Deconvolution of Gaussian Random Fields using the Graph Fourier Transform

AUTHOR: MATTHIEU SIMEONI[1,2,*]
SUPERVISORS: PAUL HURLEY[2], **VICTOR PANARETOS**[1], **MARTIN VETTERLI**[1]

[1] *École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[2] *IBM Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland*
[*] *Contact: matthieu.simeoni@epfl.ch*

Last Modification: July 10, 2016

Abstract:

Real-life acquisition systems are fundamentally limited in their ability to reproduce point sources. For example, a point source object, a star say, observed with an optical telescope is blurred by the imperfect lenses composing the system. Mathematically, the point sources are convolved with a device-specific kernel. This kernel, which depends only on the characteristics of the acquisition system, can to a certain extent be designed so as to minimise the effect of the convolution. But in many applications careful design of the sensing device is not good enough, and a proper deconvolution step is needed.
In this paper, we propose an efficient deconvolution algorithm for point source Gaussian random fields as sensed by phased arrays. The algorithm first obtains a continuous least-squares estimate of the random field's second order moment. The procedure is then in two subsequent steps, decoupling sources localisation and intensity recovery.
First, we sample the continuous estimate at a high enough resolution and use the covariance function to construct a weighted graph. We then define a signal on this graph by assigning to each of the sample locations in the field their corresponding intensity (variance of the field at this location). The Graph Fourier Transform (GFT) is then used in order to filter out the convolution artifacts within the estimate. Candidate locations of the sources can then be identified with local maxima of the filtered estimate. From these locations a deconvolution problem is solved by means of weighted linear regression and the intensities of the sources within the field recovered.
Finally, a multi-scale approach based on the filtering of the leading eigenvalues of the covariance operator is discussed, and its benefits in terms of efficiency and accuracy are highlighted.

Keywords: *Point Source Random Fields, Graph Signal Processing, Deconvolution, Integral White Noise, PCA, Sensor Array*

1. Introduction

Gaussian random fields are widely used in scientific applications to model spatial fluctuations of natural phenomena. In radio astronomy for example [1], the sky is modelled as a random field composed of uncorrelated point sources at unknown locations, emitting narrowband signals with random complex amplitudes [2, 3]. Similarly, in positron emission tomography (PET), the *bra*in's metabolic activity is indirectly mapped by estimating the intensity field of positron emissions, modelled as a Poisson random point field [4].

Acquisition systems to sense these random fields take different forms depending on the application, but often consist in an array of sensors, measuring the incoming field at different locations. As proposed in [5], the action of the acquisition system on the unknown field can conveniently be formulated in terms of a *sampling operator* $\Phi^* : \mathcal{H} \to \mathbb{C}^L$, linear mapping from an infinite dimensional Hilbert space $\mathcal{H}$ to a finite dimensional vector space $\mathbb{C}^L$, where L denotes the number of sensors. By then constructing an *interpolation operator* $\widetilde{\Phi} : \mathbb{C}^L \to \mathcal{H}$ consistent with this sampling operator, it is possible to recover a *continuous* estimate of the original random field (see fig. 1 and [5, 6] for more details). Typically, the interpolation operator is chosen [5] to be the generalised pseudo-inverse of the sampling operator Φ^*:

$$\widetilde{\Phi} = \Phi(\Phi^*\Phi)^{-1}, \tag{1}$$

where $\Phi : \mathbb{C}^L \to \mathcal{R}(\Phi)$ is the *synthesis operator*, adjoint of Φ^*, and $\Phi^*\Phi \in \mathbb{C}^{L\times L}$ the Gram matrix (following the same naming conventions as in [6]). Details on how to compute these quantities in practice are provided in [5], where the interpolation operator (1)

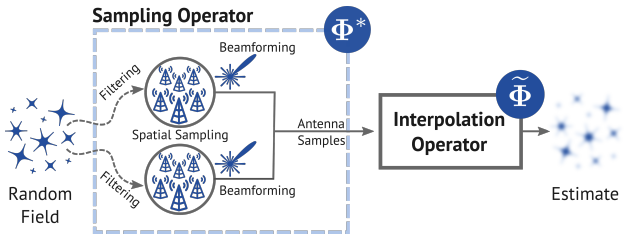
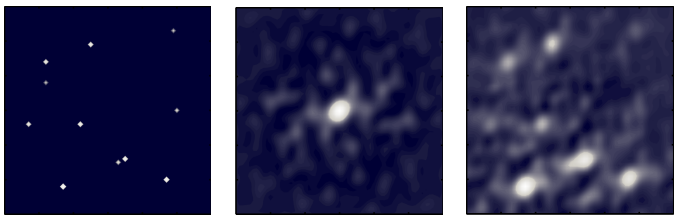


Figure 1 | Block diagram of the processing chain. For sensor arrays, the sampling operator can in general be decomposed in three subsequent operations: filtering, sampling and beamforming. The interpolation operator is then chosen consistently with the sampling operator.



(a) True sky. (b) Convolution kernel. (c) Dirty image.

Figure 2 | In radio astronomy, the least-squares estimate of the sky **(c)** is the result of the convolution between the true underlying sky **(a)** and a layout-dependent convolution kernel **(b)** (called *dirty beam*). Because of the numerous convolution artifacts polluting it, the least-squares estimate is called the *dirty image*.

is efficiently implemented for the specific example of beamformed radio interferometric data.

While mathematically elegant, computationally efficient and numerically stable, this method is unfortunately of limited use without an adequate post-processing of the estimate. Indeed, the recovered field can be shown to be in general the convolution of the true random field with some application-dependent kernel, entirely specified by the specifications of the acquisition system in use. This convolution can introduce serious artifacts within the estimate (see fig. 8), hence forbidding direct identification of features of interest (e.g. point sources in radio astronomy or audio source localisation). In this paper, we propose an efficient deconvolution algorithm to remove these convolution artifacts. We limit ourselves to the case of Gaussian point source random fields. The procedure is in two subsequent steps, decoupling the problem s of sources localisation and intensities recovery. First, we sample the field of view at a fine enough resolution. Each pixel in the field is then interpreted as a node in a fully connected graph. Edges are weighted according to the covariance existing between the different pixels, inferred from a least-squares estimate of the random field's covariance function (see [5]). We then define a signal on this graph by assigning to each of the sample locations their corresponding intensity (variance of the field at this location). This construction permits us to leverage the Graph Fourier Transform (GFT) to filter out the convolution artifacts within the estimate. An ideal high-pass filter is hence constructed, to separate the smooth artifacts from the actual point sources. Candidate locations of the sources are then directly identified from local maxima of the filtered estimate. From these locations a deconvolution problem is solved by means of weighted linear regression and the intensities of the sources within the field recovered. A multi-scale approach based on the filtering of the leading eigenvalues of the covariance operator is finally discussed, and its benefits (both in terms of efficiency and accuracy) are highlighted. In particular, we show that such method permits to super-resolve the sources within the field.

2. Background Concepts on Random Fields

In this section, we define random fields as collections of random variables indexed over some set. We then consider a few specific classes of random fields, particularly used for modelling natural phenomena. More specifically, we introduce Gaussian random fields, white noises, and integral white noises. A reader already familiar with these notions and spatial statistics in general can skip this section and go directly to section 3. The definitions and theorems summarized in this section are based on the work of M. Lifshits in his books [7, 8].

2.1. Random Functions & Random Fields

Random functions are defined as collections of random variables indexed over some set $\mathcal{X}$. Depending on the nature of the indexing set, random functions are denominated differently in the literature: for $\mathcal{X} \subset \mathbb{N}$ we speak of *random sequences*, while for $\mathcal{X} \subset \mathbb{R}^n$ we speak of *random processes* for $n = 1$ and *random fields* for $n \geq 2$.

