On Sparse Representation in Fourier and Local Bases

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

Abstract—We consider the classical problem of finding the sparse representation of a signal in a pair of bases. When both bases are orthogonal, it is known that the sparse representation is unique when the sparsity K of the signal satisfies $K < 1/\mu(D)$, where $\mu(D)$ is the mutual coherence of the dictionary. Furthermore, the sparse representation can be obtained in polynomial time by basis pursuit (BP), when $K < 0.91/\mu(D)$. Therefore, there is a gap between the unicity condition and the one required to use the polynomial-complexity BP formulation. For the case of general dictionaries, it is also well known that finding the sparse representation under the only constraint of unicity is NP-hard. In this paper, we introduce, for the case of Fourier and canonical bases, a polynomial complexity algorithm that finds all the possible K-sparse representations of a signal under the weaker condition that $K < \sqrt{2}/\mu(D)$. Consequently, when $K < 1/\mu(D)$, the proposed algorithm solves the unique sparse representation problem for this structured dictionary in polynomial time. We further show that the same method can be extended to many other pairs of bases, one of which must have local atoms. Examples include the union of Fourier and local Fourier bases, the union of discrete cosine transform and canonical bases, and the union of random Gaussian and canonical bases.

Index Terms—Sparse representation, union of bases, Prony's method, harmonic retrieval, basis pursuit, mutual coherence.

I. INTRODUCTION

CONSIDER the problem of finding the sparse representation of a signal in the union of two orthogonal bases. Specifically, let y be an N-dimensional vector given by the linear combination of K atoms of the dictionary $D = [\Psi, \Phi]$, where Ψ and Φ are two $N \times N$ orthogonal matrices. Given the synthesis model

$$y = Dx, \qquad (1)$$

we study the problem of finding the K nonzero entries of x from y.

One way to retrieve the sparse vector x is to solve the following problem:

$$(P_0) : \arg \min_{\widetilde{x}} \|\widetilde{x}\|_0 \quad \text{s.t.} \quad y = D\widetilde{x},$$

where the ℓ_0 "norm" is given by $\|\widetilde{x}\|_0 = \#\{i : |\widetilde{x}_i| \neq 0\}$. The (P_0) problem is clearly daunting since the ℓ_0 norm is nonconvex. Therefore it might be convenient to consider the following convex relaxation:

$$(P_1) : \arg \min_{\widetilde{x}} \|\widetilde{x}\|_1 \quad \text{s.t.} \quad y = D\widetilde{x},$$

where $\|\widetilde{x}\|_1 = \sum_{i=1}^{M} |\widetilde{x}_i|$ is the ℓ_1 norm. We note that (P_1), also known as Basis-Pursuit (BP) [1], can be solved using polynomial complexity algorithms.

The sparse representation problem was first posed in the above forms by Donoho and Huo in [2] for the union of Fourier and canonical bases. Specifically, let $\mu(D)$ denote the mutual coherence of D, defined as

$$\mu(D) = \max_{1 \le k, \ell \le 2N, k \neq \ell} \frac{|d_k^* d_\ell|}{\|d_k\|_2 \|d_\ell\|_2},$$

where d_k is the kth column of D and $(\cdot)^*$ denotes the conjugate transpose of a vector. They first showed that the original K-sparse vector x is the unique solution of (P_0) when

$$K < \frac{1}{\mu(D)} = \sqrt{N}, \qquad (2)$$

where we have used the fact that for the case of Fourier and canonical bases $\mu(D) = 1/\sqrt{N}$. They then went on showing that (P_0) and (P_1) are equivalent when

$$K < \frac{\sqrt{N}}{2}. \qquad (3)$$

This fact has important implications since it indicates that under the constraint (3), the sparse representation problem has a unique solution and, more importantly, it can be solved exactly using algorithms with polynomial complexity.

The findings of Donoho and Huo were extended to generic orthogonal pairs of bases by Elad and Bruckstein in [3], where the bound in (3) was also improved. Specifically, (P_0) has a unique solution, which is also equal to x, when

$$K < \frac{1}{\mu(D)}. \qquad (4)$$

Moreover, if the signal y is made of K_p atoms of Ψ and K_q atoms of Φ, with $K = K_p + K_q$, then it was shown in [3] that (P_1) is equivalent to (P_0) when

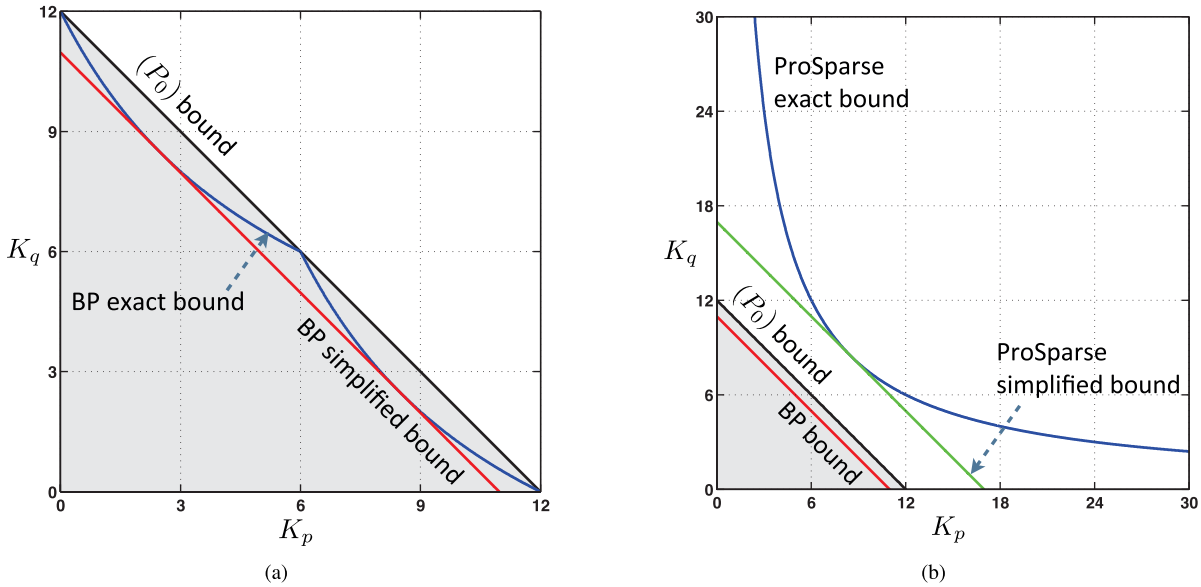$$2\mu(D)^2 K_p K_q + \mu(D) \max\{K_p, K_q\} - 1 < 0. \qquad (5)$$

Fig. 1. Comparing different bounds for sparse signal representation in a union of orthogonal bases. (a) Uniqueness of (P_0) and the two ℓ_1 bounds for a dictionary with $\mu(\boldsymbol{D}) = 1/12$. (b) The two ProSparse bounds (7) and (8) plotted against the (P_0) bound and the BP simplified bound, for the specific case when $\boldsymbol{D}$ is the union of Fourier and canonical bases. We set $N = 144$, so $\mu(\boldsymbol{D}) = 1/12$.

This bound is tight as demonstrated in [4] (see, also, Appendix A), but it is a bit obscure. For this reason, a simpler but slightly more restrictive version is usually adopted:

$$K = K_p + K_q < \frac{\sqrt{2} - 0.5}{\mu(\boldsymbol{D})}. \tag{6}$$

Figure 1(a) presents a comparison between the (P_0) bound (4), the tight (P_1) bound (5) and its simplified version (6). We immediately note that (4) poses a weaker condition than (5) or (6), as there is still a (small) gap between the (P_0) and (P_1) bounds. While we know that (P_1) can be solved with polynomial complexity algorithms, we cannot conclude from existing results whether (P_0) has the same complexity, unless the sparsity level is further reduced to satisfy (5). For arbitrary redundant dictionaries, it is well known that (P_0) is NP-hard [5], [6]. However, this general result does not address the case of structured dictionaries which we will be considering in this work. Moreover, another open issue is whether we can still reconstruct the vector $\boldsymbol{x}$ when its sparsity level K is beyond the (P_0) bound (4).

The main contribution of this paper is to show that, when $\boldsymbol{D}$ is the union of Fourier and canonical bases, there exists a polynomial complexity algorithm that can recover $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x}$, provided that

$$K_p K_q < N/2. \tag{7}$$

The proposed algorithm is based around Prony's method which is commonly used in spectral estimation theory [7]. For this reason we name it *ProSparse*—Prony's based sparsity—in honour of Baron de Prony who first invented the method that goes under his name.

Using the inequality $2\sqrt{K_p K_q} \leq K_p + K_q$, we see that a more restrictive version of (7) is to require

$$K = K_p + K_q < \sqrt{2N}, \tag{8}$$

which imposes a simple constraint on the total sparsity of $\boldsymbol{x}$. In Figure 1(b), we compare the ProSparse bounds (7) and (8) against the (P_0) and (P_1) bounds. To compute the latter two, we use the fact that $\mu(\boldsymbol{D}) = 1/\sqrt{N}$ for the case of Fourier and canonical bases. Consequently, the (P_0) problem has a unique solution when the constraint (2) is met and (P_1) and (P_0) are equivalent when

$$K < (\sqrt{2} - 0.5)\sqrt{N}. \tag{9}$$

We see from the figure that the ProSparse bounds are much weaker, meaning that the proposed algorithm can recover a larger class of sparse signals. In particular, since the uniqueness bound for (P_0) falls entirely within the ProSparse bounds, our results imply that, for the union of Fourier and canonical bases, the nonconvex problem (P_0) can be solved with polynomial complexity under the uniqueness bound $K < \sqrt{N}$. To our knowledge, no other polynomial complexity algorithm has been known in the literature to achieve this task.

