

Sampling Sparse Signals on the Sphere: Algorithms and Applications

Ivan Dokmanić, *Member, IEEE*, and Yue M. Lu, *Senior Member, IEEE*

Abstract—We propose a sampling scheme that can perfectly reconstruct a collection of spikes on the sphere from samples of their lowpass-filtered observations. Central to our algorithm is a generalization of the annihilating filter method, a tool widely used in array signal processing and finite-rate-of-innovation (FRI) sampling. The proposed algorithm can reconstruct K spikes from $(K + \sqrt{K})^2$ spatial samples. For large K , this sampling requirement improves over previously known FRI sampling schemes on the sphere by a factor of four. We showcase the versatility of the proposed algorithm by applying it to three problems: 1) sampling diffusion processes induced by localized sources on the sphere, 2) shot noise removal, and 3) sound source localization (SSL) by a spherical microphone array. In particular, we show how SSL can be reformulated as a spherical sparse sampling problem.

Index Terms—Annihilation filter, diffusion sampling, finite rate of innovation, shot noise removal, sparse sampling, sphere, spherical harmonics, sound source localization.

I. INTRODUCTION

NUMEROUS signals live on a sphere. Take, for example, any signal defined on Earth's surface [1], [2]. Signals from space measured on Earth [3] also have a spherical domain. In acoustics, spherical microphone arrays output a time-varying signal supported on a sphere [4], [5], while in diffusion weighted magnetic resonance imaging fiber orientations live on a sphere [6]. In practice, we only have access to a finite number of samples of such signals. Thus, sampling and reconstruction of spherical signals is an important problem.

Just as signals in Euclidean domains can be expanded via sines and cosines, one can naturally represent spherical signals in the Fourier domain via *spherical harmonics* [7]. A signal is *bandlimited* if it is a linear combination of finitely many spherical harmonics. Sampling bandlimited signals on the sphere has

been studied extensively: for signals bandlimited to spherical harmonic degree B , Driscoll and Healy [7] proposed a sampling theorem that requires $4B^2$ spherical samples. The best exact general purpose sampling theorem due to McEwen and Wiaux uses $2B^2$ samples [8]. Recently, Khalid, Kennedy, and McEwen devised a stable sampling scheme that requires the optimal number of samples, B^2 [9].

In this paper, we study the problem of sampling localized spikes on the sphere; in the limit, the spikes become Dirac delta functions. Such sparse signals on the sphere are encountered in many problems. For example, various acoustic sources are well-approximated by point sources; the directional distribution of multiple sources is then a finite collection of spikes. Stars in the sky observed from Earth are angular spikes, and so are plume sources on Earth.

Localized spikes are not bandlimited, so the bandlimited sampling theorems [7]–[9] do not apply. In this paper, we propose an algorithm to perfectly reconstruct collections of spikes from their lowpass-filtered observations. Our algorithm efficiently reconstructs K spikes when the bandwidth of the lowpass filter is at least $K + \sqrt{K}$.

A. Prior Art

Our work is in the same spirit as finite rate-of-innovation (FRI) sampling, introduced by Vetterli, Marziliano, and Blu [10]. They showed that a stream of K Diracs on the line can be efficiently recovered from $2K + 1$ samples. Initially developed for 1D signals, the original FRI sampling was extended to 2D and higher-dimensional signals in [11], [12], and its performance was studied in noisy conditions [13], [14].

In a related work [15], [16], Deslauriers-Gauthier and Marziliano proposed an FRI sampling scheme for signals on the sphere, reconstructing K Diracs from $4K^2$ samples. Their motivating application is the recovery of the fiber orientations in diffusion weighted magnetic resonance imaging [6], [17]. They further show that if only $3K$ spectral bins are active, the required number of samples can be reduced to $3K$. Sampling at this lower rate, however, relies on the assumption that we can apply arbitrary spectral filters to the signal before sampling. This is known as *spatial anti-aliasing*—a procedure that is generally challenging or impossible to implement in most applications involving spherical signals, where we only have access to finite samples of the underlying continuous signals.¹

In many applications, the sampling kernels (i.e., the lowpass filters) through which we observe the spikes are provided by

¹This is not to be confused with spatial anti-aliasing in image downsampling, where we do have access to all pixels.

Manuscript received March 07, 2015; revised August 13, 2015; accepted August 21, 2015. Date of publication September 14, 2015; date of current version December 08, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antonio Napolitano. The work of I. Dokmanić was supported by an ERC Advanced Grant—Support for Frontier Research—SPARSAM Nr: 247006, and a Google Ph.D. Fellowship. The work of Y. M. Lu was supported in part by the U.S. National Science Foundation under grant CCF-1319140. This work appeared in part in the *Proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April 19–24, 2015.

I. Dokmanić is with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: ivan.dokmanic@epfl.ch@epfl.ch).

Y. M. Lu is with the Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: yuelu@seas.harvard.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2015.2478751

some underlying physical process (e.g., point spread functions and Green's functions). These kernels are often approximately bandlimited, but we cannot further control or design their spectral selectivity. The impossibility of arbitrary spatial filtering suggests that our goal is to reduce the required bandwidth, or more practically, to maximize the number of spikes that we can reconstruct at a given bandwidth.

Recently, Bendory, Dekel and Feuer proposed a spherical super-resolution method [18], [19], extending the results of Candès and Fernandez-Granda [20] to the spherical domain. They showed that an ensemble of Diracs on the sphere can be reconstructed from projections onto a set of spherical harmonics by solving a semidefinite program, provided that the Diracs satisfy a minimal separation condition. When the Diracs are constrained to a discrete set of locations, their formulation allows them to bound the recovery error in the presence of noise. Our algorithm based on FRI is non-iterative and thus very fast. We also allow the weights to be complex, which may be important in applications (for an example in sound source localization, see Section IV.C). On the other hand, although our algorithm does not require any separation between the Diracs in the noiseless setting, a theoretical analysis of the stability in the presence of noise is not yet available (see Section III.F for some numerical experiments). We also need to assume that the number of Diracs is known up to a range, whereas in [18], [19] no such assumption is necessary.

B. Outline and Main Contributions

We start by reviewing some basic notions of harmonic analysis on the sphere in Section II. We then present the main result of this work in Section III: A collection of K Diracs on the sphere can be reconstructed from its lowpass filtered version, provided that the bandwidth of the sampling kernel is at least $K + \sqrt{K}$. This bandwidth requirement also implies that $(K + \sqrt{K})^2$ spatial samples taken at generic locations suffice to reconstruct the K Diracs. We establish this result by constructing a new algorithm for spherical FRI sampling. Compared to $4K^2$ samples as required in a previous work [15], our algorithm reduces the numbers of samples via a more efficient use of the available spectrum. For large K , the required number of samples is reduced by a factor of up to 4. The proposed algorithm is first developed for the noiseless case. Procedures to improve the robustness of the algorithm in noisy situations are presented in Section III.E, and we compare the performance of the algorithm with the Cramér-Rao lower bound [21] in Section III.F. Section IV presents the applications of the proposed algorithm to three problems: 1) sampling diffusion processes on the sphere, 2) shot noise removal, and 3) sound source localization. These diverse applications demonstrate the usefulness and versatility of our results. We conclude in Section V.

This paper follows the philosophy of reproducible research. All the results and examples presented in the paper can be reproduced using the code available at <http://lcav.epfl.ch/ivan.dokmanic>.

II. HARMONIC ANALYSIS ON THE SPHERE AND PROBLEM FORMULATION

A. Spherical Harmonics

We briefly recall the definitions of spherical harmonics and spherical convolution. The 2-sphere is defined as the locus of points in \mathbb{R}^3 with unit norm,

$$\mathbb{S}^2 \stackrel{\text{def}}{=} \{ \mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x}^\top \mathbf{x} = 1 \}.$$

In what follows, we often use ξ to represent a generic point on the sphere. In addition to the standard Euclidean representation $\xi = [x, y, z]^\top$, points on \mathbb{S}^2 can also be conveniently parameterized by angles of colatitude and azimuth, i.e., $\xi = (\theta, \phi)$, with θ measured from the positive z -axis, and ϕ measured in the xy plane from the positive x -axis. The two equivalent representations are related by the following conversion,

$$\begin{aligned} x &= \sin(\theta) \cos(\phi), \\ y &= \sin(\theta) \sin(\phi), \\ z &= \cos(\theta). \end{aligned} \quad (1)$$

The Hilbert space of square-integrable functions on the sphere, $L^2(\mathbb{S}^2)$, is defined through the corresponding inner product. For two functions $f, g \in L^2(\mathbb{S}^2)$ we have

$$\langle f, g \rangle \stackrel{\text{def}}{=} \int_{\mathbb{S}^2} f(\xi) \overline{g(\xi)} d\xi, \quad (2)$$

where $d\xi = \sin(\theta) d\theta d\phi$ is the usual rotationally invariant measure on the sphere. With respect to this inner product, spherical harmonics form a natural orthonormal Fourier basis for $L^2(\mathbb{S}^2)$. They are defined as [7]

$$Y_\ell^m(\theta, \phi) = N_\ell^m P_\ell^{|m|}(\cos \theta) e^{jm\phi}, \quad (3)$$

where the normalization constant is

$$N_\ell^m = (-1)^{(m+|m|)/2} \sqrt{\frac{(2\ell+1)(\ell-|m|)!}{4\pi(\ell+|m|)!}}, \quad (4)$$

and $P_\ell^m(x)$ is the associated Legendre polynomial of degree ℓ and order m . Note that different communities sometimes use different normalizations and sign conventions in the definitions of spherical harmonics and associated Legendre polynomials. As long as applied consistently, the choice of convention does not affect our results.²

In this paper, we adopt the following definition

$$P_\ell^m(x) \stackrel{\text{def}}{=} (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_\ell(x), \text{ for } m \geq 0, \quad (5)$$

where $P_\ell(x)$ is the Legendre polynomial of degree ℓ [22].

Any square integrable function on the sphere, $f \in L^2(\mathbb{S}^2)$, can be expanded in the spherical harmonic basis,

$$f(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{|m| \leq \ell} \hat{f}_\ell^m Y_\ell^m(\theta, \phi). \quad (6)$$

²It is common to write the spherical harmonic order m in the superscript. We will keep this convention for the associated Legendre polynomials $P_\ell^{|m|}$, spherical harmonics Y_ℓ^m , normalization constants N_ℓ^m and the spherical Fourier coefficients \hat{f}_ℓ^m . It is not to be confused with integer powers such as x^ℓ .

The Fourier coefficients are computed as

$$\widehat{f}_\ell^m = \langle f, Y_\ell^m \rangle = \int_{\mathbb{S}^2} f(\xi) \overline{Y_\ell^m(\xi)} d\xi. \quad (7)$$

The coefficients $[\widehat{f}_\ell^m, (\ell, m) \in \mathcal{I}]$ form a countable set supported on an infinite triangle of indices,

$$\mathcal{I} = \{(\ell, m) \in \mathbb{Z}^2 \mid \ell \geq 0, |m| \leq \ell\}. \quad (8)$$

We say that f is *bandlimited* with bandwidth B if $\widehat{f}_\ell^m = 0$ for $\ell \geq B$. Often we think of B as the smallest integer such that this holds. For a bandlimited function, the triangle \mathcal{I} is cut off at $\ell = B$. In what follows, we use

$$\mathcal{I}_B \stackrel{\text{def}}{=} \{(\ell, m) \in \mathbb{Z}^2 \mid 0 \leq \ell < B, |m| \leq \ell\} \quad (9)$$

to represent the spectral support of a bandlimited function with bandwidth B . The set \mathcal{I}_B contains B^2 indices, so we can represent the spectrum as an B^2 -dimensional column vector

$$\widehat{\mathbf{f}} \stackrel{\text{def}}{=} [\widehat{f}_0^0, \widehat{f}_1^{-1}, \widehat{f}_1^0, \widehat{f}_1^1, \dots, \widehat{f}_{B-1}^{-B+1}, \dots, \widehat{f}_{B-1}^{B-1}]^\top. \quad (10)$$

B. Rotations and Convolutions on the Sphere

Let $\mathbb{S}\mathbb{O}_3$ denote the group of rotations in \mathbb{R}^3 ; any rotation $\varrho \in \mathbb{S}\mathbb{O}_3$ is parameterized by three angles that specify rotations about three distinct axes. Thus we can write $\varrho = \varrho(\alpha, \beta, \gamma)$. The most common parameterization is called *Euler angles* [23].

