for practical purposes. Practitioners could consider the necessary precision of estimation and the maximum expected k for an application, increase D as needed, and worry little about the influence of noisy measurements in low dimensions.

V. DISCUSSION

We have found that the structure in the MMVP problem can be easily exploited for substantial improvements in signal recovery. While compressed sensing of arbitrary integer signals has proven challenging in the past, Poisson constraints not only make the recovery problem tractable, but even significantly

easier. Most inverse problems necessitate constraints that make the signal-to-measurement transformations nearly isometric: in compressed sensing, these manifest as restrictions on $\mathbf{\Phi}$, noise, and the relationship between M, N, and k. In MMVP, recovery of $\boldsymbol{\lambda}^*$ is theoretically possible under very lax conditions on $\mathbf{\Phi}$ (Theorem 3.2) and practically achievable as shown in our simulations.

In practice, our new SPoRe algorithm exhibits high performance even under deliberately challenging circumstances of high noise and $M < k$. Here, recovery with $G = 1$ appears particularly feasible even with noisy measurements if $\sum_n \lambda_n^*$ is small. Because the log-likelihood is not concave, SPoRe's gradient ascent approach is not theoretically guaranteed to find a global optimum since local optima may exist. However, if they exist, we speculate that SPoRe is naturally poised to evade these traps due to stochasticity in its gradient steps from both batch draws and MC integrations.

We noted a few scenarios in which SPoRe's MC sampling appears to cause issues with convergence or early termination that are generally associated with increases in k and $\sum_n \lambda_n^*$. We anticipate that further increases in N may also contribute to these effects. While k and N are entirely determined by the application, system designers can reduce $\sum_n \lambda_n^*$ by increasing the spatial or temporal sampling rate. In microfluidics, this adjustment translates to either generating more (smaller) partitions D given a fixed sample volume or diluting a sample prior to partitioning. Our initial implementation of SPoRe uses $S = 1000$, can easily run on personal computers, and is sufficient for systems with $N < 10^2$. This scale is appropriate for most applications in biosensing, and future work with parallelized or adaptive sampling strategies could improve the reliability of recovery for larger systems. Moreover, we found that increasing M and G appear to mitigate poor performance due to excessive sampling noise.

To date, CS implementations with low M and high G, based on our notation, reformulate the acquired measurements into a canonical CS problem with GM measurements [10]. As we discussed in Section IV-C, this reformulation implicitly assumes that the average signal in every group ($\bar{\mathbf{x}}^{(g)}$) is approximately $\boldsymbol{\lambda}^*$. However, this assumption fails in biosensing in two common scenarios: the total number of analytes in a typical sample may be small, or the sensing mechanism may impose restrictions on D. In such cases, SPoRe is the superior approach allowing $M < k$, achieving high performance even with $M = 1$ under high noise. In this work, we found practical limits on the sparsity level k that we could accurately solve, hence the "S" in SPoRe, but future research should explore the conditions needed for the reliable recovery of non-sparse signals. High performance in this regime would inspire new methods in biosensing for the analysis of heterogeneous samples with many types of cells.

The ability to recover signals in MMVP, even in the extreme case of $M = G = 1$, is unprecedented in CS and offers a new paradigm for sensor-constrained applications. The current state-of-the-art method for achieving similar efficiency in microfluidics is to essentially guarantee single-analyte capture for classification by substantially increasing the sampling rate,

whereas our MMVP framework is not reliant on such an intervention. Increasing G can make SPoRe reliable under harsh conditions, is straightforward with microfluidics, and offers a potent alternative to increasing the sampling rate. For example, diluting to a tolerable concentration is challenging with samples of unknown content such as in diagnostics applications, and increasing D either delays results or necessitates high throughput measurement acquisition which may not be feasible depending on the nature of the sensors. Our group is continuing research in CS-based microbial diagnostics [18] by working towards an *in vitro* demonstration of MMVP-based diagnostics with microfluidics.

