Sparse Representation in Fourier and Local Bases Using ProSparse: A Probabilistic Analysis

Yue M. Lu, *Senior Member, IEEE*, Jon Oñativia, *Member, IEEE*, and Pier Luigi Dragotti, *Fellow, IEEE*

Abstract—**Finding the sparse representation of a signal in an overcomplete dictionary has attracted a lot of attention over the past years. This paper studies ProSparse, a new polynomial complexity algorithm that solves the sparse representation problem when the underlying dictionary is the union of a Vandermonde matrix and a banded matrix. Unlike our previous work which establishes deterministic (worst-case) sparsity bounds for ProSparse to succeed, this paper presents a probabilistic average-case analysis of the algorithm. Based on a generating-function approach, closed-form expressions for the exact success probabilities of ProSparse are given. The success probabilities are also analyzed in the high-dimensional regime. This asymptotic analysis characterizes a sharp phase transition phenomenon regarding the performance of the algorithm.**

I. INTRODUCTION

Let $y \in \mathbb{C}^N$ be a complex signal that admits a sparse representation in an overcomplete dictionary. That is,

$$y = D x, \qquad (1)$$

where $D = [\Psi, \Phi]$ is the union of two bases or frames and x is a sparse coefficient vector. Given y, the sparse vector x can be recovered by greedy algorithms such as orthogonal matching pursuit and its variants (e.g., [1]–[4]) or by convex relaxation techniques such as basis pursuit (BP) [5], [6]. The performance of these algorithms and the corresponding requirements on the dictionary D are now well understood [6]–[14]. For a comprehensive overview of this topic, we refer the reader to the book [15].

Recently, a new algorithm, named *ProSparse* in [16], has been presented for the case where one of the sub-dictionaries, $\Psi \in \mathbb{C}^{N \times M}$, is a Vandermonde matrix (with $M \geq N$), and the other sub-dictionary Φ is an $N \times N$ banded matrix. We note that a canonical example of Ψ is the Fourier basis (or frame), whereas typical examples of Φ include the standard

Y. M. Lu is with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA (e-mail: yuelu@seas.harvard.edu). His work was supported in part by the US National Science Foundation under grant CCF-1319140, and by the US Army Research Office under contract W911NF-16-1-0265.

J. Oñativia was with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK. He is now with Egile, 20850, Spain (e-mail: jon.onativia@ieee.org). P. L. Dragotti is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK (e-mail: p.dragotti@imperial.ac.uk). J. Onativia and P. L. Dragotti were supported by the European Research Council Starting Investigator Grant through the RecoSamp Project under Award 277800.

This work was presented in part at the Signal Processing with Adaptive Sparse Structured Representations (SPARS 2015) Workshop.

basis ($\Phi = I_N$), the short-time Fourier transform, and pulse signals that are often encountered in radar and ultrasound applications. It was shown in [16] that ProSparse, an algorithm with polynomial complexity, can recover x from (1), provided that

$$2(P + b_\Phi)(K + \tau) < N + \tau(2b_\Phi + 1). \qquad (2)$$

Here, P and K denote the sparsity level of x in Ψ and Φ, respectively: the overall sparsity of x is then equal to $P + K$; b_Φ denotes the bandwidth of the banded matrix Φ; and τ is a constant that takes two possible values: $\tau = 0$ if Ψ is a circulant matrix and $\tau = 1$ otherwise. It is interesting to point out that the performance bound in (2) does not depend on M, the size of the sub-dictionary Ψ. This is in fact a consequence of a useful property of ProSparse: unlike BP, the performance of ProSparse does not degrade when Ψ is highly overcomplete, *i.e.*, when M is much larger than N. (More discussions on this point can be found in Section IV-B.)

For the prototypical case where Ψ is the Fourier basis and $\Phi = I_N$, we have $\tau = 0$ and $b_\Phi = 0$. The bound in (2) can then be simplified as

$$PK < N/2. \qquad (3)$$

We note that this sparsity requirement is much weaker than the corresponding tight BP bound [6] and the unicity bound [5] in the literature (see Figure 1 in [16] for a comparison of these different bounds.) It also improves the theoretical recovery threshold given in [13, Theorem 7] (when specialized to the case of Fourier and identity matrices) by a factor of two. Interestingly, the same sparsity requirement (3) was previously established in an early work [17], under the assumption that the support pattern of x in the sub-dictionary Φ is *known*. ProSparse achieves the same performance bound *without* the knowledge of the support of x. Finally, we also point out that (3) still suffers from the so-called *squared-root bottleneck* [11].

As a deterministic and worst-case bound, the condition in (2) guarantees the success of ProSparse in recovering every single x for which it holds. The bound is also tight in the sense that one can construct counterexamples for which (2) is not satisfied and the algorithm fails. However, these counterexamples have special support structures, the occurrence of which can be rare. Motivated by our desire to understand the typical performance of ProSparse, we present in this paper an average-case analysis of ProSparse within a probabilistic setting. Our contributions are as follows:

1. *Exact success probabilities.* We provide closed-form expressions for the exact success probabilities of ProSparse in recovering a sparse signal x, when the sparsity pattern of x is

(a) $K = \lfloor \alpha N \rfloor$ (b) $K = \lfloor \alpha N^{0.6} \rfloor$
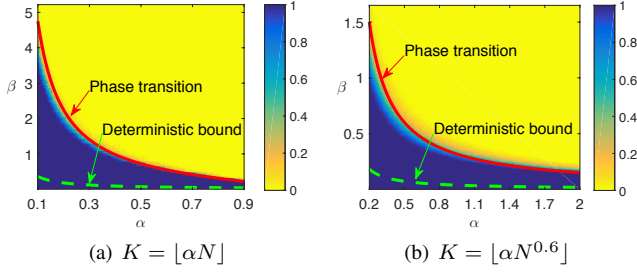
Fig. 1. Success probabilities of ProSparse: simulations vs. theory. The figures show empirical success probabilities of ProSparse in recovering a sparse signal when the dictionary is the union of a Fourier frame and the canonical basis. The signal dimension is set to $N = 10^6$. In (a), the number of Fourier atoms is $P = \lfloor \beta \log N \rfloor$, and the number of "spikes" is $K = \lfloor \alpha N \rfloor$. A different scaling is shown in (b), where $P = \lfloor \beta N^{0.4} \log N \rfloor$ and $K = \lfloor \alpha N^{0.6} \rfloor$. For each pair of (α, β), 100 different spike patterns have been generated uniformly at random. In both (a) and (b), the solid lines show the predicted phase transition boundaries given in (5), and the dashed lines show the deterministic bound given in (2).

drawn uniformly at random. This model turns out to be equivalent to the problem of discrete circle covering [18], [19] in geometric probability. The success probabilities of ProSparse can be evaluated by known formulas in the literature, but we provide an alternative expression in Proposition 1. Based on a simple generating-function approach, our new expression can be efficiently evaluated via the Fourier transform, and its simple form also plays a key role in our subsequent analysis in the high-dimensional setting.

