Efficient Multidimensional Diracs Estimation With Linear Sample Complexity

Hanjie Pan 🅾, *Student Member, IEEE*, Thierry Blu 🅾, *Fellow, IEEE*, and Martin Vetterli 🅾, *Fellow, IEEE*

Abstract—**Estimating Diracs in continuous two or higher dimensions is a fundamental problem in imaging. Previous approaches extended one-dimensional (1-D) methods, like the ones based on finite rate of innovation (FRI) sampling, in a separable manner, e.g., along the horizontal and vertical dimensions separately in 2-D. The separate estimation leads to a sample complexity of $\mathcal{O}\left(K^D\right)$ for K Diracs in D dimensions, despite that the total degrees of freedom only increase linearly with respect to D. We propose a new method that enforces the continuous-domain sparsity constraints simultaneously along all dimensions, leading to a reconstruction algorithm with *linear* sample complexity $\mathcal{O}(K)$, or a gain of $\mathcal{O}\left(K^{D-1}\right)$ over previous FRI-based methods. The multidimensional Dirac locations are subsequently determined by the intersections of hypersurfaces (e.g., curves in 2-D), which can be computed algebraically from the common roots of polynomials. We first demonstrate the performance of the new multidimensional algorithm on simulated data: multidimensional Dirac location retrieval under noisy measurements. Then, we show results on real data: radio astronomy point source reconstruction (from LOFAR telescope measurements) and the direction of arrival estimation of acoustic signals (using Pyramic microphone arrays).**

Index Terms—**Finite rate of innovation (FRI), continuous-domain sparsity, multidimension, point source reconstruction.**

I. INTRODUCTION

CONTINUOUS-DOMAIN sparse recovery is a classic problem at the heart of various applications, like point source estimation in radio astronomy [1], and direction of arrival estimation (DOA) with sensor arrays [2]. Typically, a low-pass filtered version of the sparse signal (or equivalently its low frequency spectrum) is observed in the measurement process. It is often the case that only a limited number of measurements is available—either due to the bandwidth of the instruments (as in microscopic imaging) or because of the finite number of sensors (as in array signal processing).

Therefore, it is essential to devise an efficient algorithm that estimates the sparse signal robustly from a set of discrete samples. The sampling framework for signals with finite rate of innovation (FRI) [3] is a natural candidate for this task. For instance, it has been shown that a periodic stream of Diracs in 1D, which consists of K Diracs within each period, is reconstructed exactly from $2K + 1$ ideal samples of the Dirac stream. The key to the FRI-based reconstruction is the annihilating filter method [3], [4], which allows the enforcement of the continuous-domain sparsity constraint with a discrete convolution equation (see Section II-A). Recently, the FRI framework has been generalized to cope with *non-uniform* measurements [5], making the framework applicable to many practical problems, such as radio astronomy [6] and DOA estimation for arbitrary array layouts [7]. Previous attempts to extend the FRI-based sparse recovery to two dimensions led to various sampling schemes for polygons [8], [9], and curves [10]–[12], which can be reliably reconstructed with the recently proposed generalized FRI recovery framework [5].

The focus in this work is on efficiently sampling and reconstruction of a specific type of sparse signals, namely multidimensional Diracs. This is directly related to point source estimation in various fields. Previous FRI-based approaches for two-dimensional Diracs [13] recasts the problem as two sub-problems, which estimate the Dirac locations along horizontal and vertical directions separately. The separate reconstruction approach requires an additional pairing step to combine the estimated Dirac coordinates. Additionally, the sample complexity is quadratic in 2D [5], [13].

Alternatively, as we propose in this paper, we reconstruct multi-dimensional Diracs by enforcing the annihilation constraints in all dimensions simultaneously. The Dirac locations are then given by the intersections of hypersurfaces, which are the zero-crossings of multivariate polynomials with coefficients specified by the reconstructed annihilating filter coefficients. For instance, in 2D, the estimation of Dirac locations amounts to finding the intersections of two curves, which can be computed algebraically by finding the common roots of two polynomials (see Fig. 2 and Section II-B2). Our main contributions in this work are:

i) The extension of the generic FRI sampling and reconstruction framework [5] to higher dimensions (two dimensions and above)—the resulting sampling scheme has *linear* sample complexity in signal sparsity (see e.g., Corollary 1). The reduced sample complexity is a significant improvement over our previous approach, where the
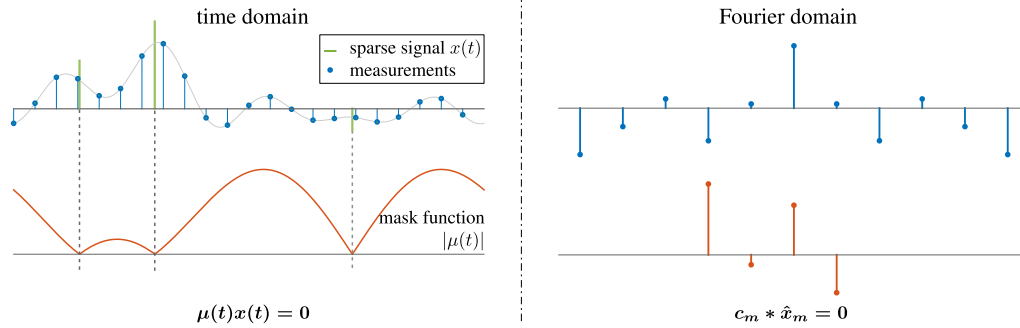
Fig. 1.   Sensing sparsity through annihilation: instead of recovering the sparse signal $x(t)$ directly, the FRI-based approach estimates a smooth function $\mu(t)$, which vanishes at the non-zero locations of the sparse signal. The continuous-domain constraint $\mu(t)x(t) = 0$, can be equivalently enforced with a discrete convolution equation in Fourier domain: $c_m * \hat{x}_m = 0$, where c_m is known as the "annihilating filter" (see text in Section II-A1 for details).
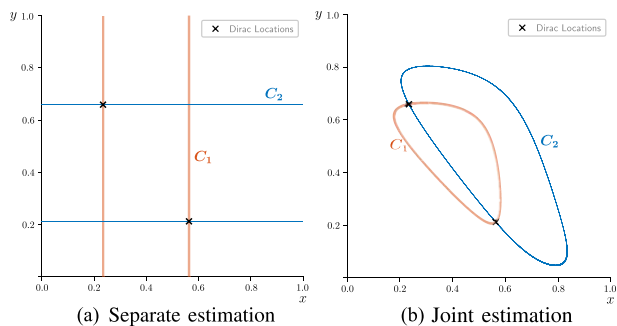

Fig. 2.   Joint estimation of 2D Diracs based on curve intersections. The Dirac locations are given by the intersections of two curves C_1 and C_2 (see text in Section II-B2). (a) The zero-crossings of two univariate polynomials specified by two 1D annihilating filters in the separate estimation (Section II-B1). (b) The zero-crossings of two (bivariate) polynomials specified by general shaped 2D annihilating filters in the joint estimation.

TABLE I
SAMPLE COMPLEXITIES OF SEPARATE AND JOINT ESTIMATION APPROACHES TO
RECONSTRUCT K D-DIMENSIONAL DIRACS

Approach	Separate estimation (Section II-B1)	Joint estimation with hypercubic sample shape (Corollary 1 and 4)	Joint estimation with flexible sample shape (Corollary 2 and 5)
minimum number of samples	$(K+1)^D$	$2^D K$	$(D+1)K$
sample complexity	$\mathcal{O}(K^D)$	$\mathcal{O}(K)$	$\mathcal{O}(K)$
degrees of freedom	$(D+1)K$		

multi-dimensional sparse recovery was reduced to several one-dimensional sub-problems and where, for a fixed number of Diracs, the minimum required sample size increases exponentially with respect to the dimension. We summarize sample complexities of the previous separate estimation and the new joint estimation approaches in Table I.

ii) An iterative algorithm (similar to the one in [5]) to solve the multi-dimensional FRI reconstruction[1] (Section II-C3). The algorithm makes it possible to apply the multi-dimensional FRI-based sparse recovery to practical problems, where the measurements are not

necessarily taken uniformly (see Section IV-A for a concrete example in radio astronomy).

Before proceeding, we point to literature on sparse recovery in high-dimensions. Prony's method [4], which is at the root of the FRI-based sparse recovery, also plays an essential role in high-resolution spectrum estimation. Various attempts have been made to generalize the approach to high dimensions such as matrix enhancement and matrix pencil (MEMP) [14] and ESPRIT-like subspace methods [15]–[18]. One drawback with MEMP is the necessity to have an additional pairing step in order to obtain the final multi-dimensional reconstruction. The sample complexity is quadratic in terms of signal sparsity in 2D [14]. In comparison, ESPRIT-like approaches exploit the shift-invariance property of several sub-matrices, which are jointly diagonalized, and achieve linear sample complexity [16], [18] (albeit the requirement of having at least 4 samples along each dimension). Recent efforts have been made to extend Prony's method to multi-dimensional settings. Notably [19] exploited a similar idea, which treats multi-dimensional Diracs as common roots of several polynomials. However, the sample complexity with the proposed approach therein scales exponentially in terms of the signal dimension [20]. Other sampling schemes with linear sample complexity have been proposed in [21], [22]. Finally, recent work on spectral compressed sensing (e.g., [23]) is not directly related with the multi-dimensional Dirac reconstruction. Instead, the focus is on recovering the Fourier spectrum of the sparse signal, which is partially sub-sampled on a grid: the sparse signal parameters are retrieved with standard spectrum estimation algorithm, such as MEMP [14], once the full spectrum has been recovered.

The rest of the paper is organized as follows. First, we briefly review both the methodology and reconstruction algorithm of the FRI-based continuous-domain sparse recovery framework in Section II-A. Then, the generic FRI reconstruction framework is adapted to address the two-dimensional sparse recovery in Section II-B. Sample complexity as well as the reconstruction algorithm details are discussed in Section II-C. The FRI-based 2D sparse recovery is further generalized to higher dimensions in Section II-D. Next, we validate the proposed approach with simulations in Section III. Finally, the significance of the new approach is highlighted with two applications on point source

[1]The Python code will be available at https://github.com/hanjiepan

reconstruction in radio astronomy and direction of arrival estimation with microphone arrays in Section IV.

II. METHODS

A. Continuous-domain Sparse Recovery with FRI Sampling

In this section, we review the general framework of FRI-based sparse recovery. First, we highlight the methodology that FRI adopts: a sparse signal is recovered by finding a complementary mask function that "kills" or annihilates the sparse signal (Section II-A1). Then, we overview a generic algorithm that recovers FRI signals robustly (Section II-A2). We will devise a similar algorithm to solve the multi-dimensional FRI recovery later in Section II-B.

1) Sensing Sparsity Through Annihilation: Finite rate of innovation (FRI) sampling is a framework to reconstruct continuous-domain sparse signals, which have finite degrees of freedom per unit time or space. A typical FRI signal is a τ-periodic stream of Diracs:

$$x(t) = \sum_{k' \in \mathbb{Z}} \sum_{k=1}^{K} \alpha_k \delta(t - t_k - k'\tau). \tag{1}$$

The goal is to estimate the Dirac locations t_k and amplitudes α_k from the measurements of the sparse signal $x(t)$, e.g., its ideally low-pass filtered samples. Instead of directly estimating where the sparse signal is different from zero, the FRI-based approach aims at reconstructing a signal-dependent smooth function $\mu(t)$ (typically a polynomial), which vanishes at the non-zero locations of the sparse signal (see Fig. 1). From this perspective, $\mu(t)$ serves as a mask that *annihilates* the sparse signal $x(t)$:

$$\mu(t)x(t) = 0. \tag{2}$$

The zero-crossings of the function $\mu(t)$ then give the Dirac locations. Once the Dirac locations have been reconstructed, estimating the amplitudes of the Diracs is a linear problem, namely a least square minimization.