Our initial theoretical and empirical results show the promise of MMVP, but there are many directions for further research. For instance, theoretical results that precisely relate M, N, D, k, $\boldsymbol{\lambda}^*$ and noise such as in a recovery guarantee would be highly valuable. Moreover, SPoRe can accept any signal-to-measurement model $p(\mathbf{y}|\mathbf{x})$. While we have proven identifiability under linear mappings with common noise models, SPoRe can be easily applied with any application-specific model[2] even if proving identifiability of the Poisson mixture is difficult. With growing interest in microfluidics, SPoRe's promising performance in the MMVP problem warrants further research to ensure that the statistical assumptions underlying these new technologies are leveraged to their full potential.

APPENDIX

A. Proof of Theorem 3.2

We use a direct proof, assuming $P(\mathcal{C}_\mathbf{u}|\boldsymbol{\lambda}) = P(\mathcal{C}_\mathbf{u}|\boldsymbol{\lambda}') \; \forall \mathbf{u}$ and proving the resulting implication $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$. Let $\mathbf{z}^{(\mathbf{x})} \in \mathbb{Z}^N$ be such that $\mathbf{x} + \mathbf{z}^{(\mathbf{x})} \in \mathcal{C}_\mathbf{x}$. By Definition 3.4, $\mathbf{x} + \mathbf{z}^{(\mathbf{x})} \in \mathbb{Z}_+^N$ and $\mathbf{z}^{(\mathbf{x})} \in \mathcal{N}(\mathbf{\Phi})$.

Lemma A.1: If $\mathcal{N}(\mathbf{\Phi}) \cap \mathbb{R}_+^N = \{\mathbf{0}\}$, and $P(\mathcal{C}_\mathbf{0}|\boldsymbol{\lambda}) = P(\mathcal{C}_\mathbf{0}|\boldsymbol{\lambda}')$, then $\sum_n \lambda_n = \sum_n \lambda_n'$.

Proof: The null space condition on $\mathbf{\Phi}$ means that $\mathcal{C}_\mathbf{0} = \{\mathbf{0}\}$; there is no vector $\mathbf{z}^{(\mathbf{0})}$ that satisfies $\mathbf{0} + z^{(\mathbf{0})} \in \mathbb{Z}_+^N$ other than $\mathbf{z}^{(\mathbf{0})} = \mathbf{0}$. Therefore, $P(\mathcal{C}_\mathbf{0}|\boldsymbol{\lambda}) = P(\mathcal{C}_\mathbf{0}|\boldsymbol{\lambda}') \Rightarrow e^{-\sum_n \lambda_n} = e^{-\sum_n \lambda_n'} \Rightarrow \sum_n \lambda_n = \sum_n \lambda_n'$. $\square$

We now turn our attention to the *one-hot* collision sets. Let $\mathbf{e}_j$ denote the jth standard basis vector. By Definition 3.4, $\mathcal{C}_{\mathbf{e}_j} = \{\mathbf{x} : \mathbf{\Phi}\mathbf{x} = \phi_j, \mathbf{x} \in \mathbb{Z}_+^N\}$. For $\mathcal{C}_{\mathbf{e}_j}$ that contain *only* $\mathbf{e}_j$, we have the following result:

Lemma A.2: If $\mathcal{N}(\mathbf{\Phi}) \cap \mathbb{R}_+^N = \{\mathbf{0}\}$ and $\mathcal{C}_{\mathbf{e}_j} = \{\mathbf{e}_j\}$, then $\lambda_j = \lambda_j'$.

Proof: The restriction on $\mathcal{C}_{1j}$ means $P(\mathcal{C}_{1j}|\boldsymbol{\lambda}) = P(\mathcal{C}_{1j}|\boldsymbol{\lambda}') \Rightarrow \lambda_j e^{-\sum_n \lambda_n} = \lambda_j' e^{-\sum_n \lambda_n'}$. Applying Lemma A.1 yields $\lambda_j = \lambda_j'$. $\square$

By similar arguments to Lemmas A.1 and A.2, we can prove Corollary III.3 under the assumption that there are no collisions instead of the null space condition. When the elements of $\mathbf{\Phi}$ are independently drawn from continuous distributions, the collision of any particular $\mathbf{x}$ and $\mathbf{x}'$ occurs with probability zero.