2. Asymptotic analysis. As the main technical contribution of this work, we analyze the above-mentioned success probabilities in the asymptotic, high-dimensional regime. Our asymptotic analysis reveals a phase transition phenomenon regarding the performance of ProSparse. Specifically, suppose that the support of $\boldsymbol{x}$ involves $K = \lfloor \alpha N^{\delta} \rfloor$ atoms from $\boldsymbol{\Phi}$ and $P = \lfloor \beta N^{1-\delta} \log N \rfloor$ atoms from $\boldsymbol{\Psi}$. Here, $0 < \delta \leq 1$, $\alpha > 0$ and $\beta > 0$ are three positive constants. (When $\delta = 1$, the constant α is further required to be less than 1.) We show in Proposition 2 that, if the K atoms from $\boldsymbol{\Phi}$ are chosen uniformly at random, then

$$\mathbf{P}(\text{ProSparse recovers } \boldsymbol{x}) \xrightarrow{N \to \infty} \begin{cases} 1, & \text{if } \beta < \beta_c(\alpha, \delta) \\ 0, & \text{if } \beta > \beta_c(\alpha, \delta), \end{cases} \tag{4}$$

where the critical threshold is given by

$$\beta_c(\alpha, \delta) = \begin{cases} \frac{\delta}{2\alpha}, & \text{if } 0 < \delta < 1 \\ \frac{-1}{2 \log(1-\alpha)}, & \text{if } \delta = 1. \end{cases} \tag{5}$$

We illustrate this phase transition phenomenon in Figure 1, for two different values of δ. The two axes of the figures correspond to the parameters α and β, respectively. The red, solid lines show the phase transition boundaries [$\beta = \frac{-1}{2 \log(1-\alpha)}$ for $\delta = 1$ in Figure 1(a) and $\beta = \frac{\delta}{2\alpha}$ for $\delta = 0.6$ in Figure 1(b)]. The theoretical predictions match very well with results from Monte Carlo simulations (for $N = 10^6$), delineating the two regions where ProSparse succeeds or fails with high probability. For comparison, we also plot the deterministic bound given in (2), which is much more pessimistic than the actual average-case performance of the algorithm.

The rest of the paper is organized as follows. Section II provides a brief overview of ProSparse, highlighting its key ideas. In Section III, we present an average-case performance analysis of ProSparse by deriving closed-form expressions for the exact success probabilities of ProSparse in sparse recovery. Section IV studies the asymptotic regime, where we prove the phase transition result shown in (4) and compare it with related results in the literature. We conclude in Section V.

II. OVERVIEW OF PROSPARSE

In this section, we briefly recall the key ideas behind ProSparse [16]. This discussion serves as the basis for our probabilistic performance analysis carried out in later sections.

Let $\boldsymbol{D} = [\boldsymbol{\Psi}, \boldsymbol{\Phi}]$ be the union of two sub-dictionaries. Of the two, $\boldsymbol{\Psi} \in \mathbb{C}^{N \times M}$ is a Vandermonde matrix with entries

$$\Psi_{n,m} = \xi_m^n, \quad \text{for } 0 \leq n < N, 0 \leq m < M,$$

where $\{\xi_m \in \mathbb{C} : 0 \leq m < M\}$ is a set of distinct numbers. For simplicity, we assume that the other sub-dictionary $\boldsymbol{\Phi}$ is an $N \times N$ identity matrix. (The more general case involving banded matrices can be handled by following the same idea. See Remark 2 below for discussions.) Let $\boldsymbol{x}_1 \in \mathbb{C}^M$ and $\boldsymbol{x}_2 \in \mathbb{C}^N$ be two vectors containing P and K nonzero entries, respectively. Our goal is to recover the (P, K)-sparse vector $\boldsymbol{x} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T]^T$ from the measurement $\boldsymbol{y} = [\boldsymbol{\Psi}, \boldsymbol{\Phi}] \boldsymbol{x}$.

Given the specific structure of the two sub-dictionaries, the nth entry of $\boldsymbol{y}$ can be written as follows:

$$y_n = \sum_{p=1}^{P} a_p \, \xi_{m_p}^n + \sum_{k=1}^{K} b_k \, \delta_{n-n_k}, \qquad n = 0, 1, \ldots, N-1.$$

Here, δ_n is the Kronecker delta function; the indices $\{m_p\}_{p=1}^{P}$ and $\{n_k\}_{k=1}^{K}$ specify the sparsity patterns of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively; $\{a_p\}$ and $\{b_k\}$ denote the corresponding nonzero coefficients.

ProSparse is based on two simple observations:

First, since the basis elements from $\boldsymbol{\Phi}$ are local, many of the entries in $\boldsymbol{y}$ are only due to the Vandermonde components $\boldsymbol{\Psi x}_1$. Specifically, we can find "clean" windows of consecutive entries of $\boldsymbol{y}$ that can be expressed as merely a sum of exponentials

$$y_n = \sum_{p=1}^{P} a_p \xi_{m_p}^n, \tag{6}$$

where we have restricted the index n to a "clean" window. See Figure 2 for an illustration.

Second, as long as we have at least $2P$ consecutive entries of the form (6), we can exactly recover the unknown parameters $\{a_p, \xi_{m_p}\}_{p=1}^{P}$ from these entries. A classical approach is to use Prony's method, first proposed by Baron de Prony in 1795 [20]. Since then, this algebraic method for parameter estimation has been used in many different contexts [21].

Exploiting the two observations stated above, the ProSparse algorithm operates by performing an exhaustive search over all possible sliding windows $y_\ell, y_{\ell+1}, \ldots, y_{\ell+2P-1}$ of length $2P$, for every $P < N/2$. Note that, since these sliding windows are sequential, the exhaustive search has polynomial complexity.
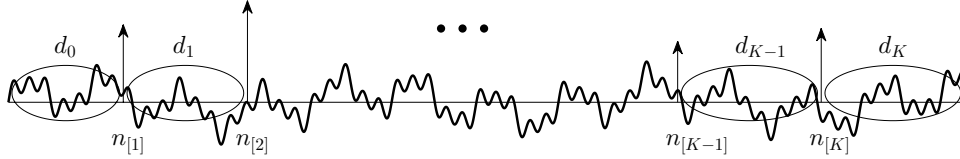
Fig. 2. Samples of y_n located between consecutive spikes are only due to the Vandermonde components, and they can be written in the form of (6). The sum of all the gaps $\{d_k\}$ is equal to $N - K$. ProSpare sequentially searches through all sliding windows of size $2P$. It is able to recover the unknown parameters $\{a_p, \xi_{m_p}\}_p$ in (6) [and subsequently the sparse signal $\boldsymbol{x}$] if there is a gap of length at least $2P$.

For each candidate sliding window, the algorithm uses Prony's method to try to estimate the parameters $\{a_p, \xi_{m_p}\}_{p=1}^{P}$, from which we can build a residual vector $\boldsymbol{r}$ with entries $r_n = y_n - \sum_{p=1}^{P} a_p \xi_{m_p}^n$. If the residual is sparse, it then directly corresponds to the spikes $\sum_{k=1}^{K} b_k \delta_{n-n_k}$. When the sparsity levels satisfy (2) (with $\tau = 0$ and $b_{\boldsymbol{\Phi}} = 0$), we can show that there always exist clean intervals of length at least $2P$, and thus ProSparse is guaranteed to recover the original signal $\boldsymbol{x}$. We refer the reader to [16, Proposition 2] for a formal proof. Strategies for further reducing computational complexity and discussions on issues involving multiple solutions can also be found in [16].