We conclude by noting that recently a generalized version of the uncertainty principle of Elad-Bruckstein was presented in [8] leading to more general uniqueness bounds. Those bounds converge to (2) for the case of Fourier and canonical bases. We also note that, while finding the sparse representation of a signal is an interesting theoretical problem, modeling signals as sparse in a certain domain has proved very useful also in many signal processing applications and we refer to the paper [9] and the book [10] for a comprehensive review of both theoretical as well as applied aspects of this topic. Finally, we also refer to the recent paper [11] where the importance of developing methods producing a list of K-sparse representations is highlighted.

The rest of the paper is organized as follows: After a brief overview of Prony's method in Section II, we present in Section III the main results of this work: There we introduce

ProSparse and show that it solves the sparse representation problem under the bound given in (7), when D is the union of the Fourier and canonical bases. These results are then generalized in Section IV, where we show that ProSparse works for many other pairs of bases. In general, it is only required that one of the bases have local atoms and the other basis be such that it allows for the efficient reconstruction of sparse signals from any small blocks of consecutive elements in the transform domain (see Proposition 3 for details). Examples of such pairs include the union of Fourier and local Fourier bases, the union of discrete cosine transform (DCT) and canonical bases, and the union of random Gaussian and canonical bases. We conclude in Section V. Unless stated otherwise, we assume throughout the paper that the basis matrices and the signals are all complex-valued, *i.e.*, $\mathbf{\Psi}, \mathbf{\Phi} \in \mathbb{C}^{N \times N}$, $\mathbf{y} \in \mathbb{C}^N$, and $\mathbf{x} \in \mathbb{C}^{2N}$.

II. OVERVIEW OF PRONY'S METHOD

Consider the case when the signal $\mathbf{y}$ is made only of K Fourier atoms, *i.e.*, $\mathbf{y} = \mathbf{F}\mathbf{c}$, where $\mathbf{F}$ is the N-point DFT matrix and $\mathbf{c}$ is some K-sparse vector in $\mathbb{C}^N$. The algebraic structure of the Fourier matrix makes it possible to reconstruct the sparse vector $\mathbf{c}$ from only $2K$ *consecutive* entries of $\mathbf{y}$.

One classical algorithm for such reconstruction is a method by Baron de Prony, developed in 1795 for the original purpose of estimating the frequency, phase, and amplitude parameters of a finite sum of sinusoids [12]. In the last several decades, Prony's method has been rediscovered and extended many times in different fields: it has been used in error correcting codes (*e.g.*, Berlekamp-Massey algorithm [13], [14]), in array signal processing [7], to solve some inverse problems [15]–[17], and more recently, in parametric sampling theory [18], [19].

In what follows, we present a simple derivation of the basic Prony's method, with emphasis on key results that will be used in later sections. We refer readers to the book [7] and to the insightful overview [17] for more details on this intriguing nonlinear estimation algorithm and its various extensions (*e.g.*, noisy measurements and multidimensional signals).

To start, we observe that $\mathbf{y}$ is the sum of K exponentials: its nth entry is of the form

$$y_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} c_{m_k} e^{j2\pi m_k n / N}, \qquad (10)$$

where m_k is the index[1] of the kth nonzero element of $\mathbf{c}$, and c_{m_k} is the corresponding weight. Writing $\alpha_k \stackrel{\text{def}}{=} c_{m_k}/\sqrt{N}$ and $u_k \stackrel{\text{def}}{=} e^{j2\pi m_k/N}$, we can simplify (10) as

$$y_n = \sum_{k=0}^{K-1} \alpha_k u_k^n. \qquad (11)$$

Assuming that K is known, we aim to retrieve the coefficients $\{\alpha_k\}$ and the exponentials $\{u_k\}$ from $2K$ consecutive elements $\{y_n : \ell \le n < \ell + 2K\}$. The original K-sparse vector $\mathbf{c}$ can then be reconstructed from $\{\alpha_k\}$ and $\{u_k\}$.

[1] In this paper we use a zero-based indexing scheme. So the first element of $\mathbf{c}$ is assigned the index 0.

The key to Prony's method is a clever use of the algebraic structure of the expression in (11). Let

$$\begin{aligned} P(x) &= \prod_{k=1}^{K} (x - u_k) \\ &= x^K + h_1 x^{K-1} + h_2 x^{K-2} + \cdots + h_{K-1}x + h_K \end{aligned} \qquad (12)$$

be a Kth order polynomial whose roots are $\{u_k\}$. Then, it is easy to verify that

$$\begin{aligned} y_{n+K} &+ h_1\, y_{n+K-1} + h_2\, y_{n+K-2} + \cdots + h_K\, y_n \\ &= \sum_{1 \le k \le K} \alpha_k u_k^n P(u_k) = 0. \end{aligned}$$

Writing this identity in matrix-vector form for all indices n such that $\ell \le n < \ell + K$, we get

$$\mathbf{0} = \begin{bmatrix} y_{\ell+K} & y_{\ell+K-1} & \cdots & y_{\ell} \\ y_{\ell+K+1} & y_{\ell+K} & \cdots & y_{\ell+1} \\ \vdots & \ddots & \ddots & \vdots \\ y_{\ell+2K-2} & \ddots & \ddots & \vdots \\ y_{\ell+2K-1} & y_{\ell+2K-2} & \cdots & y_{\ell+K-1} \end{bmatrix} \begin{bmatrix} 1 \\ h_1 \\ h_2 \\ \vdots \\ h_K \end{bmatrix} \stackrel{\text{def}}{=} \mathbf{T}_{K,\ell}\mathbf{h}, \quad (13)$$

where, by construction, $\mathbf{T}_{K,\ell}$ is a Toeplitz matrix of size $K \times (K+1)$.

The above equation reveals that the vector of polynomial coefficients $\mathbf{h} = [1, h_1, \ldots, h_K]^T$ is in the null space of $\mathbf{T}_{K,\ell}$. In fact, this condition is sufficient to uniquely identify $\mathbf{h}$, as guaranteed by the following proposition.

Proposition 1: Suppose that $\alpha_k \ne 0$ for all k and that the K parameters $\{u_k\}$ are distinct. Then

$$\text{rank } \mathbf{T}_{K,\ell} = K. \qquad (14)$$

Proof: See, [20, Appendix B]. ∎

Since $\mathbf{T}_{K,\ell}$ has full row rank, its null space is of dimension one. We can therefore conclude that the vector $\mathbf{h}$ is the unique vector satisfying the identity (13).

In light of the above derivations, we summarize Prony's method as follows:

(1) Given the input y_n, build the Toeplitz matrix $\mathbf{T}_{K,\ell}$ as in (13) and solve for $\mathbf{h}$. This can be achieved by taking the SVD of $\mathbf{T}_{K,\ell}$ and choosing as $\mathbf{h}$ the (scaled) right-singular vector associated with the zero singular value. The scaling is done so that the first element of $\mathbf{h}$ is equal to 1.

(2) Find the roots of $P(x) = 1 + \sum_{n=1}^{K} h_k x^{K-k}$. These roots are exactly the exponentials $\{u_k\}_{k=0}^{K-1}$.

(3) Given the parameters $\{u_k\}_{k=0}^{K-1}$, find the corresponding weights $\{\alpha_k\}_{k=0}^{K-1}$ by solving K linear equations as given in (11). This is a Vandermonde system of equations which yields a unique solution for the weights $\{\alpha_k\}_{k=0}^{K-1}$ since $\{u_k\}_{k=0}^{K-1}$ are distinct.

Remark 1: Building the Toeplitz matrix $\mathbf{T}_{K,\ell}$ in (13) requires $2K$ elements $\{y_n : \ell \le n < \ell + 2K\}$. Therefore, Prony's method allows us to reconstruct $\{\alpha_k, u_k\}$ and, equivalently, the K-sparse vector $\mathbf{c}$ from *any* $2K$ *consecutive* elements of $\mathbf{y}$. Moreover, due to the periodicity of the Fourier matrix, these elements of $\mathbf{y}$ just need to have

indices that are consecutive modulo (N). For example, $\{y_{N-2}, y_{N-1}, y_0, \ldots, y_{2K-3}\}$ is also a valid choice.

Remark 2: We have assumed in the above discussions that the sparsity level K is known. In fact, to apply Prony's method, we just need to know an upper bound on the sparsity level. To see this, assume that the true sparsity of c is $\widetilde{K}$, for some unknown $\widetilde{K} < K$. Following the same steps in the proof of Proposition 1, we can show that the Toeplitz matrix $T_{K,\ell}$ in this case is rank-deficient and that its rank is equal to $\widetilde{K}$. Therefore, checking the rank of $T_{K,\ell}$ allows us to obtain the true sparsity level.