Definition 2.1 — ***Random Functions.*** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space and $\mathbb{K}$ some field. Then, a $\mathbb{K}$-**valued random function** $\mathcal{S} \subset \mathbb{K}^\Omega$ is a collection of random variables $S(x) : \Omega \to \mathbb{K}$, indexed by a variable x taking value in some set $\mathcal{X}$:*

$$\mathcal{S} = \{S(x) : \Omega \to \mathbb{K},\ x \in \mathcal{X}\}.$$

*The set $\mathcal{X}$ is called the **parametric set** or **indexing set** for $\mathcal{S}$.*

Instead of defining random functions as collections of univariate functions from Ω to $\mathbb{K}$, some may find more convenient to think of them as bivariate functions with domain $\Omega \times \mathcal{X}$ and codomain $\mathbb{K}$,

$$\mathcal{S} : \Omega \times \mathcal{X} \to \mathbb{K}.$$

This naturally leads to the notion of **sample functions**, obtained by keeping fixed the first argument to some $\omega_0 \in \Omega$ and letting the second argument $x \in \mathcal{X}$ range over the entire parametric set:

Definition 2.2 — ***Sample Functions.*** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space, $\mathbb{K}$ some field and $\mathcal{X}$ some set. Let $\mathcal{S} : \Omega \times \mathcal{X} \to \mathbb{K}$ be a random function, with parametric set $\mathcal{X}$. Then, for a given $\omega_0 \in \Omega$ the function $s_{\omega_0} \in \mathbb{K}^\mathcal{X}$ given by*

$$s_{\omega_0} : \begin{cases} \mathcal{X} \to \mathbb{K}, \\ x \mapsto s_{\omega_0}(x) = \mathcal{S}(\omega_0, x), \end{cases}$$

*is called a **sample function** of $\mathcal{S}$ (or sample path when $\mathcal{X} \subset \mathbb{R}$).*

In practice, summary statistics of the random function $\mathcal{S}$ are estimated from a finite number of sample functions. The goal is to characterise the stochastic behaviour of the random function through a set of sufficient statistics. For example, Gaussian random functions, extensively used in the modelling of spatial fluctuations of natural phenomena, can be fully characterised by their first and second order moments [7, 8]. A Gaussian random field $\mathcal{S}$ is such that any finite subset of $\mathcal{S}$ follows a multivariate Gaussian distribution:

Definition 2.3 — ***Gaussian Random Functions.*** *A random function $\mathcal{S} = \{S(x) : \Omega \to \mathbb{K},\ x \in \mathcal{X}\}$ is called **Gaussian** if all its finite dimensional distributions are multivariate Gaussian:*

$$[S(x_{i_1}), \dots, S(x_{i_n})]^T \overset{d}{\sim} \mathcal{N}_n, \quad \forall \{i_1, \dots, i_n\} \subset \mathcal{X},\ \forall n \in \mathbb{N}.$$

For *centered* Gaussian random functions ($\mathbb{E}[S(x)] = 0,\ \forall x \in \mathcal{X}$), the set of sufficient statistics for $\mathcal{S}$ reduces to its **covariance function** $\kappa : \mathcal{X}^2 \to \mathbb{K}$ alone, measuring the *spatial coherency* of the random function:

$$\kappa : \begin{cases} \mathcal{X} \times \mathcal{X} \to \mathbb{K}, \\ (x, y) \mapsto \kappa(x, y) := \mathrm{Cov}[S(x), S(y)]. \end{cases}$$

For such random functions, we can hence trade the analysis of the random object for the analysis of its covariance function. This analysis, although considerably simplified by the *deterministic* nature of κ, may require occasional switches in perspective on the covariance function, sometimes seen as a bivariate function and

sometimes seen as a linear operator between Hilbert spaces. Indeed, in the same way that a matrix $A \in \mathbb{K}^{n\times m}$ can be considered either as a vector in $\mathbb{K}^{nm}$ or as a linear operator from $\mathbb{K}^m$ to $\mathbb{K}^n$, the covariance function κ can also be seen as the kernel of an associated **covariance operator** $T_\kappa : \mathcal{L}^2(\mathcal{X}) \to \mathcal{L}^2(\mathcal{X})$,

$$T_\kappa : \begin{cases} \mathcal{L}^2(\mathcal{X}) \to \mathcal{L}^2(\mathcal{X}), \\ f \mapsto (T_\kappa f)(y) := \int_{\mathcal{X}} \kappa(y,x) f(x) dx, \ \forall y \in \mathcal{X}. \end{cases}$$

By analogy with the vec(·) operator for matrices [9], κ is sometimes written as

$$\kappa = \mathrm{vec}(T_\kappa).$$

This inherent duality in the representation of the covariance function, will allow us to resort to a variety of different tools when analysing it, ranging across the fields of calculus, functional analysis and signal processing.

2.2. White Noises

In the modelling of natural phenomena with point source emissions, it is usually assumed that signals from different sources are uncorrelated, no matter how close the sources may be from one another. This assumption, although physically unrealistic, is a convenient mathematical simplification, which still retains some validity in many scenarios. In order to construct random functions with such a property, we need to introduce the concept of Gaussian white noises. Roughly speaking, Gaussian white noises can be seen as finite random measures on the set of ν-finite measurable sets for some control measure ν.

Definition 2.4 — *Gaussian White Noise.* *Let $(\mathcal{X}, \mathcal{A}, \nu)$ be a measure space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Define $\mathcal{A}_0 \subset \mathcal{A}$ as the set of ν-finite measurable sets:*

$$\mathcal{A}_0 := \{A \in \mathcal{A} : \ \nu(A) < \infty\}.$$

Let further assume a $\mathbb{K}$-valued Gaussian random function $\mathcal{W} \subset \mathbb{K}^\Omega$ with parametric set $\mathcal{A}_0$:

$$\mathcal{W} = \{W(A) : \Omega \to \mathbb{K}, \ A \in \mathcal{A}_0\}.$$

*Then, $\mathcal{W}$ is called a **Gaussian white noise** with **control measure** ν if:*

- $\mathbb{E}[W(A)] = 0, \ \forall A \in \mathcal{A}_0,$
- $\mathbb{E}[W(A)W^*(B)] = \nu(A \cap B), \ \forall A, B \in \mathcal{A}_0.$

From the definition 2.4, two properties of Gaussian white noises immediately follow:

- The intensity function (variance) of a Gaussian white noise is given by its control measure:

$$\mathrm{Var}[W(A)] = \nu(A), \ \forall A \in \mathcal{A}_0. \tag{2}$$

  Restricting the parametric set to ν-finite measurable sets hence guarantees $\mathcal{W}$ to have finite second order moment.
- $W(\emptyset) = 0$ almost surely. Indeed, as ν is a measure, $\nu(\emptyset) = 0 < \infty$ and hence $\emptyset \in \mathcal{A}_0$. Thus $\mathbb{E}[W(\emptyset)] = 0$. Moreover, from (2) we have that $\mathrm{Var}[W(\emptyset)] = \nu(\emptyset) = 0$. Hence $W(\emptyset)$ is almost surely equal to 0.
- A Gaussian white noise verifies the σ-additivity property almost surely [8]. Indeed, for any disjoint sets $A_1, \dots, A_n \in \mathcal{A}_0$, the random variables $W(A_1), \dots, W(A_n)$ are independent and

$$\sum_{i=1}^n W(A_i) = W\left(\bigcup_{i=1}^n A_i\right), \quad \text{almost surely.}$$

From the last two points, we deduce that sample functions of Gaussian white noises are almost surely finite measures on $\mathcal{A}_0$. In that sense, we can hence think of Gaussian white noises as random measures as claimed earlier. Within the scope of this work, we will restrict our attention to a specific class of Gaussian white noises, that we call Gaussian point source random functions. Roughly speaking, Gaussian point source functions can be seen as streams of Diracs with random amplitudes.