Remark 1 (The Fourier basis): Consider the special case when $\boldsymbol{\Psi}$ is the Fourier basis, that is, when $M = N$ and $\Psi_{n,m} = \omega^{mn}$ with $\omega = e^{-2\pi j/N}$. The problem becomes circulant due to the periodicity of the exponential function ω^n. This means that the search for the clean windows can be performed on a ring, with the last entry $n = N-1$ immediately followed by the first entry $n = 0$. The parameter τ in the general bound (2) is introduced to exactly take into account this special case: $\tau = 0$ if the Vandermonde matrix $\boldsymbol{\Psi}$ is the Fourier matrix, and $\tau = 1$ for other generic Vandermonde matrices for which we cannot exploit the periodicity.

Remark 2 (Banded matrices): The idea described above is applicable to more general cases when $\boldsymbol{\Phi}$ is a banded matrix, *i.e.*,

$$\Phi_{n,m} = 0 \quad \text{if} \quad m < n - b_{\boldsymbol{\Phi}} \quad \text{or} \quad m > n + b_{\boldsymbol{\Phi}},$$

where $b_{\boldsymbol{\Phi}}$ is the bandwidth of $\boldsymbol{\Phi}$. Let $\{n_k : 1 \le k < K\}$ denote the indices corresponding to the sparsity pattern of $\boldsymbol{x}_2$. These indices break the interval $[0, N-1]$ into $K+1$ segments (or K segments if $\boldsymbol{\Psi}$ is periodic.) It is easy to see that a sufficient condition for ProSparse to work is to have at least one segment whose length is greater than or equal to $2P + 2b_{\boldsymbol{\Phi}}$. This is the origin of the first term on the left-hand side of (2).

Remark 3 (Further generalizations): Throughout this paper, we assume that $\boldsymbol{\Psi}$ is a Vandermonde matrix. However, it is possible to loosen this requirement by considering a larger class of $\boldsymbol{\Psi}$-matrices satisfying the so-called *local sparse reconstruction property* [16]. In short, those are matrices that allow for the efficient reconstruction of sparse signals from *any* small blocks of consecutive elements in the transform domain. Beyond Vandermonde matrices considered in this work, other examples of such matrices include the DCT transform (and its variants) and random Gaussian dictionaries. More details on this possible generalization, including the precise definition of

the local sparse reconstruction property, can be found in [16, Sec. IV].

Finally, we note that the bounds in (2) and (3) are essentially tight: we can construct sparse signals for which the bounds do not hold and ProSparse fails to recover the signals. One such counterexample is the "picket-fence" signal, with the spikes (or local atoms) from $\boldsymbol{\Phi}$ placed at equally-spaced locations. However, such signals are rare, if the sparsity patterns of $\boldsymbol{x}$ are drawn uniformly at random. The next section provides an exact probability analysis of the performance of ProSparse. We will show that, indeed, the deterministic bounds (2) and (3) are overly conservative.

III. EXACT PROBABILISTIC PERFORMANCE ANALYSIS

From the discussions in the previous section, we see that the success of ProSparse crucially depends on whether there exists a gap of sufficient length after we break the interval $[0, 1, \ldots, N - 1]$ into a given number of pieces. In what follows, we study this question in a probabilistic setting, with randomly chosen "breaking points".

A. Probability Model and Analysis

Consider a set of N integers $0, 1, \ldots, N-1$ arranged along a line. We draw K distinct numbers, denoted by $n_1, n_2, \ldots, n_K$, by using random sampling without replacement from the set. Now consider the order statistics obtained by arranging the n_k's in increasing order:

$$0 \le n_{[1]} < n_{[2]} < \ldots < n_{[K]} \le N - 1.$$

We define the gaps between consecutive points as

$$d_k \overset{\text{def}}{=} n_{[k+1]} - n_{[k]} - 1, \quad \text{for } 0 \le k \le K, \tag{7}$$

where two additional end-points $n_{[0]} = -1$ and $n_{[K+1]} = N$ are introduced. By construction, $d_k \ge 0$, and it must hold that

$$\sum_{k=0}^{K} d_k = N - K. \tag{8}$$

Discrete random variables $\{d_k\}$ satisfying the above sum constraint arise in statistical physics, and they are often referred to as the Bose-Einstein model; see [22, pp. 20–21].

Figure 2 illustrates a typical realization of this probabilistic model. We note that, in our analysis of ProSparse, the numbers $\{n_k\}$ represent the support of a K-sparse signal whose sparsity pattern is chosen at random, whereas the gaps $\{d_k\}$ are the lengths of those clean intervals within which (6) holds.

Denote the maximum gap by

$$\Delta_{N,K} \stackrel{\text{def}}{=} \max_{0 \le k \le K} \{d_k\}.$$

A central problem in our analysis is to understand the probability distribution of $\Delta_{N,K}$, *i.e.*, to compute

$$h_{N,K}(s) \stackrel{\text{def}}{=} \mathbf{P}(\Delta_{N,K} < s)$$

for any positive integer s. For the special case when $\mathbf{\Psi}$ is the unitary Fourier matrix (see Remark 1 in Section II), we need to modify $\Delta_{N,K}$ and study the probability distribution of a related quantity

$$\Gamma_{N,K} \stackrel{\text{def}}{=} \max \left\{ d_0 + d_K, \max_{1 \le k < K} \{d_k\} \right\},$$

which takes into account the circulant nature of the problem.

Note that the model we consider here is the discrete version of the classical spacing problem (see, *e.g.*, [23]–[26]), in which a unit circle is broken at K randomly chosen points and the question of interest is to study the various statistical properties of the lengths of the intervals between consecutive points. Compared to the long history and vast literature of the continuous spacing problem, discussions on the discrete version of the problem appeared in the literature much later [18], [27], [28], [19]. It was first introduced by Holst [18], who computed, among other quantities, the probability distribution of the following random variable[1]

$$V_{N,K} \stackrel{\text{def}}{=} \sum_{k=0}^{K} (d_k - (s-1))_+ ,$$

where $(x)_+ = \max\{x, 0\}$. Clearly, the maximum gap $\Delta_{N,K} < s$ if and only if $V_{N,K} = 0$. It then follows that we can use the probability distribution of $V_{N,K}$ presented in [18] (see [19, p. 123] for a correction) to get

$$h_{N,K}(s) = \mathbf{P}(V_{N,K} = 0) = \frac{\sum_{\ell=0}^{K+1} (-1)^\ell \binom{K+1}{\ell} \binom{N-\ell s}{K}}{\binom{N}{K}}. \tag{9}$$

As an alternative way to reach (9), we can interpret $h_{N,K}(s)$ as the distribution of the largest order statistic of the random variables $\{d_0, d_1, \ldots, d_K\}$. Although d_i are not independent, they form an exchangeable family. Thus, (9) can also be seen as a special case of a general recursive formula for the distribution of orders statistics of exchangeable random variables (see, *e.g.*, [29, p. 46]).

The right-hand side of (9) involves an alternating sum of products of binomial coefficients. Therefore, evaluating (9) numerically becomes challenging even for moderate values of N and K. In what follows, we prove an alternative expression for $h_{N,K}(s)$, which, to our knowledge, has not been presented in the literature before. Based on integer powers of a certain generating function, our new expression can be efficiently and accurately evaluated by the discrete Fourier transform. More importantly, this new expression plays a key role in the asymptotic analysis of $h_{N,K}(s)$ to be carried out in Section IV.

---

[1]We have adapted the definition of $V_{N,K}$ and the subsequent formula (9) to reflect the fact that we are using slightly different notational conventions from those in [18], [19].