---

[2]Our full code base is available at https://github.com/pavankkota/SPoRe with instructions on how to implement SPoRe with custom models.

Since $\mathbb{Z}_+^N \times \mathbb{Z}_+^N$ is countably infinite, there are no collisions almost surely. As such, $\mathcal{C}_{\mathbf{e}_j} = \{\mathbf{e}_j\} \forall j$, and therefore $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$.

For the conditions in Theorem III.2, the following Lemma states the existence of at least one j satisfying $\mathcal{C}_{\mathbf{e}_j} = \{\mathbf{e}_j\}$:

Lemma A.3: If $\mathcal{N}(\boldsymbol{\Phi}) \cap \mathbb{R}_+^N = \{\mathbf{0}\}$ and $\boldsymbol{\phi}_n \neq \boldsymbol{\phi}_{n'} \, \forall n, n' \in \{1, \ldots, N\}$ with $n \neq n'$, then $\exists \, j$ such that $\mathcal{C}_{\mathbf{e}_j} = \{\mathbf{e}_j\}$.

Proof: If $\mathcal{C}_{\mathbf{e}_j} = \{\mathbf{e}_j\}$, then $\nexists \mathbf{z}^{(\mathbf{e}_j)} \neq \mathbf{0}$. Define P as the number of one-hot collision sets that contain more than just $\mathbf{e}_j$, and note that $P \leq N$. Without loss of generality, let us say that $\mathcal{C}_{\mathbf{e}_j}$ for $j \in \{1, \ldots, P\}$ meet this condition. Lemma A.3 effectively says that $P < N$, such that $N - P > 0$ one-hot collision sets contain only $\mathbf{e}_j$. We proceed with a proof by contradiction by assuming $P = N$.

By our null space condition, $\mathbf{z}^{(\mathbf{e}_j)}$ must contain both positive and negative integers. There are two additional conditions on nontrivial $\mathbf{z}^{(\mathbf{e}_j)}$. First, because $\mathbf{e}_j + \mathbf{z}^{(\mathbf{e}_j)} \in \mathbb{Z}_+^N$, the only negative component of $\mathbf{z}^{(\mathbf{e}_j)}$ is $z_j^{(\mathbf{e}_j)} = -1$. To see this, if $z_i^{(\mathbf{e}_j)}$ for $i \neq j$ were negative, then $\mathbf{e}_j + \mathbf{z}^{(\mathbf{e}_j)}$ would be negative at index i, and if $\mathbf{z}_j^{(\mathbf{e}_j)}$ were less than -1, then $\mathbf{e}_j + \mathbf{z}^{(\mathbf{e}_j)}$ would be negative at index j. Second, $\mathbf{z}^{(\mathbf{e}_j)}$'s positive elements must total to at least 2. A single positive element of $z_i^{(\mathbf{e}_j)} = 1$ would imply that $\boldsymbol{\phi}_i = \boldsymbol{\phi}_j$, violating a condition on $\boldsymbol{\Phi}$.

With $P = N$, let us concatenate the $\mathbf{z}^{(\mathbf{e}_j)}$ column vectors into a matrix for visualization.

$$\mathbf{Z} = \begin{bmatrix} -1 & z_1^{(\mathbf{e}_2)} & \ldots & z_1^{(\mathbf{e}_N)} \\ z_2^{(\mathbf{e}_1)} & -1 & \ldots & z_2^{(\mathbf{e}_N)} \\ \vdots & \vdots & \ddots & \vdots \\ z_N^{(\mathbf{e}_1)} & z_N^{(\mathbf{e}_2)} & \ldots & -1 \end{bmatrix}. \tag{22}$$

Note that each column $\mathbf{z}^{(\mathbf{e}_j)}$ in this matrix is symbolic for any vector that satisfies the conditions we described. All columns of $\mathbf{Z}$ are in $\mathcal{N}(\boldsymbol{\Phi})$. Any linear combination of vectors in $\mathcal{N}(\boldsymbol{\Phi})$ are in $\mathcal{N}(\boldsymbol{\Phi})$. Let $\mathcal{S}$ represent a subset of indices of the columns of $\mathbf{Z}$ and let $\mathbf{z}^{\mathcal{S}} \triangleq \sum_{j \in \mathcal{S}} \mathbf{z}^{(\mathbf{e}_j)}$.