Definition 2.5 — *Gaussian Point Source Random Function.* *Let $(\mathcal{X}, \mathcal{A}, \nu)$ be a measure space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Let $\mathcal{W} \subset \mathbb{K}^\Omega$ be a Gaussian white noise with parametric space $\mathcal{A}_0$ and control measure ν, as in 2.4. Then, this random function is called a **Gaussian point source random function** if the indexed random variable in $\mathcal{W}$ can be written as*

$$W(A) = \sum_{q=1}^Q \xi_q \delta_{x_q}(A), \quad \forall A \in \mathcal{A}_0.$$

In the above equation, δ_{x_q} denotes the Delta measure, $Q > 0$ denotes the number of sources, $\{x_1, \dots, x_Q\}$ are the sources locations in $\mathcal{X}$ and ξ_q are random amplitudes, independent and identically distributed:

$$\xi_q \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_q^2), \ \sigma_q^2 > 0, \ \forall q = 1, \dots, Q.$$

It is easy to see that Gaussian point source random functions are indeed Gaussian white noise. For this specific type of white noise, the control measure is given by

$$\nu(A) = \sum_{Q=1}^Q \sigma_q^2 \delta_{x_q}(A), \quad \forall A \in \mathcal{A}_0.$$

2.3. White Noise Integrals

Aside from their use for modelling purposes, Gaussian white noises are also very important theoretically, as many Gaussian random functions can be expressed in terms of white noise integrals [8]. In particular, and of interest for this work, the filtering of a white noise with a suitable kernel results in a Gaussian random function of a special kind, collection of integral white noises.

Definition 2.6 — *White Noise Integral.* *Let $(\mathcal{X}, \mathcal{A}, \nu)$ be a measure space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Take a Gaussian white noise $\mathcal{W} \subset \mathbb{K}^\Omega$ with parametric space $\mathcal{A}_0$ and control measure ν, as in 2.4. Then, for any $f \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu)$, the random variable $\mathcal{J} : \Omega \to \mathbb{K}$ given by*

$$\mathcal{J} = \int_{\mathcal{X}} f(x) \mathcal{W}(dx), \tag{3}$$

*is called an **integral white noise**.*

When defining the sampling operator in section 3, it will help to think of $\mathcal{J}$ in eq. (3) as the result of an *integral operator* associated with the white noise $\mathcal{W}$, acting on the function $f^* \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu)$

$$\mathcal{J} = \langle \mathcal{W} | f^* \rangle := \int_{\mathcal{X}} f(x) \mathcal{W}(dx) \tag{4}$$

In the above equation, we borrowed the convenient *bra-ket* notation, very popular in quantum mechanics, introduced by Paul Dirac in [10]. With this formalism, the *bra* $\langle \mathcal{W}|$ is interpreted as an element of the dual $\mathcal{H}$ of $\mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu)$. As such, it acts linearly on the *ket*

$|f^*\rangle \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu)$, as described in eq. (4). Remember that, as it is well know, to every *ket* corresponds a *bra*, but the converse does not necessarily hold. Indeed, for a *bra* $|g\rangle \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu)$ we can define a corresponding *ket* $\langle g| \in \mathcal{H}$, defined as

$$\langle g|f\rangle = \int_{\mathcal{X}} g(x) f^*(x)\, dx, \quad \forall\, |f\rangle \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu).$$

Observe that section 2.3 also defines an inner product between g and f. For a white noise $\mathcal{W} : \Omega \times \mathcal{A}_0 \to \mathbb{K}$, we have a natural interpretation of it as a *bra* $\langle \mathcal{W}| \in \mathcal{H}$ (see eq. (4)) but the existence of an associated *ket* is not guaranteed. A necessary and sufficient condition is for $\mathcal{W}$ to be absolutely continuous with respect to the Lebesgue measure. In which case, we can associate the *bra* $\langle \mathcal{W}|$ with a *ket* $|\omega\rangle$, where $\omega : \Omega \times \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu) \to \mathbb{R}_+$ is the Radon-Nikodym derivative of the measure $\mathcal{W}$ with respect to the Lebesgue measure (also called *density* of $\mathcal{W}$). Indeed, we have

$$\langle \mathcal{W}|f\rangle = \int_{\mathcal{X}} f^*(x)\mathcal{W}(dx) = \int_{\mathcal{X}} f^*(x)\omega(x)dx = \langle \omega|f\rangle.$$

Notice that for absolutely continuous white noises, the linear operation $\langle \mathcal{W}|f\rangle$ coincides with the inner product between the density ω of $\mathcal{W}$ and the function f. Borrowing some intuition from this specific case, we can hence informally think of a white noise integral (3) as the "inner product" between the white noise $\mathcal{W}$ and some function f^* (as suggested by the *bra-ket* notation in eq. (4)).

Using the *isometric property* (see [8]), one can compute the covariance between two integral white noises as

$$\mathbb{E}\left[\left(\int_{\mathcal{X}} f\, d\mathcal{W}\right)\left(\int_{\mathcal{X}} g\, d\mathcal{W}\right)^*\right] = \int_{\mathcal{X}} f g^*\, d\nu, \ \forall f, g \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu).$$

This yields in particular

$$\mathrm{Var}\left[\int_{\mathcal{X}} f\, d\mathcal{W}\right] = \int_{\mathcal{X}} |f|^2\, d\nu = \|f\|^2_{2,\nu}, \quad \forall f \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu).$$

Moreover, it is easy to see that $\mathbb{E}\left[\int_{\mathcal{X}} f\, d\mathcal{W}\right] = 0$, and the distribution of a white noise integral is finally:

$$\int_{\mathcal{X}} f\, d\mathcal{W} \overset{d}{\sim} \mathcal{N}\left(0, \int_{\mathcal{X}} |f|^2\, d\nu\right), \quad \forall f \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu).$$

Many Gaussian random functions admit a representation in terms of a white noise integral. This representation is called an **integral representation**, or **spectral representation**.

Definition 2.7 — ***Integral Representation.*** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{X}, \mathcal{A}, \nu)$ be a measure space. Let $\mathcal{A}_0 = \{A \in \mathcal{A} : \nu(A) < \infty\}$ and $\mathcal{W} = \{W(A) : \Omega \to \mathcal{K},\ A \in \mathcal{A}_0\}$ be a Gaussian white noise with control measure ν. Assume further $\mathcal{S} = \{S(x) : \Omega \to \mathbb{K},\ x \in \mathcal{X}\}$ to be a Gaussian random function. Then, $\mathcal{S}$ admits an **integral representation** (or spectral representation) if the indexed random variables $S(x) : \Omega \to \mathbb{K}$ can be written as*

$$S(x) = \int_{\mathcal{X}} \zeta(x, y)\mathcal{W}(dy), \quad \forall x \in \mathcal{X},$$

*where $\zeta(x, \cdot) \in \mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu)$. The control measure ν is called the **spectral measure** of the random function $\mathcal{S}$.*

Assuming a random field $\mathcal{S}$ with finite first and second moment and admitting an integral representation, then the covariance function takes the form:

$$\begin{aligned}\kappa(x, y) &= \mathbb{E}[S(x)S^*(y)] \\ &= \mathbb{E}\left[\left(\int_{\mathbb{R}} \zeta(x, u)\mathcal{W}(du)\right)\left(\int_{\mathbb{R}} \zeta(x, v)\mathcal{W}(dv)\right)^*\right] \\ &= \int_{\mathcal{X}} \zeta(x, u)\zeta^*(y, u)\, \nu(du), \quad \forall (x, y) \in \mathcal{X}^2, \end{aligned} \tag{5}$$

where we have used the isometry property for the third equality. One of the most famous random process admitting an integral representation is most probably the *Wiener process*, also called *Brownian motion*. In this work, integral representations will appear particularly useful to represent both the Gaussian white noise filtered by the instrument and the least-squares estimate obtained with the algorithm in [5].

3. Sampling & Interpolation

As proposed in [5], the action of the acquisition system on the unknown random field can be conveniently modelled in terms of a sampling operator. In this section we describe this sampling operator and relate it to the characteristics of the sensor array in use. Following the derivations in [5], we then construct an interpolation operator consistent with the specified sampling operator. For this work, we restrict our attention to the ideally matched interpolation operator, pseudo-inverse of the sampling operator. However the analysis in the subsequent sections could easily be extended to more general consistent interpolation operators.