Before presenting our result, we first introduce the notation

$$\text{coef}\left\{(c_0 + c_1 x + \ldots + c_n x^n), x^\ell\right\} = c_\ell,$$

to refer to the ℓth coefficient of a polynomial. (Note that $c_\ell \equiv 0$ for $\ell > n$.)

Proposition 1: For any positive integer s,

$$h_{N,K}(s) = \frac{\text{coef}\left\{f_s(x)^{K+1}, x^{N-K}\right\}}{\binom{N}{K}}, \tag{10}$$

where $f_s(x) \stackrel{\text{def}}{=} \sum_{i=0}^{s-1} x^i$ is a polynomial of degree $s-1$.

Proof: Consider the experiment of drawing K unique numbers from $0, 1, \ldots, N-1$ by sampling without replacement. We are interested in computing the probability that the maximum gap between consecutive numbers is smaller than s. This probability is given by the ratio between the number of outcomes with $\Delta_{N,K} < s$ and the total number of possible outcomes. The denominator is clearly given by the binomial coefficient $\binom{N}{K}$. For the numerator, we note that the number of outcomes associated with $\Delta_{N,K} < s$ is equal to the cardinality of the following set

$$\left\{ (d_0, d_1, \ldots, d_K) \in \{0, 1, \ldots, s-1\}^{K+1} : \sum_{k=0}^{K} d_k = N-K \right\}. \tag{11}$$

This set enumerates all possible configurations of gaps such that the maximum gap is always smaller than s, subject to the additional sum constraint (8). This counting problem can be solved by a generating function approach: consider the polynomial

$$(1 + x + x^2 + \ldots + x^{s-1})^{K+1}. \tag{12}$$

It is easy to verify that the cardinality of the set is exactly equal to the $(N-K)$th coefficient of the above polynomial. ∎

Remark 4: Using the equivalence between polynomial multiplication and convolution, the numerator in (10) can be efficiently computed by two applications of the discrete Fourier transform (DFT): one forward DFT to compute the Fourier transform of $f_s(x)$; and an inverse DFT to obtain the $(N-K)$th coefficient of $f_s(x)^{K+1}$.

Remark 5: The generating-function formula given in (10) also has an interesting probabilistic interpretation, which is what initially led us to considering (12). To count the number of elements in the set (11), we first construct $K+1$ independent random variables $d_0, d_1, \ldots, d_K$, each of which is uniformly distributed on $\{0, 1, \ldots, s-1\}$. (Note that this distribution is introduced for counting purposes; it is different from the actual distribution of the gaps $\{d_k\}$, which have to satisfy (8) and are thus *not* independent of each other.) Since all configurations are equally likely, the cardinality of (11) is equal to $s^{K+1}\mathbf{P}(\sum_k d_k = N-K)$. The probability generating function of d_k (for any k) is

$$g(x) = \sum_{i=0}^{s-1} \mathbf{P}(d_k = i) x^i = f_s(x)/s.$$

Since the random variables are i.i.d., the probability generating function of their sum $\sum_k d_k$ is then the $(K+1)$th power of $g(x)$.

Next, we consider the probability distribution of $\Gamma_{N,K}$, the maximum gap in the circulant setting. As shown in the following corollary, the distribution of $\Gamma_{N,K}$ can be obtained from that of $\Delta_{N,K}$.

Corollary 1 (The circular case): For any positive integer s,

$$\mathbf{P}(\Gamma_{N,K} < s) = \mathbf{P}(\Delta_{N-1,K-1} < s) = h_{N-1,K-1}(s).$$

Proof: The proof is based on a simple conditioning argument. Due to the circulant structure of the problem, we have

$$\mathbf{P}(\Gamma_{N,K} < s \mid n_1 = n) = \mathbf{P}(\Delta_{N-1,K-1} < s) = h_{N-1,K-1}(s).$$

Since the above conditional probability does not depend on n,

$$\mathbf{P}(\Gamma_{N,K} < s) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{P}(\Gamma_{N,K} < s \mid n_1 = n)$$
$$= h_{N-1,K-1}(s),$$

and this completes the proof. ∎

Remark 6: The deterministic bound given in (3) can be viewed as a simple consequence of our probabilistic analysis. By definition, $h_{N-1,K-1}(s)$ is the $(N-K)$th coefficient of $f_s(x)^K$, a polynomial of degree $(s-1)K$. Now set $s = 2P$. We can see that (3) holds if and only if $N-K$ is larger than the polynomial degree $(2P-1)K$, in which case $h_{N-1,K-1}(2P) = \mathbf{P}(\Gamma_{N,K} < 2P) = 0$. Consequently, there must exist at least one gap of size greater than or equal to $2P$, and this is all what ProSparse needs to recover $\boldsymbol{x}$ successfully.

B. Success Probabilities of ProSparse

In what follows, we express the success probabilities of ProSparse in terms of the function $h_{N,K}(s)$ given in (10). We consider three different settings as examples.

Example 1: $\boldsymbol{\Psi}$ is an $N \times M$ Vandermonde matrix (for which circularity cannot be exploited) and $\boldsymbol{\Phi}$ is the identity matrix. Let $\boldsymbol{x} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T]^T$, where $\boldsymbol{x}_1 \in \mathbb{C}^M$ is P-sparse and $\boldsymbol{x}_2 \in C^N$ is K sparse. Moreover, the sparsity pattern of $\boldsymbol{x}_2$ is drawn uniformly at random using sampling without replacement. In this case, ProSparse can recover $\boldsymbol{x}$ if and only if there exists at least one gap of size greater than or equal to $2P$. We thus have the following result:

$$\mathbf{P}(\text{ProSparse recovers } \boldsymbol{x}) = \mathbf{P}(\Delta_{N,K} \geq 2P) = 1 - h_{N,K}(2P).$$

Example 2: We assume that $\boldsymbol{\Psi}$ is the same as in the previous example, but $\boldsymbol{\Phi}$ is a banded matrix with bandwidth $b_{\boldsymbol{\Phi}}$. Moreover, the diagonal elements of $\boldsymbol{\Phi}$ are nonzero. We see that $\Delta_{N,K} \geq 2P$ is a necessary condition for ProSparse to recover $\boldsymbol{x}$, and $\Delta_{N,K} \geq 2P + 2b_{\boldsymbol{\Phi}}$ is a sufficient condition. It follows that

$$1 - h_{N,K}(2P + 2b_{\boldsymbol{\Phi}}) \leq \mathbf{P}(\text{ProSparse recovers } \boldsymbol{x}) \leq 1 - h_{N,K}(2P). \tag{13}$$

Example 3: Finally, we consider the case where $\boldsymbol{\Psi} = \boldsymbol{F}_N$ is the $N \times N$ unitary Fourier matrix and $\boldsymbol{\Phi} = \boldsymbol{I}_N$. The problem becomes circulant since the elements in the vector $\boldsymbol{y}$ have an
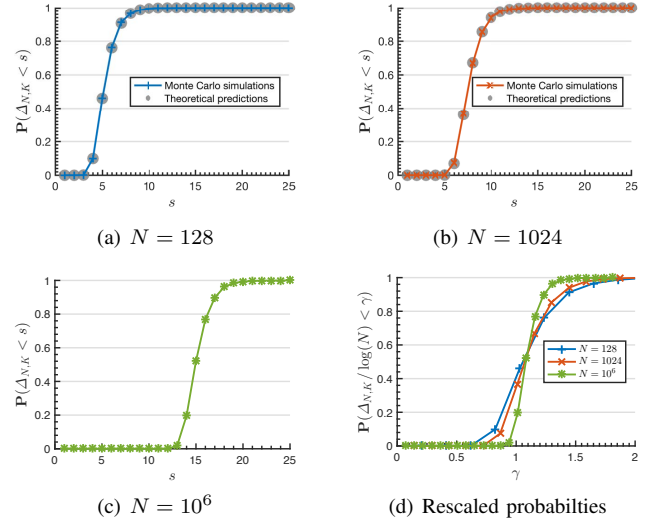