First, let $\mathcal{S} = \{1, \ldots, N\}$. Because all off-diagonal components in $\mathbf{Z}$ are nonnegative and because $\mathbf{z}^S$ must have one negative value, one of the rows of $\mathbf{Z}$ must be entirely zero except for the -1 on the diagonal. Note the ordering of the columns in $\mathbf{Z}$ is arbitrary, so without loss of generality, let this be the first row. Now, let's say that $\mathcal{S} = \{2, 3, \ldots, N\}$. The same logic holds: at least one row must contain all zeros except for the -1. Without loss of generality, we can set $[\mathbf{Z}]_{2,3}, [\mathbf{Z}]_{2,4}, \ldots [\mathbf{Z}]_{2,N} = 0$. Repeating this process, we get a lower triangular matrix:

$$\mathbf{Z} = \begin{bmatrix} -1 & 0 & 0 & \ldots & 0 \\ z_2^{(\mathbf{e}_1)} & -1 & 0 & \ldots & 0 \\ z_3^{(\mathbf{e}_1)} & z_3^{(\mathbf{e}_2)} & -1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_N^{(\mathbf{e}_1)} & z_N^{(\mathbf{e}_2)} & z_N^{(\mathbf{e}_3)} & \ldots & -1 \end{bmatrix}. \tag{23}$$

However, examining the final column, we see that $\mathbf{z}^{(\mathbf{e}_N)}$ is a vector of all zeros and one -1, such that it cannot be in $\mathcal{N}(\boldsymbol{\Phi})$, proving Lemma A.3 by contradiction. $\square$

Proof: Proof of Theorem 3.2: Lemma A.3 confirms $P < N$, meaning that we can form the concatenated matrix of $\mathbf{z}^{\mathbf{e}_j}$ vectors:

$$\mathbf{Z} = \begin{bmatrix} -1 & 0 & \ldots & 0 \\ z_2^{(\mathbf{e}_1)} & -1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ z_P^{(\mathbf{e}_1)} & z_P^{(\mathbf{e}_2)} & \ldots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ z_N^{(\mathbf{e}_1)} & z_N^{(\mathbf{e}_2)} & \ldots & z_N^{(\mathbf{e}_P)} \end{bmatrix}. \tag{24}$$

Let us now apply $P(\mathcal{C}_{\mathbf{e}_P} | \boldsymbol{\lambda}) = P(\mathcal{C}_{\mathbf{e}_P} | \boldsymbol{\lambda}')$. For all $\mathbf{x} \in \mathcal{C}_{\mathbf{e}_P}$,

$$\prod_{n=1}^N \frac{\lambda_n^{x_n}}{x_n!} - \prod_{n=1}^N \frac{{\lambda'_n}^{x_n}}{x_n!} = 0, \tag{25}$$

$$\left(\prod_{\forall i > P} \frac{\lambda_i^{x_i}}{x_i!} \right) \left(\frac{\lambda_P^{x_P}}{x_P!} - \frac{{\lambda'_P}^{x_P}}{x_P!} \right) = 0, \tag{26}$$

where Lemma A.1 ($\sum_n \lambda_n = \sum_n \lambda'_n$) yields the first equality, and Lemma A.2 (all $\lambda_i = \lambda'_i \, \forall i > P$) yields the second equality when combined with the fact that $x_i = 0$ for $i < P$ due to (24). The only $\mathbf{x} \in \mathcal{C}_{\mathbf{e}_P}$ with $x_P \neq 0$ is $\mathbf{e}_P$ which simplifies (26) to $\lambda_P = \lambda'_P$.

Now we have $\lambda_i = \lambda'_i \, \forall i > P - 1$. Following the same arguments, we can start from $P(\mathcal{C}_{\mathbf{e}_{P-1}} | \boldsymbol{\lambda}) = P(\mathcal{C}_{\mathbf{e}_{P-1}} | \boldsymbol{\lambda}')$ and arrive at $\lambda_{P-1} = \lambda'_{P-1}$. Applying this repeatedly ultimately yields $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$, proving Theorem 3.2.