3.1. Acquisition System & Sampling Operator

In experimental setups, evidence about the random field of interest is collected with an acquisition system [6]. Real-life acquisition systems act on the random field in two subsequent stages: *filtering* and *sampling* (see fig. 3). These two steps can be summarised by means of a sampling operator, measuring the resemblance of the unknown random field with some canonical basis functions, tailored to the acquisition system in use. In all that follows, we will assume an incoming Gaussian point source random field $\mathcal{S} \subset \mathbb{K}^\Omega$ as in definition 2.5, with control measure ν. We will also used the *bra-ket* formalism, introduced in section 2.3 (see eq. (4)). We note $\mathcal{H}$ the dual space of $\mathcal{L}^2(\mathcal{X}, \mathcal{A}, \nu)$.

First, the random field $\mathcal{S}$ is *spatially filtered*, yielding a Gaussian random field in integral representation. This random field is called the **measurement field** $\mathcal{Y}$, and has parametric set $\mathcal{P}$:

$$\mathcal{Y} = \{Y(p) : \Omega \to \mathbb{K}, \quad p \in \mathcal{P}\},$$

where

$$Y(p) = \int_{\mathcal{X}} \phi^*(p, x)\, \mathcal{S}(dx) = \langle \mathcal{S}|\phi(p, \cdot)\rangle \quad \forall p \in \mathcal{P}, \tag{6}$$

and $\mathcal{P} \subset \mathbb{R}^n$ is the **measuring manifold**. This manifold is often related to the geometry of the acquisition system. For example,

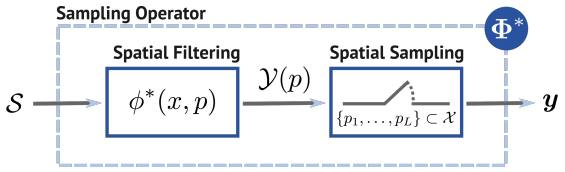


Figure 3 | The acquisition system can be modelled as a sampling operator Φ^*, acting of the random field $\mathcal{S}$ in two subsequent steps: *filtering* and *sampling*.

if one is measuring a sound field with an array of microphones disposed on a sphere, then the measuring manifold will be a sphere. The filtering kernel $\phi : \mathcal{X} \times \mathcal{P} \to \mathbb{K}$ is known and depends only on the physics of the phenomenon of interest and the characteristics of the acquisition system in use. In the case of radio astronomy for example, the kernel is a complex plane-wave [5, 3]

$$\phi(x,p) = e^{j\frac{2\pi}{\lambda_0}\langle x,p\rangle}, \quad \forall (x,p) \in \mathcal{X}\times\mathcal{P}, \tag{7}$$

where $\lambda_0 \in \mathbb{R}$ is the wavelength of the observed signal. Note that for eq. (6) to be well-defined we need $\phi(p,\cdot) \in \mathcal{L}^2(\mathcal{X},\mathcal{A},\nu)$. For point source random fields, a sufficient condition for this is that $\phi(p,\cdot)$ be bounded, which is obviously always the case in practice.

The second stage of the acquisition consists in a *sampling* of the measurement field $\mathcal{Y}$, at locations $\{p_1,\dots,p_L\} \subset \mathcal{P}$. These sampling locations are again often related to the underlying geometry of the measuring device, and can correspond for example to the positions of the various sensors composing the tool in the case of sensor arrays. We finally obtain a random vector $\boldsymbol{y} : \Omega \to \mathbb{C}^L$

$$\boldsymbol{y} = (Y(p_1),\dots,Y(p_L))^T.$$

Notice that the entries $Y(p_i)$ of the random vector $\boldsymbol{y}$ are white noise integrals, given by

$$Y(p_i) = \int_{\mathcal{X}} \phi_i^*(x)\,\mathcal{S}(dx) = \langle \mathcal{S}|\phi_i\rangle, \quad \forall i = 1,\dots,L, \tag{8}$$

where $\phi_i(x) = \phi(x,p_i)\ \forall x \in \mathcal{X}$ for $i = 1,\dots,L$. The equations in (8) can be more compactly rewritten as

$$\boldsymbol{y} = \Phi^*\mathcal{S}, \tag{9}$$

where $\Phi^* : \mathcal{H} \to \mathbb{C}^L$ is the *analysis operator* [6] associated to the family $\{\phi_i(\boldsymbol{r}),\ i = 1,\dots,J\} \subset \mathcal{L}^2(\mathcal{X},\mathcal{A},\nu)$:

$$\Phi^* : \begin{cases} \mathcal{H} \to \mathbb{C}^L \\ \mathcal{S} \mapsto [\langle \mathcal{S}|\phi_1\rangle,\dots,\langle \mathcal{S}|\phi_L\rangle]^T. \end{cases}$$

We will call this operator the **sampling operator** in all that follows. In more advanced experimental setups, the sampling operator can be refined a little by the use of *beamforming* [5]. Beamforming aims at both reducing the amount of data sent to the central processor and creating virtual super antennas, with improved sensitivity patterns [3, 11, 12]. This is typically the case for the *LOw Frequency ARray (LOFAR)*, the current largest radio telescope in the world, which leverages matched beamforming in its hierarchical design [13].

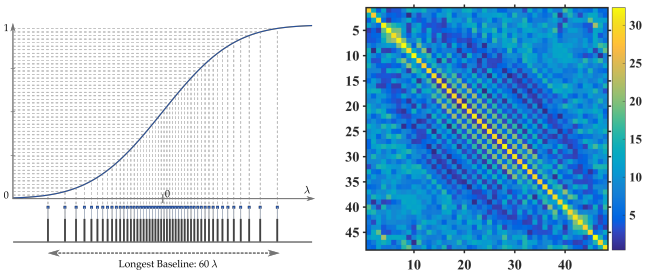
The data collected by the instrument during a certain period of time can hence be seen as various realisations of the J-dimensional random vector $\boldsymbol{y}$, with distribution

$$\boldsymbol{y} \overset{d}{\sim} \mathbb{K}\mathcal{N}_L(\boldsymbol{0}_L,\Sigma).$$

Its covariance matrix $\Sigma = \Phi^*\nu\Phi \in \mathbb{K}^{L\times L}$ (see fig. 4) is given by

$$(\Sigma)_{i,j} = \int_{\mathcal{X}} \phi_i^*(x)\phi_j(x)\,\nu(dx), \quad \forall i,j = 1,\dots,L.$$

The goal will then be to interpolate observations from $\boldsymbol{y}$ with an appropriate interpolation operator, so as to estimate some summary statistics of the underlying random field $\mathcal{S}$.



(a) Gaussian layout of the linear phased array. The array is composed of 48 antennas, distributed according to the Gaussian quantile function, with diameter 60 λ.

(b) Covariance matrix Σ estimated with 800 noisy antenna samples, for the underlying point source random function considered in fig. 6.

Figure 4 | Example of a linear phased array, and estimate of the associated covariance matrix Σ for the Gaussian point source in fig. 6. The array has a Gaussian layout and a diameter of 60 λ (λ being the wavelength of the incoming signal). The antenna samples are corrupted by an independent Gaussian white noise, with a PSNR of approximately 0 dB.