III. FINDING SPARSE REPRESENTATIONS IN FOURIER AND CANONICAL BASES

A. ProSparse: A Polynomial Complexity Algorithm

We now return to our original problem of finding the sparse representation of a signal in a pair of bases. The observed signal is $y = [F, I]x$, where F and I are two orthogonal matrices corresponding to the Fourier and canonical bases, respectively. We want to retrieve x from y, knowing that x has a small number of nonzero elements.

We begin by noting that the problem is trivial when y is made only of spikes, *i.e.*, when the first N entries of x are exactly zero. In this case, we can directly retrieve x by observing the support set of y. Likewise, by examining the support of the Fourier transform of y, we can trivially retrieve the sparse representation of y when it is made only of Fourier atoms. Let us assume now that y is made of a combination of K_p Fourier atoms and K_q spikes, for some $K_p, K_q \geq 1$. The total sparsity is then defined as $K = K_p + K_q$.

Our proposed algorithm on sparse representation is based on a simple idea: The observation y is a mixture of Fourier atoms and spikes, the latter of which are *local*. If we can find an interval of $2K_p$ consecutive entries of y that are only due to the Fourier atoms, we can then apply Prony's method presented in Section II on these entries to retrieve the K_p Fourier atoms. Once this has been achieved, the spikes can be obtained by removing from y the contribution due to the Fourier atoms. Moreover, when both K_p and K_q are small, such "nice" intervals should always exist and there might even be a large number of them.

To quantify the above intuition, denote by $0 \leq n_1 < n_2 < \cdots < n_{K_q} \leq N - 1$ the set of indices corresponding to the K_q spikes. We can count the number of all length-$2K_p$ intervals that are not "corrupted" by these spikes as

$$\mathcal{N}(n_1, n_2, \ldots, n_{K_q})$$
$$\stackrel{\text{def}}{=} \#\Big\{\ell : 0 \leq \ell < N \text{ and } \{\ell, \ell+1, \ldots, \ell+2K_p-1\}$$
$$\cap \{n_1, n_2, \ldots, n_{K_q}\} = \emptyset\Big\}. \quad (15)$$

Note that, due to the periodicity of the Fourier exponential $e^{j2\pi n/N}$, we should view indices through the modulo (by N) operator. This means that $N = 0 \pmod{N}$ and thus the entry $n = N - 1$ is immediately followed by the entry $n = 0$, and so on.

Lemma 1: Let y be a mixture of K_p Fourier atoms and K_q spikes, for some $K_p, K_q \geq 1$. Then, for any choice of spike locations $0 \leq n_1 < n_2 < \cdots < n_{K_q} \leq N - 1$,

$$\mathcal{N}(n_1, n_2, \ldots, n_{K_q}) \geq N - 2K_p K_q. \quad (16)$$

Proof: Let d_i, for $1 \leq i \leq K_q$, denote the number of consecutive entries of y that are "sandwiched" between two neighboring spikes n_i and n_{i+1}. (Here, n_{K_q+1} is defined to be equal to n_1.) Then $d_i = (n_{i+1} - n_i) \pmod{N} - 1$. Clearly, $d_i \geq 0$, and

$$\sum_{1 \leq i \leq K_q} d_i = N - K_q. \quad (17)$$

By construction, each of these K_q intervals are "uncorrupted" by the spikes. For the ith interval, if its length $d_i < 2K_p$, then that particular interval does not contain enough entries for building the Toeplitz matrix in (13) as required in Prony's method; if however, $d_i \geq 2K_p$, then we can find $d_i - 2K_p + 1$ (overlapping) subintervals, each of length $2K_p$. It follows that the quantity in (15) can be computed as

$$\mathcal{N}(n_1, n_2, \ldots, n_{K_q}) = \sum_{1 \leq i \leq K_q} \max\{0, d_i - 2K_p + 1\}$$
$$\geq \sum_{1 \leq i \leq K_q} (d_i - 2K_p + 1)$$
$$= \Big(\sum_{1 \leq i \leq K_q} d_i\Big) - K_q(2K_p - 1). \quad (18)$$

Substituting (17) into (18) leads to the bound (16). ∎

Remark 3: Lemma 1 implies that

$$K_p K_q < N/2 \quad (19)$$

is a sufficient condition for $\mathcal{N}(n_1, n_2, \ldots, n_{K_q}) \geq 1$, *i.e.*, for at least one interval of $2K_p$ consecutive entries containing only Fourier atoms to exist. The requirement in (19) is also essentially necessary: Suppose that K_q divides N. Let the K_q spikes be evenly spaced to form a "picket-fence" signal. In this case, we can have at most $N/K_q - 1$ consecutive entries of y that contain only Fourier atoms, before running into another spike. If (19) is not satisfied, *i.e.*, if $K_p K_q \geq N/2$, the length of such "clean" intervals will be strictly smaller than $2K_p$, and thus $\mathcal{N}(n_1, n_2, \ldots, n_{K_q}) = 0$.

Based on the above analysis, we are now able to state the following result:

Proposition 2: Assume $D = [F, I]$, where F and I are, respectively, the $N \times N$ Fourier and identity matrices. Let $y \in \mathbb{C}^N$ be an arbitrary signal. There exists an algorithm, with a worst-case complexity of $\mathcal{O}(N^3)$, that finds *all* (K_p, K_q)-sparse signals x such that

$$y = Dx \quad \text{and} \quad K_p K_q < N/2. \quad (20)$$

Proof: We provide a constructive proof of this proposition by introducing the ProSparse algorithm. We will show that ProSparse finds all x satisfying (20), with the additional constraint that $K_p \leq K_q$. The remaining cases, *i.e.*, those x satisfying (20) but with $K_p > K_q$, can be obtained through the duality of the Fourier and canonical bases: Suppose that a signal $y = [F, I]x$ is made of K_p Fourier atoms and K_q spikes. Denoting by $(\cdot)^*$ and $\overline{(\cdot)}$ the

---

Algorithm 1 *ProSparse*—Prony's Based Sparsity

Input: A dictionary $D = [F, I]$ and an observed vector $y \in \mathbb{C}^N$.

Output: A set $\mathcal{S}$, containing all (K_p, K_q)-sparse signal x that satisfies (20), with $K_p \leq K_q$.

Initialize $\mathcal{S} = \{[\mathbf{0}^T, y^T]^T\}$. This is a trivial solution, corresponding to $K_p = 0$ and $K_q = \|y\|_0$.

for $K_p = 1, 2, \ldots, \lceil \sqrt{N/2} - 1 \rceil$ **do**

    for $\ell = 0, 1, \ldots, N - 1$ **do**

        Build the Toeplitz matrix $T_{K_p,\ell}$ as in (13).

        Apply Prony's method on $T_{K_p,\ell}$ to find the parameters $\{\alpha_k, u_k\}$, where $0 \leq k < K_p$.

        if $\{u_k\}$ contains K distinct values, with each $u_k \in \{e^{j2\pi m/N} : m \in \mathbb{Z}\}$ **then**

            Compute the estimated Fourier contribution $\widehat{y}_n = \sum_{k=0}^{K_p-1} \alpha_k u_k^n$, for $0 \leq n < N$.

            Compute the residual $r = y - \widehat{y}$ and let $K_q = \|r\|_0$.

            if $K_p \leq K_q$ and $K_p K_q < N/2$ **then**

                Obtain the sparse signal x from the Fourier contribution $\widehat{y}$ and the residue r.

                $\mathcal{S} \Leftarrow \mathcal{S} \cup \{x\}$.

            end if

        end if

    end for

end for

Hermitian and complex conjugate operators, respectively, we can then easily verify that a "dual signal", $\overline{F^* y} = [I, F]\overline{x}$, is made of K_q Fourier atoms and K_p spikes. Consequently, to recover all x satisfying (20), we just need to run ProSparse twice, with y and $\overline{F^* y}$ being the input each time.

Next, we present ProSparse and verify that it indeed has the stated properties. The algorithm, summarized in the insert Algorithm 1 operates as follows: Let $\mathcal{S}$ be the set of solutions the algorithm will return. After initializing $\mathcal{S}$ with the trivial solution that $x = [\mathbf{0}^T, y^T]^T$ (corresponding to $K_p = 0$ and $K_q = \|y\|_0$), set $K_p = 1$. For each $\ell = 0, 1, \ldots, N - 1$, build the Toeplitz matrix $T_{K_p,\ell}$ using a sliding window $[y_\ell, y_{\ell+1}, \ldots, y_{\ell+2K_p-1}]$ of size $2K_p$. Apply Prony's method on the $K_p \times (K_p + 1)$ Toeplitz matrix $T_{K_p,\ell}$ in order to retrieve the K_p potential locations $\{u_k\}$ and amplitudes $\{\alpha_k\}$ of the Fourier atoms. If the parameters $\{u_k\}$ do not have K_p different values or if they are not in the expected form, *i.e.*, $u_k = e^{j2\pi m/N}$ for some integer m, set $\ell \Leftarrow \ell + 1$ and repeat the process. Otherwise, compute the contribution due to the Fourier atoms as $\widehat{y}_n = \sum_{k=0}^{K_p-1} \alpha_k u_k^n$ for $n = 0, 1, \ldots, N - 1$. Remove this contribution from y and check whether K_q, the number of nonzero entries of the residual, satisfies $K_p K_q < N/2$ and $K_p \leq K_q$. If these two conditions are satisfied, use the estimated Fourier atoms and the nonzero entries of the residual as one solution, and add it to the set $\mathcal{S}$. Set $K_p \Leftarrow K_p + 1$ and repeat the process up to $K_p = \lceil \sqrt{N/2} - 1 \rceil$, where $\lceil c \rceil$ denotes the smallest integer that is greater than or equal to a real number c.