In general, it is challenging to enforce the continuous-domain annihilation constraint (2) directly. However, many FRI signals are or can be transformed into a weighted sum of sinusoids, whose frequencies are related to the sparse signal parameters. Thanks to a result known for more than two centuries [4], the continuous-domain annihilation constraint can equivalently be enforced with a discrete convolution equation: there exists a finite-length discrete filter such that its convolution with the uniform sinusoidal samples are zero (hence the filter is also known as "annihilating filter"). In fact, the polynomial, whose coefficients are specified by the discrete filter (e.g., $\mu(t)$ in (3)), is the mask function that annihilates the sparse signal.

More concretely, take the periodic stream of Diracs (1) as an example. Its Fourier series coefficients are

$$\hat{x}_m = \frac{1}{\tau} \sum_{k=1}^{K} \alpha_k e^{-j\frac{2\pi m}{\tau}t_k},$$

which are annihilated (i.e., $c_m * \hat{x}_m = 0$) by a $(K+1)$-tap filter $[c_0, \cdots, c_{K+1}]$ with z-transform

$$C(z) = \sum_{k=0}^{K+1} c_k z^{-k} = c_0 \prod_{k=1}^{K} \left(1 - e^{-j\frac{2\pi}{\tau}t_k} z^{-1}\right).$$

Note that $C(z)$ is a polynomial of degree $K+1$ and the mask function $\mu(t) = C\left(e^{-j\frac{2\pi}{\tau}t}\right)$ vanishes precisely at the Dirac locations: $\mu(t_k) = 0$ for $k = 1, \cdots, K$:

$$\mu(t)x(t) \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad c_m * \hat{x}_m. \tag{3}$$

The ideally low-pass filtered samples of the sparse signal $x(t)$ have a one-to-one correspondence with its Fourier series coefficients $\hat{x}_m$. Consequently, the annihilating filter coefficients (or equivalently the mask function) are reconstructed by enforcing a discrete convolution equation. We illustrate the annihilating filter based sparse recovery in Fig. 1.

2) A generic FRI Reconstruction Algorithm: In [5], a generic FRI reconstruction problem was formulated as a constrained optimization. The fitting error of the re-synthesized measurements (based on the estimated sparse signal model) to the given measurements is minimized subject to the annihilation constraint:

$$\min_{\mathbf{c} \in \mathcal{C}, \mathbf{b}} \quad \|\mathbf{a} - \mathbf{Gb}\|_2^2 \tag{4}$$
$$\text{subject to} \quad \mathbf{c} * \mathbf{b} = \mathbf{0}.$$

Therefore, the goal is to approximate the measurements optimally (in the least square sense) given that the signal follows the sparsity model. Here

- $\mathbf{a}$ is the vector of the sparse signal measurements to be reconstructed, e.g., the low-pass filtered samples.
- $\mathbf{b}$ is the vector of the unknown uniform sinusoidal samples, e.g., the Fourier series coefficients of (1). The choice of $\mathbf{b}$ is problem-dependent. By imposing the discrete convolution equation in (4), the estimated signal is guaranteed to follow the sparse signal model.
- $\mathbf{G}$ is the exact or approximate linear mapping from the uniform sinusoidal samples $\mathbf{b}$ to the measurements $\mathbf{a}$, which typically has a full column rank. In fact, an inexact mapping is not detrimental to the final reconstruction quality [5], [7] although, in general, we should not expect to achieve perfect reconstruction with noiseless measurements, when an approximate linear mapping is used. However, see [6] for a strategy to obtain the exact solution by refining the linear mapping based on the previous reconstructions.
- $\mathbf{c}$ is the vector of the annihilating filter coefficients. The feasible set $\mathcal{C}$ specifies a proper normalization in order to avoid a trivial solution $\mathbf{c} \equiv \mathbf{0}$. One choice of the feasible set $\mathcal{C}$, which leads to robust performance (see [5]), is $\mathcal{C} = \{\mathbf{c} | \mathbf{c}_0^H \mathbf{c} = 1\}$, where $\mathbf{c}_0$ is a random initialization of the annihilating filter. The randomness allows the algorithm to explore different feasible sets over several random initializations.

Because of the annihilation constraint, (4) is non-convex with respect to the $(\mathbf{b}, \mathbf{c})$-pair. An iterative strategy in [5] finds a valid solution, which (i) satisfies the annihilation constraint; and (ii) has a fitting error $\|\mathbf{a} - \mathbf{Gb}\|_2^2$ within the noise level. This

strategy exploits the "bi-linearity" of the annihilation constraint: for a fixed $\mathbf{c}$ (respectively $\mathbf{b}$), the annihilation constraint is linear in terms of $\mathbf{b}$ (respectively $\mathbf{c}$). Hence, with a fixed $\mathbf{c}$, we have a closed-form solution $\mathbf{b}$ as a function of $\mathbf{c}$. Then, the original bivariate optimization (4) reduces to a minimization with respect to $\mathbf{c}$ alone. The simplified formulation inspired the iterative algorithm in [5].

Finally, it has been shown that the iterative algorithm can be efficiently implemented: at each iteration, the updates of $\mathbf{b}$ and $\mathbf{c}$ amount to solving two linear systems of equations. Readers are referred to [5] for detailed discussions. We will adopt a similar strategy to solve the multi-dimensional FRI reconstruction in Section II-C3.

B. FRI Reconstruction in 2D

In our previous approach [5], the multi-dimensional problem was reduced to several sub-problems by enforcing the annihilation constraints along each dimension separately (Section II-B1). We discuss the limitations associated with the separate estimation approach and propose an alternative formulation, where the annihilation constraints are enforced along all dimensions simultaneously (Section II-B2).

1) Separate Estimation Formulation: In [5], the multi-dimensional FRI reconstruction problem was recast as separate annihilation problems along each dimension. Take the reconstruction of K two-dimensional Diracs as an example. The Fourier transform on a uniform grid can be annihilated by two discrete filters of size $K+1$ along rows and columns, respectively:

$$\min_{\substack{\mathbf{c}_1 \in \mathcal{C}_1, \mathbf{b}}} \|\mathbf{a} - \mathbf{Gb}\|_2^2 \quad \text{and} \quad \min_{\substack{\mathbf{c}_2 \in \mathcal{C}_2, \mathbf{b}}} \|\mathbf{a} - \mathbf{Gb}\|_2^2 .$$
$$\text{subject to} \quad \mathbf{c}_1 * \mathbf{b} = \mathbf{0} \qquad \text{subject to} \quad \mathbf{c}_2 * \mathbf{b} = \mathbf{0}$$
$$(5)$$

Here
- $\mathbf{a}$ is the vector of the given measurements of the 2D Diracs, e.g., the ideally low-pass filtered samples.
- $\mathbf{b}$ is the vector of the (unknown) uniformly sampled Fourier transform, e.g., the vectorized Fourier series coefficients of the 2D Diracs in a column-by-column order.
- $\mathbf{G}$ is the linear transformation that links the sinusoidal samples $\mathbf{b}$ to the ideally low-pass filtered samples $\mathbf{a}$, e.g., the inverse DFT transformation.
- $\mathbf{c}_1$ and $\mathbf{c}_2$ are the annihilating filters along the horizontal and vertical directions, respectively. They each belong to a certain space, e.g., $\mathcal{C}_1 = \{\mathbf{c}_1 \in \mathbb{C}^{K+1} | (\mathbf{c}_1^{(0)})^{\mathrm{H}} \mathbf{c}_1 = 1\}$ where $\mathbf{c}_1^{(0)}$ is a random initialization of $\mathbf{c}_1$ in the iterative algorithm (see [5] for details). $\mathcal{C}_2$ is similarly defined.

The x-coordinates (respectively y-coordinates) are then given by the roots of a univariate polynomial whose coefficients are specified by $\mathbf{c}_1$ (respectively $\mathbf{c}_2$).

All Diracs that are located on the K^2 intersections of reconstructed x and y coordinates (see Fig. 2(a)), satisfy the annihilation constraints in both problems in (5). Therefore, it is necessary to find the correct correspondence among the possible x and y locations in order to reconstruct the 2D Diracs.

The exhaustive approach amounts to considering $\binom{K^2}{K}$ possible combinations, which becomes computationally infeasible with a large number of Diracs. In [5], we showed experimentally that we can identify the (x, y)-pairs by reconstructing amplitudes of Diracs, which are located at K^2 possible x and y intersections, all at once with least square minimization. Then, the coordinates of the K Diracs (among the K^2 possibilities) that have the largest amplitudes give the Dirac locations in 2D. Finally, we reconstruct the Dirac amplitudes with the K correctly identified 2D Dirac locations.

Despite the empirical success of such a simple strategy, there are two noticeable drawbacks:
  i) The restrictive shape of the 1D filters used in the separate estimation approach, requires that the uniform sinusoidal samples $\mathbf{b}$ should be at least $K+1$ along each dimension. Therefore, the total number of samples increases *quadratically* with respect to the number of Diracs. However, an optimal algorithm should only use $\mathcal{O}(K)$ number of measurements. More generally, in D dimensions, the sample complexity of the separate estimation approach is $\mathcal{O}(K^D)$, i.e., a penalty of $\mathcal{O}(K^{D-1})$ over the optimal linear sample complexity.
  ii) The two 1D FRI reconstructions in (5) are solved independently, without enforcing the estimated uniform sinusoidal samples $\mathbf{b}$ to be the same. But we know that there is always a feasible solution (i.e., the noiseless $\mathbf{b}$) that satisfies both annihilation constraints in the two 1D problems and fits the given measurements up to the noise level. Additionally, since $\mathbf{b}$ should be the same in both problems, any linear combination of the two filters $\mathbf{c}_1$ and $\mathbf{c}_2$ is also a valid solution. Therefore, instead of estimating two annihilating filters separately, what we should reconstruct is a set of (two) linearly independent vectors that spans the null space of the convolution matrix associated with $\mathbf{b}$.

Because of the constraints above, we now propose a new joint estimation approach.

2) Joint Estimation Formulation: Instead of solving two 1D Dirac estimations separately, a more natural choice for the 2D FRI reconstruction is:

$$\min_{\substack{\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}, \\ \mathbf{b}}} \|\mathbf{a} - \mathbf{Gb}\|_2^2$$
$$\text{subject to} \quad \mathbf{c}_1 * \mathbf{b} = \mathbf{0}$$
$$\mathbf{c}_2 * \mathbf{b} = \mathbf{0}.$$
$$(6)$$

The goal is to estimate a set of uniform sinusoidal samples $\mathbf{b}$, which fits the noisy measurements (up to the noise level) and two annihilating filters $\mathbf{c}_1, \mathbf{c}_2$ that form bases of the null space of the convolution matrix specified by $\mathbf{b}$. Note that here $\mathbf{c}_1$ and $\mathbf{c}_2$ are, in general, 2D non-separable filters. Here, $\mathcal{C}$ specifies a few[2] linear constraints such that the solution of (6) corresponds to one (out of many possible) set of annihilating filters that spans the null space of the convolution matrix specified by $\mathbf{b}$. One possible choice is $\mathcal{C} = \{\mathbf{C} | \mathbf{C}_0^{\mathrm{T}} \mathbf{C} + \mathbf{C}^{\mathrm{T}} \mathbf{C}_0 = 2\mathbf{I}\}$, where $\mathbf{C}$ and

---

[2]In general, $(D+1)D/2$ constraints are needed to uniquely specify a set of orthogonal vectors that spans a vector space of dimension D.

$\mathbf{C}_0$ are 2-column matrices, with columns specified by the two annihilating filters $\mathbf{c}_1$, $\mathbf{c}_2$ and their corresponding initializations $\mathbf{c}_1^{(0)}$, $\mathbf{c}_2^{(0)}$, respectively.