3.2. Interpolation

The problem of estimating the random field $\mathcal{S}$ from the measurements $\boldsymbol{y} \in \mathbb{C}^L$ is referred to as *interpolation*. Mathematically speaking, we would like to find a linear operator $\widetilde{\Phi} : \mathbb{C}^L \to \mathcal{H}$, that maps the *finite* measurements set $\boldsymbol{y}$ onto an *infinite*-dimensional object $\mathcal{S}_D \in \mathcal{H}$, hopefully constituting in a good approximation of the true random field $\mathcal{S}$ from which the measurements were obtained. This can indeed be interpreted as an interpolation step, in the sense that a discrete set of samples from the measuring field is mapped to a continuous function [6]. A usual requirement for the interpolation operator is for it to be *consistent* with the sampling operator [6]:

$$\Phi^*\widetilde{\Phi} = I. \tag{10}$$

The consistency condition (10) requires interpolation followed by sampling to be equal to the identity, which seems like a sensible enough condition to impose on our interpolation operator. Mathematically speaking, this is equivalent to saying that $\widetilde{\Phi}$ is a right-inverse for Φ^*, and hence $\widetilde{\Phi}\Phi^*$ is a projection operator [6] (orthogonal or oblique). To obtain an orthogonal projection operator, we must have $\mathcal{R}(\widetilde{\Phi}) = \mathcal{R}(\Phi)$, which is guaranteed when choosing $\widetilde{\Phi}$ to be equal to the pseudo-inverse of Φ^*

$$\widetilde{\Phi} = \Phi\left(\Phi^*\Phi\right)^{-1} = \Phi G_\Phi^{-1}.$$

In which case, we say that the interpolation operator $\widetilde{\Phi}$ is *ideally matched* with the sampling operator Φ^* [6]. The operator $\Phi : \mathbb{K}^L \to \mathcal{H}$, adjoint of Φ^*, is called the *synthesis* operator and is defined as

$$\Phi : \begin{cases} \mathbb{K}^L \to \mathcal{H}, \\ \alpha \mapsto (\Phi\alpha)(x) = \sum_{k=1}^L \alpha_k\phi_k(x), \quad \forall x \in \mathcal{X}. \end{cases}$$

The matrix $G_\Phi = \Phi^*\Phi \in \mathbb{C}^{L\times L}$ is called the **Gram matrix** and is defined as

$$(G_\Phi)_{ij} = \langle \phi_i|\phi_j\rangle = \int_{\mathcal{X}} \phi_i(x)\phi_j^*(x)\,dx, \ \forall i,j \in \{1,\dots,L\}.$$

This matrix can either be computed numerically or, in some specific scenarios, analytically. For complex plane waves (7) defined over the circle $\mathcal{X} = \mathbb{S}^1$, we have for example

$$\langle \phi_i|\phi_j\rangle = \int_{\mathbb{S}^1} e^{j2\pi\langle x,p_i-p_j\rangle}\,dx = 2\pi J_0(2\pi\|p_i-p_j\|), \quad \forall p_i,p_j \in \mathbb{R}^2,$$

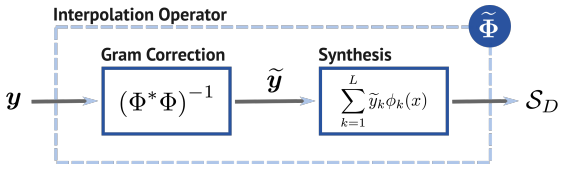


Figure 5 | The sampling operator $\widetilde{\Phi}$, acts of the antenna samples $\boldsymbol{y} : \Omega \to \mathbb{K}^L$ in two subsequent steps: first, the samples are corrected by the Gram matrix and then fed to the synthesis operator Φ.

where J_0 denotes the Bessel function of degree 0. Similar analytical results exist also for the sphere and more generally for $\mathbb{S}^{n-1}$, $n \in \mathbb{N}$ (see [5] for an example). The invertibility of the Gram matrix comes from the assumption that the family of functions $\{\phi_k,\ k = 1, \dots, L\}$ is linearly independent (often verified in practice if the array layout is properly chosen). To prevent G_Φ from being ill-conditioned (and hence facilitate its numerical inversion), we can further require $\{\phi_k,\ k = 1, \dots, L\}$ to be a *Riesz basis* for its span $\mathcal{R}(\Phi) = \text{span}\{\phi_k,\ k = 1, \dots, L\}$ (see [6]).

The interpolation step can hence be decomposed in two subsequent steps (see fig. 5). First, a Gram correction is applied to the samples $\boldsymbol{y} : \Omega \to \mathbb{K}^L$

$$\widetilde{\boldsymbol{y}} = G_\Phi^{-1} \boldsymbol{y}.$$

This correction compensates for the lack of orthogonality of the family $\{\phi_k,\ k = 1, \dots, L\}$. Indeed, if the family were orthogonal, the Gram matrix would simply reduce to the identity and the above correction would be trivial.

Then, the corrected samples $\widetilde{\boldsymbol{y}} : \Omega \to \mathbb{K}^L$ are acted on by the synthesis operator Φ. This yields a Gaussian random function $\mathcal{S}_D \subset \mathbb{K}^\Omega$, given by

$$\mathcal{S}_D = \left\{ S_D(x) = \sum_{k=1}^{L} \widetilde{y}_k \phi_k(x) : \Omega \to \mathbb{K}, \quad x \in \mathcal{X} \right\}. \tag{11}$$

We can show that $\mathcal{S}_D$ is the *least-squares* estimate of $\mathcal{S}$ in the sense that

$$\mathcal{S}_D = \underset{\varphi \in \mathcal{R}(\Phi)}{\text{argmin}}\, \mathbb{E}\|\mathcal{S} - \varphi\|_2^2, \quad \text{and} \quad \mathcal{S} - \mathcal{S}_D \in \mathcal{R}(\Phi)^\perp.$$

The covariance operator $T_{\kappa_D} : \mathcal{L}^2(\mathcal{X}) \to \mathcal{L}^2(\mathcal{X})$ of the random function $\mathcal{S}_D$ is given by

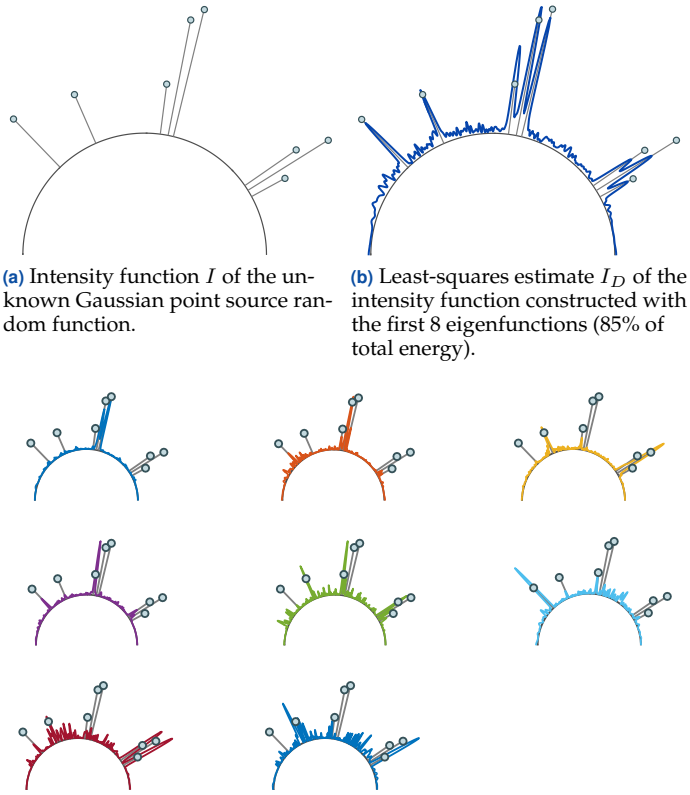
$$T_{\kappa_D} = \Phi G_\Phi^{-1} \Sigma G_\Phi^{-1} \Phi^*. \tag{12}$$

It may appear easier to some to interpret T_{κ_D} in terms of its associated kernel $\kappa_D : \mathcal{X} \times \mathcal{X} \to \mathbb{K}$, given by

$$\begin{aligned} \kappa_D &= \text{vec}\,(T_{\kappa_D}) \\ &= \text{vec}\left(\Phi G_\Phi^{-1} \Sigma G_\Phi^{-1} \Phi^*\right) \\ &= \left(\overline{\Phi} \otimes \Phi\right)\left(G_\Phi^{-1} \otimes G_\Phi^{-1}\right) \text{vec}(\Sigma), \end{aligned} \tag{13}$$

where $\otimes$ denotes the tensor product. The last equality in eq. (13), was obtained by leveraging well-known properties of the the tensor product (see [14, 9, 1] for more details on the tensor product and the Kronecker product in the finite dimensional case). If we define $\widetilde{\boldsymbol{\sigma}} = \left(G_\Phi^{-1} \otimes G_\Phi^{-1}\right) \text{vec}(\Sigma)$, then eq. (13) can be re-written as a weighted sum of separable kernels:

$$\kappa_D(x, y) = \sum_{i,j=1}^{L} \widetilde{\sigma}_{i,j}\ \phi_i(x) \phi_j^*(y), \quad \forall (x, y) \in \mathcal{X}^2. \tag{14}$$



(a) Intensity function I of the unknown Gaussian point source random function.