Fig. 3. Illustrations of the probability $\mathbf{P}(\Delta_{N,K} \leq s)$ for different values of N, with $K = \lfloor \alpha N \rfloor$ for $\alpha = 0.6$. The exact probabilities shown in (a) and (b) are computed using the expression in (10), whereas the estimated probabilities from Monte Carlo simulations are obtained by averaging over 5000 independent trials.

underlying periodicity of N samples. Moreover, the problem can also be solved in a dual form

$$\overline{\boldsymbol{F}_N^* \boldsymbol{y}} = [\boldsymbol{I}_N, \boldsymbol{F}_N]\,\overline{\boldsymbol{x}},$$

where $\overline{(\cdot)}$ and $(\cdot)^*$ denote the complex conjugate and Hermitian operators, respectively. In the above dual representation, the original spikes become Fourier atoms, and vice versa. Let $\boldsymbol{x} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T]^T$ be a (P, K)-sparse signal. We assume that the support patterns of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are drawn independently by sampling without replacement. Exploiting the duality and using the result of Corollary 1, we have

$$\mathbf{P}(\text{ProSparse recovers } \boldsymbol{x}) = 1 - h_{N-1,K-1}(2P)\,h_{N-1,P-1}(2K).$$

IV. ASYMPTOTIC ANALYSIS AND PHASE TRANSITIONS

A. Asymptotic Analysis

Proposition 1 provides an analytical formula to compute the success probabilities of the ProSparse algorithm. In Figures 3(a) and 3(b), we verify the analytical results given in (10) by comparing them against Monte Carlo simulations, for $N = 128$ and $N = 1024$, respectively. For much larger values of N, evaluating the analytical formula in (10) becomes numerically intractable. Thus, we only present results from Monte Carlo simulations in Figure 3(c) for $N = 10^6$. In all three experiments, the number of spikes are set to be a constant fraction of N, *i.e.*, $K = \lfloor \alpha N \rfloor$ for $\alpha = 0.6$.

Figures 3(a)–3(c) show that the probability $\mathbf{P}(\Delta_{N,K} < s)$ grows rapidly from 0 to 1 within a narrow range of s, although the point where this transition takes place shifts to the right as N increases. We divide $\Delta_{N,K}$ by $\log(N)$ and overlay the three rescaled probability curves $\mathbf{P}(\Delta_{N,K}/\log(N) < \gamma)$ as functions of γ in Figure 3(d). We can see that, after this rescaling, the probability curves are aligned. Moreover, the transitions from regions of low probability to those of high probability become sharper as N increases. This suggests a

phase transition phenomenon that takes place in the asymptotic regime where $N \to \infty$. We confirm and characterize this observation in the following proposition.

Proposition 2: Let $\boldsymbol{y} = [\boldsymbol{\Psi}, \boldsymbol{\Phi}]\boldsymbol{x}$, where $\boldsymbol{\Psi}$ is an $N \times M$ Vandermonde matrix with $M \geq N$, $\boldsymbol{\Phi}$ is an $N \times N$ banded matrix with bandwidth $b_{\boldsymbol{\Phi}}$, and $\boldsymbol{x} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T]^T$ is a (P, K)-sparse signal. Furthermore, the support of $\boldsymbol{x}_1$ is arbitrary whereas the support of $\boldsymbol{x}_2$ is drawn uniformly at random by sampling without replacement. Suppose that the bandwidth $b_{\boldsymbol{\Phi}}$ is fixed, and let $K = \lfloor \alpha N^{\delta} \rfloor$ and $P = \lfloor \beta N^{1-\delta} \log N \rfloor$ for three constants $0 < \delta \leq 1$, $\alpha > 0$ and $\beta > 0$. (When $\delta = 1$, we also require $\alpha < 1$.) It holds that

$$\mathbf{P}(\text{ProSparse recovers } \boldsymbol{x}) \xrightarrow{N \to \infty} \begin{cases} 1, & \text{if } \beta < \beta_c(\alpha, \delta), \\ 0, & \text{if } \beta > \beta_c(\alpha, \delta), \end{cases}$$
(14)

where the critical threshold is given by

$$\beta_c(\alpha, \delta) = \begin{cases} \frac{\delta}{2\alpha}, & \text{if } 0 < \delta < 1 \\ \frac{-1}{2 \log(1-\alpha)}, & \text{if } \delta = 1. \end{cases}$$
(15)

Proof: We see from (13) that the desired probability is sandwiched between a lower bound and an upper bound. The proof then consists of two parts. The first part deals with the lower bound in (13). We show that, when β is below the phase transition threshold in (14), this lower bound converges to 1 as $N \to \infty$. The second part of the proof provides an upper bound for $1 - h_{N,K}(2P)$, which is shown to converge to 0 when β is above the same threshold. Combined, these two parts verify that a phase transition in the probability of success indeed takes place according to the prediction in (14).

The lower bound: Let s be a positive integer. We use Proposition 1 to write

$$h_{N,K}(s) = \frac{\text{coef}\left\{ f_s(x)^{K+1}, x^{N-K} \right\}}{\binom{N}{K}}.$$
(16)

Since all the coefficients of the polynomial $f_s(x)^{K+1} = (1 + x + \ldots + x^{s-1})^{K+1}$ are positive, it is clear that

$$\text{coef}\left\{ f_s(x)^{K+1}, x^{N-K} \right\} \leq \frac{f_s(x)^{K+1}}{x^{N-K}}$$

for all $x > 0$. Replacing $f_s(x)$ by its equivalent expression $(1 - x^s)/(1 - x)$ and setting $x = 1 - K/N$, we then get

$$\text{coef}\left\{ f_s(x)^{K+1}, x^{N-K} \right\} \leq \frac{\left(1 - (1 - K/N)^s \right)^{K+1}}{(1 - K/N)^{N-K} (K/N)^{K+1}}.$$
(17)

Using the following standard estimate on factorials

$$\sqrt{2\pi}\, n^{n+1/2} e^{-n} \leq n! \leq n^{n+1/2} e^{1-n},$$

which holds for all positive integers n, we can verify that

$$\binom{N}{K} \geq C K^{-1/2} \frac{N^N}{K^K (N-K)^{N-K}},$$
(18)

where C is a constant that does not depend on N or K. Substituting (17) and (18) into (16) leads to

$$h_{N,K}(s) \leq CN \left(1 - (1 - K/N)^s \right)^{K+1}.$$
(19)