A necessary condition in order to satisfy the annihilation equations is that the DTFT of each of the two annihilating filters $\mathbf{c}_1$ and $\mathbf{c}_2$

$$\mu(x, y) = \sum_{k,l} c_{k,l} (e^{-j2\pi x/\tau_1})^k (e^{-j2\pi y/\tau_2})^l, \qquad (7)$$

should vanish at the Dirac locations (x_k, y_k) for $k = 1, \cdots, K$:

$$\mu(x_k, y_k) = 0.$$

Here $c_{k,l} = [\mathbf{c}_i]_{k,l}$ for $i = 1$ or 2 and τ_1, τ_2 are the periods of the 2D Dirac stream along x and y directions, respectively.

The corresponding FRI signal (i.e., the zero-crossings of a bivariate polynomial (7) with respect to $e^{-j2\pi x/\tau_1}$ and $e^{-j2\pi y/\tau_2}$) is a curve [10]. Any Dirac located on this curve satisfies the annihilation constraints in (6). The 2D Dirac locations are then obtained from the intersections of the two curves (Fig. 2).

A straight-forward way to determine the intersections of curves, is to evaluate both bivariate polynomials (with coefficients specified by $\mathbf{c}_1$ and $\mathbf{c}_2$) on a grid: with a finer grid, more accurate Dirac locations can be determined. Alternatively, we will discuss in details an algebraic approach in Section II-C2, which treats the curve intersections as the common roots of two polynomials. This approach allows us to determine the curve intersections reliably without resorting to grid search.

C. Algorithm and Implementation

We give algorithmic details related to the joint estimation introduced in the previous section. We first derive the sample complexity with respect to the number of Diracs in Section II-C1. Next, an algebraic approach is outlined in Section II-C2, which finds curve intersections based on the Bézout resultant of a polynomial system. Finally, an iterative algorithm to solve the joint annihilation problem (6) is presented in Section II-C3.

1) Sample Complexity of 2D Dirac Estimation: Conventionally, a $(K+1)$-tap annihilating filter is used in order to recover K Diracs in 1D [3]. In higher dimensions, e.g., 2D, it is not always possible to have a general 2D filter with a size matched to the number of Diracs K (except in the degenerated cases where the 2D filter reduces to a 1D filter). We then need to specify the support size (i.e., the number of non-zero entries) for a given choice of a 2D filter shape.

Proposition 1 (Filter support size): Suppose the two annihilating filters in (6) have the same support, then the minimum support size is $K + 2$ for the estimation of K Diracs.

Proof: See Appendix A. ∎

Because the annihilating filter is scale-invariant with respect to any non-zero scalar, each filter has one degree of freedom less than the total number of non-zero elements. Consequently, at least $2(K + 1)$ annihilation equations are needed to reconstruct K Diracs.

In a typical setup, an equal number of samples is available along each direction. Under this assumption, a filter shape, which is as square as possible, e.g., $\lceil \sqrt{K+2} \rceil \times \lceil \sqrt{K+2} \rceil$,
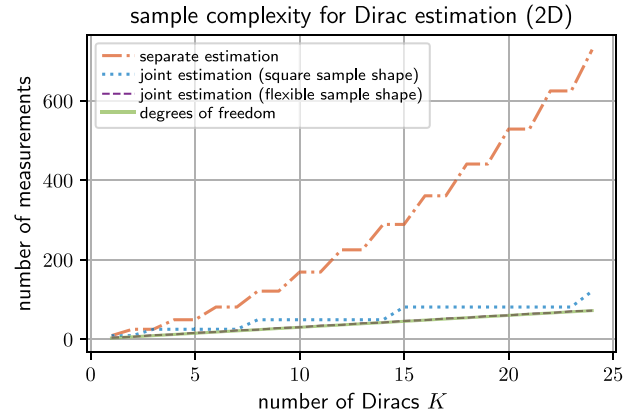

Fig. 3. Sample complexities (ideal low-pass filtering with equal sizes along each dimension) of the previous separate and the proposed joint estimation approaches (see (5) and (6), respectively). The joint estimation approach is able to estimate K Diracs from $3K$ measurements if the sample sizes are not restricted to be the same along each dimension (see Corollary 1 and Corollary 2).

allows the maximum number of annihilation equations to be built. Since the ideally low-pass filtered samples have a direct correspondence with the uniform sinusoidal samples, we can derive the sample complexity by comparing the total number of equations (from the annihilation constraint and the normalization of the filter coefficients $\mathcal{C}_1$ and $\mathcal{C}_2$) with the unknowns (i.e., the filter coefficients).

Corollary 1: Assume that the ideally low-pass filtered sample size is $M \times M$, where M is an odd number[3]. In order to reconstruct K 2D Diracs, the sample size should satisfy

$$M \geq \sqrt{K+2} + \sqrt{K} - 1. \qquad (8)$$

Proof: See Appendix B. ∎

From (8), the minimum number of samples M^2 has a *linear* complexity $\mathcal{O}(K)$ with respect to the number of Diracs. In comparison, the sample complexity is $\mathcal{O}(K^2)$ in the previous approach, where the annihilation constraints were enforced separately along each direction (see Section II-B1). For K 2D Diracs, the degrees of freedom are $3K$: $2K$ for the Dirac locations and K for the amplitudes. The total degrees of freedom give the lowest possible sample complexity of *any* reconstruction algorithm. It is possible to reach this lower bound with the proposed joint estimation if the sample sizes are not required to be the same along each dimension. We summarize the result with a simplified setup, where uniformly sampled Fourier transform of the Diracs are directly available.

Corollary 2: For simplicity, let the number of Diracs K be an even number, then K 2D Diracs can be reconstructed from $2 \times (K/2 + K)$ uniform Fourier transform samples of the Diracs.

Proof: The proof follows similarly to that of Corollary 1 by choosing annihilating filters of size $2 \times (K/2 + 1)$. ∎

We show the sample complexity with the separate and the proposed joint estimation approaches against the total degrees of freedom in Fig. 3.

---

[3]The requirement that M should be an odd number is a technicality of the ideal low-pass filtering, which has symmetric cut-off frequencies. For cases, where direct Fourier domain measurements are available, e.g., in array signal processing, then the number of measurements is not required to be odd.

The merits of an algorithm with lower sample complexity are two-fold: Not only can more Diracs be reconstructed for a given sample size, but also a more robust reconstruction may be expected—Because of more efficient data utilization, more annihilation equations are available to further constrain the reconstruction. We demonstrate both improvements with simulations in Section III-A and Section III-B.

Finally, we point out that the proposed joint annihilation approach can also cope with the singular cases where some Diracs have the same x or y coordinates. In these cases, the minimum size of the annihilating filter is further constrained by the number of Diracs that have common x or y locations. Please refer to Appendix C for detailed discussions and Section III-C for an example.

2) Polynomial Common Root Finding: Instead of resorting to grid search for finding the curve intersections, an alternative approach is to re-cast the geometric problem as finding the common roots of two polynomials. Since the mask function (7) is a bivariate polynomial with respect to $u = e^{-j2\pi x/\tau_1}$ and $v = e^{-j2\pi y/\tau_2}$, the curve intersections are the solutions of:

$$\begin{cases} 0 = \mu_1(u,v) = \sum_{k,l}[\mathbf{c}_1]_{k,l}u^k v^l, \\ 0 = \mu_2(u,v) = \sum_{k,l}[\mathbf{c}_2]_{k,l}u^k v^l. \end{cases} \quad (9)$$

One way to determine the polynomial common roots is to compute the Bézout resultant [24], [25] of the polynomial system of equations. It is based on the following two observations:

(i) The bivariate polynomials in (9) can be rearranged with respect to one variable, e.g., u, where the polynomial coefficients are functions of the other variable, e.g., v:

$$\mu_1(u,v) = \sum_{k=0}^{K_1-1} p_k(v)u^k \quad \text{and} \quad \mu_2(u,v) = \sum_{k=0}^{K_2-1} q_k(v)u^k.$$

Here $p_k(\cdot)$ and $q_k(\cdot)$ are some polynomials of v, and the filters $\mathbf{c}_1$ and $\mathbf{c}_2$ in (9) are of size $L_1 \times K_1$ and $L_2 \times K_2$, respectively.

(ii) For any $n \geq 0$, we have

$$u^n \mu_1(u,v) = u^n \mu_2(u,v) = 0.$$

From (i) and (ii), we can build a linear system of equations in terms of a vector $\left[u^0, \cdots, u^{K_1+K_2-3}\right]^{\mathrm{T}}$:

$$\underbrace{\begin{bmatrix} p_0 & \cdots & p_{K_1-1} & 0 & \cdots \\ 0 & p_0 & \cdots & p_{K_1-1} & \ddots \\ \vdots & \ddots & \ddots & & \ddots \\ q_0 & \cdots & q_{K_2-1} & 0 & \cdots \\ 0 & q_0 & \cdots & q_{K_2-1} & \ddots \\ \vdots & \ddots & \ddots & & \ddots \end{bmatrix}}_{\mathbf{A}(v)} \begin{bmatrix} u^0 \\ \vdots \\ u^{K_1+K_2-3} \end{bmatrix} = \mathbf{0}. \quad (10)$$

A necessary and sufficient condition for the polynomial systems (9) to have common roots is that the Bézout resultant should vanish (see [26] Theorem 7):

$$\det(\mathbf{A}(v)) = 0. \quad (11)$$

Since (11) is a polynomial equation of v with constant coefficients, the common roots $v_k \overset{\text{def}}{=} e^{-j\frac{2\pi}{\tau_2}y_k}$ are then given by the roots of this univariate polynomial. Once we have computed v_k, $u_k \overset{\text{def}}{=} e^{-j\frac{2\pi}{\tau_1}x_k}$ is obtained from (9) with $v = v_k$. Unlike the separate estimation cases, an extra pairing step (see text in Section II-B1) is no longer needed: indeed, from the common-root-finding procedure, the Dirac locations in 2D are directly reconstructed with the correct association.

From (10), the degree of the polynomial in (11) is:

$$(K_2 - 1)\max_k(\deg(p_k(v))) + (K_1 - 1)\max_k(\deg(q_k(v))). \quad (12)$$

Consequently, there are at least[4] a number of common roots given by (12). In general, this may be larger than the number of Diracs K. The reason is that in the root-finding procedure, we proceed as if the common roots were arbitrary complex numbers. However, since our goal is to reconstruct multi-dimensional Diracs (which have real-valued coordinates, i.e., $x_k, y_k \in \mathbb{R}$), the roots are necessarily located on the unit circle[5]. One way to eliminate invalid solutions for the Dirac reconstruction (i.e., the u_k, v_k that leads to complex-valued x_k, y_k) is based on the moduli of the common roots. However, it is not always obvious how to choose an adequate threshold level on the moduli of the roots in noisy scenarios. Alternatively, an approach that works empirically is as follows: all roots are first projected onto the unit circle $u_k/|u_k|$, $v_k/|v_k|$, from which the Dirac locations are reconstructed. Then, amplitudes for all Diracs are estimated by minimizing the discrepancies between the re-synthesized and given measurements in the least square sense. The valid Dirac locations then correspond to the ones that have the K largest amplitudes. We then solve the least square minimization once more with the extracted K Dirac locations (see [5] for a similar treatment in the separate annihilation of 2D Dirac cases).