By construction of the algorithm, any solution vector x in $\mathcal{S}$ must satisfy (20), subject to the additional condition that $K_p \leq K_q$. The opposite direction is also true, *i.e.*, $\mathcal{S}$ contains all such vectors. To see this, we first note that the two constraints, $K_p K_q < N/2$ and $K_p \leq K_q$, imply that $K_p < \sqrt{N/2}$. The trivial case, when $K_p = 0$, leads to a solution $x = [\mathbf{0}^T, y^T]^T$, which is added to $\mathcal{S}$ at the beginning of the algorithm. Now suppose that y can be written as a combination of K_p Fourier atoms and K_q spikes, such that $1 \leq K_p < \sqrt{N/2}$, $K_p K_q < N/2$ and $K_p \leq K_q$. Such a solution will always be found by ProSparse, because when $K_p K_q < N/2$, we know from Lemma 1 that an interval with $2K_p$ consecutive entries due only to Fourier atoms exists. Prony's method will then estimate the correct Fourier atoms from these entries and, in this case, the residual will have K_q nonzero entries, satisfying the required conditions.

Finally, we show that ProSparse has a worst-case complexity of $\mathcal{O}(N^3)$. We note that the algorithm has two nested iterations, over K_p and ℓ, respectively. Within the iterations, we apply Prony's method on a matrix of size $K_p \times (K_p + 1)$. Finding the polynomial coefficients h as in (13) through SVD has complexity $\mathcal{O}(K_p^3)$. Polynomial root finding in the algorithm has complexity up to $\mathcal{O}(K_p N)$. This is due to the fact that the correct roots in this case can only have N possible choices in the form of $\{e^{j2\pi m/N}, m = 0, 1, \ldots, N - 1\}$ [see also (11) and (12)]. Therefore, we just need to evaluate $P(x)$ of (12) at $x = e^{j2\pi m/N}, 0 \leq m < N$, to check if this is really a root of the polynomial, whose degree is up to K_p. After Prony's, the steps where the Fourier contribution is re-synthesized and where we compute the residue and check its sparsity have complexity $\mathcal{O}(K_p N)$. Therefore, for any fixed K_p and ℓ, the complexity of the algorithm is $\mathcal{O}(K_p^3 + K_p N)$. Since ProSparse loops over $1 \leq K_p \leq \lceil \sqrt{N/2} - 1 \rceil$ and $0 \leq \ell < N$, its overall complexity can thus be estimated as $\sum_{1 \leq K_p < \lceil \sqrt{N/2} \rceil} \mathcal{O}(K_p^3 N + K_p N^2) \lesssim \mathcal{O}(N^3)$. ∎

Remark 4: The reason that we consider $K_p \leq K_q$ (by using duality) in the ProSparse algorithm is to reduce the computational complexity. Note that, in this way, K_q just needs to iterate from 1 to $\lceil \sqrt{N/2} - 1 \rceil$ leading to an overall complexity of $\mathcal{O}(N^3)$. Without the constraint $K_p < K_q$, we should consider all K_p up to N, and this would yield a higher overall complexity.

In Proposition 2, the condition for successful sparse recovery, $K_p K_q < N/2$, is given in terms of the individual sparsity levels on the Fourier and canonical bases. It is often convenient to have a condition that only depends on the total sparsity level $K = K_p + K_q$. The following result serves this purpose.

Corollary 1: Assume $D = [F, I]$ and let $y \in \mathbb{C}^N$ be an arbitrary signal. There exists an algorithm, with a worst-case complexity of $\mathcal{O}(N^3)$, that finds *all* K-sparse signals x such that $y = Dx$ and

$$K < \sqrt{2N}. \tag{21}$$

In particular, this implies that, if $y = Dx$ for some K-sparse signal x with $K < \sqrt{N}$, the nonconvex problem (P_0), which is known to admit a unique solution in this case, can be solved by an algorithm with polynomial complexity.
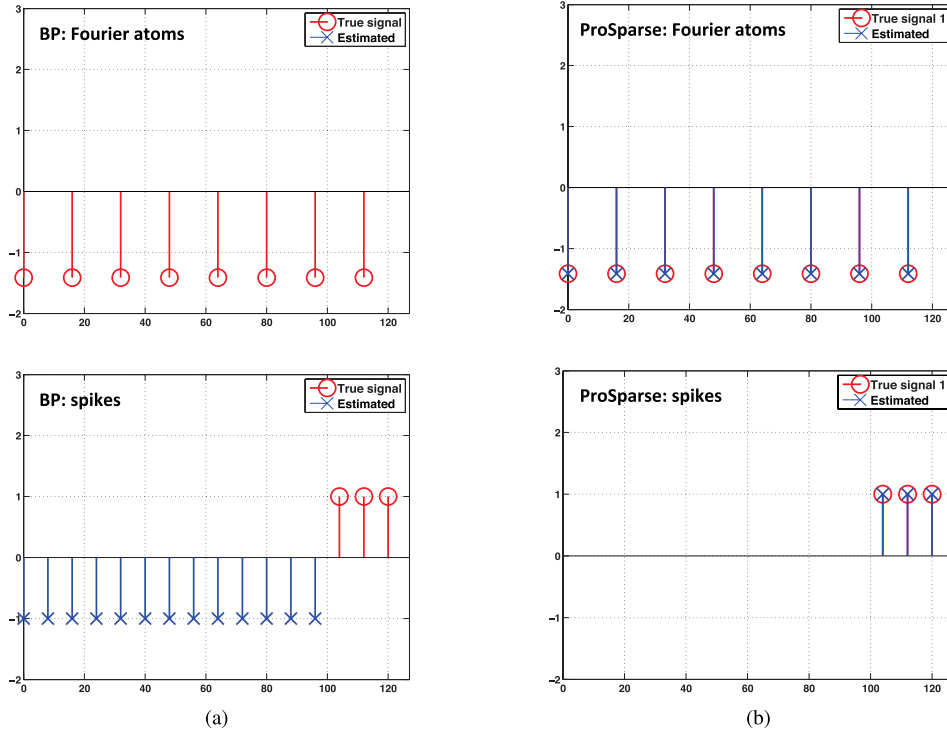
Fig. 2. ProSparse vs BP. The observed signal y is of length $N = 128$, made of $K_p = 8$ Fourier atoms and $K_q = 3$ spikes. For comparisons, both the reconstructed signals (in blue) and the ground truth signals (in red) are shown in the figures. (a): The Fourier atoms and spikes recovered by BP. In this case, BP does not find the correct sparse representation. In particular, none of the Fourier atoms has been recovered. (b): ProSparse perfectly retrieves the Fourier atoms and spikes from y. See Appendix V-A for details on how this example is constructed.

Proof: For any $K_p, K_q \geq 0$, we have $K = K_p + K_q \geq 2\sqrt{K_p K_q}$. Using this inequality, we can easily see that (21) poses a more restrictive condition than $K_p K_q < N/2$, meaning that the former implies the latter. The result then follows from Proposition 2. ∎

B. Numerical Validation

To visualize the results of Proposition 2 and Corollary 1, we refer the reader to Figure 1(b), where we plot the exact ProSparse bound $K_p K_q < N/2$ and its simplified version in (21). In that same figure, we also show the (P_0) bound (2) and the BP bound (9). It is evident that, compared with (P_0) and BP, ProSparse provides performance guarantees over a much wider range of sparsity levels. In what follows, we further validate these theoretical results by presenting two numerical examples where the (P_1) or BP formulation fails to retrieve the original sparse vector while ProSparse remains effective.

Example 1 (Beyond the BP Bound): In Figure 2, we show the results of applying BP and ProSparse, respectively, to find the sparse representation of a signal y. The length of y is $N = 128$, and it is made of $K_p = 8$ Fourier atoms and $K_q = 3$ spikes. This example has been constructed by adapting the methodology proposed in [4] and our construction is explained in more details in Appendix A.

We note that the sparsity levels are such that the uniqueness condition (2) for (P_0) holds but the BP bound (5) is not satisfied. Figure 2(a) shows the reconstruction results by using BP. In this case, BP fails to find the original sparse representation. In comparison, ProSparse retrieves the correct

Fourier atoms and spikes from y, as shown in Figure 2(b). That ProSparse works is expected, since the sparsity levels in this case stay well-within the ProSparse bound ($K_p K_q < N/2$) for successful recovery.

Example 2 (Beyond the (P_0) Uniqueness Bound): We consider an example where two different K-sparse signals lead to the same y. Clearly, this can be achieved only when $K > \sqrt{N}$, *i.e.*, when K is beyond the (P_0) uniqueness bound. In Appendix B, we construct one such y, with parameters $N = 128$ and $K = 12$. As shown in Figure 3(b) and Figure 3(c), ProSparse recovers both sparse solutions exactly, whereas the BP approach fails to find either of the two [see Figure 3(a)].