Counter-clockwise rotation of a vector $\mathbf{x} \in \mathbb{R}^3$ about the z -axis is achieved by multiplying \mathbf{x} by the corresponding rotation matrix,

$$\mathbf{R}_z(\alpha) = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where α is the rotation angle. Rotation matrices around axes x and y can be defined analogously.

We use $\Lambda(\varrho)$ to represent the rotation operator corresponding to ϱ , that acts on spherical functions. Thus for f a function on the sphere, $\Lambda(\varrho)f$ represents the rotated function, defined as

$$[\Lambda(\varrho)f](\xi) \stackrel{\text{def}}{=} f(\varrho^{-1} \circ \xi), \quad (11)$$

where ϱ^{-1} is the inverse rotation of ϱ , and by $\varrho^{-1} \circ \xi$ we mean pre-multiplying by $\mathbf{R}(\varrho^{-1})$ the unit column vector corresponding to ξ , cf. (1). Compare this definition with the Euclidean case where shifting the argument to the left (subtracting a positive number) results in the shift of the function to the right.

There are various definitions of convolution on the sphere, all being non-commutative. One function, call it f , provides the weighting for the rotations of the other function h . A standard definition is then [7], [24]

$$\begin{aligned} [f * h](\xi) &\stackrel{\text{def}}{=} \left[\left(\frac{1}{2\pi} \int_{\mathbb{S}\mathbb{O}_3} d\varrho \cdot f(\varrho \circ \eta) \cdot \Lambda(\varrho) \right) h \right](\xi) \\ &= \frac{1}{2\pi} \int_{\mathbb{S}\mathbb{O}_3} f(\varrho \circ \eta) h(\varrho^{-1} \circ \xi) d\varrho, \end{aligned} \quad (12)$$

where $\eta \in \mathbb{S}^2$ is the north pole. It is easy to verify that this definition generalizes the standard convolution in Euclidean spaces, with the rotation operator ϱ playing the same role as translations

do on the line. Because the spherical convolution is not commutative, it is important to fix the ordering of the arguments. In our case, the second argument— h in (12)—will always be the filter, i.e., the observation kernel.

The familiar convolution-multiplication rule in standard Euclidean domains holds for spherical convolutions too. It can be shown ([7], Theorem 1) that for any two functions $f, h \in L^2(\mathbb{S}^2)$, the Fourier transform of their convolution is a pointwise product of the transforms

$$(f * h)_\ell^m = \sqrt{\frac{4\pi}{2\ell + 1}} \widehat{f}_\ell^m \widehat{h}_\ell^0. \quad (13)$$

We note that f can also be a generalized function (a distribution). In particular, we consider spherical Dirac delta functions, which may be defined by their action on functions as

$$\langle f, \delta(\cdot; \xi_0) \rangle = \int_{\mathbb{S}^2} f(\xi) \delta(\xi; \xi_0) d\xi = f(\xi_0). \quad (14)$$

Where appropriate, we will explicitly use the colatitude-azimuth notation $\delta(\theta, \phi; \theta_0, \phi_0)$.

C. Problem Formulation

Consider a collection of K Diracs on the sphere

$$f(\xi) = \sum_{k=1}^K \alpha_k \delta(\xi; \xi_k), \quad (15)$$

where the weights $\{\alpha_k \in \mathbb{C}\}_{k=1}^K$ and the locations of the Diracs $\{\xi_k = (\theta_k, \phi_k)\}_{k=1}^K$ are all unknown parameters. Let $y(\xi)$ be a filtered version of $f(\xi)$, i.e.,

$$y(\xi) = [f * h](\xi),$$

where the filter (or sampling kernel) $h(\xi)$ is a bandlimited function with bandwidth B . We further assume that the spherical Fourier transform of $h(\xi)$ is nonzero within its spectral support, i.e., $\widehat{h}_\ell^m \neq 0$ for all $\ell < B$. Given spatial samples of $y(\xi)$, we would like to reconstruct $f(\xi)$, or equivalently, recover the unknown parameters $\{(\alpha_k, \xi_k)\}_{k=1}^K$.

Since the filtered signal $y(\xi)$ is bandlimited, we can use bandlimited sampling theorems on the sphere (e.g., [7], [8]) or direct linear inversion (see Section III.A) to recover its Fourier spectrum \widehat{y}_ℓ^m from its spatial samples of sufficient density. Using the convolution-multiplication identity in (13), we can then recover the lowpass subband of $f(\xi)$ as

$$\widehat{f}_\ell^m = [(2\ell + 1)/(4\pi)]^{1/2} \cdot (\widehat{y}_\ell^m / \widehat{h}_\ell^0),$$

for $0 \leq \ell < L$ and $|m| \leq \ell$. Being a collection of Diracs, $f \notin L^2(\mathbb{S}^2)$, but its Fourier transform \widehat{f}_ℓ^m can still be computed via (7) and (14) in the sense of distributions:

$$\begin{aligned} \widehat{f}_\ell^m &= \sum_{k=1}^K \alpha_k \overline{Y_\ell^m(\theta_k, \phi_k)} \\ &= N_\ell^m \sum_{k=1}^K \alpha_k P_\ell^{|m|}(\cos \theta_k) e^{-jm\phi_k}. \end{aligned} \quad (16)$$

The problems we address in this paper can now be stated as follows: Can we reconstruct a collection of K Diracs on the

sphere from its Fourier coefficients \widehat{f}_ℓ^m in the lowpass subband \mathcal{I}_B as defined in (9)? If so, then what is the minimum bandwidth B that allows us to do it? In practice, the sampling kernel is often given and not subject to our control. In this case, the previous question can be reformulated as determining the maximum number of spikes that we can reconstruct at a given bandwidth B .

We assume that the number of Diracs K is given as part of the input. If K is unknown, a simple strategy to determine it is to iterate over the admissible range of values, and select the one that gives the smallest estimation residual. Alternatively, we may use heuristics based on the decay of the singular values of the annihilation matrix (see Section III.C).

III. SAMPLING SPHERICAL FRI SIGNALS

In this section we address the questions stated above. Our main result can be summarized in the following theorem:

Theorem 1: Let f be a collection of K Diracs on the sphere \mathbb{S}^2 as in (15), with complex weights $\{\alpha_k\}_{k=1}^K$ at locations $\{\xi_k = (\theta_k, \phi_k)\}_{k=1}^K$, such that θ_k are distinct, ϕ_k are distinct, and $\theta_k \notin \{0, \pi\}$ for all k . Convolve f with a bandlimited sampling kernel h_B , where the bandwidth $B \geq K + \sqrt{K}$, and sample the resulting signal $[f * h_B](\xi)$ at B^2 points $\{\psi_n \in \mathbb{S}^2\}_{n=1}^{B^2}$ which form a B -admissible sampling grid on \mathbb{S}^2 . Then the samples

$$f_n = [f * h_B](\psi_n), \quad n = 1, \dots, B^2$$

are a sufficient characterization of f .

Remark: The definition of an admissible grid is given in the following section; in short, samples taken on a B -admissible grid allow to reconstruct any bandlimited signal of bandwidth B . Almost all sampling grids containing at least B^2 samples are B -admissible.

Remark: It is clear from the statement that the proposed scheme is coordinate-system-dependent. As will be detailed in Section III.C, a simple strategy to ensure that all θ_k and all ϕ_k are distinct, and that $\theta_k \notin \{0, \pi\}$, is to apply a random rotation to the coordinate system prior to reconstruction.

We provide a constructive proof of Theorem 1 by presenting an algorithm that can efficiently recover K localized spikes from B^2 samples, where $B \geq K + \sqrt{K}$. Before presenting the algorithm and the proof, we first define some relevant notation and state two lemmas.

A. From Samples to the Fourier Transform

Our algorithms perform computation with spectral coefficients. In practice, we have access to spatial samples of the function, so we need a procedure to convert between the spatial and the Fourier representations. We first describe a method to compute the Fourier transform from samples taken at generic sampling points.

Let the function $f \in L^2(\mathbb{S}^2)$ have bandwidth B ; then we can express it as

$$f(\theta, \phi) = \sum_{\ell=0}^{B-1} \sum_{m=-\ell}^{\ell} \widehat{f}_\ell^m Y_\ell^m(\theta, \phi). \quad (17)$$

Choose a set of sampling points $\mathcal{P} = \{\psi_n \in \mathbb{S}^2\}_{n=1}^N$, and let $\mathbf{Y}_{B,\mathcal{P}} \stackrel{\text{def}}{=} [y_{n,(\ell,m)}]$ where $y_{n,(\ell,m)} = Y_\ell^m(\psi_n)$,

$1 \leq n \leq N$, $0 \leq \ell < B$, and $|m| \leq \ell$. Furthermore, let $\mathbf{f} = [f(\psi_1), \dots, f(\psi_N)]^\top$ be the vector of samples of f . We can then write

$$\mathbf{f} = \mathbf{Y}_{B,\mathcal{P}} \widehat{\mathbf{f}}, \quad (18)$$

where $\widehat{\mathbf{f}}$ is the B^2 -dimensional vector of spectral coefficients as defined in (10). The goal is to recover the spectral coefficients $\widehat{\mathbf{f}}$. We can recover $\widehat{\mathbf{f}}$ from \mathbf{f} as soon as the matrix $\mathbf{Y}_{B,\mathcal{P}}$ has full column rank. In that case, we compute

$$\widehat{\mathbf{f}} = \mathbf{Y}_{B,\mathcal{P}}^\dagger \mathbf{f}, \quad (19)$$

where $\mathbf{Y}_{B,\mathcal{P}}^\dagger$ denotes the Moore-Penrose pseudoinverse of $\mathbf{Y}_{B,\mathcal{P}}$. In relation to the invertibility of $\mathbf{Y}_{B,\mathcal{P}}$, we can define *admissible* sampling grids as follows:

Definition 1: A sampling grid $\mathcal{P} = \{\psi_n\}_{n=1}^N$ is B -admissible if the corresponding matrix $\mathbf{Y}_{B,\mathcal{P}}$ has full column rank.

In particular, if we draw the samples uniformly at random on the sphere, we can show that $\mathbf{Y}_{B,\mathcal{P}}$ is regular with probability one:

Proposition 1: Draw N sampling points $\mathcal{P} = \{\psi_n\}_{n=1}^N$ from any absolutely continuous probability measure on the sphere (e.g., uniformly at random). Then \mathcal{P} is B -admissible almost surely if $N \geq B^2$, that is, if $\mathbf{Y}_{B,\mathcal{P}}$ has at least as many rows as columns.

The proof of this proposition is very similar to that of Theorem 3.2 in [25], and is thus omitted; the key insight is that the zero sets of linear combinations of spherical harmonics have measure zero on the sphere. The same technique is used in the proof of Lemma 3 in Appendix C.

The above result indicates that we can recover the spectral coefficients \widehat{f}_ℓ^m in the lowpass region \mathcal{I}_B from B^2 samples taken at generic points on the sphere. The reconstruction requires a matrix inversion as in (19).

Much faster reconstruction is possible when the function is sampled on certain regular grids. In that case, we can leverage the structure of $\mathbf{Y}_{B,\mathcal{P}}$ to accelerate the matrix inversion. Such efficient schemes were proposed by Driscoll and Healy [7], requiring $4B^2$ samples; by McEwen and Wiaux [8], requiring $2B^2$ samples; and most recently, by Khalid, Kennedy and McEwen [9], requiring B^2 samples.

B. The Data Matrix

Using the definition of associated Legendre polynomials in (5), we rewrite the spherical harmonics (3) as

$$Y_\ell^m(\theta, \phi) = \widetilde{N}_\ell^m (\sin \theta)^{|m|} \left[\frac{d^{|m|}}{d(\cos \theta)^{|m|}} P_\ell(\cos \theta) \right] e^{jm\phi}, \quad (20)$$

where $\widetilde{N}_\ell^m = (-1)^m N_\ell^m$.