3) Iterative Algorithm for the Joint Annihilation: Following [5], let $\mathbf{T}(\cdot)$ be the operator that builds the Toeplitz matrix from the input data and the associated right dual $\mathbf{R}(\cdot)$ such that $\mathbf{R}(\mathbf{c})\mathbf{b} = \mathbf{T}(\mathbf{b})\mathbf{c}, \forall \mathbf{c}, \mathbf{b}$. Then the joint annihilation constraints in (6) amount to vertically stacking two right dual matrices:

$$\begin{bmatrix} \mathbf{R}(\mathbf{c}_1) \\ \mathbf{R}(\mathbf{c}_2) \end{bmatrix} \mathbf{b} = \mathbf{0}. \quad (13)$$

For a set of uniform sinusoidal samples $\mathbf{b}$ with a sufficiently large size (compared with that of the filters), it may appear that we could have *more* equations than unknowns. For instance, if the shape of both filters $\mathbf{c}_1$ and $\mathbf{c}_2$ is 2×2 and that of the sinusoidal samples $\mathbf{b}$ is 5×5, then the total number of annihilation equations is 32 (i.e., $4 \cdot 4 + 4 \cdot 4$), which is larger than the size of $\mathbf{b}$ (i.e., 25 in this example). If we solve (6) with the annihilation constraint (13) for any fixed $\mathbf{c}_1$ and $\mathbf{c}_2$ different from zero, then

---

[4]There could be multiple roots for u with a given $v = v_k$.

[5]With the strategy used here, any damping factors are eliminated, which could be important in another related (but different) spectrum estimation problem [2]. For dampened sinusoid estimation, all the common roots, which are not required to have unitary norm, are valid solutions. Therefore, we do not need to follow the root-eliminating procedure here, which is specific to the Dirac estimation. Further investigation is required to apply the proposed approach to general spectrum estimation in future work.

it appears that $\mathbf{b}$ would be forced to be identically zero. But as we will show next, the vertically stacked convolution matrices in (13) do not have full rank.

Corollary 3: Define the joint annihilation matrix as

$$\mathbf{R}_{\text{joint}} \overset{\text{def}}{=} \begin{bmatrix} \mathbf{R}(\mathbf{c}_1) \\ \mathbf{R}(\mathbf{c}_2) \end{bmatrix},$$

then the null space dimension of $\mathbf{R}_{\text{joint}}$ is at least $2(K_0 - 1)(L_0 - 1)$ for filters $\mathbf{c}_1$ and $\mathbf{c}_2$ of size $L_0 \times K_0$.

Proof: See Appendix D. ∎

In terms of implementation, this requires the extraction of the independent rows of $\mathbf{R}_{\text{joint}}$ when (6) is minimized with respect to $\mathbf{b}$ (for a certain choice of $\mathbf{c}_1$ and $\mathbf{c}_2$) at each iteration. One possibility is to resort to the QR-decomposition[6] of the joint annihilation matrix. Let $\mathbf{R}_{\text{joint}} = \mathbf{Q}\mathbf{U}$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{U}$ is an upper triangle matrix, and $\mathbf{Q}_{\text{sub}}$ extracts all but the last $2(K_0 - 1)(L_0 - 1)$ columns of $\mathbf{Q}$. Then, the joint annihilation matrix $\mathbf{R}_{\text{joint}}$ is replaced with $\mathbf{Q}_{\text{sub}}^{\text{H}}\mathbf{R}_{\text{joint}}$, which is an orthogonal projection onto the row space of $\mathbf{R}_{\text{joint}}$.

Similarly, we can express the two *linearly independent* annihilating filters with their *effective* degrees of freedom, which are less than their total sizes. Specifically, the iterative algorithm (see details in Appendix E) involves the convolution between $\boldsymbol{\beta} \overset{\text{def}}{=} (\mathbf{G}^{\text{H}}\mathbf{G})^{-1}\mathbf{G}^{\text{H}}\mathbf{a}$ and the two annihilating filters. Let the QR-decomposition of $\mathbf{T}(\boldsymbol{\beta})^{\text{H}}$ be $\widetilde{\mathbf{Q}}\widetilde{\mathbf{U}}$, then the convolutions can be equivalently written as:

$$\begin{cases} \mathbf{T}(\boldsymbol{\beta})\mathbf{c}_1 = \mathbf{0} \\ \mathbf{T}(\boldsymbol{\beta})\mathbf{c}_2 = \mathbf{0} \end{cases} \Leftrightarrow \begin{cases} \mathbf{T}(\boldsymbol{\beta})\widetilde{\mathbf{Q}}_1\boldsymbol{\gamma}_1 = \mathbf{0} \\ \mathbf{T}(\boldsymbol{\beta})\widetilde{\mathbf{Q}}_2\boldsymbol{\gamma}_2 = \mathbf{0} \end{cases},$$

for two linearly independent vectors[7] $\mathbf{c}_1 = \widetilde{\mathbf{Q}}_1\boldsymbol{\gamma}_1$ and $\mathbf{c}_2 = \widetilde{\mathbf{Q}}_2\boldsymbol{\gamma}_2$. Here $\widetilde{\mathbf{Q}}_1 = \widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{Q}}_2$ is a sub-matrix of $\widetilde{\mathbf{Q}}$ that consists of all but one column of $\widetilde{\mathbf{Q}}$. In the algorithm, only $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are reconstructed.

An equivalent formulation of (6) (expressed in terms of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$) is derived in Appendix E, based on which an iterative algorithm (similar to that in [5]) is proposed. For numerical stability, we may introduce auxiliary variables such that no nested matrix inverses are involved at each iteration. It can be shown[8] that at each iteration, the updated $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are solutions of an extended linear system of equations:

$$\begin{bmatrix} \mathbf{0} & \mathbf{T}_{\text{joint}}(\boldsymbol{\beta})^{\text{H}}\mathbf{Q}_{\text{sub}} & \mathbf{0} & \boldsymbol{\Gamma}_0^* \\ \mathbf{Q}_{\text{sub}}^{\text{H}}\mathbf{T}_{\text{joint}}(\boldsymbol{\beta}) & \mathbf{0} & -\mathbf{Q}_{\text{sub}}^{\text{H}}\mathbf{R}_{\text{joint}} & \mathbf{0} \\ \mathbf{0} & -\mathbf{R}_{\text{joint}}^{\text{H}}\mathbf{Q}_{\text{sub}} & \mathbf{G}^{\text{H}}\mathbf{G} & \mathbf{0} \\ \boldsymbol{\Gamma}_0^{\text{T}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\ell} \\ \mathbf{v} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\rho} \end{bmatrix},$$

(14)

where $\boldsymbol{\ell}$, $\mathbf{v}$, $\boldsymbol{\lambda}$ are auxiliary variables, $\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}_1^{\text{T}}, \boldsymbol{\gamma}_2^{\text{T}} \end{bmatrix}^{\text{T}}$ and the constant vector $\boldsymbol{\rho} = [1, 1, 0]^{\text{T}}$. Here $\boldsymbol{\Gamma}_0$ is a 3-column matrix

---

[6]Alternatively, we can use the singular value decomposition, which is computationally more demanding.

[7]In D dimensions, D linearly independent annihilating filters are represented as $\mathbf{c}_i = \widetilde{\mathbf{Q}}_i\boldsymbol{\gamma}_i$ for $i = 1, \cdots, D$, where $\widetilde{\mathbf{Q}}_i$ is a sub-matrix built from all but $i - 1$ columns of $\widetilde{\mathbf{Q}}$.

[8]We omit the derivations for brevity, which are very similar to the ones in Appendix A and B of [5].

Algorithm 1: Joint estimation of 2D Diracs

Input : Measurements of the 2D Diracs $\mathbf{a}$, the linear transformation matrix $\mathbf{G}$, (*optional*) noise level ε.

Output: Dirac locations r and amplitudes $\boldsymbol{\alpha}$.

if ε *not given* **then**
  | $\varepsilon = 0$
end

for *loop* $\leftarrow 1$ **to** *max. initializations* **do**

1   Initialize $\mathbf{c}_1$ with $\mathbf{c}_1^{(0)}$, $\mathbf{c}_2$ with $\mathbf{c}_2^{(0)}$;
     for $n \leftarrow 1$ **to** *max. iterations* **do**

2   Build $\mathbf{R}_{\text{joint}}$ from $\mathbf{c}_1^{(n-1)}$ and $\mathbf{c}_2^{(n-1)}$ and computes its QR-decomposition;

3   Build the linear system (14) and update $\mathbf{c}_1^{(n)}$ and $\mathbf{c}_2^{(n)}$ as $\widetilde{\mathbf{Q}}_1\boldsymbol{\gamma}_1$ and $\widetilde{\mathbf{Q}}_2\boldsymbol{\gamma}_2$;

4   Build the linear system (15) and update $\mathbf{b}^{(n)}$ from its solution;

       if $\left\| \mathbf{a} - \mathbf{G}\mathbf{b}^{(n)} \right\|_2^2 \leq \varepsilon^2$ **then**
5 | Terminate both loops;
       end
     end
end

6 $\mathbf{b} \leftarrow \mathbf{b}^{(n)}$, $\mathbf{c}_1 \leftarrow \mathbf{c}_1^{(n)}$ and $\mathbf{c}_2 \leftarrow \mathbf{c}_2^{(n)}$;
7 Dirac locations $r \leftarrow$ common roots of $\mathbf{c}_1$ and $\mathbf{c}_2$;
8 Dirac amplitudes $\boldsymbol{\alpha} \leftarrow$ least square fitting.

---

specified by initializations of the two annihilating filters:

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix} \widetilde{\mathbf{Q}}_1^{\text{T}}\mathbf{c}_1^{(0)} & \mathbf{0} & \widetilde{\mathbf{Q}}_1^{\text{T}}\mathbf{c}_2^{(0)} \\ \mathbf{0} & \widetilde{\mathbf{Q}}_2^{\text{T}}\mathbf{c}_2^{(0)} & \widetilde{\mathbf{Q}}_2^{\text{T}}\mathbf{c}_1^{(0)} \end{bmatrix},$$

and

$$\mathbf{T}_{\text{joint}}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{T}(\boldsymbol{\beta})\widetilde{\mathbf{Q}}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}(\boldsymbol{\beta})\widetilde{\mathbf{Q}}_2 \end{bmatrix}.$$

The uniform sinusoidal samples $\mathbf{b}$ are updated as the solution of

$$\begin{bmatrix} \mathbf{G}^{\text{H}}\mathbf{G} & \mathbf{R}_{\text{joint}}^{\text{H}}\mathbf{Q}_{\text{sub}} \\ \mathbf{Q}_{\text{sub}}^{\text{H}}\mathbf{R}_{\text{joint}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\ell} \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{\text{H}}\mathbf{a} \\ \mathbf{0} \end{bmatrix}, \quad (15)$$

where $\boldsymbol{\ell}$ is an auxiliary variable.

We emphasize that we do not aim at finding a global optimum of (6). Instead, we only use the iterative algorithm to find a valid solution $(\mathbf{c}_1, \mathbf{c}_2, \mathbf{b})$, which satisfies the annihilation constraints and has a fitting error $\|\mathbf{a} - \mathbf{G}\mathbf{b}\|_2^2$ within the noise level (see Section II-A2 for the same strategy used in the generic FRI reconstruction [5]). We claim that any solution that satisfies both criteria is equivalent up to the uncertainty due to noise. Since any intermediate solution at each iteration already satisfies the annihilation constraints (because of (15)), we can terminate the iteration as soon as the objective function value is below the noise level.

We summarize the joint estimation algorithm in Algorithm 1. The computational complexity is determined by that of solving linear systems of equations (14) and (15), which is cubic [27]

with respect to the size of the uniform sinusoidal samples $\mathbf{b}$. In the joint estimation, the size of $\mathbf{b}$ scales linearly with respect to the number of Diracs (compared with the quadratic scaling in the separate estimation (5)). Therefore, we also improve the computational complexity over the previous separate estimation approach.