IV. GENERALIZATIONS: OTHER PAIRS OF BASES

In this section, we generalize the result of Proposition 2 and show that ProSparse-like algorithms can be used to solve the sparse representation problem for a larger family of dictionaries. In what follows, let $D = [\Psi, \Phi]$ be a dictionary consisting of a pair of bases. Unlike in our previous discussions, here we no longer require the two bases to be orthogonal. Let x_p, x_q be two N-dimensional vectors containing K_p and K_q nonzero entries, respectively. Our goal is to recover the (K_p, K_q)-sparse vector $x = [x_p^T, x_q^T]^T$ from the measurement $y = Dx$.

We note that the ProSparse algorithm presented in Section III utilizes two fundamental properties of the Fourier and identity matrices: First, each column of the identity matrix has only one nonzero entry so that Ix_q leaves only a sparse "footprint" on the observation vector y. Most entries

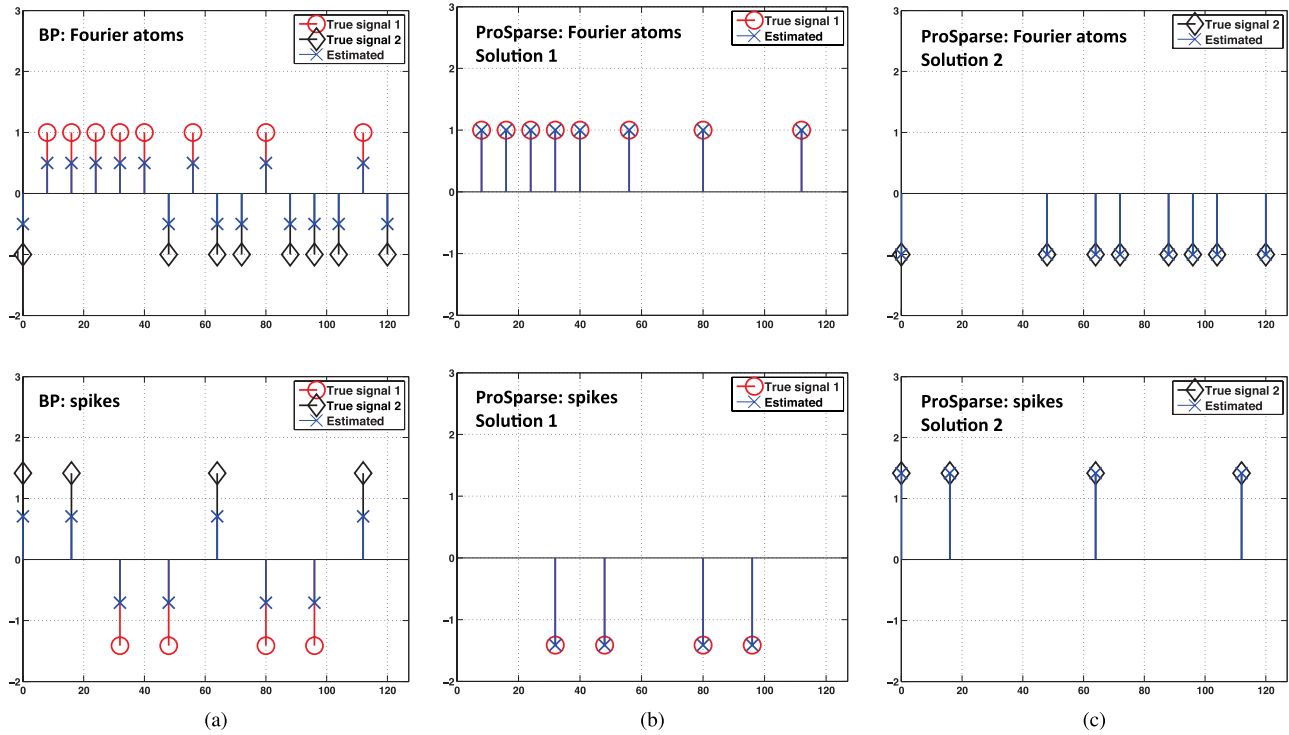Fig. 3. A case when the constraint $y = Dx$ admits two K-sparse solutions. The observed signal y is of length $N = 128$ and the sparsity level is $K = 12$. See Appendix V-B for details. In all the figures, the reconstructed signals are shown in blue, whereas the two ground truth signals are shown in red and black, respectively. (a): The BP approach fails to retrieve either of the two solutions. (b) and (c): ProSparse retrieves the two sparse solutions exactly.

of y are solely due to the Fourier component Fx_p. Second, the algebraic structure of the Fourier matrix allows us to reconstruct the sparse vector x_p from only a small number of consecutive entries of Fx_p.

We first generalize Φ from the canonical basis to *local bases*. For our purpose, we define the *support length* of a vector v as $\ell(v) \overset{\text{def}}{=} \max\{n : v_n \neq 0\} - \min\{n : v_n \neq 0\} + 1$. Essentially, $\ell(v)$ is the length of the shortest continuous interval that can cover the support of v, and it holds that $\ell(v) \geq \|v\|_0$. We call Φ a local basis if all the basis vectors $\{\Phi_i\}_{0 \leq i < N}$ have small support lengths, *i.e.*, the quantity

$$L_\Phi \overset{\text{def}}{=} \max_i \ell(\Phi_i) \qquad (22)$$

is small. For example, when Φ is the canonical basis, we have $L_\Phi = 1$; When Φ is a banded matrix, L_Φ is equal to the bandwidth of that matrix.

Next, we generalize the Fourier basis Ψ to those satisfying the *local sparse reconstruction* property.

Definition 1: Let Ψ be a basis and $z = \Psi c$ for some K-sparse vector c. The basis Ψ is said to satisfy the local sparse reconstruction property, if there exists a *polynomial complexity* algorithm that can reconstruct c from any $S_\Psi(K)$ consecutive entries

$$\{z_n, z_{n+1}, \dots, z_{n+S_\Psi(K)-1}\}, \qquad (23)$$

where $S_\Psi(K)$ is the minimum number of measurements required at the sparsity level K. In what follows, we shall refer to $S_\Psi(K)$ as the *sparse sampling factor* of Ψ.

From our previous discussions, we know that $S_\Psi(K) = 2K$ for Fourier matrices, and the reconstruction can be done by Prony's method. In Appendix C, we present a more general family of matrices, characterized by

$$\Psi = \Lambda V B, \qquad (24)$$

where $\Lambda \in \mathbb{C}^{N \times N}$ is a diagonal matrix, $V \in \mathbb{C}^{N \times M}$ is a Vandermonde matrix with $M \geq N$, and $B \in \mathbb{C}^{M \times N}$ is a matrix whose columns have sparse supports. There, we show that, under mild additional conditions on V and B, Prony's method can be used to recover a sparse vector c from $S_\Psi(K) = 2DK$ number of consecutive entries of $y = \Psi c$, where D is some positive integer (see Proposition 4 for details.) In particular, it is shown that the DCT matrix can be written in the form of (24) and that, in this case, we can reconstruct a K-sparse vector c from any $S_\Psi(K) = 4K$ consecutive entries of y.

To state our next result, we need to distinguish two cases: For those matrices (*e.g.*, the Fourier matrix) that have periodic rows, the indices in (23) should be viewed through the modulo (by N) operator. In this case, the starting index n can be arbitrarily chosen from $[0, N-1]$, and thus there is a total of N intervals in the form of (23). However, general basis matrices, such as those characterized in Appendix C, do not have the periodic property. Consequently, the starting index n in (23) can only be chosen from a smaller set, *i.e.*, $[0, N - S_\Psi(K)]$. In what follows, we refer to matrices in the former case as periodic matrices.

Proposition 3: Let Φ be a local basis with a maximum support length L_Φ as defined in (22), and Ψ be a basis satisfying the local sparse reconstruction property with a sparse

sampling factor $S_\Psi(K)$ as given in Definition 1. Assume $D = [\Psi, \Phi]$ and let $y \in \mathbb{C}^N$ be an arbitrary signal. There exists a polynomial complexity algorithm that finds *all* (K_p, K_q)-sparse signals x such that $y = Dx$ and

$$\left(S_\Psi(K_p) + L_\Phi - 1\right)(K_q + \tau) < N + \tau L_\Phi, \qquad (25)$$

where $\tau = 0$ if Ψ is a periodic matrix and $\tau = 1$ otherwise.

Proof: See Appendix D, where we provide a constructive proof by presenting a generalized version of the ProSparse algorithm. ∎

Example 3 (Fourier and Local Fourier Bases): Let us assume that D is the union of the Fourier basis F_N and the local Fourier basis H, defined as

$$H = \begin{bmatrix} F_L & 0 & \dots & 0 \\ 0 & F_L & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & F_L \end{bmatrix},$$

where the subscripts in F_N and F_L indicate that they are the Fourier matrices of size $N \times N$ and $L \times L$, respectively. Note that when $L = 2$, the matrix H can also be seen as the Haar wavelet basis with one level of decomposition. The mutual coherence of the dictionary is $\mu(D) = \sqrt{L/N}$, and thus, from (4), the uniqueness condition for (P_0) is

$$K_p + K_q < \sqrt{N/L}.$$

To apply the result of Proposition 3, we substitute $S_\Psi(K_p) = 2K_p$, $\tau = 0$, and $L_\Phi = L$ into (25) and get the ProSparse bound as

$$(2K_p + L - 1)K_q < N. \qquad (26)$$

For an easier comparison between the above two bounds, we can verify that a sufficient condition for (26) to hold is[2]

$$K_p + K_q < \sqrt{2N} - (L - 1)/2. \qquad (27)$$

If we choose, for example, $L = \sqrt{N}$, then the (P_0) problem is unique when the total sparsity is below $N^{1/4}$. In contrast, (27) implies that the generalized ProSparse algorithm can handle a much wider range of sparsity levels, recovering all signals whose sparsity level is below $(\sqrt{2} - 0.5)\sqrt{N} + 0.5$.