B. Proof of Proposition 3.5

Proof: By straightforward integration,

$$\mathbb{E}_{\mathbf{y}|n^*}[p(\mathbf{y}|n)] \propto \underbrace{\exp \left\{ -\frac{1}{4\sigma^2} \|\boldsymbol{\phi}_n - \boldsymbol{\phi}_{n^*}\|_2^2 \right\}}_{\triangleq \kappa(n, n^*)}. \tag{27}$$

Therefore, given the constraints $\|\boldsymbol{\lambda}\|_1 \leq 1$ and $\lambda_n \geq 0$, the first-order KKT condition is

$$\mathbb{E}_{n^*} \left[\frac{\boldsymbol{\kappa}(n^*)}{\langle \boldsymbol{\kappa}(n^*), \boldsymbol{\lambda} \rangle} \right] = c\mathbf{s} - \boldsymbol{\mu}, \tag{28}$$

where $\boldsymbol{\kappa}(n^*) = (\kappa(n, n^*))_{n=0}^N$, $\mathbf{s} \in \partial \|\boldsymbol{\lambda}\|_1$, $c \geq 0$, and $\mu_n \geq 0$. By complementary slackness, $\mu_n \lambda_n = 0$ for all n. Because $\mathbf{K}$ is symmetric, we can rewrite the above as

$$\mathbf{K} \left(\frac{\boldsymbol{\lambda}^*}{\mathbf{K}\boldsymbol{\lambda}} \right) = c\mathbf{s} - \boldsymbol{\mu}, \tag{29}$$

where the fraction represents element-wise division. Solving for $\boldsymbol{\lambda}$ and rescaling $\boldsymbol{\mu}$, we obtain the desired expression.

C. Proof of Theorem 3.6

Proof: Let κ_g be defined the same as κ from (27) with $\boldsymbol{\phi}_n^{(g)}$. Then again by straightforward integration,

$$\tilde{\kappa}(n, n^*) \triangleq \mathbb{E}_g [\kappa_g(n, n^*)] \tag{30}$$

$$= \begin{cases} 1 & n = n^*, \\ \left(\frac{2\sigma^2}{2\sigma^2+1}\right)^{M/2} & n \neq n^*, 0 \in \{n, n^*\}, \\ \left(\frac{\sigma^2}{\sigma^2+1}\right)^{M/2} & n \neq n^*, 0 \notin \{n, n^*\}. \end{cases} \quad (31)$$

Let $\widetilde{\mathbf{K}} = (\tilde{\kappa}(n, n'))_{n,n'=0}^{N}$, and let $\widehat{\mathbf{K}} = (\tilde{\kappa}(n, n'))_{n,n'=1}^{N}$ be the sub-matrix of $\widetilde{\mathbf{K}}$ excluding $n = 0$. Then

$$\widehat{\mathbf{K}} = (1-a)\mathbf{I} + a\mathbf{J}, \quad (32)$$

where $a = \left(\frac{\sigma^2}{\sigma^2+1}\right)^{M/2}$ and $\mathbf{J}$ is a matrix of all ones. Leveraging the block matrix inverse, we observe that we have the form

$$\widetilde{\mathbf{K}}^{-1} = \begin{bmatrix} \frac{1}{1-a}\mathbf{I} + b\mathbf{J} & c\mathbf{1} \\ c\mathbf{1}^T & d \end{bmatrix}, \quad (33)$$

assuming the final column corresponds to $n = 0$, for some scalars b, c, and d. Using the formula from Proposition 3.5 and the fact that $\mathbf{s} - \boldsymbol{\mu} = \mathbf{1}$ by assumption, we conclude that for $n > 0$, $\tilde{\lambda}_n \propto \lambda_n^* + C$ for some C. Rote algebra verifies that the constant of proportionality is non-negative. $\qquad \square$

ACKNOWLEDGMENT

REFERENCES

[1] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[2] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] A. Cohen, W. Dahmen, and R. Devore, "Compressed sensing and best *k*-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, Jan. 2009.