D. Generalization to Higher Dimensions

In the previous two sections, we detailed the problem formulation and reconstruction algorithm for the joint annihilation problem in 2D. The approach can be generalized to higher dimensions following the same line of reasoning. Specifically, for Diracs reconstruction in D dimensions, the joint estimation amounts to solving

$$
\begin{array}{cl}
\min\limits_{\substack{\mathbf{c}_1,\cdots,\mathbf{c}_D \in \mathcal{C}, \\ \mathbf{b}}} & \|\mathbf{a} - \mathbf{G}\mathbf{b}\|_2^2 \\
\text{subject to} & \mathbf{c}_i * \mathbf{b} = \mathbf{0}, \quad \text{for } i = 1,\cdots,D,
\end{array}
\tag{16}
$$

where $\mathcal{C}$ is similarly defined as in the 2D cases. The D-dimensional Dirac locations are the common roots of D polynomials whose coefficients are specified by the annihilating filters $\mathbf{c}_i$ for $i = 1,\cdots,D$. Geometrically, the Dirac locations are the intersections of D hypersurfaces, which are the zero-crossings of these D-variate polynomials.

One way to compute the common roots of multiple polynomials is to collect all polynomials with respect to one variable and compute the Bézout resultants of polynomial pairs, which themselves are polynomials of the other variables. The elimination process is then repeated by computing the resultant of resultants (which depends on one less variable) until the resultant reduces to an univariate polynomial, whose roots can be computed directly.

Concretely, take the 3D case as an example. Let $\mu_1(x,y,z)$, $\mu_2(x,y,z)$, $\mu_3(x,y,z)$ be the polynomials specified by the annihilating filters and $\text{res}_s(\cdot,\cdot)$ be the resultant of two polynomials with respect to s. Then $\text{res}_z(\mu_1,\mu_2)$ and $\text{res}_z(\mu_2,\mu_3)$ are bivariate polynomials of x and y. The common roots in x are given by the roots of the univariate polynomial $\text{res}_y\big(\text{res}_z(\mu_1,\mu_2),\text{res}_z(\mu_2,\mu_3)\big)$. Such a recursive approach may not be numerically stable, especially for polynomials of higher degree (i.e., more Diracs K). Further work is needed to find alternatives in determining the common roots of multiple polynomials. One possible idea could be based on [28], where the canonical polynomial basis is replaced with Chebyshev polynomials.

For simplicity, if we assume that the sample size is the same along each dimension, then by choosing annihilating filters of size[9] $\lceil (K+D)^{1/D} \rceil$ along each dimension, the maximum number of annihilation equations can be built from the given samples. The sample complexity compared with the number of Diracs is summarized as follows.
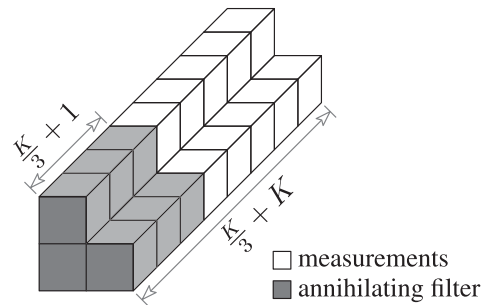

Fig. 4. Illustration of sample and annihilating filter shapes in the critical sampling cases in 3D (see Corollary 5).

Corollary 4: Assuming that the ideally low-pass filtered sample size is $\underbrace{M \times \cdots \times M}_{D \text{ terms}}$, then M, which is an odd number, should satisfy

$$
M \geq (K+D)^{1/D} + K^{1/D} - 1,
\tag{17}
$$

in order to reconstruct K D-dimensional Diracs.

The minimum total number of required samples M^D is *linear* in terms of the number of Diracs K. For $K \gg D$, the sample complexity can be approximated as $M^D \geq 2^D K$. Even though the sample complexity is still proportional to the number of Diracs, the proportion factor grows exponentially with respect to the dimension. In comparison, the total degrees of freedom of K D-dimensional Diracs are[10] $(D+1)K$. The gap is closed as soon as sample sizes are not required to be the same along all dimensions. We summarize the result as follows.

Corollary 5: For simplicity, let the number of Diracs K be an integer multiple of the dimension D, then K D-dimensional Diracs can be reconstructed from $(D+1)K$ uniformly sampled Fourier transform of the Diracs, which are supported at indices $\mathbf{k} = (k_1,\cdots,k_D)$ such that

- $k_1 = 0,\cdots,K/D + K - 1$;
- $k_j = 0,1$ for $j \neq 1$ and $\sum_{j=2}^{D} k_j \leq 1$ (i.e., at most one j such that k_j can be non-zero).

Proof: The proof is straightforward by choosing a filter that has the same support with the Fourier samples k_j for $j = 2,\cdots,D$ and $k_1 = 0,\cdots,K/D$. ∎

Pictorially, the samples and filters consist of D parallel "lines" along dimension k_1. We illustrate this with an example in 3D (Fig. 4). Further investigation is required to design a sampling scheme such that these uniform Fourier transforms are accessible from a given set of spatial domain samples of the same size.

The joint estimation approach in higher dimensions is exemplified with a 3D Dirac reconstruction example with noiseless and noisy measurements in Section III-D.

---

[9]This is because the filter support size is $K+D$ in D dimensional Dirac reconstruction. See Appendix A for details.

[10]More precisely, it is the complex exponentials that are estimated from the FRI reconstruction instead of the Dirac locations directly, which are real-valued. It may be possible to enforce the Dirac locations to be real-valued in the reconstruction by using e.g., Hermitian symmetry.
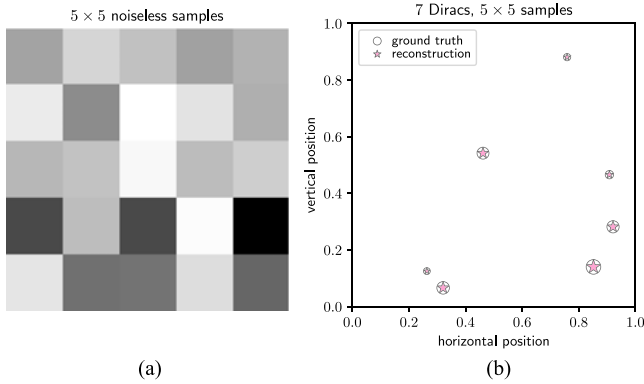
Fig. 5.   Exact reconstruction of 7 Diracs on a 2D plane from 5×5 ideally low-pass filtered samples. (a) The noiseless ideally low-pass filtered samples. (b) The exact reconstruction with the proposed joint estimation. The joint annihilation approach utilizes the given measurements efficiently. In comparison, the separate annihilation approach (5) fails to reconstruct any Dirac, because it requires a minimum sample size 8×8 in this case.
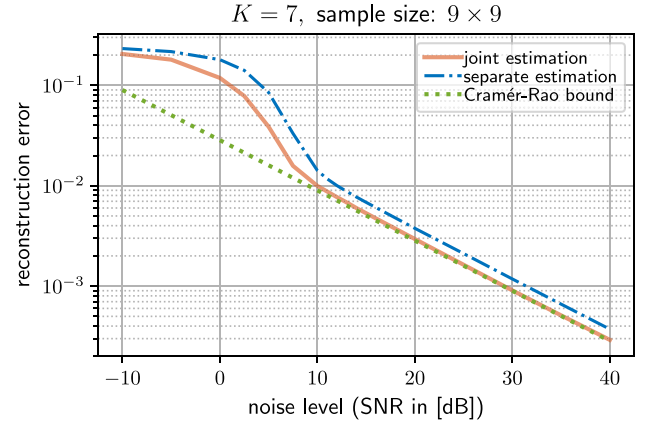

Fig. 6.   Average reconstruction error of 2D Diracs against different noise levels with the joint and separate estimation approaches (see (6) and (5), respectively). The joint estimation approache reaches the Cramér-Rao lower bound down to around 10 dB. The reconstruction error at each noise level is averaged over 3000 different noise realizations (number of Diracs: 7, ideally low-pass filtered sample size: 9×9).

III. SIMULATION RESULTS

In this section, we validate the proposed joint annihilation algorithm with synthetic experiments for the reconstruction of multi-dimensional Diracs. Diracs are ideally low-pass filtered and uniformly sampled with equal sample sizes along each dimension. We first demonstrate the improvements of the joint estimation approach in 2D (Section III-A to Section III-C). The generality of the proposed method is further exemplified with a 3D simulation in Section III-D.

In all simulations, 50 random initializations and 20 maximum iterations per initialization are used in the joint estimation algorithm (Algorithm 1).

A. Efficient Data Utilization

In our previous approach, where the annihilation constraints were enforced separately, the sample size along each dimension should at least be the size of the 1D annihilating filter. The restrictive filter shape limits the applicability of the separate estimation approach. Such an artificial constraint is lifted with the new approach: as long as the sample size satisfies (8) in 2D or more generally (17) in D dimensions, Diracs can be successfully estimated.

We demonstrate the efficiency of the joint estimation approach with a periodic stream of Diracs in 2D. Within each period $[0, 1) \times [0, 1)$, 7 Diracs of random amplitudes and locations are ideally low-pass filtered and uniformly sampled. Even with sample size as small as 5×5, the joint annihilation approach manages to reconstruct Diracs exactly (Fig. 5). Both annihilating filters used here are of size 3×3. In comparison, the separate estimation (5) has to use two 1D filters of length 8 and hence requires a minimum sample size 8×8. Because of the restriction on the filter shapes, the previous separate approach fails to reconstruct any Dirac from such a limited number of samples.

B. Robust Reconstruction

Another improvement, as a consequence of the flexible annihilating filter shapes, is that more annihilation equations can be built from the given measurements. These extra equations further constrain the reconstruction, leading to a more robust estimation in the presence of noise.

To quantify these improvements experimentally, we consider a periodic stream of Diracs, which consists of 7 Diracs with random locations and amplitudes within each period. The ideally low-pass filtered and uniformly sampled measurements (sample size: 9×9) are contaminated with Gaussian white noise such that the signal to noise ratio (SNR) varies from -10 dB to 40 dB. We measure the reconstruction quality with the average distance[11] between the estimated and ground truth Dirac locations. The performances of both the joint and separate estimation approaches are averaged over 3000 different noise realizations at each noise level. The joint estimation approach reaches the Cramér-Rao lower bound (see e.g., [29]) down to around 10 dB. In comparison, the separate estimation does not reach the lower bound even in very high SNR cases (Fig. 6).

C. Dirac With Common x/y Coordinates

In specific cases where some Diracs share common x or y coordinates, the annihilating filter shapes are further constrained as a consequence of the fundamental theorem of algebra (see details in Appendix C). Similar to examples in the previous two sections, we consider 2D periodic stream of Diracs with unit period. Within each period, 5 Diracs of unitary amplitudes are simulated such that they are arranged in a dagger-shape (the right column in Fig. 7). From our analysis in Section II-C1, the minimum annihilating filter shape is $\lceil \sqrt{7} \rceil \times \lceil \sqrt{7} \rceil$ (i.e., 3×3)

---

[11]The pairing between the reconstruction and the ground truth is identified by permuting the Dirac locations such that the distance between the permuted reconstruction and the ground truth Dirac locations is minimized.
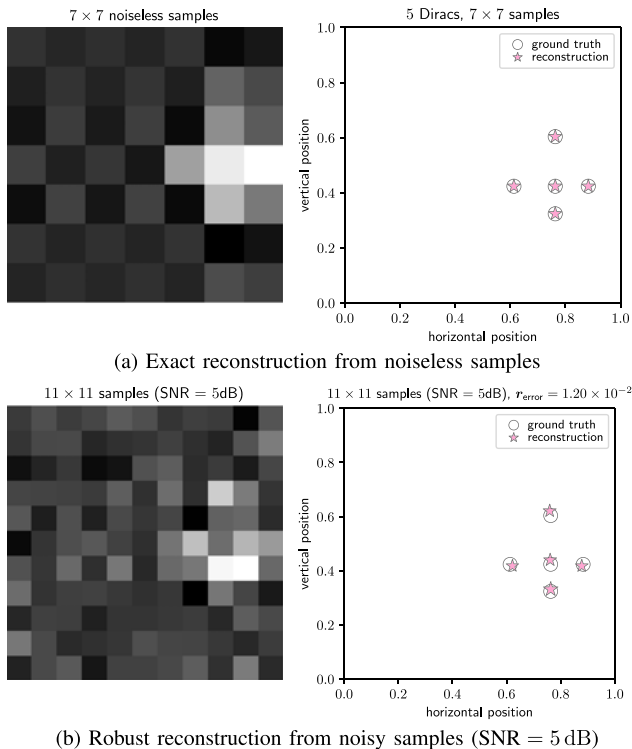
(a) Exact reconstruction from noiseless samples

(b) Robust reconstruction from noisy samples (SNR = 5 dB)

Fig. 7. Reconstruction of 2D Diracs, some of which have common x or y coordinates, with the joint estimation (6). (a) The exact reconstruction from the noiseless ideally low-pass filtered samples (sample size: 7×7). (b) Robust estimation of 2D Diracs from noisy samples (sample size: 11×11, SNR = 5 dB, reconstruction error on the Dirac locations: 1.2×10^{-2}).

in the generic cases where no Diracs share common coordinates along any direction. However, in this specific case, there could be up to 3 Diracs that share the same x (or y) coordinates. From our analysis in Appendix C, the annihilating filter shape, therefore, should be at least 4×4. The joint estimation of the 2D Diracs from 7×7 noiseless ideally low-pass filtered samples, are shown in Fig. 7(a). Additionally, we also considered the case with noisy samples (size: 11×11). Gaussian white noise is added to the noiseless samples such that the signal-to-noise ratio is 5 dB. The reconstruction error on the Dirac locations is 1.2×10^{-2}.