Let β_1, β_2 be two arbitrary constants such that

$$0 < \beta_1 < \beta_2 < \beta_c(\alpha, \delta),$$
(20)

where $\beta_c(\alpha, \delta)$ is the critical threshold given in (15). We show next that, for $P = \lfloor \beta_1 N^{1-\delta} \log N \rfloor$ and $s = 2P + 2b_{\boldsymbol{\Phi}}$, the right-hand side of (19) goes to 0 as $N \to \infty$. To that end, we first note that $s < 2\beta_2 N^{1-\delta} \log N$ for all sufficiently large N. Substituting this bound into the right-hand side of (19) gives us

$$h_{N,K}(s) \leq CN (1 - (1 - K/N)^{2\beta_2 N^{1-\delta} \log N})^{K+1}$$
$$= CN \exp\left\{ (K+1) \log \left(1 - N^{2\beta_2 N^{1-\delta} \log(1-K/N)} \right) \right\}$$
$$\leq CN \exp\left\{ -\alpha N^{\delta + 2\beta_2 N^{1-\delta} \log(1-K/N)} \right\},$$
(21)

where (21) follows from the inequality $\log(1 + x) \leq x$ which holds for all $x > -1$ and from the fact that $K + 1 > \alpha N^{\delta}$. For any $0 < \delta \leq 1$, it is easy to verify that

$$\lim_{N \to \infty} (\delta + 2\beta_2 N^{1-\delta} \log(1 - K/N)) = \delta \left(1 - \beta_2 / \beta_c(\alpha, \delta) \right)$$
$$> 0,$$

where the inequality is due to (20). It then follows from (21) that $h_{K,N}(2P + 2b_{\boldsymbol{\Phi}}) \xrightarrow{N \to \infty} 0$. This establishes the first part of (14).

Upper bound: It remains to show the second part of (14). To do so, we consider the upper bound in (13). Let $s = 2P$, where $P = \lfloor \beta_3 N^{1-\delta} \log N \rfloor$ for some fixed constant

$$\beta_3 > \beta_c(\alpha, \delta).$$
(22)

A simple union bound yields

$$1 - h_{N,K}(s) = \mathbf{P}(\Delta_{N,K} \geq s)$$
$$= \mathbf{P}\left(\cup_{k=0}^{K} \{d_k \geq s\} \right)$$
$$\leq \sum_{k=0}^{K} \mathbf{P}(d_k \geq s),$$
(23)

where $\{d_k\}$ are the gaps defined in (7). Due to symmetry, the random variables $\{d_k\}$ are exchangeable, and thus they have the same marginal distributions. It follows that

$$\sum_{k=0}^{K} \mathbf{P}(d_k \geq s) = (K+1) \mathbf{P}(d_0 \geq s).$$
(24)

For any integer $s \in [0, N - K]$, the probability $\mathbf{P}(d_0 \geq s)$ is given by $\binom{N-s}{K} / \binom{N}{K}$, since the event of having the first gap of size at least s is equivalent to having the K breaking points all appearing in the last $N - s$ locations. Substituting this expression into (24), we get

$$\sum_{k=0}^{K} \mathbf{P}(d_k \geq s) = (K+1) \frac{(N-s)!\,(N-K)!}{N!\,(N-K-s)!}$$
$$= (K+1) \frac{(N-K)\ldots(N-K-s+1)}{N(N-1)\ldots(N-s+1)},$$
(25)

for $s \leq N - K$. Next, we consider the two cases $0 < \delta < 1$ and $\delta = 1$ separately.

Case 1: $0 < \delta < 1$. From (25),

$$\sum_{k=0}^{K} \mathbf{P}(d_k \geq s) = (K+1)\frac{(N-K)^s}{N^s} \prod_{i=0}^{s-1} \frac{1 - i/(N-K)}{1 - i/N}$$

$$\leq (K+1)(1 - K/N)^s \qquad (26)$$

$$\leq \frac{K+1}{N^\delta} N^\delta \exp\{-sK/N\}, \qquad (27)$$

where (26) and (27) follow from the inequalities $i/(N-K) \geq i/N$ and $\log(1+x) \leq x$ for all $x > 1$, respectively. Substituting $s = 2\lfloor \beta_3 N^{1-\delta} \log N \rfloor$ and $K = \lfloor \alpha N^\delta \rfloor$ into (27) gives us

$$\sum_{k=0}^{K} \mathbf{P}(d_k \geq s) \leq (\alpha + 1)N^{\delta - 2\alpha\beta_3}, \qquad (28)$$

for all sufficiently large N. Since $\beta_c(\alpha, \delta) = \delta/(2\alpha)$ when $\delta < 1$, the condition (22) implies that the right-hand side of (28) tends to zero as $N \to \infty$. The second part of (14) for the case of $0 < \delta < 1$ then follow from (23) and (13) immediately.

Case 2: $\delta = 1$. We first rewrite (25) as

$$\sum_{k=0}^{K} \mathbf{P}(d_k \geq s) = (K+1)\prod_{i=0}^{s-1} \frac{1 - \frac{K}{N} - \frac{i}{N}}{1 - \frac{i}{N}}.$$

To bound the above expression, we introduce a function $r(x) \stackrel{\text{def}}{=} \log \frac{\eta - x}{1 - x}$, where η is a constant satisfying

$$1 - \alpha < \eta < e^{-1/(2\beta_3)}. \qquad (29)$$

Note that, due to (22), the above interval is nonempty and we can always find a suitable η. Since $K/N \to \alpha$, it holds for all sufficiently large values of N that

$$\sum_{k=0}^{K} \mathbf{P}(d_k \geq s) \leq (K+1) \exp\left\{\sum_{i=0}^{s-1} r(i/N)\right\}$$

$$\leq (K+1) \exp\left\{N \int_0^{\frac{s-1}{N}} r(x)dx\right\}, \qquad (30)$$

where the second inequality is due to the fact that the function $r(x)$ is negative and monotonically decreasing on the interval $x \in [0, \eta)$.

Let $g(x) \stackrel{\text{def}}{=} \int_0^x r(y)\, dy$. It is easy to verify that $g(x)$ is a smooth function in a neighborhood around $x = 0$. Moreover, $g'(0) = \log \eta$ and $g''(0) = 1 - \frac{1}{\eta}$. Recall that $s = 2P = 2\lfloor \beta_3 \log N \rfloor$ and thus $(s-1)/N = o(1)$. A Taylor expansion at $x = 0$ leads to

$$g\left(\frac{s-1}{N}\right) = (\log \eta)\frac{s-1}{N} + \frac{\eta - 1}{2\eta}\frac{(s-1)^2}{N^2}(1 + o(1)). \qquad (31)$$

Inserting (31) into (30) and setting $s = 2\lfloor \beta_3 \log N \rfloor$, we get

$$\sum_{k=0}^{K} \mathbf{P}(d_k \geq s) \leq \frac{K+1}{\eta} \exp\{2\beta_3 \log \eta \log N\}(1 + o(1))$$

$$= \frac{\alpha}{\eta} N^{1 + 2\beta_3 \log \eta}(1 + o(1)).$$

By the second inequality in (29), the exponent $1 + 2\beta_3 \log \eta < 0$. Thus, the above upper bound tends to 0 as $N \to \infty$. Now applying (23) and (13), we have shown the second part of (14) for the case of $\delta = 1$. ∎

B. Comparison with Basis Pursuit

A popular way to solve the sparse representation problem discussed in this paper is to use basis pursuit (BP), namely, by solving:

$$\arg \min_{\widetilde{x}} \|\widetilde{x}\|_1 \quad \text{s.t.} \quad y = D\widetilde{x}. \qquad (32)$$

The performance of BP is well understood and, in particular, probabilistic recovery guarantees for BP have been derived in [10]–[12].