The essential observation is that the bracketed term in (20) is a polynomial in $x = \cos \theta$. At bandwidth B , the largest spherical harmonic degree is $B - 1$, so the largest power of x in (20) is $B - 1$ as well. It follows that we can rewrite the derivative term as a linear combination of powers of x , i.e.,

$$\widetilde{N}_\ell^m \frac{d^{|m|}}{d(\cos \theta)^{|m|}} P_\ell(\cos \theta) = \mathbf{c}_{\ell m}^\top \mathbf{x}, \quad (21)$$

where $\mathbf{x} \stackrel{\text{def}}{=} [x^{B-1}, x^{B-2}, \dots, x, 1]^\top$, $x = \cos \theta$ and $\mathbf{c}_{\ell m} \in \mathbb{R}^B$ contains the corresponding polynomial coefficients.

Using the dot-product formulation (21), the spectrum of f , as given by (16), can be expressed as

$$\hat{f}_\ell^m = \mathbf{c}_{\ell m}^\top \sum_{k=1}^K \alpha_k \mathbf{x}_k (\sin \theta_k)^{|m|} e^{-jm\phi_k}, \quad (22)$$

where $\mathbf{x}_k \stackrel{\text{def}}{=} [x_k^{B-1}, x_k^{B-2}, \dots, x_k, 1]^\top$ with $x_k = \cos \theta_k$, and we factored $\mathbf{c}_{\ell m}^\top$ out of the summation as it does not depend on k .

A key ingredient in our proposed algorithm is what we call the *data matrix* Δ , formed as a product of three matrices,

$$\Delta \stackrel{\text{def}}{=} \mathbf{X} \mathbf{A} \mathbf{U}, \quad (23)$$

where

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K] \in \mathbb{R}^{B \times K}, \quad (24)$$

is a Vandermonde matrix with roots $\cos \theta_k$, $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_K)$ is the diagonal matrix of Dirac magnitudes, and we define

$$\mathbf{U} = [u_{km}] \in \mathbb{R}^{K \times (2B-1)}, \quad (25)$$

with $u_{km} \stackrel{\text{def}}{=} (\sin \theta_k)^{|m|} e^{-jm\phi_k}$.

It is convenient to keep a non-standard indexing scheme for the rows and columns of Δ , as illustrated in Fig. 1(B). Rows of Δ , indexed by p , correspond to decreasing powers of $\cos \theta_k$, from $p = B - 1$ at the top, to $p = 0$ at the bottom; columns correspond to u_{km} , with m increasing from $-B + 1$ on the left, to $B - 1$ on the right. We see from (22) and (23) that computing any spectral coefficient \hat{f}_ℓ^m amounts to applying a linear functional on Δ as follows

$$\hat{f}_\ell^m = \mathbf{c}_{\ell m}^\top \Delta \mathbf{e}_m = \langle \mathbf{c}_{\ell m} \mathbf{e}_m^\top, \Delta \rangle_F, \quad (26)$$

where $\mathbf{e}_m \in \mathbb{R}^{2B-1}$ is the vector with one in position m for $-B < m < B$, and zeros elsewhere, and $\langle \cdot, \cdot \rangle_F$ denotes the standard inner product between two matrices, defined as $\langle \mathbf{A}, \mathbf{B} \rangle_F = \sum_{ij} \bar{a}_{ij} b_{ij} = \text{trace}(\mathbf{A}^H \mathbf{B})$.

The last expression in (26) implies that the spectral coefficient \hat{f}_ℓ^m can be obtained as an inner product between the data matrix Δ and a mask $\mathbf{c}_{\ell m} \mathbf{e}_m^\top$ that is overlaid over Δ . One can verify that the support of this mask for \hat{f}_ℓ^m is on the column corresponding to m , and on the rows corresponding to $0 \leq p < B - |m|$. That means that certain parts of the data matrix are not involved in the creation of any spectral coefficient; consequently, they cannot be recovered from the spectrum. Nevertheless, we can recover a large part:

Lemma 1: There is a one-to-one linear mapping between the spherical harmonic coefficients in the lowpass subband, $[\hat{f}_\ell^m, (\ell, m) \in \mathcal{I}_B]$, and the triangular part of the data matrix Δ indexed by $\mathcal{J}_B = \{(p, m) | 0 \leq |m| \leq p < B\}$ (with indexing as illustrated in Fig. 1).

Proof: We see from (26) that the m th spectral column maps to the m th column of the data matrix; that is, we can decouple

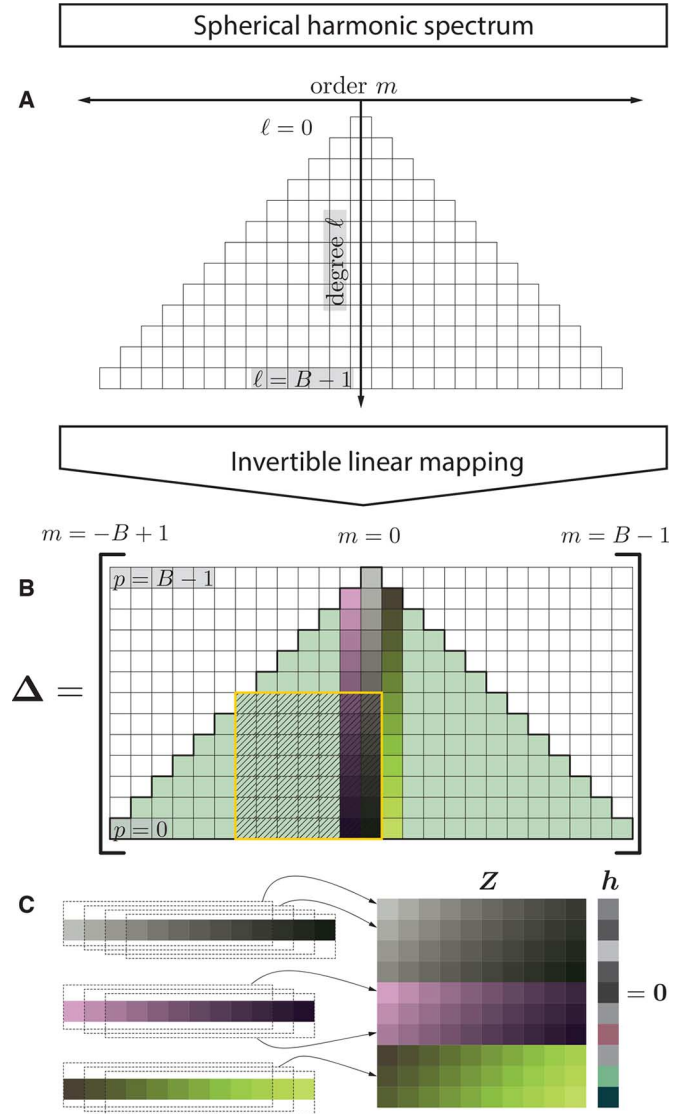


Fig. 1. Illustration of Algorithm 1. Spherical harmonic spectrum (A) is linearly mapped onto the shaded triangular part of the data matrix Δ (B). Columns of the data matrix are indexed from left to right by m , $-(B-1) \leq m \leq (B-1)$, corresponding to spherical harmonic order. Rows are indexed from bottom to top by p , $0 \leq p \leq (B-1)$ corresponding to powers of $\cos \theta$. Note that the triangular part of the data matrix *does not* coincide with the spherical harmonic spectrum, although there is a one-to-one linear mapping between the two (see Lemma 1). Existing results on 2D harmonic retrieval can exploit only a small part of the data matrix, for example the hatched square (see Section III.D). Finally, sufficiently long columns of Δ are rearranged in the block-Hankel-structured annihilation matrix \mathbf{Z} , whose nullspace contains exactly the sought annihilation filter, \mathbf{h} (C).

the mapping over m . In particular, by writing out (26) for all \hat{f}_ℓ^m of a given degree m , we obtain

$$\mathbf{f}_m = \mathbf{C} \Delta \mathbf{e}_m, \quad (27)$$

where $\mathbf{f}_m \stackrel{\text{def}}{=} [\hat{f}_{|m|}^m, \hat{f}_{|m|+1}^m, \dots, \text{Point P}, \dots, \hat{f}_{B-1}^m]^\top$ and $\mathbf{C} \stackrel{\text{def}}{=} [\mathbf{c}_{|m|,m}, \mathbf{c}_{|m|+1,m}, \dots, \mathbf{c}_{B-1,m}]^\top \in \mathbb{R}^{(B-|m|) \times B}$. From the definition of $\mathbf{c}_{\ell m}$ (21), we see that $\mathbf{C} = [\mathbf{0}_{(B-|m|) \times |m|}, \mathbf{C}_\square]$, where $\mathbf{C}_\square \in \mathbb{R}^{(B-|m|) \times (B-|m|)}$ is a square lower-antitriangular matrix with no zeros on the main antidiagonal (because

$\frac{d^{|m|}}{dx^{|m|}}P_\ell(x)$ is a polynomial of degree $\ell - |m|$. Thus \mathbf{C}_\square is an invertible matrix and we can write

$$[\Delta \mathbf{e}_m]_\square = \mathbf{C}_\square^{-1} \mathbf{f}_m, \quad (28)$$

where by $[\Delta \mathbf{e}_m]_\square$ we denoted the recoverable part of the m th column of the data matrix (shaded in green in Fig. 1(B)).

C. Reconstruction by Generalized Annihilating Filtering

An element of the data matrix Δ at the position (p, m) (with reference to Fig. 1(B)) can be expanded as

$$d_{pm} = \sum_{k=1}^K \alpha_k x_k^p (\sin \theta_k)^{|m|} e^{-jm\phi_k}, \quad (29)$$

where p varies from 0 to $B-1$, and m from $-(B-1)$ to $(B-1)$. For either positive or negative m , the sum (29) is a sum of 2D exponentials. Lemma 1 implies that we can recover the shaded triangular part of the data matrix in Fig. 1 from the spectrum. In what follows, we propose a new algorithm to recover the parameters of the Diracs from that triangular part.

The vector $\mathbf{d}_m \stackrel{\text{def}}{=} \Delta \mathbf{e}_m$ is a linear combination of columns of \mathbf{X} , i.e., it is a linear combination of K exponentials with bases x_k ,

$$d_{pm} = \sum_{k=1}^K (\alpha_k u_{km}) x_k^p, \quad (30)$$

where $x_k = \cos(\theta_k)$. Similarly to standard Euclidean FRI sampling [10], we can use the *annihilating filter* technique to estimate the roots $\{x_k = \cos \theta_k\}_{k=1}^K$ of these exponentials.

An annihilating filter is a finite impulse response (FIR) filter with zeros positioned so that it annihilates signals of the form (30). Consider an FIR filter $H(z)$ with the transfer function

$$H(z) \stackrel{\text{def}}{=} \prod_{k=1}^K (1 - x_k z^{-1}) \stackrel{\text{def}}{=} \sum_{n=0}^K h_n z^{-n}, \quad (31)$$

where $\mathbf{h} = [h_0, h_1, \dots, h_K]^\top$ is the vector of filter coefficients. It holds that $\mathbf{h} * \mathbf{d}_m \equiv \mathbf{0}$ (see Appendix A) for any m , provided that \mathbf{d}_m is of length at least $K+1$. Equivalently,

$$[d_{n,m}, d_{n-1,m}, \dots, d_{n-K,m}] \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_K \end{bmatrix} = 0, \quad (32)$$

for $n \geq K$. In our scenario, we do not know the bases of the exponentials $\{x_k\}_{k=1}^K$ —they are exactly the parameters we aim to estimate. Thus we do not know the filter $H(z)$ either.

Up to a scaling factor, there is a unique $(K+1)$ -tap filter $H(z)$ with the sought property. The orthogonality relation (32) says that \mathbf{h} lives in the nullspace of $[d_{n,m} \ d_{n-1,m} \ \dots \ d_{n-K,m}]$. We need at least K such vectors to make their joint nullspace one-dimensional, thus to pinpoint \mathbf{h} . Once the filter coefficients are found, we can obtain the unknown parameters $\{x_k\}$ by root finding and using the factorization in (31).