D. Dirac Reconstruction in 3D

To demonstrate the generality of the joint estimation approach to higher dimensions (see Section II-D), we show an example of a periodic stream of Diracs in 3D. Within each period (1 along each dimension), 5 Diracs with random locations and amplitudes are simulated.

In the first case, the Diracs are ideally low-pass filtered and sampled. The joint estimation reconstructs 5 Diracs exactly from the minimum number of noiseless samples based on Corollary 4, i.e., $3 \times 3 \times 3$ (Fig. 8(a)).

Next, we consider the same 3D Diracs but the samples are contaminated by Gaussian white noise such that the signal-to-noise ratio is 5 dB. In this noisy scenario, the Diracs are over-sampled (compared with the minimum sample sizes).

The average reconstruction error on the Dirac locations from $7 \times 7 \times 7$ noisy samples is 1.35×10^{-2} (Fig. 8(b)).

IV. REAL APPLICATION RESULTS

Because of more efficient data utilization as well as a more robust performance, the joint estimation approach of FRI signals may have significant implications for many real applications. In this section, we demonstrate the robustness of the new approach in the challenging point source estimation problem in radio astronomy (Section IV-A) and direction of arrival estimation problem in acoustics (Section IV-B).

A. Point Source Estimation in Radio Astronomy

Source estimation in radio astronomy aims at reconstructing celestial sources from the electromagnetic waves collected with antenna arrays. For example, the Low Frequency Array (LOFAR) consists of around 20 000 dipole antennas, which are located primarily in Netherland in addition to a few remote sites across Europe. One commonly used source model in radioastronomy is the point source model [30], which represents the sky image as a weighted sum of Diracs. In a far-field context, all sources are assumed to be located on a hypothetical celestial sphere (as the depth information is lost in the far-field setting). Consequently, the electromagnetic waves arrive as parallel wavefronts at the antenna array. The cross-correlations of the received signals at the two antennas are determined by the separation of the antenna pair (i.e., the baseline) as well as the point source locations. In narrow field-of-view cases, we can approximate the source estimation on the sphere by projecting the sky image in the neighborhood of the telescope focus to the tangential plane. It can be shown that the cross-correlations sample the Fourier transform of the two-dimensional sky image projection at irregular frequencies (see [31], Chapter 3). The Fourier transform frequencies are specified by the radio telescope antenna layout.

To summarize, the far-field point source reconstruction within a narrow field-of-view amounts to estimating 2D Dirac locations from the irregularly sampled Fourier transform of the sky image. Readers are referred to [6] for full accounts of various complications encountered when applying the joint estimation approach to the actual observation of a radio telescope.

In Fig. 9, we demonstrate the effectiveness of the joint estimation approach in reconstructing point sources accurately with actual LOFAR observation from the Boötes field. The reconstructed source locations are compared with the catalog [32]. The average error on the reconstructed source locations is $1'10.63''$. We also include the image reconstructed with the standard CLEAN algorithm, which is widely used in the radio astronomy community (see [30] and many of its variants [33], [34]), from the same measurements. Contrary to the FRI-based approach, in CLEAN-based methods, additional source detection has to be applied to the CLEAN image in order to obtain the source locations and intensities. Consequently, CLEAN-based methods may not always distinguish closely located sources, which are merged within one bright blob in the estimated sky

(a) Exact reconstruction from noiseless samples

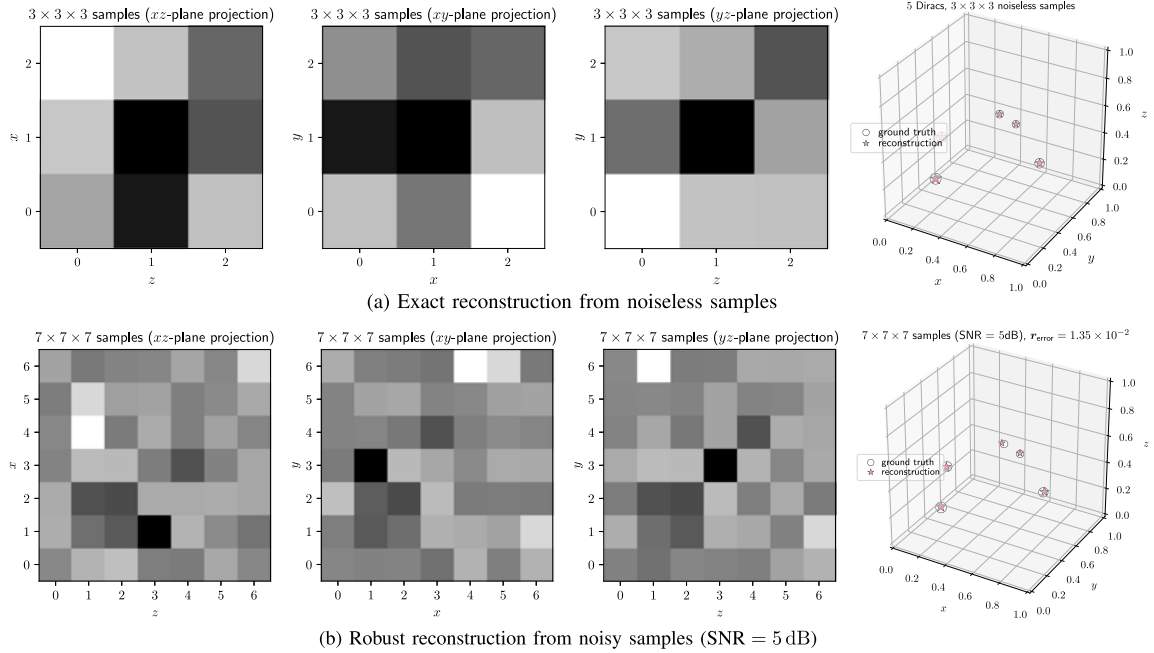(b) Robust reconstruction from noisy samples (SNR = 5 dB)

Fig. 8.   Reconstruction of 3D periodic stream of Diracs from ideally low-pass filtered samples. (a) Exact reconstruction from the noiseless samples (sample size: $3 \times 3 \times 3$). (b) Robust reconstruction from noisy samples (sample size: $7 \times 7 \times 7$, SNR = 5 dB, average reconstruction error on the Dirac locations: 1.35×10^{-2}). Maximum intensity projection is used to project the 3D samples onto relevant planes for display. (a) Exact reconstruction from noiseless samples. (b) Robust reconstruction from noisy samples (SNR = 5 dB).
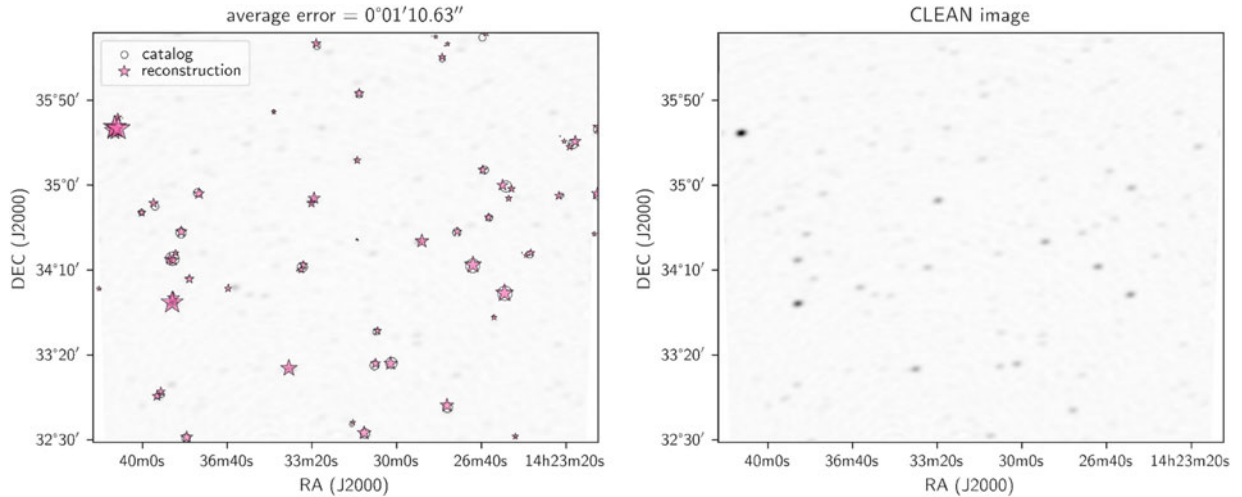

Fig. 9.   Point source estimation with the joint estimation approach from actual LOFAR observation from the Boötes field (field of view: $4° \times 4°$, number of sources reconstructed: 60, average reconstruction error on source locations: $1'10.63''$). The background CLEAN image (also shown separately in the right column) is overlaid with the reconstructed point sources and the catalog data in the J2000 coordinate (horizontal axis: right ascension, vertical axis: declination). Please refer to [6] for details related to the application of the joint estimation approach in radio astronomy.

image. The blob size is related to the telescope angular resolution limit, which is $3'23.1''$ with the setting used here.

B. Direction of Arrival Estimation with Microphone Arrays

The same principle, which was used in radio astronomy point source reconstruction, applies to the direction of arrival (DOA) estimation in acoustics. The goal here is to locate the direction of arrival of sound sources in 3D with microphone arrays. In the far-field context, this boils down to identifying the azimuths and latitudes of the acoustic sources. However, unlike

the radio astronomy problem above, which typically considers a small portion of the sky around the telescope focus, we can no longer approximate the spherical reconstruction within a tangential plane for the DOA problem—sources can be located in a much wider range on the sphere. The cross-correlations of the received signals at different microphones are related to the Fourier domain samples at non-uniform frequencies, where the Fourier expansion basis are specified by the spherical harmonics (see e.g., [5], [35], [36]). We can apply the same joint estimation technique to reconstruct sound source locations on the sphere.
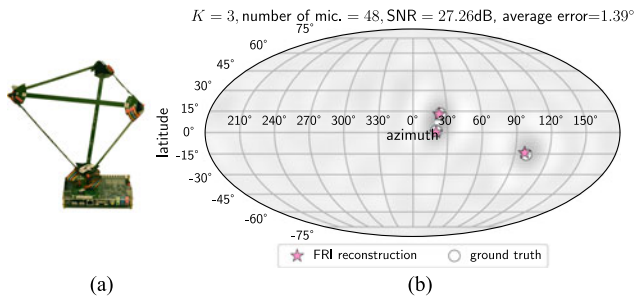
Fig. 10. Direction of arrival estimation of sound sources from actual recordings in an anechoic chamber (estimated SNR: 27.26 dB). The noise level is estimated from the silence recording. (a) The Pyramic array used to collect sound waves emitted by loudspeakers located at different locations. (b) The reconstructed source location on the sphere (average reconstruction error on source locations: $1.39°$, background image: the least square estimate of the sound field). We used the Mollweide projection to map the spherical image to a plane for display.