Example 4 (DCT and Canonical Bases): Let $D = [\Psi, I]$ be the union of the DCT and canonical bases. The mutual coherence in this case is $\mu(D) = \sqrt{2/N}$. Consequently, unicity of (P_0) is guaranteed when

$$K_p + K_q < \sqrt{N/2}.$$

We have shown in Appendix C that $S_\Psi(K_p) = 4K_p$ for DCT matrices. Substituting this quantity, together with $\tau = 1$ (since Ψ is not periodic) and $L_\Phi = 1$ into (25), we conclude that ProSparse can retrieve all (K_p, K_q)-sparse signals when $4K_p(K_q + 1) < N + 1$. A sufficient condition for this bound to hold is

$$K_p + K_q < \sqrt{N + 1} - 1.$$

---

[2]Here, we have used again the inequality $x + y \geq 2\sqrt{xy}$ with $x = K_p + (L - 1)/2$ and $y = K_q$.

Therefore, in this case, the ProSparse bound is again a superset of the (P_0) bound.

Example 5 (Random Gaussian and Canonical Bases): In this example, we consider $D = [\Psi, I]$, where the entries $\{\Psi_{i,j}\}$ of the first basis matrix are independent realizations of Gaussian random variables, *i.e.*, $\Psi_{i,j} \sim \mathcal{N}(0, 1)$. Adapting standard results in compressed sensing [21]–[23], we verify in Appendix E the following result: Define

$$S_\Psi(K) = \max\{p(N), \quad \min\{N, c_1 K \log(N/K)\}\}, \qquad (28)$$

where $p(N)$ is some positive function of N and c_1 is some constant. Then there exist constants $c_1, c_2 > 0$, which do not depend on N or K, such that, with probability at least

$$1 - 2N^2 e^{-c_2 p(N)}, \qquad (29)$$

the random matrix Ψ will satisfy the following property: Let c be any K-sparse vector. One can efficiently reconstruct c from any $S_\Psi(K)$ consecutive (modulo N) entries of the vector $z = \Psi c$. By properly choosing $p(N)$, the probability in (29) can be made arbitrarily close to one for sufficiently large N, and thus, the matrix Ψ will satisfy the desired property with high probabilities. It then follows from Proposition 3 that, for those suitable Ψ, we can find all (K_p, K_q)-sparse signals x from $y = [\Psi, I]x$ if

$$S_\Psi(K_p)K_q < N,$$

where $S_\Psi(\cdot)$ is the function defined in (28).

Finally, we make the following observation. Denote by $D = [\Psi, \Phi]$ a dictionary for which ProSparse can be used successfully. Namely, D can be the union of any pair of bases discussed so far. Let A be an arbitrary $N \times N$ invertible matrix. Then, ProSparse can also be used on the dictionary $\widetilde{D} = [A\Psi, A\Phi]$. This fact can be trivially demonstrated by noting that, given $y = ADx$, we can return to the original dictionary by working with $\widetilde{y} = A^{-1}y$.

V. CONCLUSIONS

We considered the problem of finding the sparse representations of a signal in the union of two bases. We introduced a new polynomial complexity algorithm *ProSparse* which, for the case of Fourier and canonical bases, is able to find all the sparse representations of the signal under the condition that $K_p K_q < N/2$ (or, in terms of the total sparsity level, $K_p + K_q < \sqrt{2N}$). The new algorithm provides deterministic performance guarantees over a much wider range of sparsity levels than do existing algorithms such as the nonconvex ℓ_0-minimization or BP. In particular, our results imply that the ℓ_0-minimization problem for sparse representation is not NP-hard under the unicity condition and when the dictionary is the union of Fourier and canonical bases. Furthermore, we have shown that the proposed algorithm can be extended to other relevant pairs of bases, one of which must have local atoms. Examples include the Fourier and local Fourier bases, the DCT and canonical bases, and the random Gaussian and canonical bases.

APPENDIX

A. Constructing Counterexamples

In [4], Feuer and Nemirovsky constructed an example showing that (P_0) and (P_1) are not equivalent, for the case where $D = [H, I]$. Here, H is the scaled Hadamard matrix, with $H^T H = I$. In this appendix, we summarize the construction in [4] and show how to adapt it to the case where D is the union of Fourier and canonical bases.

We first note that the uniqueness condition (4) for the union of Hadamard and canonical bases is the same as that for Fourier and canonical bases. In both cases, the solution to (P_0) is unique when $K < \sqrt{N}$. In what follows, we set $N = 2^{2d-1}$ for some positive integer d, and let $K = \lfloor \sqrt{N} \rfloor$.

The key idea behind Feuer and Nemirovsky's construction [4] is to find a vector $z \in \mathbb{R}^{2N}$ such that $Dz = 0$ and that

$$\|z\|_1 = \sum_{n=0}^{N-1} |z_n| < 2 \sum_{k=0}^{K-1} |z_{n_k}|, \qquad (30)$$

where the indices $n_0, n_1, \ldots, n_{K-1}$ correspond to the K largest absolute values of z. We will provide explicit constructions of z a little later. For now, assume that such a vector z has already been found.

Given z, one then builds two vectors $x, \widetilde{x} \in \mathbb{R}^{2N}$ as follows: a K-sparse vector x whose nonzero entries satisfy $x_{n_k} = -2z_{n_k}$, where the indices $\{n_k\}$ are same as those in (30); and a second vector $\widetilde{x} = z + x$. Given $y = Dx$, we know that x is the unique solution of (P_0), since the bound $K < \sqrt{N}$ is satisfied here. However, x is not the solution of (P_1). To see this, we note that, since $Dz = 0$, we must have $y = Dx = D\widetilde{x}$. Meanwhile, by construction,

$$\|\widetilde{x}\|_1 = \|z\|_1 < \|x\|_1,$$

where the inequality is due to (30). Consequently, given y, the solution to (P_1) will not be x, since there is at least one alternative vector, $\widetilde{x}$, satisfying the same synthesis equation $y = D\widetilde{x}$ but with a smaller ℓ_1 norm.

Next, we present explicit constructions of the vector z with the desired properties. A suitable z was found in [4] for the case of Hadamard and identity matrices. Here, we modify that construction so that it is suitable to the case of Fourier and canonical bases. Recall that $N = 2^{2d-1}$. Define $m = 2^d$, and let $v \in \mathbb{R}^N$ be a "picket-fence" signal, containing exactly N/m uniformly spaced nonzero entries, all of which are equal to $\sqrt{2}$. More precisely, v is equal to the following Kronecker product $\sqrt{2}(1 \otimes e_0)$, where $1 \in \mathbb{R}^{N/m}$ is a vector of all 1's, and $e_0 = [1, 0, 0, \ldots, 0]^T$ is the first canonical basis vector in $\mathbb{R}^m$. Let

$$z = \begin{bmatrix} v \\ -Fv \end{bmatrix}. \qquad (31)$$

By construction, $Dz = [F, I]z = 0$. Meanwhile, it can be verified that, just like v, the vector Fv is also a "picket-fence" signal, containing m uniformly spaced nonzero entries, all of which are equal to 1. Since z is a concatenation of v and $-Fv$, it contains exactly $2^{d-1} + 2^d$ nonzero entries, the first 2^{d-1} of which are equal to $\sqrt{2}$ and the remaining 2^d of which are

equal to 1. Now we just need to verify that z satisfies (30). To that end, we first note that $K = \lfloor \sqrt{N} \rfloor = \lfloor 2^{d-0.5} \rfloor \geq 2^{d-1}$. It follows that

$$2 \sum_{k=0}^{K-1} |z_{n_k}| - \sum_{n=0}^{N-1} |z_n|$$
$$= 2 \left(2^{d-1}\sqrt{2} + \lfloor 2^{d-0.5} - 2^{d-1} \rfloor \right) - \left(2^{d-1}\sqrt{2} + 2^d \right)$$
$$= 2 \left(2^{d-1}(\sqrt{2}/2 - 1) + \lfloor 2^{d-1}(\sqrt{2} - 1) \rfloor \right). \qquad (32)$$

It is easy to show that, for all $d \geq 4$, the right-hand side of (32) is strictly positive. Therefore, the vector z as constructed above satisfies the condition (30) for all $d \geq 4$.