Let D be a dictionary with column vectors $\{d_i\}$. Its mutual coherence is defined as

$$\mu(D) \stackrel{\text{def}}{=} \max_{k \neq \ell} \frac{|d_k^H d_\ell|}{\|d_k\|\|d_\ell\|}.$$

The case when $D = [\Psi, \Phi]$ is the union of two orthonormal bases in $\mathbb{C}^N$ was considered in [10], [11]. Let $x \in \mathbb{C}^{2N}$ be a (P, K)-sparse signal, with an arbitrarily-chosen sparsity pattern of P nonzero coefficients corresponding to Ψ and a randomly-chosen sparsity pattern of K nonzero coefficients corresponding to Φ. It has been shown that BP can recover x from $y = Dx$ with high probability if

$$P + K < C\frac{\mu(D)^{-2}}{\log(2N)} \qquad (33)$$

for some constant C. Later, this result was extended in [12], [14] to more general cases where D is the union of two arbitrary sub-dictionaries. It was shown in [12, Theorem 6] that BP can recover x with high probability if, in addition to (33), the sparsity levels P and K satisfy some extra constraints given by the mutual coherences computed on the sub-dictionaries Ψ and Φ. This recovery guarantee is further improved in [14], where the authors exploit the individual coherence parameters of the sub-dictionaries as well as the mutual coherence between Ψ and Φ. (See [14, Theorem 5] for details.)

To compare the probabilistic ProSparse bound given in Proposition 2 with the above BP bounds, we consider the special case when the sub-dictionary $\Phi \in \mathbb{R}^{N \times N}$ is the identity matrix and the sub-dictionary $\Psi \in \mathbb{C}^{N \times M}$ (with $M \geq N$) is a Fourier frame with entries

$$\Psi_{n,m} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/M}.$$

When $M = N$, both sub-dictionaries are orthonormal bases, and the mutual coherence of the entire dictionary becomes $\mu(D) = 1/\sqrt{N}$. It follows that the BP bounds in (33) becomes

$$P + K < C\frac{N}{\log N}. \qquad (34)$$

In comparison, the result of Proposition 2 shows that ProSparse can recover x with high probability if

$$K \leq \lfloor \alpha N^\delta \rfloor \quad \text{and} \quad P < \beta_c(\alpha, \delta)N^{1-\delta} \log N \qquad (35)$$

for $\delta \in (0, 1]$ and $\alpha \in (0, 1)$. ProSparse requires an unbalanced distribution of atoms between the local basis Φ and the Fourier basis Ψ. Moreover, the requirement that P is of order $N^{1-\delta} \log N$ is much worse than what the BP bounds in (34) allows.

The situation becomes different when $\mathbf{\Psi}$ is a *redundant* Fourier frame, *i.e.*, when $M > N$. It is easy to show that the mutual coherence parameters in this case satisfy

$$\mu(\mathbf{D}) \geq \mu(\mathbf{\Psi}) \geq \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{j2\pi n/M} \right| = \frac{1}{N} \left| \frac{1 - e^{j2\pi N/M}}{1 - e^{j2\pi/M}} \right|, \quad (36)$$

where $\mu(\mathbf{\Psi})$ is the mutual coherence parameter of the sub-dictionary $\mathbf{\Psi}$. Let $d = M/N$ be the redundancy factor. Fixing $d > 1$ and letting $N \to \infty$, the right-hand side of (36) converges to a constant

$$\frac{d\sqrt{1 - \cos(2\pi/d)}}{\sqrt{2}\pi}$$

that does not depend on N. This implies that $\mu(\mathbf{\Psi}) = \mathcal{O}(1)$. The BP recovery guarantee given in [14, Theorem 5] requires, among other conditions, that $P\mu^2(\mathbf{\Psi}) < C$ for some constant C, and thus, P needs to be of $\mathcal{O}(1)$. In contrast, since the ProSparse bound does not depend on the mutual coherence, the same bound in (35) still holds even when $\mathbf{\Psi}$ is very redundant. This suggests that ProSparse might have an advantage over BP when dealing with redundant sub-dictionaries.

We need to highlight at this point that, although the ProSparse bound is asymptotically tight, the same cannot be claimed for the existing BP bounds. They are merely sufficient conditions for BP to succeed. The actual performance of BP can be better than what the bounds predict. Nevertheless, we observe from numerical simulations that ProSparse can indeed outperform BP when $\mathbf{\Psi}$ is redundant. To illustrate this point, we show in Figure 4 the comparison of ProSparse and BP for sparse recovery from a dictionary consisting of a Fourier frame and a canonical basis. The Fourier frame has size $N \times M$ with $N = 128$ and $M = dN$, where d is the redundancy factor. We consider three settings: $d = 4, 8$ and 16, corresponding to increasingly redundant frames. Empirical success probabilities of both algorithms are computed by generating 50 different realizations of the sparse vector $\boldsymbol{x}$ for each sparsity combination (P, K). The locations of the nonzero entries of $\boldsymbol{x}$ are drawn uniformly at random from the sets $\{1, 2, ..., M\}$ and $\{1, 2, ..., N\}$. The amplitudes of these elements are complex, with the real and imaginary parts drawn from $\mathcal{N}(0, 1)$.

The BP results in Figure 4(a)–Figure 4(c) are obtained by solving the ℓ_1-minimization problem in (32) with CVX, a package for specifying and solving convex programs [30], [31]. The success of the algorithm is measured by computing $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2 / \|\boldsymbol{x}\|^2$ and by checking that it is below 10^{-3}. Here $\tilde{\boldsymbol{x}}$ is the reconstructed sparse vector. The ProSparse results in Figure 4(d) are obtained by checking that the maximum gap between consecutive spikes is at least equal to $2P$. When this is the case, the Fourier atoms can be recovered and therefore perfect reconstruction of the vector $\boldsymbol{x}$ is achieved. Since the performance of ProSparse does not change with the redundancy factor d of the Fourier frame, only one figure is shown for ProSparse.

As expected, both algorithms show a phase transition behavior since the (K, P)-plane is clearly split in two regions, one where the algorithms achieve perfect reconstruction with
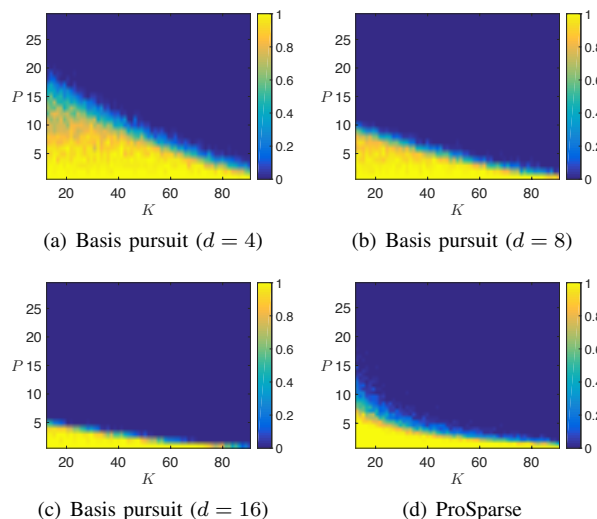

Fig. 4. Comparison between ProSparse and basis pursuit for sparse recovery from a dictionary consisting of a Fourier frame and a canonical basis. The Fourier frame is of size $N \times M$, with $N = 128$. We test three redundancy factors: $d = M/N$ is set to $4, 8$ and 16, respectively. Shown in the figures are the empirical success probabilities of the two algorithms. Since the performance of ProSparse does not change with the redundancy factor d, only one figure is shown for ProSparse.

very high probability, and the other where the algorithms fail in most of the cases. BP outperforms ProSparse when $d = 4$. However, as we increase the redundancy ratio, which leads to increased mutual coherence between the dictionary elements, the performance of BP degrades. In contrast, the performance of ProSparse is not affected by the increasing redundancy. At $d = 16$, ProSparse outperforms BP by recovering sparse vectors in a larger region of the (K, P)-plane. This suggests that, in the case where $\mathbf{\Phi}$ is a highly redundant frame, ProSparse can be the method of choice.