[7] J. Chen and X. Huo, "Sparse representations for multiple measurement vectors (MMV) in an over-complete dictionary," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2005, pp. 257–260.

[8] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[9] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[10] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[11] R. M. Willett, R. F. Marcia, and J. M. Nichols, "Compressed sensing for practical optical imaging systems: A tutorial," *Opt. Eng.*, vol. 50, no. 7, Jul. 2011.

[12] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[13] S. Li, L. D. Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2177–2186, Nov. 2013.

[14] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 454–465, Feb. 2011.

[15] T. Ragheb, J. N. Laska, H. Nejati, S. Kirolos, R. G. Baraniuk, and Y. Massoud, "A prototype hardware for random demodulation based compressive analog-to-digital conversion," in *Proc. 51st Midwest Symp. Circuits Syst.*, 2008, vol. 4, pp. 37–40.

[16] B. Cleary, L. Cong, A. Cheung, E. S. Lander, and A. Regev, "Efficient generation of transcriptomic profiles by random composite measurements," *Cell Theory*, vol. 171, no. 6, pp. 1424–1436, Nov. 2017.

[17] D. Koslicki, S. Foucart, and G. Rosen, "WGSQuikr: Fast whole-genome shotgun metagenomic classification," *PLoS One*, vol. 9, no. 3, Mar. 2014, Art. no. e 91784.

[18] A. Aghazadeh *et al.*, "Universal microbial diagnostics using random DNA probes," *Sci. Adv.*, vol. 2, no. 9, Sep. 2016, Art. no. e1600025.

[19] S. Ghosh *et al.*, "A compressed sensing approach to group-testing for COVID-19 detection," *IEEE Open J. Signal Process.*, vol. 2, pp. 248–264, 2021.

[20] M. T. Guo, A. Rotem, J. A. Heyman, and D. A. Weitz, "Droplet microfluidics for high-throughput biological assays," *Lab Chip*, vol. 12, pp. 2146–2155, Feb. 2012.

[21] M. Hosokawa, Y. Nishikawa, M. Kogawa, and H. Takeyama, "Massively parallel whole genome amplification for single-cell sequencing using droplet microfluidics," *Sci. Rep.*, vol. 7, Jul. 2017, Art. no. 5199.

[22] L. Mazutis, J. Gilbert, W. L. Ung, D. A. Weitz, A. D. Griffiths, and J. A. Heyman, "Single-cell analysis and sorting using droplet-based microfluidics," *Nature Protoc.*, vol. 8, no. 5, pp. 870–891, Apr. 2013.

[23] B. Vogelstein and K. W. Kinzler, "Digital PCR," *Proc. Nat. Acad. Sci. USA*, vol. 96, no. 16, pp. 9236–9241, Aug. 1999.

[24] S. H. Kim, S. Iwai, S. Araki, S. Sakakihara, R. Iino, and H. Noji, "Large-scale femtoliter droplet array for digital counting of single biomolecules," *Lab Chip*, vol. 12, no. 23, pp. 4986–4991, Dec. 2012.

[25] V. Yelleswarapu, J. R. Buser, M. Haber, J. Baron, E. Inapuri, and D. Issadore, "Mobile platform for rapid sub-picogram-per-milliliter, multiplexed, digital droplet detection of proteins," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 10, pp. 4489–4495, Feb. 2019.

[26] A. S. Basu, "Digital assays part I: Partitioning statistics and digital PCR," *SLAS Technol.*, vol. 22, no. 4, pp. 369–386, Aug. 2017.

[27] S. Moon, E. Ceyhan, U. A. Gurkan, and U. Demirci, "Statistical modeling of single target cell encapsulation," *PLoS One*, vol. 6, no. 7, Jul. 2011, Art. no. e 21580.

[28] D. O. Velez *et al.*, "Massively parallel digital high resolution melt for rapid and absolutely quantitative sequence profiling," *Sci. Rep.*, vol. 7, Feb. 2017, Art. no. 42326.