For the annihilating filter technique to be applicable, we need to ensure that all the colatitude angles θ_k are distinct. Furthermore, the form of our equations reveals that for $\theta_k \in \{0, \pi\}$, $u_{km} = 0$ for all m . In the parameterization (23), this is equivalent to setting $\alpha_k = 0$, and it prevents us from recovering the corresponding Dirac. This behavior is undesirable, but we can guarantee that no Dirac sits on a pole by first applying a random rotation. This fact is formalized in the following lemma, which follows immediately from the absolute continuity of the Haar measure.

Lemma 2: Consider a collection of Dirac delta functions on the sphere, $f(\xi) = \sum_{k=1}^K \alpha_k \delta(\xi; \xi_k)$, and a random rotation ϱ drawn from the Haar measure on $\mathbb{S}\mathbb{O}_3$ (i.e., uniformly over the elements of the group). Then with probability 1, $\Lambda(\varrho)f$ contains Diracs with distinct colatitude angles, $\theta_i \neq \theta_j$ for $i \neq j$, and no Dirac is on a pole, $\theta_k \notin \{0, \pi\}$ for all k .

We are now well-equipped to prove the main result.

Proof of Theorem 1: We provide a constructive proof, summarized in Algorithm 1. First observe that B^2 random samples almost surely suffice to compute the spectral coefficients \hat{f}_ℓ^m in the lowpass subband \mathcal{I}_B with bandwidth B , as detailed in Section III.A (Proposition 1). By Lemma 1, we can then compute the shaded part of Δ given the spectrum $\hat{\mathbf{f}}$.

Our aim is to construct the annihilating matrix \mathbf{Z} , structured as follows:

$$\mathbf{Z} = \begin{bmatrix} d_{B-1,0} & d_{B-2,0} & \cdots & d_{B-K-1,0} \\ d_{B-2,0} & d_{B-3,0} & \cdots & d_{B-K-2,0} \\ \vdots & \vdots & & \vdots \\ d_{K,0} & d_{K-1,0} & \cdots & d_{0,0} \\ d_{B-2,1} & d_{B-3,1} & \cdots & d_{B-K-2,1} \\ d_{B-3,1} & d_{B-4,1} & \cdots & d_{B-K-3,1} \\ \vdots & \vdots & & \vdots \end{bmatrix}. \quad (33)$$

\mathbf{Z} is constructed by stacking segments of length $(K+1)$ extracted from the columns of Δ . From the annihilation property (32), it follows that the nullspace of \mathbf{Z} contains the sought annihilating filter.

The trick now is to count how many such segments we can get from the shaded part of Δ . For $m=0$, p varies from 0 to $B-1$. Therefore, we can construct $B-K$ rows of the matrix \mathbf{Z} . For $m=1$, p varies from 0 to $B-2$, so we can construct $B-K-1$ rows of \mathbf{Z} , and the same goes for $m=-1$. This process is illustrated in Figs. 1(B) and (C). Summing up, we get the total number of rows of \mathbf{Z} that we can construct from the available spectrum,

$$\begin{aligned} \# &= (B-K) + 2 \times (B-K-1) + \cdots + 2 \times 1 \\ &= (B-K)^2. \end{aligned} \quad (34)$$

\mathbf{Z} needs at least K rows, as we need a 1D nullspace. Thus

$$\begin{aligned} (B-K)^2 &\geq K \\ \Rightarrow B &\geq K + \sqrt{K}. \end{aligned} \quad (35)$$

In Appendix C we show that \mathbf{Z} has rank K as soon as it has K or more rows. In other words, it has a one-dimensional nullspace,

and thus the annihilating filter coefficients are uniquely determined, up to a scaling factor.

We find the parameters $[\theta_k]_{k=1}^K$ by taking the arc cosine of the roots of $H(z)$. This procedure is well-posed because arc cosine is one-to-one on $[0, \pi]$. To ensure that the roots are distinct, we apply a random rotation before the estimation, and the inverse of this random rotation after recovering all the parameters of the Diracs (invoking Lemma 2).

In order to recover the azimuths $\{\phi_k\}_{k=1}^K$, note that after recovering the colatitudes, we can construct the matrix \mathbf{X} , and compute $\mathbf{A}\mathbf{U}\mathbf{e}_m$ for $|m| \leq B - K$. The azimuths are then given as the phase difference between $\mathbf{A}\mathbf{U}\mathbf{e}_0$ and $\mathbf{A}\mathbf{U}\mathbf{e}_1$. The magnitudes α_k are obtained simply as $\mathbf{A}\mathbf{U}\mathbf{e}_0$. \square

Algorithm 1: Spherical Sparse Sampling

Input: Spatial samples of $f \in L^2(\mathbb{S}^2)$ with bandwidth B , number of Diracs K

Output: Colatitudes, azimuths and magnitudes $\{(\alpha_k, \theta_k, \phi_k)\}_{k=1}^K$ of the K Diracs

- 1: Sample a random rotation $\varrho \sim \text{Haar}(\mathbb{S}\mathbb{O}_3)$
- 2: Apply ϱ to f , $f \leftarrow \Lambda(\varrho)f$ (relabel sampling points)
- 3: Compute the spectrum \hat{f} from the rotated samples of f
- 4: Form the recoverable part of $\mathbf{\Delta}$ from \hat{f} using (28)
- 5: Form \mathbf{Z} from $\mathbf{\Delta}$ according to (33)
- 6: $\mathbf{h} \leftarrow$ Right singular vector of \mathbf{Z} for smallest sing. val.
- 7: Compute the colatitudes, $(\theta_k)_{k=1}^K \leftarrow \arccos[\text{Roots}(\mathbf{h})]$
- 8: Construct \mathbf{X} from $x_k = \cos \theta_k$ according to (24)
- 9: Using \mathbf{X} in (23), compute $\mathbf{A}\mathbf{U}\mathbf{e}_0$ and $\mathbf{A}\mathbf{U}\mathbf{e}_1$
- 10: $(\phi_k)_{k=1}^K \leftarrow \text{Angle}[(\mathbf{A}\mathbf{U}\mathbf{e}_0) \oslash (\mathbf{A}\mathbf{U}\mathbf{e}_1)] \quad \triangleright$ See the note
- 11: $(\alpha_k)_{k=1}^K \leftarrow \mathbf{A}\mathbf{U}\mathbf{e}_0$
- 12: Apply the inverse of ϱ , $\xi_k = (\theta_k, \phi_k) \leftarrow \varrho^{-1} \circ \xi_k, \forall k$

\triangleright Note: we use the symbol \oslash to denote element-wise division of vectors.

D. Sampling Efficiency and Relation to Prior Work

Our proposed sampling scheme and the spherical FRI sampling theorem by Deslauriers-Gauthier and Marziliano [15] are both naturally expressed in terms of the bandwidth B of the sampling kernel required to recover K Diracs. In our case, the bandwidth requirement is that it be at least $K + \sqrt{K}$. This implies that we need at least $(K + \sqrt{K})^2$ spatial samples to recover the K Diracs. For comparison, the FRI sampling theorem of Deslauriers-Gauthier and Marziliano [15] requires $B \geq 2K$, and thus their algorithm recovers K Diracs given $4K^2$ samples. This is asymptotically four times the number of samples required by Algorithm 1.

The difference in sampling efficiency can be explained by spectrum usage. Fig. 2 illustrates the portion of the spectrum used by the two algorithms. We can see that the proposed algorithm is more efficient in that it uses a larger portion of the available spectrum to reconstruct the Diracs.

Similar problems have been considered in the literature on 2D harmonic retrieval [26]. However, these earlier works assume that the entire data matrix is known. In our case, $\mathbf{\Delta}$ is known

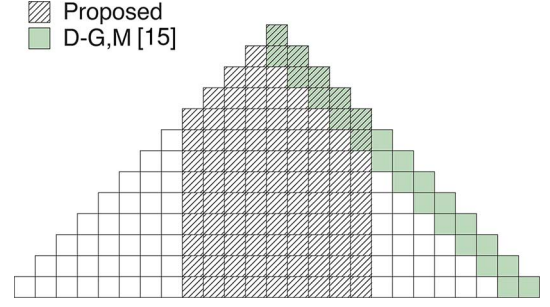


Fig. 2. Spectrum usage for different algorithms. Spectral coefficients used by our algorithms are shown hatched. Spectrum used by the algorithm of Deslauriers-Gauthier and Marziliano [15] is shaded green. In the example, the bandwidth is set to $B = 12$, so the maximum number of Diracs that can be recovered by Algorithm 1 is $K = 9$. The algorithm in [15] recovers $K = 6$ Diracs.

only partially, as illustrated in Fig. 1(B). To apply the existing results on 2D harmonic retrieval, we could use a square portion that falls strictly inside a half of the triangle, either for $m \geq 0$ or for $m \leq 0$. However, we can see in Fig. 1(B) that this is an inefficient use of available spectrum, and it requires an unnecessarily high sampling density.

As mentioned earlier, in most situations we do not get to choose L as it is fixed by the underlying physical process. Then the question is how many Diracs we can reconstruct given a kernel with a fixed bandwidth L . By solving $B \geq K + \sqrt{K}$ for K , we get that

$$K \leq B - \left(B + \frac{1}{4}\right)^{1/2} + \frac{1}{2}. \quad (36)$$

In contrast, the algorithm in [15] can reconstruct up to $K = B/2$ Diracs.

E. Denoising Strategies

Theorem 1 and Algorithm 1 provide a tool to recover sparse signals on the sphere in the noiseless case. We may apply several procedures to improve the robustness of the algorithm in the presence of noise.

In general, if the samples are noisy then the annihilating matrix \mathbf{Z} in (33) will not have a nontrivial nullspace. A simple and robust approach is to use the right singular vector corresponding to the smallest singular value of \mathbf{Z} as the annihilation filter. Let $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ be the SVD of \mathbf{Z} ; then we set $\mathbf{h} = \mathbf{v}_{K+1}$.

To further improve the algorithm performance, we can use the output of Algorithm 1 to initialize a local search for the minimizer of the ℓ^2 error between the spectrum generated by the estimated Diracs, and the measured spectrum,

$$\underset{(\tilde{\alpha}_k, \tilde{\theta}_k, \tilde{\phi}_k)_{k=1}^K}{\text{minimize}} \sum_{\ell=0}^{B-1} \sum_{m=-\ell}^{\ell} \left| \hat{f}_{\ell}^m - \sum_{k=1}^K \tilde{\alpha}_k Y_{\ell}^m(\tilde{\theta}_k, \tilde{\phi}_k) \right|^2. \quad (37)$$

We note that directly solving (37) with a random starting point is hopeless due to a multitude of local minima.

F. Cramér-Rao Lower Bound

We evaluate the Cramér-Rao lower bound (CRLB) for the estimation problem. For simplicity we treat the $K = 1$ case,

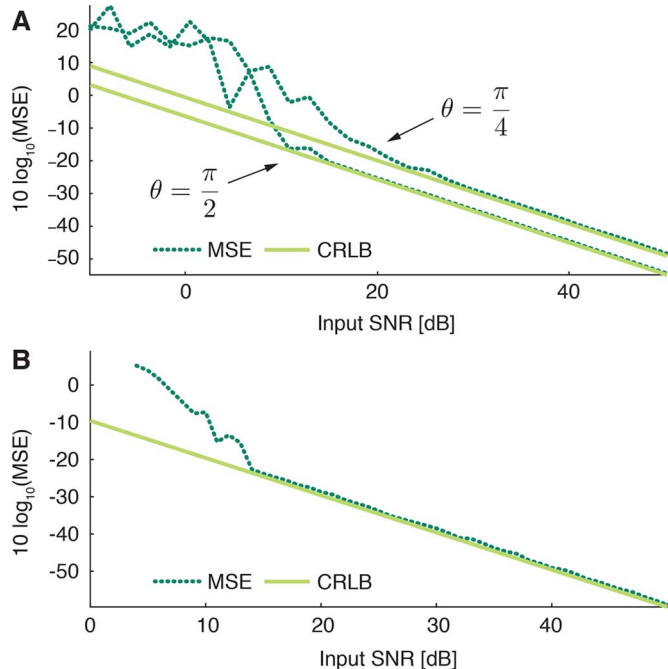


Fig. 3. Comparison between the mean squared error (MSE) of the proposed algorithm in estimating the spherical location (θ_k, ϕ_k) and the Cramér-Rao lower bound (CRLB). The location of a single spike ($K = 1$) was estimated for two different spike colatitudes (A). Note that the bound is different for different colatitudes of the spike, due to the dependence on the parameterization. Estimating the locations of two spikes ($K = 2$) at $(\pi/2, 1)$ and $(\pi/4, 2)$ (B) yields similar output MSE as estimating either of them separately. In both cases, the MSE is shown for the output of Algorithm 1 followed by the minimization of (37) using Matlab's `fminsearch` function.

so that the minimal bandwidth is $L = 2$, and $\ell \in \{0, 1\}$. We assume that the spatial samples are taken on the sampling grid defined by McEwen-Wiaux [8], given at this bandwidth as

$$\begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\phi} \end{bmatrix} = \begin{bmatrix} \pi/3 & \pi/3 & \pi/3 & \pi & \pi & \pi \\ 0 & 0 & 2\pi/3 & 2\pi/3 & 4\pi/3 & 4\pi/3 \end{bmatrix}. \quad (38)$$

Resulting expressions for elements of the Fisher information matrix are complicated, and there is no need to exhibit them explicitly. We give the details of the computation in Appendix B, and we compute the CRLB numerically. The resulting bound is plotted in Fig. 3 for $K = 1$, $K = 2$, and two different spike colatitudes, together with the MSE achieved by Algorithm 1 followed by the descent (37). As pointed out before, because our scheme is coordinate-system-dependent, the bound depends on the colatitude of the Dirac.