B. Constructing Examples Where $y = Dx$ Admits Two Sparse Solutions

We show how to construct two K-sparse vectors x_0, x_1 such that $Dx_0 = Dx_1$, where $D = [F, I]$. We start with the vector z defined in (31) in Appendix A. By construction, $Dz = 0$ and z has exactly $L = 2^{d-1} + 2^d$ nonzero entries, where d is a positive integer satisfying $2^{2d-1} = N$. We set $d \geq 2$ so that L is even. The two vectors x_1, x_0 are then easily built by assigning $K = L/2$ randomly chosen nonzero entries of z to x_0 and then setting $x_1 = x_0 - z$. Since $Dz = 0$, we must have $Dx_0 = Dx_1$. Meanwhile, both vectors have the same sparsity level $K = L/2$, which is beyond the (P_0) uniqueness bound (2) but still within the ProSparse bound given in (21).

C. Generalizing Prony's Method

In Section II, we showed that Prony's method provides an efficient way to reconstruct a sparse signal c from a small number of consecutive entries of the observation vector $y = Fc$, where F is the DFT matrix. Here, we generalize Prony's method for sparse recovery to a larger class of bases, all of which have the following form:

$$\Psi = \Lambda V S, \qquad (33)$$

where $\Lambda \in \mathbb{C}^{N \times N}$ is a diagonal matrix; $V \in \mathbb{C}^{N \times M}$ is a Vandermonde matrix whose rows are the powers of a vector $p = [p_0, p_1, \ldots, p_{M-1}]$ with *distinct* elements, i.e., $[V]_{n,m} = p_m^n$ for $0 \leq n < N, 0 \leq m < M$; and $S \in \mathbb{C}^{M \times N}$ is a matrix whose columns are all sparse. In particular, we assume that, for all $0 \leq n < N$, the nth column of S, denoted by s_n, satisfies

$$\|s_n\|_0 \leq D,$$

for some $D > 0$.

Proposition 4: Let $y = \Psi c$, where c is a K-sparse vector and Ψ is an invertible matrix in the form of (33). If Λ, V and S satisfy the conditions stated above, we can use Prony's method to recover c from any $2DK$ consecutive entries of y.

Proof: The case when $2DK \geq N$ is trivial: since the entire vector y is available, we can reconstruct c by a direct linear inversion, i.e., $c = \Psi^{-1} y$. In what follows, we assume that $2DK < N$.

The basis matrix Ψ being invertible implies that the diagonal matrix $\Lambda = \text{diag}\{\lambda_0, \lambda_1, \ldots, \lambda_{N-1}\}$ must also be invertible.

Introducing two vectors $z \overset{\text{def}}{=} \Lambda^{-1} y$ and $x \overset{\text{def}}{=} Sc$, we can simplify the relationship $y = \Lambda V S c$ as

$$z = Vx.$$

Since x is a linear combination of K vectors, each of which has a sparsity level bounded by D, we must have $\|x\|_0 \le DK < N/2$. Let $\{m_0, m_1, \ldots, m_{DK-1}\}$ denote indices of the nonzero elements of x. It follows from the Vandermonde structure of V that the nth entry of z can be written as

$$z_n = \sum_{k=0}^{DK-1} x_{m_k} p_{m_k}^n,$$

which has the same "sums of exponentials" form as in (11). Consequently, by following the same derivations in Section II, we can show that Prony's method[3] can be used to reconstruct $\{x_{m_k}\}$ and $\{p_{m_k}\}$, and therefore x, from any $2DK$ consecutive entries of z. Since $z_n = y_n / \lambda_n$, this is equivalent to requiring $2DK$ consecutive entries of y. Finally, since Ψ is invertible, the matrix S must necessarily have full column-rank. Thus, the K-sparse vector c can be obtained from x through a simple linear inversion $c = (S^T S)^{-1} S^T x$. ∎

Example 5 (Discrete Cosine Transform): Let Ψ be the DCT matrix, whose (n, m)th entry is

$$\psi_{n,m} = b(n) \sqrt{\frac{2}{N}} \cos \frac{\pi n (m + 0.5)}{N}, \quad \text{for} \quad 0 \le n, m, < N$$

with

$$b(n) = \begin{cases} 1/\sqrt{2} & \text{if } n = 0, \\ 1 & \text{if } 1 \le n < N. \end{cases}$$

Using the identity

$$2 \cos \frac{\pi n (m + 0.5)}{N} = e^{j\pi n(m+0.5)/N} + e^{-j\pi n(m+0.5)/N},$$

we can factor Ψ in the form of (33): The diagonal matrix is $\Lambda = \frac{1}{\sqrt{2N}} \text{diag} \{b_0, b_1, \ldots, b_{N-1}\}$; the Vandermonde matrix V is generated by powers of the row vector $p = [p_0, p_1, \ldots, p_{2N-1}]^T$, where $p_m = e^{-j\pi(m+0.5)/N}$ for $0 \le m \le N - 1$ and $p_m = -e^{j\pi(m+0.5)/N}$ for $N \le m < 2N$; the third matrix $S = [1, 1]^T \otimes I_N$, where $\otimes$ denotes the matrix Kronecker product and I_N is the $N \times N$ identity matrix.

We can easily verify that the entries of p are all distinct and that each column of S has exactly two nonzero entries (thus, $D = 2$). It follows from Proposition 4 that Prony's method is applicable in this case: We can recover a K-sparse vector c from any $4K$ consecutive entries of y.

D. Generalized ProSparse Algorithm

In this appendix, we provide a constructive proof of Proposition 3. To do that, we first need to establish the following result, which is a more general version of Lemma 1. Let $\Phi_{n_1}, \Phi_{n_2}, \ldots, \Phi_{n_K}$ be a set of K atoms from the local basis Φ. Similar to (15), we can count the number

[3]It is possible that the number of nonzero elements of x is less than DK. This will not cause a problem for Prony's method, since the algorithm only needs to know an upper bound on the true sparsity level. See Remark 2 in Section II for more details.

of all intervals of length S that are not "corrupted" by any of these atoms as

$$
\begin{aligned}
&\mathcal{N}_{\Phi,\tau}(S; n_1, n_2, \ldots, n_K) \\
&\overset{\text{def}}{=} \#\Big\{ \ell : 0 \le \ell \le N - 1 - \tau(S - 1) \text{ and} \\
&\quad \{\ell, \ell+1, \ldots, \ell+S-1\} \cap \bigcup_{1 \le i \le K} \text{supp } \Phi_{n_i} = \emptyset \Big\}, \quad (34)
\end{aligned}
$$

where $\text{supp } \Phi_{n_i}$ denotes the support of Φ_{n_i}, and τ is a binary value: $\tau = 0$ if the indices in (34) are periodic on the torus $[0, 1, \ldots, N - 1)$, and $\tau = 1$ otherwise.

Lemma 2: Let Φ be a local basis with a maximum support length L_Φ. For any choice of K basis vectors $\{\Phi_{n_i}\}_{1 \le i \le K}$, it holds that

$$\mathcal{N}_{\Phi,\tau}(S; n_1, n_2, \ldots, n_K) \ge N + \tau L_\Phi - (S + L_\Phi - 1)(K + \tau). \quad (35)$$

Proof: By the definition of maximum support length (22), the support of each basis vector must be fully inside of an interval of length L_Φ, i.e.,

$$\text{supp } \Phi_{n_i} \subseteq \mathcal{I}_i \overset{\text{def}}{=} [m_i, m_i + 1, \ldots, m_i + L_\Phi - 1].$$

Without loss of generality, we assume that the starting indices, $\{m_i\}$, are in ascending order, with $m_1 \le m_2 \le \cdots \le m_K$.

We first consider the case when $\tau = 1$, i.e., the indices are not periodic. Let d_i denote the length of the "uncorrupted" interval that strictly falls between two neighboring intervals $\mathcal{I}_i$ and $\mathcal{I}_{i+1}$. It is easy to verify that

$$d_i = \max \{0, m_{i+1} - m_i - L_\Phi\}, \quad (36)$$

for $0 \le i \le K$. Note that we define $m_0 = -L_\Phi$ and $m_{K+1} = N$, so that d_0 and d_K count, respectively, the number of indices in front of the first interval and the number of those after the last interval. For those intervals with $d_i \ge S$, we can find $d_i - S + 1$ (overlapping) subintervals, each of length S. It follows that

$$
\begin{aligned}
&\mathcal{N}_{\Phi,\tau=1}(S; n_1, n_2, \ldots, n_K) \\
&\ge \sum_{0 \le i \le K} \max \{0, d_i - S + 1\} \quad (37) \\
&\ge \sum_{0 \le i \le K} (m_{i+1} - m_i - L_\Phi) - (K+1)(S-1) \\
&= m_{K+1} - m_0 - (K+1)(S + L_\Phi - 1), \quad (38)
\end{aligned}
$$

which leads to the bound in (35) for $\tau = 1$.

The proof for the case when $\tau = 0$, i.e., when the indices are periodic, is similar. Setting $m_0 = m_1$ and $m_{K+1} = N + m_1$ in (36), we have $d_0 = 0$ and d_K measures the number of indices (modulo N) between the last interval $\mathcal{I}_K$ and the first interval $\mathcal{I}_1$. Unlike in (37) where we sum over $0 \le i \le K$, here, since $d_0 = 0$, we just need to sum over $1 \le i \le K$ and get

$$\mathcal{N}_{\Phi,\tau=0}(S; n_1, n_2, \ldots, n_K) \ge \sum_{1 \le i \le K} \max \{0, d_i - S + 1\}.$$

Following the same steps in reaching (38), we can show that the above inequality yields the bound (35) for $\tau = 0$. ∎

Algorithm 2 Generalized ProSparse for Sparse Signal Reconstruction

---

Input: A dictionary $D = [\Psi, \Phi]$ and an observed vector $y \in \mathbb{C}^N$.