(b) Least-squares estimate I_D of the intensity function constructed with the first 8 eigenfunctions (85% of total energy).

(c) Squared magnitude of the eigenfunctions associated with the eight largest eigenvalues of the covariance operator T_{κ_D}, in decreasing order.

Figure 6 | Estimation of the second order moments of a Gaussian point source random function on the sphere. The random function is sensed by the linear array in fig. 4. From 800 noisy observations of the antenna samples $\boldsymbol{y}$, the eigenfunctions of T_{κ_D} are estimated and the intensity field reconstructed. We observe that many convolution artefacts pollute the eigenfunctions and consequently the intensity field, complicating the identification of sources.

Finally, the diagonal part of T_{κ_D} yields the intensity field $I_D : \mathcal{X} \to \mathbb{R}$, describing the variance of the Gaussian random field $\mathcal{S}_D$ for every location $x \in \mathcal{X}$:

$$I_D(x) = \kappa_D(x, x) = \sum_{i,j=1}^{L} \widetilde{\sigma}_{i,j}\ \phi_i(x) \phi_j^*(x), \quad \forall (x, y) \in \mathcal{X}^2. \tag{15}$$

As we can observe in fig. 6b, the intensity field is severely polluted by *sidelobes*, forbidding the direct identification of sources within the function. We will show in section 3.4 that these sidelobes are actually convolution artefacts.

3.3. Functional Principal Component Analysis

In eq. (14), we have expressed κ_D as a linear combination of L^2 separable bivariate functions $\phi(x)\phi_j^*(y)$. These functions, although tailored to the instrument in use, are however not the most ideally suited for representing κ_D. Indeed, if we use the eigenfunctions of T_{κ_D} as spanning functions, it is then possible to represent κ_D with a much smaller set of function, with cardinality ranging from 0 to L depending on the desired accuracy. This effectively accounts to performing a **functional principal component analysis** on κ_D, by projecting it onto optimal finite subspaces with increasing dimensionality. As stated in theorem 3.1, these eigenfunctions can

be efficiently computed by applying the synthesis operator Φ to the eigenvectors of a generalized eigenvalue problem:

Theorem 3.1 — ***Functional PCA for Interpolated Random Functions.*** Let $\mathcal{S} \in \mathcal{H}$ be a Gaussian point source random function, sampled by a full rank operator $\Phi^* : \mathcal{H} \to \mathbb{K}^L$, yielding the random vector

$$\boldsymbol{y} = \Phi^* \mathcal{S},$$

with covariance matrix $\Sigma \in \mathbb{K}^{L \times L}$. We note $K \leq L$ the rank of Σ. Let further $\widetilde{\Phi} = \Phi G_\Phi^{-1}$ be the interpolation operator ideally matched with Φ^* and

$$\mathcal{S}_D = \widetilde{\Phi} \boldsymbol{y},$$

be the orthogonal projection of $\mathcal{S}$ onto the range of Φ. Then, the covariance operator $T_{\kappa_D} : \mathcal{L}^2(\mathcal{X}) \to \mathcal{L}^2(\mathcal{X})$ of $\mathcal{S}_D$ given in eq. (12), has a null eigenvalue with infinite multiplicity, and K non-null real eigenvalues. These eigenvalues and their associated eigenfunctions are given by:

$$\left(\lambda_i, \frac{\Phi \alpha_i}{\sqrt{\alpha_i^* G_\Phi \alpha_i}} \right) \in \mathbb{K} \times \mathcal{L}^2(\mathcal{X}), \quad i = 1, \dots, K,$$

where $\lambda_i \in \mathbb{K}$ and $\alpha_i \in \mathbb{K}^L$ are the K distinct non-null eigenpairs of the generalized eigenvalue problem:

$$\Sigma \alpha_i = \lambda_i G_\Phi \alpha_i, \quad \forall i = 1, \dots, K. \tag{16}$$

Hence, T_{κ_D} has rank $K \leq L$ and admits the following decomposition

$$T_{\kappa_D} = \sum_{i=1}^{K} \lambda_i \frac{\Phi \alpha_i \alpha_i^* \Phi^*}{\alpha_i^* G_\Phi \alpha_i}. \tag{17}$$

■ **Proof 3.1** First, notice that any element $f \in \mathcal{N}(\Phi^*)$ with $\|f\|_2 = 1$ is an eigenfunction for T_{κ_D} for the eigenvalue 0:

$$T_{\kappa_D} f = \Phi G_\Phi^{-1} \Sigma G_\Phi^{-1} \underbrace{\Phi^* f}_{=0} = 0.$$

Hence, because $\mathcal{N}(\Phi^*) = \mathcal{R}(\Phi)^\perp$ and Φ has rank L, the eigenspace associated to the eigenvalue 0 is infinite dimensional, and finally 0 has infinite multiplicity.

Assume now that $\{(\lambda_i, \alpha_i),\ i = 1, \dots, K\}$ are distinct, non-null eigenpairs of the generalized eigenvalue problem:

$$\Sigma \alpha_i = \lambda_i G_\Phi \alpha_i. \tag{18}$$

The matrices Σ and G_Φ being Hermitian and definite-positive, existence of non-trivial eigenpairs for eq. (18) is guaranteed, and the eigenvectors are G_Φ-orthogonal [15]:

$$\alpha_i^* G_\Phi \alpha_j = 0, \quad \forall i, j = 1, \dots, K. \tag{19}$$

Moreover, $\text{rank}(\Sigma) = K \leq L$ and $\text{rank}(G_\Phi) = L$ and hence we have indeed K distinct, real eigenvalues different from 0.

Define then $f_i := \Phi \alpha_i \in \mathcal{L}^2(\mathcal{X}),\ i = 1, \dots, K$. We have

$$\begin{aligned} T_{\kappa_D} f_i &= \Phi G_\Phi^{-1} \Sigma G_\Phi^{-1} \Phi^* \Phi \alpha_i \\ &= \Phi G_\Phi^{-1} \Sigma \alpha_i \\ &= \lambda_i \Phi \alpha_i \\ &= \lambda_i f_i, \end{aligned}$$

where the third equality was obtained from eq. (18). Indeed,

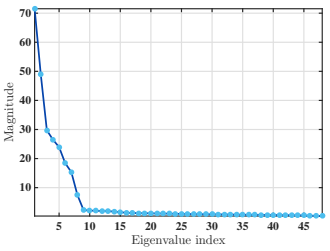
$$\Sigma \alpha_i = \lambda_i G_\Phi \alpha_i \ \Leftrightarrow\ G_\Phi^{-1} \Sigma \alpha_i = \lambda_i \alpha_i, \quad \forall i = 1, \dots, K.$$

We can verify that the functions f_i are orthogonal to one another. Indeed, we have from eq. (19):

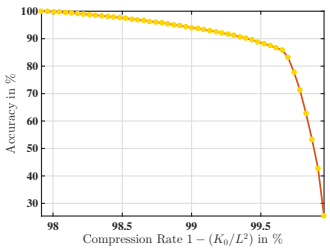
$$\langle f_i, f_j \rangle = \alpha_i^* \Phi^* \Phi \alpha_j = \alpha_i^* G_\Phi \alpha_j = 0, \quad \forall i, j = 1, \dots K.$$

Finally, we can compute the norm of the functions f_i as

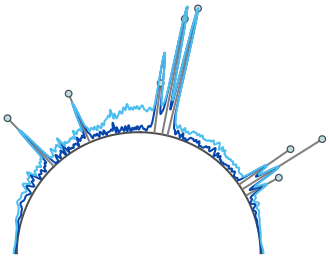
$$\|f_i\|_2 = \sqrt{\langle f_i, f_i \rangle} = \sqrt{\alpha_i^* G_\Phi \alpha_i}, \quad \forall i, j = 1, \dots K.$$



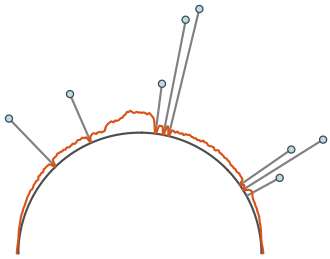
(a) Spectrum of T_{κ_D}. We observe a sharp decay (85 % of the energy contained in the 8 first eigenvalues). This can be leveraged to compress the signal with almost no loss.