IV. APPLICATIONS

To showcase the versatility of the proposed algorithm, we present three stylized applications: 1) sampling diffusion processes on the sphere, 2) shot noise removal, and 3) sound source localization with spherical microphone arrays.

A. Sampling Diffusion Processes on the Sphere

The diffusion process models many natural phenomena. Examples include heat diffusion and plume spreading from a smokestack. Often, the source of the diffusion process is localized in space and instantaneous in time. Sampling

such processes in Euclidean domains has been well studied [27]–[29].

Diffusion processes on the sphere are governed by the following equation [30]:

$$k\Delta v(\xi, t) = \frac{\partial}{\partial t} v(\xi, t), \quad (39)$$

where Δ is the Laplace-Beltrami operator on \mathbb{S}^2 , and k is the diffusion constant. In the spherical harmonic domain, this becomes

$$-k\ell(\ell+1)\widehat{v}_\ell^m(t) = \frac{\partial}{\partial t} \widehat{v}_\ell^m(t), \quad (40)$$

giving the solution

$$\widehat{v}_\ell^m(t) = e^{-\ell(\ell+1)kt} \widehat{v}_\ell^m(0), \quad (41)$$

where $\widehat{v}_\ell^m(0)$ is the spectrum of the initial distribution. Therefore, we interpret the term $e^{-\ell(\ell+1)kt} \delta_{m,0}$ as the spectrum of the Green's function of the spherical diffusion equation. In other words, it is the spectrum of the diffusion kernel on the sphere. Then (41) should be interpreted as the convolution between the kernel and the initial distribution.

We consider the case when the diffusion process is initiated by K sources localized in space and time, i.e., the initial distribution in (41) is

$$v(\xi; 0) = \sum_{k=1}^K \alpha_k \delta(\xi; \xi_k). \quad (42)$$

We show how to use the proposed sampling algorithm to estimate the locations and the strengths of the sources from spatial samples of the diffusion field taken at a later time t_0 . Recovering all parameters (locations, amplitudes and release times) of multiple diffusion sources is a challenging task [27]. To focus on the proposed sampling result, we make the simplifying assumption that the K sources are released simultaneously, and at a known time ($t = 0$). In principle, the more challenging case of unknown and different release times can be handled by adapting the techniques derived in [28], [29], but these generalizations are out of the scope of this work.

We note that the parameters of the simultaneous diffusion sources could be recovered by optimal filtering on the sphere [31]. This approach is designed to perform favorably in noise, but it assumes that the function is sampled on a dense grid. Our perspective is sampling-theoretic: we want to estimate these parameters from as few samples as possible.

In the spatial domain, the diffusion kernel at time t_0 after the release is given as

$$h_{\text{dif}}(\xi; t_0) = \sum_{\ell=0}^{\infty} e^{-\ell(\ell+1)kt_0} Y_\ell^0(\xi). \quad (43)$$

Combining (43) with (41) and the spherical convolution-multiplication rule (13), we get

$$\begin{aligned} v(\xi; t_0) &= v(\xi; 0) * h_{\text{dif}}(\xi; t_0) \\ &= \sum_{k=1}^K c_k [\Lambda(\varrho_k) h_{\text{dif}}(\cdot; t_0)](\xi). \end{aligned}$$

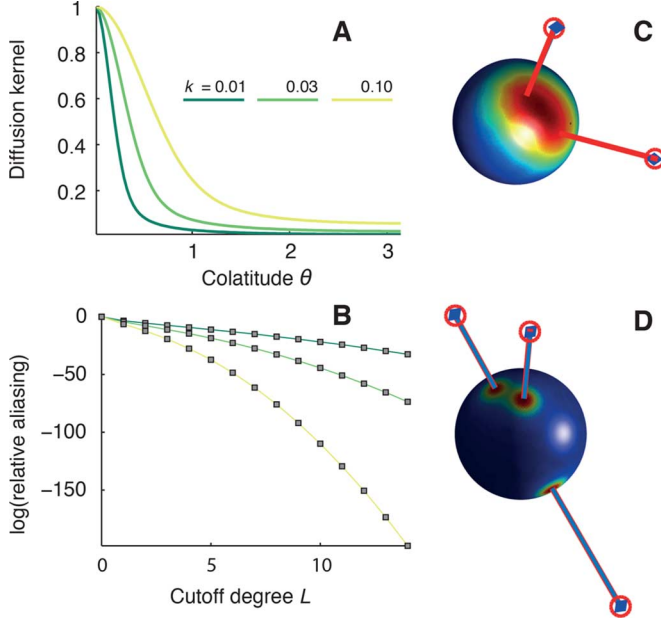


Fig. 4. Estimating the release locations and magnitudes of diffusive sources on the sphere. We assume that the diffusive sources appear at time $t = 0$ s, and that we sample the field at time $t = 1$ s. Shape of the diffusion kernel as a function of θ is shown in subfigure A for three different values of the coefficient k (in units of inverse time). The logarithm of the aliasing error (44) is plotted as a function of the cutoff degree B in subfigure B. Subfigures C and D show a typical reconstruction result for $k = 0.1$ (2 sources) and $k = 0.01$ (3 sources). Magnitudes of the sources are represented by the distance of the corresponding symbols from the sphere's center. Blue diamonds represent true source locations and magnitudes, while red circles represent estimated source locations and magnitudes. The sphere color corresponds to the value of the function induced on the sphere by the sources (red is large, blue is small). Signal-to-noise ratio in both C and D was set to 30 dB. We used the approximate bandwidth of $B = 7$, so that the number of samples taken in either case was 49.

This signal is a sum of rotations of a known template. The diffusion kernel in (43) is not exactly bandlimited, but it is approximately so. We can therefore apply the spherical FRI theory and Algorithm 1 to recover the locations and the magnitudes of the diffusive sources.

Fig. 4(A) shows the shape of the symmetric diffusion kernel as a function of the colatitude θ . The high degree of smoothness is reflected in an approximately bandlimited spectrum. This is demonstrated in Fig. 4(B), where we see that the aliasing energy due to spectral truncation, defined as

$$\varepsilon(B) = \frac{1}{\|v\|_2^2} \sum_{\ell=B}^{\infty} \frac{|\hat{v}_\ell^m|^2}{2\ell+1}, \quad (44)$$

rapidly becomes negligible as we increase the cutoff bandwidth L . Figs. 4(C) and (D) demonstrate accurate reconstruction of the localized diffusion sources at two different values of the diffusion coefficient (the detailed parameters of the numerical experiment are given in the figure caption).

B. Shot Noise Removal

Suppose that we sample a bandlimited function on the sphere, but a small number of samples are corrupted—they contain *shot noise*—due to sensor malfunction. Moreover, the identities of the malfunctioning sensors are not known *a priori*. Can we detect and correct these anomalous measurements? We show that

our sampling results can be applied to solve this problem, provided that the number of erroneous sensors is not too large and that the original sampling grid is oversampling the bandlimited function. A similar idea was used in [26] to remove shot noise in the 1D Euclidean case.

For this application we assume that the samples are taken on a uniform grid on the sphere, $\{(\theta_p, \phi_q) | p, q \in \mathbb{Z}, 0 \leq p < 2B', 0 \leq q < 2B'\}$, defined by

$$\theta_p = \frac{p\pi}{2B'}, \quad \phi_q = \frac{q\pi}{B'}. \quad (45)$$

Imagine now that we sample f on this sampling grid. Some samples are corrupted, so we measure $g(\theta_p, \phi_q) = f(\theta_p, \phi_q) + s_{pq}$, where

$$s_{pq} = \begin{cases} \text{nonzero} & (p, q) \in \mathcal{S} \\ \text{zero} & \text{otherwise,} \end{cases} \quad (46)$$

and \mathcal{S} holds the indices of the corrupted samples. We will leverage an elegant quadrature rule by Driscoll and Healy [7]:

Theorem 2 ([7], Theorem 3): Let f be a bandlimited function on \mathbb{S}^2 such that $\hat{f}_\ell^m = 0$ for $\ell \geq B'$. Then for $(\ell, m) \in \mathcal{I}_{B'}$ we have

$$\hat{f}_\ell^m = \sum_{p=0}^{2B'-1} \sum_{q=0}^{2B'-1} a_p^{(B')} f(\theta_p, \phi_q) \overline{Y_\ell^m(\theta_p, \phi_q)}, \quad (47)$$

where the weights $a_p^{(B')}$ are defined in [7].

In other words, the Fourier coefficients \hat{f}_ℓ^m can be expressed as a dot-product between weighted sample values and the basis functions evaluated at the sampling points. In analogy with the Euclidean case, we now observe that the lowpass portion of the spectrum of f coincides with the lowpass portion of the spectrum of the generalized function obtained by placing weighted Diracs at grid points. Let f be bandlimited so that $\hat{f}_\ell^m = 0$ for $\ell \geq B$. Let further $B < B'$; that is, the grid (45) oversamples f . Then the spectral coefficients can be expressed as the following inner product,

$$\hat{f}_\ell^m = \left\langle \sum_{p,q=0}^{2B'-1} a_p^{(B')} f(\theta_p, \phi_q) \delta_{\theta_p, \phi_q}, Y_\ell^m \right\rangle, \quad (48)$$

for $\ell < B$, $|m| \leq \ell$.

This is the key insight. Notice that the lowpass portion of the spectrum of g (for $\ell < B'$) can be written as

$$\hat{g}_\ell^m = \hat{f}_\ell^m + \left\langle \sum_{(p,q) \in \mathcal{S}} a_p^{(B')} s_{pq} \delta_{\theta_p, \phi_q}, Y_\ell^m \right\rangle. \quad (49)$$

But $\hat{f}_\ell^m = 0$ for $\ell \geq B$, so the portion of the spectrum for $B \leq \ell < B'$ contains only the influence of the corrupted samples,

$$\hat{g}_\ell^m = \left\langle \sum_{(p,q) \in \mathcal{S}} a_p^{(B')} s_{pq} \delta_{\theta_p, \phi_q}, Y_\ell^m \right\rangle, \quad B \leq \ell < B'. \quad (50)$$

Consequently, we can use this part of the spectrum to learn which samples are corrupted, and by how much. This is the subject of the following proposition.

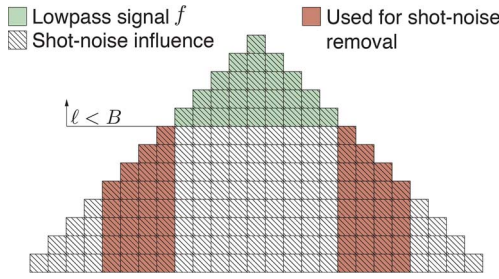


Fig. 5. Spectrum structure in shot noise removal. Green-shaded bins get contribution from the desired signal f with bandwidth B ; hatched bins are influenced by the shot noise; red-shaded columns are (i) long enough to annihilate shot noise and (ii) recoverable from the corrupted spectrum.