In the experiment, a tetrahedron-shaped microphone array (the "Pyramic array", see Fig. 10(a)) is used to record speech signals from loudspeakers that are located at three different locations in an anechoic chamber. Each side of the tetrahedron consists of 8 MEMS microphones [37]. In Fig. 10(b), we show an example of the DOA estimation with the joint estimation approach from the actual recordings. Two of the three loudspeakers are closely located (great-circle distance: $12.76°$). In the least square estimate of the sound field, both sources are contained within one bright spot. The joint estimation approach manages to resolve the two sources correctly. The average reconstruction error on the source locations is $1.39°$. Readers are referred to an on-going work [38] for detailed discussions on the exact formulations and extensive tests of the joint estimation approach in this acoustic DOA estimation problem.

V. CONCLUSION

We have proposed an efficient FRI-based sampling scheme for multi-dimensional Dirac estimation. The minimum sample size required, in order to have the exact reconstruction, scales linearly with the number of Diracs. The linear sample complexity is a substantial improvement over previous separate estimation method, which has an exponential scaling in the space dimension. We have adapted the algorithmic framework for the reconstruction of FRI signals [5] to estimate multi-dimensional Diracs reliably. The Dirac locations are determined by the intersections of hypersurfaces, which are the zero-crossings of the reconstructed mask functions. The efficiency and robustness of the proposed approach has been validated in simulations as well as in two applications with real data on point source estimation in radio astronomy and acoustics.

We are currently working on the application of the multi-dimensional Dirac estimation approach to super-resolution microscopy [39], [40]. Preliminary results indicate that we are able to improve resolution significantly. Another future research direction is on the extension of the approach to cases with damped sinusoids. This may be useful for the estimation of pulses with variable width [41] in higher dimensions.

APPENDIX A
SUPPORT SIZE OF A 2D ANNIHILATING FILTER

For brevity of derivations, we use multi-index notation, where $\mathbf{z} = (z_1, z_2) = (e^{-j2\pi x/\tau_1}, e^{-j2\pi y/\tau_2})$, $\mathbf{k} = (k, l)$, and $\mathbf{z}^\mathbf{k} = (z_1^k, z_2^l)$. Let the support size of a 2D annihilating filter $\mathbf{c}$ be S and the non-zero indices in the filter be $\mathbf{k}_s = (k_s, l_s)$ for $s = 1, \cdots, S$.

From the annihilation constraint, we have that the mask function should vanish at each one of the K Dirac locations:

$$\mu(\mathbf{z}^{(i)}) = \sum_{s=1}^{S} (\mathbf{z}^{(i)})^{\mathbf{k}_s} \mathbf{c}_{\mathbf{k}_s} = 0,$$

for $i = 1, \cdots, K$, i.e.,

$$\underbrace{\begin{bmatrix} (\mathbf{z}^{(1)})^{\mathbf{k}_1} & \cdots & (\mathbf{z}^{(1)})^{\mathbf{k}_S} \\ \vdots & \cdots & \vdots \\ (\mathbf{z}^{(K)})^{\mathbf{k}_1} & \cdots & (\mathbf{z}^{(K)})^{\mathbf{k}_S} \end{bmatrix}}_{\text{a } K \times S \text{ matrix } \mathbf{Z}} \begin{bmatrix} \mathbf{c}_{\mathbf{k}_1} \\ \vdots \\ \mathbf{c}_{\mathbf{k}_S} \end{bmatrix} = \mathbf{0}.$$

In general when $S \geq K$, the matrix $\mathbf{Z}$ is of rank K (i.e., the values $\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(K)}$ for which this does not happen is of zero measure). Therefore, a necessary (and sufficient) condition to have two linearly independent solutions $\mathbf{c}_1$ and $\mathbf{c}_2$ with the same support in (6) is $S \geq K + 2$.

The derivation generalizes to higher dimensions following the same reasoning in a straight-forward manner. Specifically, the support size of a D-dimensional annihilating filter in a joint estimation formation (16) is at least $K + D$.

APPENDIX B
SAMPLE COMPLEXITY OF DIRAC RECONSTRUCTION IN 2D

On the one hand, $2(M - \sqrt{K+2} + 1)^2$ annihilation equations can be built from $M \times M$ ideally low-pass filtered samples in addition to 3 linear constraints from the normalization $\mathcal{C} = \{\mathbf{C} | \mathbf{C}_0^\mathrm{T} \mathbf{C} + \mathbf{C}^\mathrm{T} \mathbf{C}_0 = 2\mathbf{I}\}$ (due to symmetry). On the other hand, the number of unknowns is $2(K+2) - 1$, where[12] -1 is a consequence of linear independence of $\mathbf{c}_1$ and $\mathbf{c}_2$ (see Section II-C3 for implementation details). The minimum number of samples is then obtained from

$$2\left(M - \sqrt{K+2} + 1\right)^2 + 3 \geq 2(K+2) - 1.$$

APPENDIX C
2D DIRACS WITH SHARED x OR y COORDINATES

In the particular cases, where some of the Diracs share the same x or y coordinates, the minimum size of the annihilating filter shape is further restricted by the number of Diracs with common coordinates along each direction. This can be understood from the fundamental theorem of algebra.

Suppose there are K_{same} Diracs that have the same x coordinate x_0 and the shape of the filter is $P \times Q$. From the annihilation constraint, the DTFT of the annihilating filter should vanish

---

[12]More generally in D dimensions, the degrees of freedom of D vectors are reduced by $D(D-1)/2$ because of linear independence.

at the Dirac locations:

$$\mu(x_0, y) = \sum_{l=0}^{P-1} \left(\sum_{k=0}^{Q-1} c_{k,l} (e^{-j2\pi x_0/\tau_1})^k \right) (e^{-j2\pi y/\tau_2})^l = 0.$$

(18)

Observe that (18) is a polynomial in $e^{-j2\pi y/\tau_2}$ of degree $P-1$. In order to have K_{same} such y-s within the period $[0, \tau_2)$ that satisfy the annihilation constraint, $P-1 \geq K_{\text{same}}$. Similarly, if there are L_{same} Diracs that have the same y coordinate, then $Q-1 \geq L_{\text{same}}$.

In general, this leads to annihilating filters with different shapes than $\lceil\sqrt{K+2}\rceil \times \lceil\sqrt{K+2}\rceil$ in the generic cases discussed in Section II-C1. Consequently, the minimum number of samples required in order to reconstruct K Diracs, where some of them have shared coordinates along one direction, is also larger than that in the generic cases in Corollary 1.

APPENDIX D
NULL SPACE DIMENSION OF THE JOINT ANNIHILATION MATRIX

In order to show the null space dimension of the joint annihilation matrix, we construct explicitly a non-zero vector $\mathbf{b}$ such that $\mathbf{R}_{\text{joint}}\mathbf{b} = \mathbf{0}$. The number of linearly independent $\mathbf{b}$-s gives a lower bound of the null space dimension.

Construct $b_{k,l} = u^{-k}v^{-k}$, where $u, v \in \mathbb{C}$ and are different from zero ($\mathbf{b}$ is the vectorization of the 2D data block in e.g., a column-by-column order), then

$$[c_{k,l} * b_{k,l}]_{m,n} = \sum_{k,l} c_{k,l} b_{m-k,n-l} = u^{-m} v^{-n} \sum_{k,l} c_{k,l} u^k v^l.$$

Therefore, $b_{k,l}$ that satisfies the joint annihilation constraints

$$\begin{cases} [\mathbf{c}_1]_{k,l} * b_{k,l} = 0 \\ [\mathbf{c}_2]_{k,l} * b_{k,l} = 0 \end{cases}$$

can be treated as the common roots of the two polynomial linear systems of equations (9). From (12), we know that there are at least $2(K_0-1)(L_0-1)$ such $\mathbf{b}$-s, for filters $\mathbf{c}_1$, $\mathbf{c}_2$ of size $L_0 \times K_0$. Consequently, we have proved the minimum null space dimension of the joint annihilation matrix in Corollary 3.

APPENDIX E
EQUIVALENT FORMULATION OF THE JOINT ANNIHILATION (6)

From our analysis in Section II-C3, for fixed $\mathbf{c}_1$, $\mathbf{c}_2$ at each iteration, (6) is equivalent to

$$\min_{\mathbf{b}} \quad \|\mathbf{a} - \mathbf{Gb}\|_2^2$$

$$\text{subject to} \quad \mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} \mathbf{b} = \mathbf{0}.$$

The associated Lagrangian is

$$\mathcal{L}(\mathbf{b}, \boldsymbol{\ell}) = \frac{1}{2}\|\mathbf{a} - \mathbf{Gb}\|_2^2 + \boldsymbol{\ell}^{\text{H}} \mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} \mathbf{b}.$$

From the optimality conditions, we have

$$\begin{cases} \mathbf{G}^{\text{H}}(\mathbf{Gb} - \mathbf{a}) + \mathbf{R}_{\text{joint}}^{\text{H}} \mathbf{Q}_{\text{sub}} \boldsymbol{\ell} = \mathbf{0}, \\ \mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} \mathbf{b} = \mathbf{0}. \end{cases}$$

Hence,

$$\mathbf{b} = \boldsymbol{\beta} - (\mathbf{G}^{\text{H}}\mathbf{G})^{-1} \mathbf{R}_{\text{joint}}^{\text{H}} \mathbf{Q}_{\text{sub}}$$
$$\cdot (\mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} (\mathbf{G}^{\text{H}}\mathbf{G})^{-1} \mathbf{R}_{\text{joint}}^{\text{H}} \mathbf{Q}_{\text{sub}})^{-1} \mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} \boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\mathbf{G}^{\text{H}}\mathbf{G})^{-1}\mathbf{G}^{\text{H}}\mathbf{a}$. Consequently, the objective function reduces to

$$\|\mathbf{a} - \mathbf{Gb}\|_2^2$$
$$= \boldsymbol{\beta}^{\text{H}} \mathbf{R}_{\text{joint}}^{\text{H}} \mathbf{Q}_{\text{sub}} (\mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} (\mathbf{G}^{\text{H}}\mathbf{G})^{-1} \mathbf{R}_{\text{joint}}^{\text{H}} \mathbf{Q}_{\text{sub}})^{-1} \mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} \boldsymbol{\beta}$$
$$+ \text{ terms independent of } \mathbf{c}_1 \text{ and } \mathbf{c}_2.$$

Additionally, from the "bi-linearity" of the annihilation constraint, we have:

$$\begin{bmatrix} \mathbf{R}(\mathbf{c}_1) \\ \mathbf{R}(\mathbf{c}_2) \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \mathbf{T}(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{T}(\boldsymbol{\beta}) \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$$
$$= \underbrace{\begin{bmatrix} \mathbf{T}(\boldsymbol{\beta})\widetilde{\mathbf{Q}}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}(\boldsymbol{\beta})\widetilde{\mathbf{Q}}_2 \end{bmatrix}}_{\overset{\text{def}}{=} \mathbf{T}_{\text{joint}}(\boldsymbol{\beta})} \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix}.$$

Therefore, an equivalent form of the joint annihilation (6) is

$$\min_{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2} \quad \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix}^{\text{H}} \boldsymbol{\Lambda}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix},$$
$$\text{subject to} \quad \mathbf{c}_1 = \widetilde{\mathbf{Q}}_1 \boldsymbol{\gamma}_1 \in \mathcal{C},$$
$$\mathbf{c}_2 = \widetilde{\mathbf{Q}}_2 \boldsymbol{\gamma}_2 \in \mathcal{C},$$

where

$$\boldsymbol{\Lambda}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) =$$
$$\mathbf{T}_{\text{joint}}^{\text{H}}(\boldsymbol{\beta}) \mathbf{Q}_{\text{sub}} \left(\mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{R}_{\text{joint}} (\mathbf{G}^{\text{H}}\mathbf{G})^{-1} \mathbf{R}_{\text{joint}}^{\text{H}} \mathbf{Q}_{\text{sub}} \right)^{-1} \mathbf{Q}_{\text{sub}}^{\text{H}} \mathbf{T}_{\text{joint}}(\boldsymbol{\beta}).$$

The iterative strategy proposed in [5] then amounts to building $\boldsymbol{\Lambda}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ from the *previous* reconstruction and updating $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ from the quadratic minimization.