(b) Compression rate versus accuracy. We observe that with a compression rate of 99% we still have an accuracy above 90 % (here accuracy means percent of total energy in the signal).



(c) Light blue: intensity function recovered with the K eigenvalues (100 % of the total energy). Dark blue: Compressed intensity function, recovered with the first $K_0 = 8$ largest eigenvalues of T_{κ_D} (85 % of total energy). We observe that the energy in the noisy sidelobes is significantly reduced, and hence the compression acts as a denoiser.



(d) Absolute error between compressed and uncompressed intensity function. We observe that the compression is not uniform across the field of view: the error near actual sources is close to zero, and increases in noisy regions. Hence the compression is virtually lossless as it only affects insignificant parts of the signal.

Figure 7 | Functional PCA used as both a compression scheme and denoiser, for the scenario desribed in fig. 6. Here the peak signal to noise ratio was set to PSNR=-6.67 dB. The eigenvalues associated to the 8 largest eigenvalues can be found on fig. 6.

The pairs $\{(\lambda_i, \epsilon_i),\ i = 1, \dots, K\}$, where $\epsilon_i = f_i / \|f_i\|,\ i = 1, \dots, K$, form hence K distinct eigenpairs for the self-adjoint, bounded operator T_{κ_D}. Moreover, from the expression eq. (12) of T_{κ_D}, we see that

$$\text{rank}(T_{\kappa_D}) = \min(\text{rank}(\Phi), \text{rank}(K)) = \min(L, K) = K.$$

Hence, $\{(\lambda_i, \epsilon_i),\ i = 1, \dots, K\}$ are the only K non-null eigenpairs of T_{κ_D}, and we have

$$T_{\kappa_D} = \sum_{i=1}^{K} \lambda_i \frac{\Phi \alpha_i \alpha_i^* \Phi^*}{\alpha_i^* G_\Phi \alpha_i}.$$

■

The decomposition eq. (17) provides us with a much more convenient representations of κ_D and I_D than previously derived in eqs. (14) and (15). Indeed, if we call the eigenfunctions

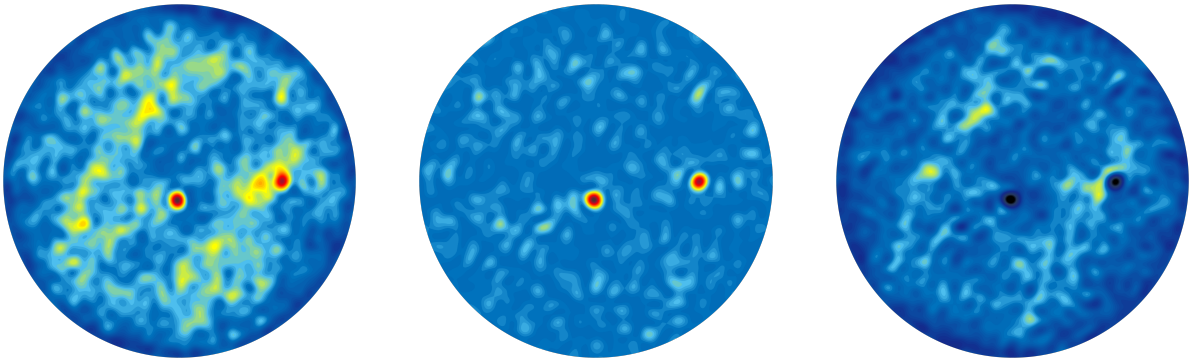
$$\epsilon_i := \frac{\Phi \alpha_i}{\sqrt{\alpha_i^* G_\Phi \alpha_i}}, \quad i = 1, \dots K,$$

then we can write

$$\kappa_D(x, y) = \sum_{i=1}^{K} \lambda_i\ \epsilon_i(x) \epsilon_i^*(y), \quad \forall (x, y) \in \mathcal{X}^2, \tag{20}$$

and

$$I_D(x) = \sum_{i=1}^{K} \lambda_i\ |\epsilon_i(x)|^2, \quad \forall x \in \mathcal{X}. \tag{21}$$



(a) Intensity field estimated using all the $K = 48$ eigenvalues.

(b) Compressed intensity field estimated using only the $K_0 = 2$ largest eigenvalues.

(c) Absolute difference between compressed and uncompressed intensity field estimates.

Figure 8 | Estimating the full sky intensity field at 50 MHz using real data from LOFAR. The estimation was carried out using 800 observations from 48 LBA antennas in station CS302. For this scenario, because of the very limited amount of samples and exposition time, only two very bright stars are distinguishable. These two sources correspond to the two largest eigenvalues, in which most of the energy is contained. By performing functional PCA and keeping only the contributions from these two leading eigenvalues, we can successfully denoise the intensity field estimate (see (b)). We observe that this compression does not introduce any bias in the estimated sources' intensities, and mainly affects noisy regions with prominent sidelobes (see (c)).

This yields an alternative algorithm for the computation and processing of κ_D and I_D in practical applications (see algorithm 1). This algorithm has numerous advantages over the algorithm proposed in [5], both in terms of computational complexity and numerical stability. Moreover, it offers much more flexibility in the processing of the estimates, allowing for example to trade accuracy and efficiency in a very controlled and systematic manner, as well as the parallel processing of each of the independent contributions of energy described by the eigenfuncitons ϵ_i. These advantages are listed below:

- **Numerical Stability:** unlike the algorithm proposed in [5], algorithm 1 does not require the inversion of the Gram matrix G_Φ, but only the resolution of the generalized eigenvalue problem eq. (16). The matrices Σ and G_Φ being respectively Hermitian symmetric and Hermitian symmetric positive definite, this problem can be solved for in a numerically stable and efficient manner using the Cholesky factorization. This improved numerical stability avoids magnifying potential noise corruptions in the samples at an early stage.
- **Computational Efficiency:** the use of eqs. (22) and (23) rather than eqs. (14) and (15) for the computation of κ_D and I_D is of course much more efficient computationally speaking. Indeed, we only need to evaluate $K_0 \ll L^2$ functions, against L^2 in eqs. (14) and (15). Moreover, both the eigenvalues λ_i and eigenfunctions ϵ_i are real-valued unlike the complex plane waves ϕ_i. In practice, we observe tremendous speed improvements when using algorithm 1 rather than the algorithm proposed in [5].
- **Trading Accuracy for Efficiency (and Denoising):** depending on the decay rate of the eigenvalues λ_i forming the spectrum of T_{κ_D}, it is possible to approximate κ_D and I_D with a reduced number of basis functions $K_0 < K$, hence further reducing the complexity with almost no loss in accuracy (see fig. 7). This trade off can be explicitly studied to find optimal compression rates. In fig. 6 for example, 8 eigenfunctions only are necessary to represent 85 % of the total energy of the signal, against the 2304 functions necessary to represent it with the algorithm proposed in [5]. This represents a reduction factor of 288, which is tremendous for practical applications. Moreover, this compression often acts as a denoiser, resulting in estimate even cleaner than if we used the entire K eigenfunctions (see figs. 7 and 8). Indeed, the eigenfunctions associated with the smallest eigenvalues are often either very corrupted by the noise or even representing the noise component only. Hence,

Algorithm 1 | Estimation of the second order moments of a Gaussian point source random function $\mathcal{S}$, from its samples $\boldsymbol{y} = \Phi^* \mathcal{S}$. The energy fraction $0 < \tau \leq 1$ trades accuracy of the estimate versus efficiency of computation.