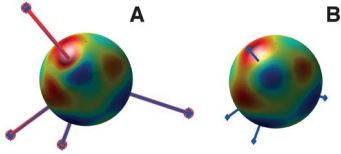


Fig. 6. Shot noise removal via spherical FRI, for $B = 6$, $B' = 12$ and $K = 4$ malfunctioning sensors. Corrupted signal is shown in subfigure A, together with the true corruption values (blue diamonds) and the estimated corruptions (red circles); same signal with the shot noise removed is shown in B, with the correct sample values at the corrupted locations denoted by blue diamonds.

Proposition 2: Let f be a signal on the sphere of bandwidth B . Then we can perfectly reconstruct f from corrupted samples taken on the grid (45), as long as the number of corruptions K satisfies

$$K \leq B' - B - \sqrt{B' - B + 1} + 1. \quad (51)$$

Proof: As discussed in Section III, we can use any $m = \text{const.}$ line in the spectrum to get the rows of the annihilation matrix. However, we first need to compute the corresponding columns of the data matrix. From Fig. 5, we see that the middle columns cannot be used for shot noise removal: we seek columns influenced only by corruptions. But the middle columns of the data matrix are obtained from the middle spectral columns (for $m < B$), so they are influenced both by the desired signal and the corruptions. This means that we can only use spectral bins for $B \leq m < B'$, as illustrated in Fig. 5. For $m = B$ and $m = -B$, the number of segments of length $K + 1$ that we can get is $B' - B - K$. For $m = B + 1$ and $m = -(B + 1)$ it is $B' - B - K - 1$, and so on. Summing up we have that the total number of segments of length $K + 1$ we can use is

$$\begin{aligned} \# &= 2(B' - B - K) + 2(B' - B - K - 1) + \dots + 2 \cdot 1 \\ &= (B' - B - K)(B' - B - K + 1). \end{aligned}$$

We need this number to be at least K , because we need K rows in the annihilation matrix. We thus obtain the claim of the proposition by solving the inequality $\# \geq K$. \square

After detecting the corrupted readings, we can use the estimated corruption values to estimate the function. Another option is to simply ignore them altogether, as we have more samples than the minimum number thanks to oversampling. A shot noise removal experiment is illustrated in Fig. 6.

C. Sound Source Localization

Spherical microphone arrays output a time-varying spherical signal. If the signal is induced by a collection of point sources, we can use the proposed spherical FRI sampling scheme to estimate the directions-of-arrival (DOAs) of the sources. For simplicity, we consider the narrowband case, i.e., the sources emit a single sinusoid.

How does this example fit into our sparse sampling framework? In spherical microphone arrays, the microphones are distributed on the surface of a sphere, either open or rigid [5]. Therefore, the microphone signals represent samples of a time-varying function on \mathbb{S}^2 . If a sound source emits a sinusoid, every microphone measures the amplitude and the phase of that sinusoid shaped by the characteristics of the propagating medium and of the spherical casing. Equivalently, for every microphone we get a complex number.

Suppose that a source of unit intensity is located at \mathbf{s} , and that the microphones are mounted on a rigid sphere of radius r with center at the origin. The response measured by the microphone at \mathbf{r} , such that $\|\mathbf{r}\| = r$, is given by the corresponding Green's function. For a wavenumber $\kappa = 2\pi\nu/c$, where ν is the frequency and c is the speed of sound, the Green's function is [5]

$$g(\mathbf{r}|\mathbf{s}, \kappa) = \frac{j\kappa}{4\pi} \sum_{\ell=0}^{\infty} b_{\ell}(\kappa r) h_{\ell}^{(1)}(\kappa r) (2\ell + 1) P_{\ell}(\cos \alpha_{\mathbf{r}\mathbf{s}}), \quad (52)$$

where $h_{\ell}^{(1)}$ is the spherical Hankel function of the first kind and of order ℓ , P_{ℓ} is the Legendre polynomial, and $\cos \alpha_{\mathbf{r}\mathbf{s}} = \frac{1}{rs} \langle \mathbf{r}, \mathbf{s} \rangle$. Mode strength $b_{\ell}(\kappa r)$ is defined as

$$b_{\ell}(\kappa r) \stackrel{\text{def}}{=} j_{\ell}(\kappa r) - \frac{j'_{\ell}(\kappa r)}{h_{\ell}^{(1)'}(\kappa r)} h_{\ell}^{(1)}(\kappa r), \quad (53)$$

where j_{ℓ} is the spherical Bessel function³ of order ℓ , and prime $(\cdot)'$ denotes the derivative with respect to the argument.

The Green's function g should be seen as a filter that describes how the point source's influence spreads over the sphere. It is shown for two different frequencies in Figs. 7(A) and (D), while the corresponding spectra are given in Figs. 7(B) and (E). We see that the absolute pressure on the sphere has a similar shape for both frequencies, but the real and imaginary parts vary faster at higher frequencies, implying higher bandwidth. In both cases we observe that the Green's function is approximately bandlimited.

Assume now that there are K sound sources at locations $\{\mathbf{s}_k\}_{k=1}^K$, with complex intensities $\{\alpha_k\}_{k=1}^K$. The resulting measurement by a microphone at point \mathbf{r} is

$$f(\mathbf{r}) = \sum_{k=1}^K \alpha_k g(\mathbf{r}|\mathbf{s}_k, \kappa). \quad (54)$$

If all the source locations \mathbf{s}_k are at the same distance from the sphere, then the Green's function (52) depends only on the angle between \mathbf{r} and \mathbf{s} . For some fixed source distance d , we can define $h_{\text{SSL}}(\xi) \stackrel{\text{def}}{=} g(\mathbf{x}_{\xi} r | \mathbf{x}_{\eta} d, \kappa)$, where \mathbf{x}_{ξ} denotes the unit

³We use the standard symbol j_{ℓ} for the spherical Bessel function. Note the subtle difference from the imaginary unit j .

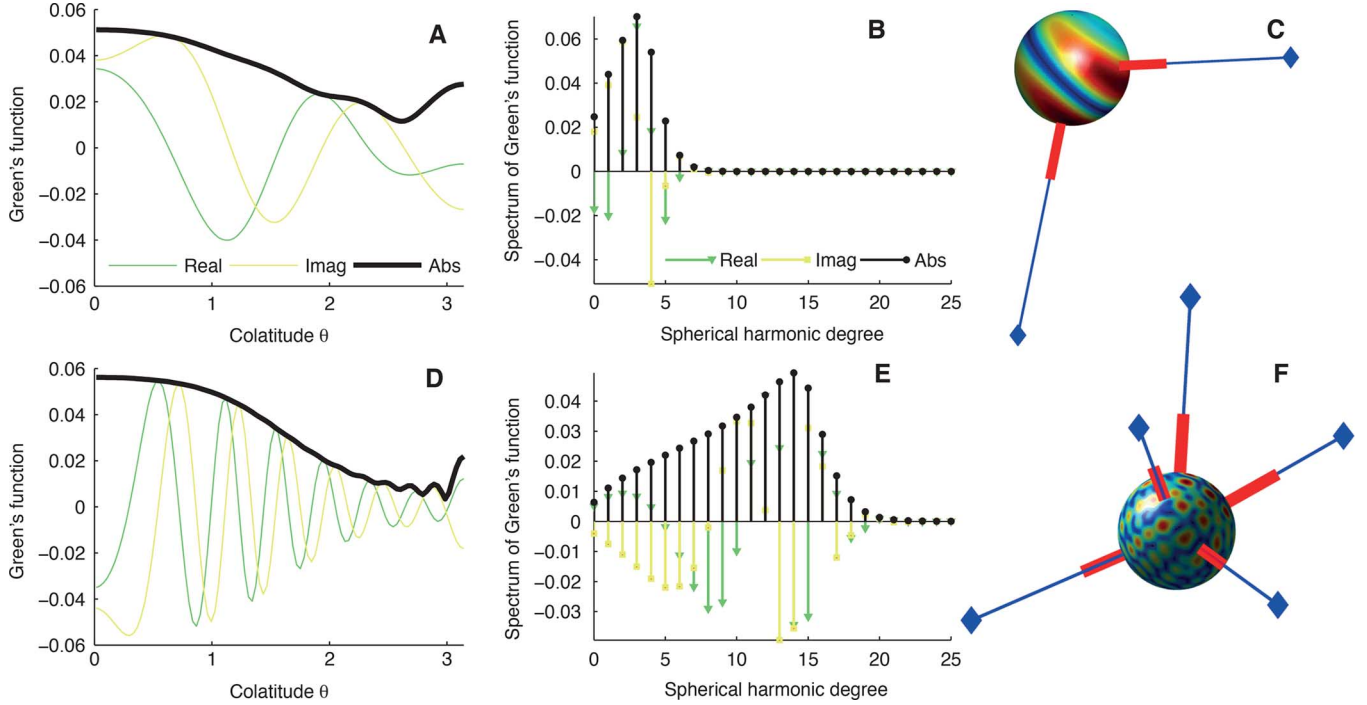


Fig. 7. Multiple DOA estimation by a spherical microphone array. First row of subfigures corresponds to $f_1 = 1000$ Hz, and second row to $f_2 = 4000$ Hz. The sphere is of radius $r = 0.2$ m, and the source is located at $[0, 0, 3]^T$ m. The real and imaginary parts, and the absolute value of the Green's function are shown in subfigures A and D. Real part, imaginary part and absolute value of the spectrum are shown in subfigures B and E. Subfigures C and F show the simulation results for $K = 2$ and $K = 5$, and random source placement. Blue diamonds represent the source locations, and thick red lines show the estimated directions. Size of the sphere is exaggerated for the purpose of illustration. The sphere color corresponds to the absolute value of the function induced on the sphere by the sources (microphones measure samples of this function). The bandwidth was set to $B = 12$ at 1000 Hz and to $B = 30$ at 4000 Hz.

vector corresponding to ξ , \mathbf{x}_η the unit vector corresponding to the north pole η , and the subscript SSL stands for *sound source localization*. Then (54) corresponds to a weighted sum of K rotations of a known template function h_{SSL} ,

$$f(\xi) = \sum_{k=1}^K \alpha_k h_{\text{SSL}}(\varrho_k^{-1} \circ \xi). \quad (55)$$

As it is unrealistic to assume that the sources are all at the same distance, we hope that the shape of $g(\mathbf{r}|\mathbf{s}, \kappa)$ does not (strongly) depend on $\|\mathbf{s}\|$. Indeed, it turns out that the shape is approximately preserved within a certain range, as illustrated in Fig. 8. We therefore suppress the dependency of g on $\|\mathbf{s}\|$ and approximate (54) as follows,

$$\begin{aligned} f(\xi) &= \sum_{k=1}^K \alpha_k g(\xi|\mathbf{s}_k, \kappa) \approx \sum_{k=1}^K \tilde{\alpha}_k h_{\text{SSL}}(\varrho_k^{-1} \circ \xi) \\ &= \left[\sum_{k=1}^K \tilde{\alpha}_k \delta(\xi; \xi_{\mathbf{s}_k}) \right] * h_{\text{SSL}}(\xi). \end{aligned} \quad (56)$$

Here, we absorbed α_k and additional (complex) scaling due to different distances into $\tilde{\alpha}_k$, and h_{SSL} is computed at some predefined *mean* distance.