Since $\boldsymbol{\Lambda}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ contains nested matrix inverses, the direct implementation may not be numerically stable. By introducing auxiliary variables as in [5], it can be shown that the update of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ amounts to solving an extended linear system of equation (14) at each iteration.

ACKNOWLEDGMENT

REFERENCES

[1] R. Perley, F. Schwab, and A. Bridle, *Synthesis Imaging in Radio Astronomy*. San Francisco, CA, USA: Astronomical Soc. of the Pacific, Jan. 1989.
[2] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[3] 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.

[4] R. 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'alkool, à différentes températures," *J. de l'Ecole Polytechnique*, vol. 1, no. 2, pp. 24–76, 1795.

[5] H. Pan, T. Blu, and M. Vetterli, "Towards generalized FRI sampling with an application to source resolution in radioastronomy," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 821–835, Feb. 2017.

[6] H. Pan, M. Simeoni, P. Hurley, T. Blu, and M. Vetterli, "LEAP: Looking beyond pixels with continuous-space EstimAtion of Point sources," *Astron. Astrophys.*, vol. 608, 2017, Art. no. A136.

[7] H. Pan, R. Scheibler, E. Bezzam, I. Dokmanić, and M. Vetterli, "FRIDA: FRI-based DOA estimation for arbitrary array layouts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2017, pp. 3186–3190.

[8] P. Shukla and P. L. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3670–3686, Jul. 2007.

[9] C. Chen, P. Marziliano, and A. C. Kot, "2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2851–2859, Jun. 2012.

[10] H. Pan, T. Blu, and P. L. Dragotti, "Sampling curves with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 458–471, Jan. 2014.

[11] S. Mulleti and C. S. Seelamantula, "Ellipse fitting using the finite rate of innovation sampling principle," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1451–1464, Mar. 2016.

[12] M. Fatemi, A. Amini, and M. Vetterli, "Sampling and reconstruction of shapes with algebraic boundaries," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 5807–5818, Nov. 2016.

[13] I. Maravić and M. Vetterli, "Exact sampling results for some classes of parametric nonbandlimited 2-D signals," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 175–189, Jan. 2004.

[14] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[15] S. Rouquette and M. Najim, "Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 237–245, Jan. 2001.

[16] T. Jiang, N. D. Sidiropoulos, and J. M. ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1849–1859, Sep. 2001.

[17] F. Roemer, M. Haardt, and G. Del Galdo, "Analytical performance assessment of multi-dimensional matrix and tensor-based ESPRIT-type algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2611–2625, May 2014.

[18] S. Sahnoun, K. Usevich, and P. Comon, "Multidimensional ESPRIT for damped and undamped signals: Algorithm, computations, and perturbation analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5897–5910, Nov. 2017.

[19] T. Peter, G. Plonka, and R. Schaback, "Prony's method for multivariate signals," in *Proc. Appl. Math. Mech.*, vol. 15, no. 1, 2015, pp. 665–666.

[20] S. Kunis, T. Peter, T. Römer, and U. von der Ohe, "A multivariate generalization of Prony's method," *Linear Algebra Appl.*, vol. 490, pp. 31–47, 2016.

[21] G. Plonka and M. Wischerhoff, "How many Fourier samples are needed for real function reconstruction?" *J. Appl. Math. Comput.*, vol. 42, no. 1/2, pp. 117–137, 2013.

[22] A. Cuyt and W.-s. Lee, "Multivariate exponential analysis from the minimal number of samples," *Adv. Comput. Math.*, pp. 1–16, 2017.

[23] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576–6601, Oct. 2014.

[24] J. J. Sylvester, "On a theory of the syzygetic relations of two rational integral functions, comprising an application to the theory of Sturm's functions, and that of the greatest algebraical common measure," *Philos. Trans. Roy. Soc. London*, vol. 143, pp. 407–548, 1853.

[25] A. Cayley, "Note sur la méthode d'élimination de Bezout." *J. Für Die Reine Und Angewandte Math.*, vol. 53, pp. 366–367, 1857.

[26] J. L. Coolidge, *A Treatise on Algebraic Plane Curves*, vol. 1. Chelmsford, MA, USA: Courier Corp., 2004.

[27] G. H. Golub and C. F. Van Loan, vol. 3. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2012.

[28] Y. Nakatsukasa, V. Noferini, and A. Townsend, "Computing the common zeros of two bivariate functions via Bézout resultants," *Numer. Math.*, vol. 129, no. 1, pp. 181–209, 2015.

[29] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008.

[30] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl. Ser.*, vol. 15, pp. 417–426, 1974.

[31] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*. New York, NY, USA: Wiley, 2001.

[32] W. Williams *et al.*, "LOFAR 150-MHz observations of the Boötes field: Catalogue and source counts," *Monthly Notices Roy. Astron. Soc.*, vol. 460, no. 3, pp. 2385–2412, 2016.

[33] S. Bhatnagar and T. Cornwell, "Scale sensitive deconvolution of interferometric images-I. Adaptive Scale Pixel (Asp) decomposition," *Astron. Astrophys.*, vol. 426, no. 2, pp. 747–754, 2004.

[34] T. J. Cornwell, K. Golap, and S. Bhatnagar, "The noncoplanar baselines effect in radio interferometry: The W-projection algorithm," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 5, pp. 647–657, Oct. 2008.

[35] S. Deslauriers-Gauthier and P. Marziliano, "Sampling signals with a finite rate of innovation on the sphere," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4552–4561, Sep. 2013.

[36] I. Dokmanić and Y. M. Lu, "Sampling sparse signals on the sphere: Algorithms and applications," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 189–202, Jan. 2016.

[37] J. Azcarreta Ortiz, "Pyramic array: An FPGA based platform for many-channel audio acquisition," Master's thesis, Dept. Comput. Commun. Sci., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Aug. 2016.

[38] R. Scheibler, H. Pan, E. Bezzam, I. Dokmanić, and M. Vetterli, "FRIDA: FRI-based direction of arrival (DoA) estimation," in preparation, 2018.

[39] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–796, 2006.

[40] E. Betzig *et al.*, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[41] G. Baechler, A. Scholefield, L. Baboulaz, and M. Vetterli, "Sampling and exact reconstruction of pulses with variable width," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2629–2644, May 2017.

Hanjie Pan (S'11) was born in Jiangsu, China, in 1988. He received the B.Eng. degree (with first honor) and the M.Phil. degree in electronic engineering from The Chinese University of Hong Kong, Shatin, Hong Kong, in 2010 and 2013, respectively, and the Ph.D. degree in computer and communication sciences from the École Polytechnique Fédétale de Lausanne, Lausanne, Switzerland, for a study on continuous sparse recovery. In the summer 2010, he was a visiting student with Imperial College, London, UK. From 2016 to 2017, he was an affiliated with IBM Research Zürich as a research engineer working on source estimation from LOFAR radio telescope measurements. His research interests include sampling theory, biomedical imaging, and sparse signal recovery.

Thierry Blu (M'96–SM'06–F'12) was born in Orléans, France, in 1964. He received the "Diplôme d'ingénieur" from École Polytechnique, Palaiseau, France, in 1986, and from Télécom Paris (ENST), Paris, France, in 1988, and the Ph.D. degree in electrical engineering from ENST for a study on iterated rational filterbanks, applied to wideband audio coding.

He is currently a Professor with the Department of Electronic Engineering, the Chinese University of Hong Kong, Shatin, Hong Kong. He was with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland, from 1998 and 2007. His research interests include wavelets, approximation and sampling theory, sparse representations, image denoising, biomedical imaging, optics, and wave propagation.

Dr. Blu was the recipient of two best paper awards from the IEEE Signal Processing Society (2003 and 2006). He is also coauthor of a paper that received a Young Author best paper award (2009) from the IEEE Signal Processing Society. He has been a member of the IEEE Signal Processing Theory and Methods Technical Committee (2008–2013), and an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2002–2006), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006–2010), and *Elsevier Signal Processing* (2008–2011). He is currently on the board of *EURASIP Journal on Image and Video Processing* (since 2010).

Martin Vetterli (M'86–SM'90–F'95) received the Dipl. El.-Ing. degree from ETH Zürich (ETHZ), Zürich, Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, California, in 1982, and the Doctorat és Sciences degree from École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 1986.

He was a Research Assistant with Stanford and EPFL, and has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University, New York, where he was last an Associate Professor of electrical engineering and the Co-Director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California at Berkeley, where he was a Professor with the Department of Electrical Engineering and Computer Sciences until 1997, and has held an Adjunct Professor position until June 2010. Since 1995, he has been a Professor of Communication Systems with EPF Lausanne, Switzerland, where he chaired the Communications Systems Division (1996/1997), and heads the Audiovisual Communications Laboratory. From 2001 to 2004, he directed the National Competence Center in Research on mobile information and communication systems. He also was a Vice-President with EPFL from October 2004 to February 2011 in charge, among others, of international affairs and computing services. He has held visiting positions with ETHZ (1990) and Stanford (1998). From March 2011 to 2012, he was the Dean of the School of Computer and Communication Sciences of EPFL. From 2013 to 2016, he led the Swiss National Science Foundation and since 2017 he is President of EPFL. He has authored and coauthored more than 170 journal papers on a variety of topics in signal/image processing and communications and holds about 50 patents and patent applications. His research interests include sampling, wavelets, multirate signal processing, computational complexity, signal processing for communications, digital image/video processing, joint source/channel coding, signal processing for sensor networks, and inverse problems like acoustic tomography.

Prof. Vetterli is a Fellow of ACM, a Fellow of EURASIP, and a member of SIAM and NAE. He is on the editorial boards of Applied and Computational Harmonic Analysis, the Journal of Fourier Analysis and Application and the IEEE Journal on Selected Topics in Signal Processing and has been elected Foreign Member of the NAE in 2015. He received the Best Paper Award of EURASIP in 1984, the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986, the IEEE Signal Processing Society's Senior Paper Awards in 1991, in 1996 and in 2006 (for papers with D. LeGall, K. Ramchandran, and Marziliano and Blu, respectively). He won the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, the IEEE Signal Processing Technical Achievement Award in 2001, the IEEE Signal Processing Society Award in 2010 for fundamental contributions to signal processing theory, technology and education and is an ISI highly cited researcher in engineering. He was a member of the Swiss Council on Science and Technology from 2000 to 2003. He was a plenary speaker at various conferences (e.g., IEEE ICIP, ICASSP, ISIT) and is the coauthor of three books with J. Kovacevic, *Wavelets and Subband Coding*, 1995, with P. Prandoni *Signal Processing for Communications*, 2008, and with J. Kovacevic and V. K. Goyal, *Foundations of Signal Processing*, 2015.