[29] B. M. Peters, M. A. Jabra-Rizk, G. A. O'May, J. W. Costerton, and M. E. Shirtliff, "Polymicrobial interactions: Impact on pathogenesis and human disease," *Clin. Microbiol. Rev.*, vol. 25, no. 1, pp. 193–213, Jan. 2012.

[30] S. Vural, X. Wang, and C. Guda, "Classification of breast cancer patients using somatic mutation profiles and machine learning approaches," *BMC Syst. Biol.*, vol. 10, no. 62, pp. 263–276, Aug. 2016.

[31] M. Raginsky, R. M. Willett, Z. T. Harmany, and R. F. Marcia, "Compressed sensing performance bounds under Poisson noise," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3990–4002, Aug. 2010.

[32] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms-theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, Mar. 2012.

[33] A. Suea-Ngam, P. D. Howes, M. Srisa-Art, and A. J. deMello, "Droplet microfluidics: From proof-of-concept to real-world utility," *Chem. Commun.*, vol. 55, no. 67, pp. 9895–9903, Jul. 2019.

[34] A. Orth *et al.*, "Super-multiplexed fluorescence microscopy via photostability contrast," *Biomed. Opt. Exp.*, vol. 9, no. 7, pp. 2943–2954, Jul. 2018.

[35] Y. Zhu and Q. Fang, "Analytical detection techniques for droplet microfluidics—A review," *Anal. Chim. Acta*, vol. 787, no. 17, pp. 24–35, Jul. 2013.

[36] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 40–44.

[37] D. Kim and J. P. Haldar, "Greedy algorithms for nonnegativity-constrained simultaneous sparse recovery," *Signal Process.*, vol. 125, pp. 274–289, Aug. 2016.

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

[39] L. Fukshansky, D. Needell, and B. Sudakov, "An algebraic perspective on integer sparse recovery," *Appl. Math. Comput.*, vol. 340, no. 1, pp. 31–42, Jan. 2019.

[40] U. Nakarmi and N. Rahnavard, "BCS: Compressive sensing for binary sparse signals," in *Proc. IEEE Mil. Commun. Conf.*, 2012, pp. 1–5.

[41] Z. Tian, G. Leus, and V. Lottici, "Detection of sparse signals under finite-alphabet constraints," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2009, pp. 2349–2352.

[42] S. C. Draper and S. Malekpour, "Compressed sensing over finite fields," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 669–673.

[43] B. Shim, S. Kwon, and B. Song, "Sparse detection with integer constraint using multipath matching pursuit," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1851–1854, Oct. 2014.

[44] J.-H. Lange, M. E. Pfetsch, B. M. Seib, and A. M. Tillmann, "Sparse recovery with integrality constraints," *Discrete Appl. Math.*, vol. 283, pp. 346–366, Sep. 2020.

[45] A. Flinth and G. Kutyniok, "PROMP: A sparse recovery approach to lattice-valued signals," *Appl. Comput. Harmon. Anal.*, vol. 45, pp. 668–708, Mar. 2017.

[46] S. Sparrer and R. F. H. Fischer, "MMSE-based version of OMP for recovery of discrete-valued sparse signals," *Electron. Lett.*, vol. 52, no. 1, pp. 75–77, Jan. 2016.

[47] M. Alam and Q. Zhang, "A survey: Non-orthogonal multiple access with compressed sensing multiuser detection for mMTC," *Wireless Commun. Mobile Comput.*, vol. 2021, 2021, Art. no. 8840519.

[48] Y. Ji, C. Bockelmann, and A. Dekorsky, "Compressed sensing based multi-user detection with modified sphere detection in machine-to-machine communications," in *Proc. 10th Int. ITG Conf. Syst., Commun. Coding*, 2015, pp. 1–6.

[49] F. Monsees, C. Bockelmann, and A. Dekorsky, "Reliable activity detection for massive machine to machine communication via multiple measurement vector compressed sensing," in *Proc. IEEE Globecom Workshops*, 2014, pp. 1057–1062.

[50] J. Liu, H.-Y. Cheng, C.-C. Liao, and A.-Y. A. Wu, "Scalable compressive sensing-based multi-user detection scheme for Internet-of-Things applications," in *Proc. IEEE Workshop Signal Process. Syst.*, 2015, pp. 1–6.