We thus reduced the sound source localization problem to a problem of finding the parameters of a weighted sum of Diracs. In order to apply our spherical FRI algorithm, we need to verify that g is bandlimited on the sphere. Figs. 7(B) and (E) show that

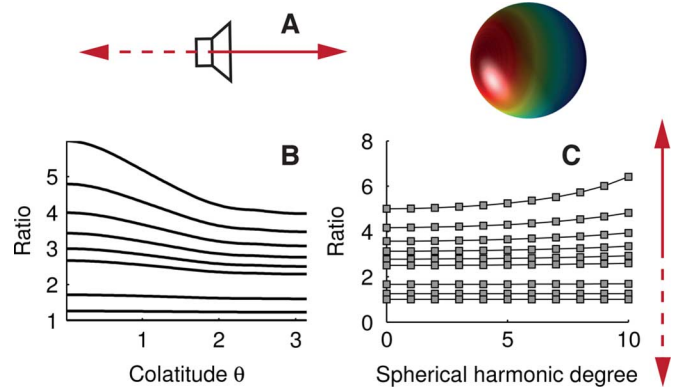


Fig. 8. Ratios of Green's functions. We computed the Green's function for nine different source distances (1.0 m, 1.2 m, 1.4 m, 1.6 m, 1.8 m, 2.0 m, 3.0 m, 4.0 m, 5.0 m). Then we plotted the magnitude of the ratio of the Green's function at each distance and the Green's function at the largest distance (5 m), both in space (B) and in the spectrum (C). The more parallel the ratio curve is with the abscissa axis, the more similar the Green's function at that distance is to the Green's function at 5 m. Curves are plotted in the order of increasing distance in the direction of the dashed arrow (up to down), as indicated in (A).

it is indeed approximately bandlimited, and that the bandwidth depends on the frequency (it also depends on the sphere radius).

Figs. 7(C) and (F) show an example of recovering two sources at 1000 Hz and five sources at 4000 Hz using the proposed spherical sparse sampling scheme. It is worth noting that the recovery is accurate in spite of the model mismatch due to varying source distances. This provides some numerical indications on the robustness of the proposed reconstruction algorithm.

V. CONCLUSION

We presented a new sampling theorem for sparse signals on the sphere. In particular, by leveraging ideas from finite rate of innovation sampling, we showed how to reconstruct sparse collections of spikes on the sphere from their lowpass-filtered observations. Compared to existing sparse sampling schemes on the sphere, we use the available spectrum more efficiently by generalizing known results on 2D harmonic retrieval. This allows us to reduce the number of samples required to reconstruct the parameters of the spikes.

We illustrated the usefulness of our algorithm by applying it to three problems: sampling diffusion processes, shot noise removal, and sound source localization. But there is a wealth of other applications, for example in astronomy. Just think about the numerous spherical signal processing challenges put forward by the square kilometer array (SKA) project [33].

We mentioned some approaches to estimation from noisy samples, but more efficient denoising schemes should be studied. One example, effective in the Euclidean setting, is the Cadzow denoising algorithm [34]. The problem seems more challenging on the sphere; in particular, the annihilating matrix is block-Hankel, rather than Hankel. It is particularly important to understand the interplay between the noise level and the achievable resolution, and to establish relevant comparisons with the results in [18], [19].

APPENDIX A

A. Annihilating Property

For the sake of completeness, we show in this Appendix that the annihilation filter annihilates linear combinations of exponentials. We compute the response of the filter $H(z)$ in (31) to a signal of the form $y_n = \sum_{k=1}^K b_k x_k^n$ as

$$\begin{aligned} (y * h)_n &= \sum_{m=0}^K y_{n-m} h_m = \sum_{m=0}^K \left(\sum_{k=1}^K b_k x_k^{n-m} \right) h_m \\ &= \sum_{k=1}^K x_k^n b_k \sum_{m=0}^K h_m x_k^{-m} = \sum_{k=1}^K x_k^n b_k \prod_{i=1}^K (1 - x_k x_i^{-1}) \\ &= 0. \end{aligned}$$

B. Computation of the Cramér-Rao Lower Bound

A lowpassed collection of K Diracs can be written as follows,

$$f(\eta, \nu) = \sum_{\ell=0}^{B-1} \sum_{m=-\ell}^{\ell} \left(\sum_{k=1}^K \alpha_k \overline{Y_{\ell}^m(\theta_k, \phi_k)} \right) Y_{\ell}^m(\eta, \nu). \quad (57)$$

We take samples on the sphere at the locations $\{(\eta_n, \nu_n)\}_{n=1}^N$. The n th sample is given by

$$\mu_n = f(\eta_n, \nu_n) + \varepsilon_n \quad (58)$$

where $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$ independently. By $\zeta = [\alpha_1, \theta_1, \phi_1, \dots, \alpha_K, \theta_K, \phi_K]^T$, we denote the vector of

parameters we estimate. To make the dependence on ζ explicit, we rewrite (58) slightly as

$$\mu_n = f_n(\zeta) + \varepsilon_n, \quad (59)$$

where $f_n(\zeta) \stackrel{\text{def}}{=} f(\eta_n, \nu_n)$.

With this notation in hand, we can write the conditional probability density function of the n th measurement as

$$p(\mu|\zeta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[\mu - f_n(\zeta)]^2 / (2\sigma^2)}, \quad (60)$$

so that the log-likelihood function is

$$\begin{aligned} L(\zeta) &\stackrel{\text{def}}{=} \ln p(\mu_1, \dots, \mu_N | \zeta) \\ &= \sum_{n=1}^N \left[-\frac{1}{2} \ln(2\pi\sigma^2) - (\mu_n - f_n(\zeta))^2 / (2\sigma^2) \right]. \end{aligned} \quad (61)$$

Consequently, differentiating L with respect to any entry ζ_i of ζ gives $\frac{\partial L}{\partial \zeta_i} = \frac{1}{\sigma^2} \sum_{n=1}^N \varepsilon_n \frac{\partial f_n(\zeta)}{\partial \zeta_i}$. The required derivatives are:

$$\begin{aligned} \frac{\partial f_n(\zeta)}{\partial \alpha_k} &= \sum_{\ell=0}^{B-1} \sum_{|m| \leq \ell} \overline{Y_{\ell}^m(\theta_0, \phi_0)} Y_{\ell}^m(\eta_n, \nu_n), \\ \frac{\partial f_n(\zeta)}{\partial \theta_k} &= \alpha_0 \sum_{\ell=0}^{B-1} \sum_{|m| \leq \ell} \left[m \cot \theta_0 \overline{Y_{\ell}^m(\theta_0, \phi_0)} \right. \\ &\quad \left. + \sqrt{(l-m)(l+m+1)} e^{j\phi_0} \overline{Y_{\ell}^{m+1}(\theta_0, \phi_0)} \right] Y_{\ell}^m(\eta_n, \nu_n), \\ \frac{\partial f_n(\zeta)}{\partial \phi_k} &= \alpha_0 \sum_{\ell=0}^{B-1} \sum_{|m| \leq \ell} (-jm) \overline{Y_{\ell}^m(\theta_0, \phi_0)} Y_{\ell}^m(\eta_n, \nu_n). \end{aligned}$$

Now $\nabla L = [\frac{\partial L}{\partial \alpha_1}, \frac{\partial L}{\partial \theta_1}, \frac{\partial L}{\partial \phi_1}, \dots, \frac{\partial L}{\partial \alpha_K}, \frac{\partial L}{\partial \theta_K}, \frac{\partial L}{\partial \phi_K}]^T$, and the Fisher information matrix is

$$\mathbf{I}(\zeta) \stackrel{\text{def}}{=} \mathbb{E}[\nabla L(\zeta) \nabla L(\zeta)^H] = \frac{1}{\sigma^2} \sum_{n=1}^N \nabla f_n(\zeta) \nabla f_n(\zeta)^H.$$

Let $\hat{\zeta}$ be any unbiased estimator of the parameters ζ . The CRLB can then be computed as $\text{cov}(\hat{\zeta}) \succeq \mathbf{I}(\zeta)^{-1}$.

C. Rank of the Annihilating Matrix

In this Appendix, we show that the rank of the annihilating matrix \mathbf{Z} (33) is K with probability one, as soon as it has at least K rows. It then follows that the annihilating filter \mathbf{h} is uniquely determined, up to a scaling factor, by solving $\mathbf{Z}\mathbf{h} = \mathbf{0}$.

Consider the factorization $\mathbf{\Delta} = \mathbf{X}\mathbf{A}\mathbf{U}$,

$$\mathbf{\Delta} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_k \\ \vdots & & \vdots \\ x_1^{B-1} & \cdots & x_k^{B-1} \end{bmatrix} \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_K \end{bmatrix} \begin{bmatrix} u_{1,1-B} & \cdots & 1 & \cdots & u_{1,B-1} \\ \vdots & & & & \vdots \\ u_{K,1-B} & \cdots & 1 & \cdots & u_{1,B-1} \end{bmatrix},$$

where $x_k = \cos \theta_k$ and $u_{k,m} = (\sin \theta_k)^{|m|} e^{-jm\phi_k}$.

To construct the annihilating matrix \mathbf{Z} as in (33), we create Hankel blocks from columns of $\mathbf{\Delta}$. The $(B - K) \times (K + 1)$ Hankel block corresponding to the middle ($m = 0$) column of $\mathbf{\Delta}$ can be factored as

$$\mathbf{B}_0 = \begin{bmatrix} x_1^{B-K-1} & \cdots & x_K^{B-K-1} \\ \vdots & & \vdots \\ x_1 & \cdots & x_K \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_K \end{bmatrix} \begin{bmatrix} x_1^K & \cdots & x_1^0 \\ \vdots & & \vdots \\ x_K^K & \cdots & x_K^0 \end{bmatrix} \stackrel{\text{def}}{=} \mathbf{X}_0 \mathbf{A} \Xi. \quad (62)$$

The second block of the annihilating matrix obtained from the column corresponding to $m = -1$ is similar: $\mathbf{B}_{-1} = \mathbf{X}_1 \mathbf{Y}_{-1} \mathbf{A} \Xi$, where $\mathbf{Y}_{-1} \stackrel{\text{def}}{=} \text{diag}(u_{1,-1}, \dots, u_{K,-1})$, and \mathbf{X}_m is obtained by removing the m leading rows from \mathbf{X}_0 . Then we can write

$$\mathbf{Z} = \left[\mathbf{B}_0^\top, \mathbf{B}_{-1}^\top, \mathbf{B}_1^\top, \dots, \mathbf{B}_{K-B+1}^\top, \mathbf{B}_{B-K-1}^\top \right]^\top = \begin{bmatrix} \mathbf{X}_0 & \cdot & \mathbf{I} & \cdot & \mathbf{A} \Xi \\ \mathbf{X}_1 & \cdot & \mathbf{Y}_{-1} & \cdot & \mathbf{A} \Xi \\ \mathbf{X}_1 & \cdot & \mathbf{Y}_1 & \cdot & \mathbf{A} \Xi \\ \vdots & & \vdots & & \vdots \\ \mathbf{X}_{B-K-1} & \cdot & \mathbf{Y}_{K-B+1} & \cdot & \mathbf{A} \Xi \\ \mathbf{X}_{B-K-1} & \cdot & \mathbf{Y}_{B-K-1} & \cdot & \mathbf{A} \Xi \end{bmatrix}, \quad (63)$$

with the $\mathbf{A} \Xi$ factor being common for all row-blocks. We want to show that the nullspace of \mathbf{Z} has dimension one. To that end, we just need to establish that the following matrix,

$$\mathbf{T} = \begin{bmatrix} \mathbf{X}_0 & \cdot & \mathbf{I} \\ \mathbf{X}_1 & \cdot & \mathbf{Y}_{-1} \\ \mathbf{X}_1 & \cdot & \mathbf{Y}_1 \\ \vdots & & \vdots \\ \mathbf{X}_{B-K-1} & \cdot & \mathbf{Y}_{K-B+1} \\ \mathbf{X}_{B-K-1} & \cdot & \mathbf{Y}_{B-K-1} \end{bmatrix} \quad (64)$$

has full column rank. To see why this is the case, let \mathbf{v} be a non-zero vector such that $\mathbf{0} = \mathbf{Z} \mathbf{v} = \mathbf{T} (\mathbf{A} \Xi \mathbf{v})$. It then follows from the full-rankness of \mathbf{T} that $\mathbf{A} \Xi \mathbf{v} = \mathbf{0}$. Since \mathbf{A} is a diagonal matrix with non-zero entries on the diagonal and Ξ is a $K \times (K + 1)$ Vandermonde matrix with distinct roots, the vector \mathbf{v} is uniquely determined up to a multiplicative factor. We now show that the matrix \mathbf{T} indeed has full column rank almost surely.