Output: A set $\mathcal{S}$, containing all (K_p, K_q)-sparse signal x that satisfies the conditions $y = Dx$ and (25).

Set $\tau = 0$ if Ψ is a periodic matrix and $\tau = 1$ otherwise.

Initialize $\mathcal{S} = \{[0^T, y^T]^T\}$. This is a trivial solution, corresponding to $K_p = 0$ and $K_q = \|y\|_0$.

for $K_p = 1, 2, \ldots, N$ **do**

  for $\ell = 0, 1, \ldots, (N - 1 - \tau(S_\Psi(K_p) - 1))$ **do**

    Use a polynomial-complexity algorithm to estimate a K_p-sparse signal x_1 from a set of consecutive measurements $[y_\ell, \ldots, y_{\ell+S_\Psi(K_p)-1}]$.

    Compute the estimated contribution from the first basis as $\widehat{y} = \Psi x_1$.

    Compute $x_2 = \Phi^{-1}(y - \widehat{y})$ and let $K_q = \|x_2\|_0$.

    if (K_p, K_q) satisfy the condition (25) **then**

      Obtain the sparse signal as $x = [x_1^T, x_2^T]^T$.

      $\mathcal{S} \Leftarrow \mathcal{S} \cup \{x\}$.

    end if

  end for

end for

Proof of Proposition 3: We provide a constructive proof of Proposition 3 by presenting in Algorithm 2 a generalized version of ProSparse. We first show that the algorithm can find every (K_p, K_q)-sparse signal $x = [x_1^T, x_2^T]^T$ satisfying $y = Dx$ and (25). To see this, we note that, if (25) holds, then Lemma 2 guarantees the existence of at least one length-$S_\Psi(K_p)$ interval that is only due to the atoms in Ψ. Since the algorithm searches over all possible values of K_p and all possible choices of the intervals, the above-mentioned interval will be examined by the algorithm. By the definition of $S_\Psi(K_p)$, such an interval is sufficient for us to reconstruct the K_p-sparse signal x_1 with polynomial complexity. Given x_1, the second half of x can then be obtained by removing from y the contributions of Ψ, *i.e.*, $x_2 = \Phi^{-1}(y - \Psi x_1)$.

For computational complexity, we note that the generalized algorithm has two nested iterations, over $1 \leq K_p \leq N$ and $0 \leq \ell \leq (N - 1 - \tau(S_\Psi(K_p) - 1))$, respectively. Within the iterations, the steps in estimating x_1 and x_2 both take polynomial time. Therefore, the overall complexity of the algorithm is polynomial in N.

E. Random Gaussian Basis

Let $\Psi_{n,S} \in \mathbb{R}^{S \times N}$ denote a submatrix constructed from S consecutive rows of the random Gaussian matrix Ψ. The subscript n in $\Psi_{n,S}$ indicates that these rows are taken at indices $\{n, n+1, \ldots, n+S-1\}$, where the indices are viewed through the modulo (by N) operator. With high probabilities, the normalized matrix $\Psi(n, S)/\sqrt{S}$ satisfies the restricted isometry property (RIP) [24] in compressed sensing [21], [22], [25], allowing one to reconstruct a sparse vector c by solving the following convex optimization problem:

$$\arg \min_{\widetilde{c}} \|\widetilde{c}\|_1 \quad \text{s.t.} \quad \Psi_{n,S} \widetilde{c} = \Psi_{n,S} c. \tag{39}$$

More precisely, it was shown in [23] that the following holds: There exist two positive constants c_1, c_2 such that, with probability $\geq 1 - 2e^{-c_2 S}$, the matrix $\Psi_{n,S}$ will satisfy the RIP with suitable parameters that are sufficient to guarantee the success of the optimization problem (39) in recovering any K-sparse vector c, for all K satisfying the condition $c_1 K \log(N/K) \leq S$.

To apply the result of Proposition 3, we need to show that the random matrix Ψ will satisfy the following property with high probabilities: One can efficiently reconstruct any K-sparse vector c from any $S_\Psi(K)$ consecutive entries of the observation $z = \Psi c$, where $S_\Psi(K)$ is the function defined in (28). To that end, it is sufficient to show that, with high probabilities, the submatrices $\Psi_{n,S_\Psi(K)}$ for all n and all K will simultaneously satisfy the RIP condition. We note that any given submatrix $\Psi_{n,S_\Psi(K)}$ will fail to satisfy the required RIP condition with probability $\leq 2e^{-c_2 S_\Psi(K)} \leq 2e^{-c_2 p(N)}$. Since there is a total of N possible starting indices (*i.e.*, $0 \leq n < N$) and up to N different values of K, we can conclude, by applying the union bound, that the matrix Ψ will satisfy the desired property with probability at least $1 - 2N^2 e^{-c_2 p(N)}$. By choosing, for example, $p(N) = (3/c_2) \log(N)$, the previous probability bound becomes $1 - 2/N$, which can be made arbitrarily close to one for sufficiently large N.

REFERENCES

[1] 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, Apr. 1998.

[2] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[3] M. Elad and A. 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.

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

[5] G. Davis, S. Mallat, and M. Avallaneda, "Adaptive greedy approximations," *Constructive Approx.*, vol. 13, no. 1, pp. 57–98, 1997.

[6] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[7] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.

[8] B. Ricaud and B. Torrésani, "Refined support and entropic uncertainty inequalities," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4272–4279, Jul. 2013.

[9] B. Ricaud and B. Torrésani, "A survey of uncertainty principles and some signal processing applications," *Adv. Comput. Math.*, vol. 40, no. 3, pp. 629–650, Jun. 2014.

[10] M. Elad, *Sparse and Redundant Representations*. New York, NY, USA: Springer-Verlag, 2010.

[11] M. A. Khamis, A. C. Gilbert, H. Q. Ngo, and A. Rudra, "Sparse approximation, list decoding, and uncertainty principles," *IEEE Trans. Inf. Theory*, to be published.

[12] G. C. F. M. R. de Prony, "Essai expérimental et analytique sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et la vapeur de l'alkool à différentes températures," *J. École Polytech.*, vol. 1, no. 2, pp. 24–76, 1795.

[13] E. R. Berlekamp, *Algebraic Coding Theory*. New York, NY, USA: McGraw-Hill, 1968.

[14] J. L. Massey, "Shift-register synthesis and BCH decoding," *IEEE Trans. Inf. Theory*, vol. 15, no. 1, pp. 122–127, Jan. 1969.

[15] P. Milanfar, G. C. Verghese, W. C. Karl, and A. S. Willsky, "Reconstructing polygons from moments with connections to array processing," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 432–443, Feb. 1995.

[16] B. Gustafsson, C. He, P. Milanfar, and M. Putinar, "Reconstructing planar domains from their moments," *Inverse Problems*, vol. 16, no. 4, pp. 1053–1070, Aug. 2000.

[17] M. Elad, P. Milanfar, and G. H. Golub, "Shape from moments—An estimation theory perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1814–1829, Jul. 2004.

[18] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[19] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[20] A. Hormati, O. Roy, Y. M. Lu, and M. Vetterli, "Distributed sampling of signals linked by sparse filtering: Theory and applications," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1095–1109, Mar. 2010.

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

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

[23] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

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

[25] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

Pier Luigi Dragotti is Professor of Signal Processing with the Electrical and Electronic Engineering Department at Imperial College London. He received the Laurea degree (summa cum laude) in electrical and electronic engineering from the University of Naples Federico II, Naples, Italy, in 1997; the masters degree in communications systems from the Swiss Federal Institute of Technology of Lausanne (EPFL), Lausanne, Switzerland, in 1998, and the Ph.D. degree from EPFL in April 2002. 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. He was Technical Co-Chair for 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). Currently he is an elected member of the IEEE Signal Processing Theory and Method Technical Committee. He is a recipient of the ERC Starting Investigator Award for the project RecoSamp. His work includes sampling theory, wavelet theory and its applications, image and video compression, image-based rendering, and image super-resolution.

Yue M. Lu received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana- Champaign, Urbana, IL, in 2007.

He was a Research Assistant at the University of Illinois at Urbana-Champaign, and was with Microsoft Research Asia, Beijing, China and Siemens Corporate Research, Princeton, NJ. From September 2007 to September 2010, he was a postdoctoral researcher at the Audiovisual Communications Laboratory at Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is currently an Assistant Professor of Electrical Engineering at Harvard University, Cambridge, MA, directing the Signals, Information, and Networks Group (SING) at the School of Engineering and Applied Sciences. His research interests include statistical signal processing, sampling theory, multiscale geometrical representations, imaging, and image processing.

He received the Most Innovative Paper Award of IEEE International Conference on Image Processing (ICIP) in 2006 for his paper (with Minh N. Do) on the construction of directional multiresolution image representations, and the Best Student Paper Award of IEEE ICIP in 2007. He also coauthored a paper (with Ivan Dokmanić and Martin Vetterli) that won the Best Student Paper Award of IEEE International Conference on Acoustics, Speech and Signal Processing in 2011.