V. CONCLUSIONS

We presented an average-case performance analysis of ProSparse to recover a sparse signal from the union of a Vandermonde matrix and a banded matrix. The underlying probability model turns out to be equivalent to the problem of discrete circle covering. Based on a simple generating-function approach, we presented a new analytical expression for the exact success probabilities of ProSparse. We also analyzed the above-mentioned success probabilities in the high-dimensional regime. Our asymptotic analysis reveals a phase transition phenomenon regarding the performance of ProSparse. Unlike BP, the average-case performance of ProSparse does not depend on the mutual coherence of the underlying dictionary. This unique property makes ProSparse a potentially more attractive choice than BP when the sub-dictionary corresponding to the Vandermonde matrix is highly redundant.

REFERENCES

[1] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing, Special Issue on Wavelets and Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.

[2] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, 1993.

[3] 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.

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

[5] 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.

[6] M. Elad and M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[7] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[8] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1579–1581, 2003.

[9] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[10] E. J. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics*, vol. 6, pp. 227–254, Dec. 2005.

[11] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harmonic Analysis*, vol. 25, no. 1, pp. 1–24, Jul. 2008.

[12] P. Kuppinger, G. Durisi, and H. Bolcskei, "Uncertainty Relations and Sparse Signal Recovery for Pairs of General Signal Sets," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 263–277, Jan. 2012.

[13] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.

[14] G. Pope, A. Bracher, and C. Studer, "Probabilistic Recovery Guarantees for Sparsely Corrupted Signals," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3104–3116, May 2013.

[15] M. Elad, *Sparse and Redundant Representations*. Springer, 2010.

[16] P. L. Dragotti and Y. M. Lu, "On sparse representation in Fourier and local bases," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7888–7899, Dec. 2014.

[17] D. Donoho and P. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49/3, pp. 906–931, June 1989.

[18] L. Holst, "On discrete spacings and the Bose-Einstein distribution," *Contributions to Probability and Statistics. Essays in honour of Gunnar Blom. Ed. by Jan Lanke and Georg Lindgren, Lund*, pp. 169–177, 1985.

[19] G. Barlevy and H. N. Nagaraja, "Properties of the vacancy statistic in the discrete circle covering problem," in *Ordered Data Analysis, Modeling and Health Research Methods*. Springer, 2015, pp. 121–146.

[20] G. C. F. M. R. Prony, "Essai expérimental et analytique sur les lois de la dilabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l' eau et de la vapeur de l' alkool, à différentes températures," *J. de l' École Polytechnique*, vol. 1, pp. 24–76, 1795.

[21] P. Stoica and R. L. Moses, *Introduction to spectral analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[22] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Edition*, 3rd ed. Wiley, 1968.

[23] D. E. Barton and F. N. David, "Some Notes on Ordered Random Intervals," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 18, no. 1, pp. 79–94, 1956.

[24] L. Flatto and A. G. Konheim, "The Random Division of an Interval and the Random Covering of a Circle," *SIAM Rev.*, vol. 4, no. 3, pp. 211–222, 1962.

[25] R. Pyke, "Spacings," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 27, no. 3, pp. 395–449, 1965.

[26] L. Holst, "On the lengths of the pieces of a stick broken at random," *J. Appl. Probab.*, pp. 623–634, 1980.

[27] K. Drakakis, "On the maximal distance between consecutive choices in the set of winning numbers in Lottery," *Applied Mathematical Sciences*, vol. 3, no. 55, pp. 2725–2738, 2009.

[28] T. E. Huillet, "A Bose–Einstein approach to the random partitioning of an integer," *J. Stat. Mech. Theor. Exp.*, vol. 2011, no. 08, p. P08021, 2011.

[29] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2003.

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

[31] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, no. 95–110.

Yue M. Lu did his undergraduate studies at Shanghai Jiao Tong University, China. He then attended the University of Illinois at Urbana-Champaign, where he received the M.Sc. degree in mathematics and the Ph.D. degree in electrical engineering, both in 2007. Following his work as a postdoctoral researcher at the Audiovisual Communications Laboratory at Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, he joined Harvard University in 2010, where he is currently an Associate Professor of Electrical Engineering at the John A. Paulson School of Engineering and Applied Sciences.

He received the Most Innovative Paper Award of IEEE International Conference on Image Processing (ICIP) in 2006, the Best Student Paper Award of IEEE ICIP in 2007, and the Best Student Presentation Award at the 31st SIAM SEAS Conference in 2007. His current research interests include theoretical and algorithmic aspects of high-dimensional signal and information processing.

He is a Member of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee since 2015, and a Member of the IEEE Signal Processing Theory and Methods Technical Committee since 2016. He received the 2015 ECE Illinois Young Alumni Achievement Award.

Jon Oñativia completed a double degree program and received the Ingeniero de Telecomunicación degree and Diplôme d'Ingénieur from the Universidad Politecnica de Madrid, Spain, and Télécom ParisTech, France, respectively in 2008. He received the M.Sc. degree (distinction) in Communications and Signal Processing and the Ph.D. degree from Imperial College London, UK, in 2011 and 2015, respectively.

During his studies he did two internships, one at Vicomtech, San Sebastian, Spain, where he worked in augmented reality applications for digital televsion and another one at General Electric Healthcare, Buc, France, where he collaborated in the development of a 3D digital mammography software. He was a software developer from 2008 to 2010 at NDS (now part of Cisco Systems), Issy les Moulineaux, France, where he participated in a variety of projects related to digital television. He is now a research engineer at Egile, Mendaro, Spain, working in the development of a surgical robot for spine surgery.

Pier Luigi Dragotti received the Laurea (summa cum laude) degree in electrical and electronic engineering from the University of Naples Federico II, Naples, Italy, in 1997, and the master's degree in communications systems and the Ph.D. from the Swiss Federal Institute of Technology of Lausanne, Lausanne, Switzerland, in 1998 and 2002, respectively. He is currently Professor of Signal Processing with the Electrical and Electronic Engineering Department at Imperial College London. He has held several visiting positions at different universities and research centers. He was a Visiting Student with Stanford University, Stanford, CA, USA, in 1996, a Researcher with the Mathematics of Communications Department, Bell Labs, Lucent Technologies, Murray Hill, NJ, USA, in 2000, and a Visiting Scientist with the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2011. His research interests include sampling theory, wavelet theory and its applications, image and video compression and image super-resolution. He was Technical Co-Chair of the European Signal Processing Conference in 2012, an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2006 to 2009 and an Elected Member of the IEEE Image, Video and Multidimensional Signal Processing Technical Committee (2008-2013). He is currently a member of the IEEE Signal Processing Theory and Methods Technical Committee and a recipient of the ERC Starting Investigator Award for the project RecoSamp.