[51] M. Alam and Q. Zhang, "Enhanced compressed sensing based multiuser detection for machine type communication," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2018, pp. 1–6.

[52] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.

[53] O. Balkan, K. Kreutz-Delgado, and S. Makeig, "Localization of more sources than sensors via jointly-sparse Bayesian learning," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 131–134, Feb. 2014.

[54] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 711–726, Feb. 2015.

[55] A. Koochakzadeh, H. Qiao, and P. Pal, "On fundamental limits of joint sparse support recovery using certain correlation priors," *IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4612–4625, Sep. 2018.

[56] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, Mar. 2006.

[57] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[58] J. Chen and R. Luss, "Stochastic gradient descent with biased but consistent gradient estimators," 2019, *arXiv:1807.11880*.

[59] H. Teicher, "Identifiability of finite mixtures," *Ann. Math. Statist.*, vol. 34, no. 4, pp. 1265–1269, Dec. 1963.

[60] G. M. Tallis, "The identifiability of mixtures of distributions," *J. Appl. Probability*, vol. 6, no. 2, pp. 389–398, 1969.

[61] S. J. Yakowitz and J. D. Spragins, "On the identifiability of finite mixtures," *Ann. Math. Statist.*, vol. 39, no. 1, pp. 209–214, 1968.

[62] L. Yang and X. Wu, "A new sufficient condition for identifiability of countably infinite mixtures," *Metrika*, vol. 77, pp. 377–387, May 2013.

[63] D. Y. Zhang, S. X. Chen, and P. Yin, "Optimizing the specificity of nucleic acid hybridization," *Nature Chem.*, vol. 4, pp. 208–214, Jan. 2012.

[64] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," Jan. 2009, *arXiv:0901.3403*.

[65] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[66] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

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

[68] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Adv. Learn. Control*, in Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds., Berlin, Germany: Springer, 2008, pp. 95–110. [Online]. Available: http://stanford.edu/boyd/graph_dcp.html

[69] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, Jul. 1960.

Pavan K. Kota received the B.S.E. degree in biomedical engineering from Case Western Reserve University, Cleveland, OH, USA, in 2017. He is currently working toward the Ph.D. degree with the Department of Bioengineering, Rice University, Houston, TX, USA. He is an NIH NLM Fellow of Biomedical Informatics and Data Science. His main research interests include the application of signal processing and machine learning techniques to biomedical diagnostics.

Daniel LeJeune (Member, IEEE) received the B.S. degree in engineering from McNeese State University, Lake Charles, LA, USA, in 2014, and the M.S. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA, in 2016. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA. His research interests include machine learning theory and adaptive algorithms.

Rebekah A. Drezek received the B.S.E. degree in electrical engineering from Duke University, Durham, NC, USA, in 1996, and the M.Sc. and Ph.D. degrees in electrical engineering from The University of Texas at Austin, Austin, TX, USA, in 1998 and 2001, respectively.

She is currently a Professor and the Associate Chair with the Department of Bioengineering, Rice University, Houston, TX, USA. Her research interests include optical imaging, nanomedicine, and diagnostics. She is a Fellow of AIMBE.

Richard G. Baraniuk (Fellow, IEEE) received the B.S. degree in electrical engineering from the University of Manitoba, Winnipeg, MB, Canada, in 1987, the M.S. degree from the University of Wisconsin-Madison, Madison, WI, USA, in 1988, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1992.

He is currently the Victor E. Cameron Professor of electrical and computer engineering with Rice University, Houston, TX, USA, and the Founding Director of OpenStax. He holds 35 U.S. and six foreign patents. His research interests include new theory, algorithms, and hardware for sensing, signal processing, and machine learning. He is a Fellow of the American Academy of Arts and Sciences, National Academy of Inventors, and American Association for the Advancement of Science. He was the recipient of the DOD Vannevar Bush Faculty Fellow Award (National Security Science and Engineering Faculty Fellow), IEEE James H. Mulligan, Jr. Education Medal, and IEEE Signal Processing Society Technical Achievement, Education, Best Paper, Best Magazine Paper, and best column awards.