Any column in \mathbf{T}^\top is of the form

$$\begin{bmatrix} (\cos \theta_1)^r (\sin \theta_1)^{|s|} e^{j\phi_1 s} \\ \vdots \\ (\cos \theta_K)^r (\sin \theta_K)^{|s|} e^{j\phi_K s} \end{bmatrix}, \quad (65)$$

where $0 \leq r < B - K - |s|$ and $-(B - K - 1) \leq s \leq B - K - 1$. If the locations of the Diracs are random, we can use the following lemma to show that the matrix \mathbf{T} will have full column rank with probability one:

Lemma 3: Draw $[\xi_k = (\theta_k, \phi_k)]_{k=1}^K$ independently at random from any absolutely continuous probability distribution on $\mathcal{R} = [0, \pi] \times [0, 2\pi]$ (w.r.t. Lebesgue measure). Let

$\mathcal{M} = \{(r_1, s_1), \dots, (r_N, s_N)\}$ be a set of distinct integer pairs and let $\mathbf{G} = [g_{pq}]$, where $g_{pq} = (\cos \theta_p)^{r_q} (\sin \theta_p)^{|s_q|} e^{j\phi_p s_q}$. Then \mathbf{G} has full rank almost surely.

Proof: This proof is parallel to that of Theorem 3.2 from [25]. Let \mathbf{G}_M be the upper left $M \times M$ minor of \mathbf{G} . We define the bad set \mathcal{B}_M as the set on which \mathbf{G}_M is singular,

$$\mathcal{B}_M = \{(\xi_1, \dots, \xi_M) \in \mathcal{R}^M \mid \det \mathbf{G}_M = 0\}. \quad (66)$$

The goal is to show that $\mu(\mathcal{B}_K) = 0$, where μ is the Lebesgue measure on \mathcal{R}^K . We proceed by induction on M ; for $M = 1$, we have that $\mathbf{G}_1 = [(\cos \theta_1)^{r_1} (\sin \theta_1)^{|s_1|} e^{j\phi_1 s_1}]$, which is non-zero almost surely, so the claim holds. Assume now that $M < \min(K, N)$ and that the bad set \mathcal{B}_M has measure zero. Let $(\xi_1, \dots, \xi_M) \notin \mathcal{B}_M$, i.e., \mathbf{G}_M is invertible. Because it is invertible, there exists a unique coefficient vector $\mathbf{b} = \mathbf{b}(\xi_1, \dots, \xi_M)$ such that

$$\mathbf{G}_M \mathbf{b} = \mathbf{g}_{M+1}, \quad (67)$$

where by \mathbf{g}_{M+1} we denote the first M entries of the last column of \mathbf{G}_{M+1} . The bigger matrix \mathbf{G}_{M+1} will be singular if and only if the same linear combination is also consistent with its $(M + 1)$ st row. In other words, \mathbf{G}_{M+1} is invertible if and only if ξ_{M+1} is not in the set

$$\begin{aligned} \mathcal{Z}_M &= \left\{ (\theta, \phi) = \xi \in \mathcal{R} \mid (\cos \theta)^{r_{M+1}} (\sin \theta)^{|s_{M+1}|} e^{j\phi s_{M+1}} \right. \\ &= \left. \sum_{i=1}^M b_i (\cos \theta)^{r_i} (\sin \theta)^{|s_i|} e^{j\phi s_i} \right\}. \end{aligned}$$

For fixed (ξ_1, \dots, ξ_M) , this is the set of zeros of a particular (generalized) trigonometric polynomial, thus it has measure zero. Note that the definition of \mathcal{Z}_M makes sense only for $(\xi_1, \dots, \xi_M) \notin \mathcal{B}_M$, as otherwise \mathbf{G}_M is not invertible. Thus, the solution \mathbf{b} to (67) may not exist.

Consider now the following two sets:

$$\begin{aligned} \mathcal{U}_{M+1} &\stackrel{\text{def}}{=} \{(\xi_1, \dots, \xi_{M+1}) \mid (\xi_1, \dots, \xi_M) \in \mathcal{B}_M, \xi_{M+1} \in \mathcal{R}\}, \\ \mathcal{V}_{M+1} &\stackrel{\text{def}}{=} \{(\xi_1, \dots, \xi_{M+1}) \mid (\xi_1, \dots, \xi_M) \in \mathcal{R}^M, \xi_{M+1} \in \mathcal{Z}_M\}. \end{aligned}$$

The bad set \mathcal{B}_{M+1} must be a subset of the set $\mathcal{U} \cup \mathcal{V}$. But we just showed that the set \mathcal{V} has measure zero; by the induction hypothesis, \mathcal{U} also has measure zero. Thus their union, too, has measure zero.

It follows that \mathcal{B}_{M+1} has measure zero. Finally, because the distributions of ξ_i are absolutely continuous w.r.t. the Lebesgue measure, so is their product distribution. Hence the probability that (ξ_1, \dots, ξ_K) lies in the zero-measure set \mathcal{B}_K is zero. \square

To complete the argument, note that the matrix \mathbf{T}^\top has the same form as the matrix \mathbf{G} in the statement of Lemma 3, with $0 \leq r < B - K - |s|$ and $-(B - K - 1) \leq s \leq B - K - 1$. Thus, the columns of \mathbf{T} are independent with probability one, provided that its number of rows is at least K .

ACKNOWLEDGMENT

The authors would like to thank M. Vetterli for his support and mentoring.

REFERENCES

- [1] K. F. Evans, "The spherical harmonics discrete ordinate method for three-dimensional atmospheric radiative transfer," *J. Atmos. Sci.*, vol. 55, no. 3, pp. 429–446, 1998.
- [2] P. Audet, "Directional wavelet analysis on the sphere: Application to gravity and topography of the terrestrial planets," *J. Geophys. Res.*, vol. 116, no. E1, pp. 1–16, 2011.
- [3] R. H. Macphie and E. H. Okongwu, "Spherical harmonics and earth-rotation synthesis in radio astronomy," *IEEE Trans. Antennas Propag.*, vol. 23, no. 3, pp. 386–391, 1975.
- [4] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE ICASSP*, Orlando, FL, USA, 2002, pp. 1781–1784, IEEE.
- [5] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: Algorithm and applications," *J. Acoust. Soc. Amer.*, vol. 132, no. 3, pp. 1462–1472, 2012.
- [6] J. D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly, "Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution," *NeuroImage*, vol. 23, no. 3, pp. 1176–1185, Nov. 2004.
- [7] J. R. Driscoll and D. M. Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Adv. Appl. Math.*, vol. 15, no. 2, pp. 202–250, Jun. 1994.
- [8] J. D. McEwen and Y. Wiaux, "A novel sampling theorem on the sphere," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5876–5887, 2011.
- [9] Z. Khalid, R. A. Kennedy, and J. D. McEwen, "An optimal-dimensionality sampling scheme on the sphere with fast spherical harmonic transforms," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4597–4610, 2014.
- [10] 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, June 2002.
- [11] 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.
- [12] 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, 2007.
- [13] I. Maravić and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.
- [14] P. L. Dragotti and F. Homann, "Sampling signals with finite rate of innovation in the presence of noise," in *Proc. IEEE ICASSP*, Taipei, Taiwan, 2009, pp. 2941–2944, IEEE.
- [15] 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, 2013.
- [16] S. Deslauriers-Gauthier and P. Marziliano, "Sampling great circles at their rate of innovation," in *Proc. SPIE Wavelets and Sparsity*, D. Van De Ville, V. K. Goyal, and M. Papadakis, Eds., San Diego, CA, USA, Sep. 2013, Int. Soc. Opt. Photon.
- [17] S. Deslauriers-Gauthier and P. Marziliano, "Spherical finite rate of innovation theory for the recovery of fiber orientations," in *Proc. IEEE EMBC*, San Diego, CA, USA, 2012, pp. 2294–2297.
- [18] T. Bendory, S. Dekel, and A. Feuer, "Super-resolution on the sphere using convex optimization," Dec. 2014.
- [19] T. Bendory, S. Dekel, and A. Feuer, "Exact recovery of Dirac ensembles from the projection onto spaces of spherical harmonics," *Constr. Approx.*, pp. 1–25, 2014.
- [20] E. J. Candès and C. F. Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, Jun. 2014.
- [21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* 1998.
- [22] M. Abramowitz and I. A. Stegun, "Handbook of mathematical functions," Nat. Bureau of Stand. 1972.
- [23] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum*. Singapore: World Scientific, 1989.
- [24] I. Dokmanić and D. Petrinović, "Convolution on the n -sphere with application to PDF modeling," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1157–1170, Mar. 2010.
- [25] R. F. Bass and K. Gröchenig, "Random sampling of multivariate trigonometric polynomials," *SIAM J. Math. Anal.*, vol. 36, no. 3, pp. 773–795, Jan. 2005.
- [26] F. Vanpoucke, M. Moonen, and Y. Berthoumieu, "An efficient subspace algorithm for 2-D harmonic retrieval," in *Proc. IEEE ICASSP*, Adelaide, SA, Australia, 1994, pp. 461–464, IEEE.
- [27] I. Dokmanić, J. Ranieri, A. Chebira, and M. Vetterli, "Sensor networks for diffusion fields: Detection of sources in space and time," in *Proc. Allerton*, 2011, pp. 1552–1558.
- [28] Y. M. Lu and M. Vetterli, "Distributed spatio-temporal sampling of diffusion fields from sparse instantaneous sources," in *Proc. IEEE CAMSAP*, Aruba, 2009.
- [29] Y. M. Lu, P. L. Dragotti, and M. Vetterli, "Localizing point sources in diffusion fields from spatiotemporal samples," in *Proc. SampTA*, Singapore, 2011.
- [30] T. Bulow, "Spherical diffusion for 3D surface smoothing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 12, pp. 1650–1654, 2004.
- [31] J. D. McEwen, M. P. Hobson, and A. N. Lasenby, "Optimal filters on the sphere," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3813–3823, Aug. 2008.
- [32] P. Marziliano, M. Vetterli, and T. Blu, "Sampling and exact reconstruction of bandlimited signals with additive shot noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2230–2233, May 2006.
- [33] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, "The square kilometre array," *Proc. IEEE*, vol. 97, no. 8, pp. 1482–1496, June 2009.
- [34] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, 1988.



Ivan Dokmanić (S'04–M'14) received a doctorate in communication sciences from Ecole Polytechnique Fédérale de Lausanne (EPFL) in 2015. Previously he was a teaching assistant at the University of Zagreb, a codec developer for Main-Concept AG, and a digital audio effects designer for Little Endian Ltd. During summer 2013 he was with Microsoft Research, Redmond, WA, USA. He works on a medley of inverse problems, audio and acoustics, signal processing for sensor networks, and fundamental signal processing. For his work on room shape reconstruction using sound he received the best student paper award at ICASSP 2011. In 2014 he received a Google Ph.D. Fellowship. He is currently a postdoctoral researcher with Laurent Daudet (Institut Langevin), Stéphane Mallat (ENS), and Martin Vetterli (EPFL), and an adjunct professor at UIUC where he will start as an assistant professor in Fall 2016.



Yue M. Lu (S'04–M'07–SM'12) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, USA, in 2007.

He was a Research Assistant at the University of Illinois at Urbana-Champaign, and has worked for Microsoft Research Asia, Beijing, China and Siemens Corporate Research, Princeton, NJ, USA. Following his work as a postdoctoral researcher at the Audiovisual Communications Laboratory at Ecole Polytechnique Fédérale de Lausanne (EPFL),

Switzerland, he joined Harvard University in 2010 as an Assistant Professor of Electrical Engineering at the John A. Paulson School of Engineering and Applied Sciences.

He received the Most Innovative Paper Award (with Minh N. Do) of IEEE International Conference on Image Processing (ICIP) in 2006, the Best Student Paper Award of IEEE ICIP in 2007, and the Best Student Presentation Award at the 31st SIAM SEAS Conference in 2007. Student papers coauthored by him won the Best Student Paper Award (with Ivan Dokmanić and Martin Vetterli) of IEEE International Conference on Acoustics, Speech and Signal Processing in 2011 and the Best Student Paper Award (with Ameya Agaskar and Chuang Wang) of IEEE Global Conference on Signal and Information Processing (GlobalSIP) in 2014.

He has been an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING since December 2014 and an Elected Member of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee since January 2015. He received the ECE Illinois Young Alumni Achievement Award in 2015.