1: **procedure** INTERP($\boldsymbol{y}, \Phi, \tau$)

2: Estimate the covariance matrix $\Sigma \in \mathbb{C}^{L \times L}$

$$\hat{\Sigma} \leftarrow \frac{1}{N_s} \sum_{i=1}^{N_s} \boldsymbol{y}_i \boldsymbol{y}_i^H,$$

3: Compute the Gram matrix $G_\Phi \in \mathbb{C}^{L \times L}$

$$G_\Phi \leftarrow \Phi^* \Phi.$$

4: Find the $K \leq L$ sorted eigenpairs of the generalized eigenvalue problem

$$\hat{\Sigma} \alpha_i = \lambda_i G_\Phi \alpha_i, \quad \forall i = 1, \ldots, K.$$

5: Select the $K_0 \leq K$ greatest eigenvalues such that

$$\sum_{i=1}^{K_0} \lambda_i > \tau \left(\sum_{i=1}^{K} \lambda_i \right).$$

6: Evaluate $\epsilon_i = \Phi \alpha_i$, $i = 1, \ldots, L$ and approximate $\kappa_D : \mathcal{X}^2 \to \mathbb{K}$ and $I_D : \mathcal{X} \to \mathbb{K}$ as

$$\kappa_D(x, y) \simeq \sum_{i=1}^{K_0} \lambda_i \, \epsilon_i(x) \epsilon_i^*(y), \quad \forall (x, y) \in \mathcal{X}^2, \tag{22}$$

$$I_D(x) \simeq \sum_{i=1}^{K_0} \lambda_i \, |\epsilon_i(x)|^2, \quad \forall x \in \mathcal{X}. \tag{23}$$

7: **end**

filtering them out results in reduced sidelobes in the estimates, with virtually no bias in relevant parts of the signal.

- **Post-processing of the eigenfunctions:** Representing independent contributions of energy, the eigenfunctions can be post-processed separately and differently according to the structures they exhibit. This not only offers an opportunity for parallelisation schemes, but also permits to design multi-scale algorithms, able to cope with wider variety of structures within the signal. One example of this will be given in section 4.3, with a multi-scale graph filtering.

3.4. A Deconvolution Problem

We have claimed earlier that the estimated random function $\mathcal{S}_D$ could be seen as a convolution between the Gaussian white noise $\mathcal{S}$ and an instrument-specific kernel $\zeta : \mathcal{X}^2 \to \mathbb{K}$, associated with the projection operator $\widetilde{\Phi}\Phi^*$:

$$\zeta = \mathrm{vec}\left(\widetilde{\Phi}\Phi^*\right).$$

This kernel can actually be written explicitly in terms of the family of the basis functions $\varphi = \{\phi_i,\ i = 1, \dots, L\} \subset \mathcal{L}^2(\mathcal{X})$. To this end, we need to define a set of basis functions associated to the interpolation operator $\widetilde{\Phi}$, that we denote $\widetilde{\varphi} = \{\widetilde{\phi}_i,\ i = 1, \dots, L\} \subset \mathcal{L}^2(\mathcal{X})$. The two bases are *dual* of one another, and form a bi-orthogonal pair of bases (see [6]). Elements of $\widetilde{\varphi}$ can be written in terms of the elements of φ as

$$\widetilde{\phi}_i(x) = \sum_{j=1}^{L} \gamma_{ji}\phi_j(x), \quad \forall x \in \mathcal{X},\ i = 1, \dots, L,$$

where γ_{ji} denotes the element of $G_\Phi^{-1} \in \mathbb{K}^{L\times L}$ at the jth row and ith column. Hence, we can re-write the indexed random variables in $\mathcal{S}_D$ (see eq. (11)) as

$$\begin{aligned} S_D(x) &= \sum_{i=1}^{L} y_i\widetilde{\phi}_i(x) \\ &= \sum_{i=1}^{L} \left(\int_{\mathcal{X}} \phi_i^*(y)\,\mathcal{S}(dy)\right)\widetilde{\phi}_i(x) \\ &= \int_{\mathcal{X}} \left(\sum_{i=1}^{L} \widetilde{\phi}_i(x)\phi_i^*(y)\right)\mathcal{S}(dy), \quad \forall x \in \mathcal{X}. \end{aligned} \tag{24}$$

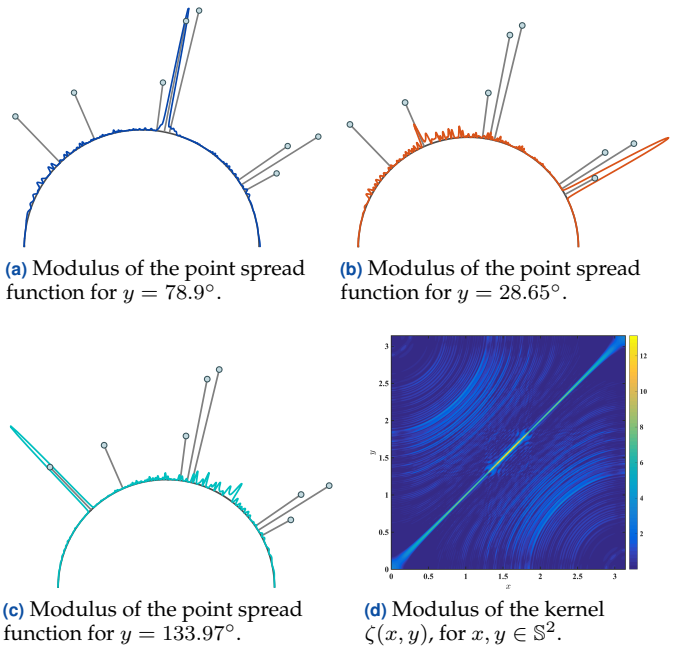
and we finally get the following expression for $\zeta : \mathcal{X}^2 \to \mathbb{K}$,

$$\zeta(x, y) = \sum_{i=1}^{L} \widetilde{\phi}_i(x)\phi_i^*(y), \quad \forall (x, y) \in \mathcal{X}^2. \tag{25}$$

The quantity $\zeta(\cdot, y_0) : \mathcal{X} \to \mathbb{K}$ is sometimes called the point spread function of the instrument at $y_0 \in \mathcal{X}$. Notice that it depends only on the functions ϕ_i, and hence on the layout of the underlying phased array. We could have derived an equivalent but more compact expression for ζ using eq. (17), but we would have then made the kernel data dependent, which is undesirable. Indeed, we would like to be able to think of ζ as a data-independent quantity, which can be optimized by properly choosing the layout of the array for example.

From eqs. (24) and (25), we can derive a spectral representation (see definition 2.7) for $\mathcal{S}_D$:

$$S_D(x) = \int_{\mathcal{X}} \zeta(x, y)\,\mathcal{S}(dy), \quad \forall x \in \mathcal{X}.$$



(a) Modulus of the point spread function for $y = 78.9°$.

(b) Modulus of the point spread function for $y = 28.65°$.

(c) Modulus of the point spread function for $y = 133.97°$.

(d) Modulus of the kernel $\zeta(x, y)$, for $x, y \in \mathbb{S}^2$.

Figure 9 | Convolution kernel ζ for the layout described in fig. 4. We observe that the point spread function is not shift invariant: the artefacts in the point spread function are stronger for off-centred locations.

As shown in eq. (5), a similar representation holds for the covariance function κ_D:

$$\kappa_D(x, y) = \int_{\mathcal{X}} \zeta(x, u)\zeta^*(y, u)\,\nu(du), \quad \forall (x, y) \in \mathcal{X}^2,$$

where ν is the control measure of the Gaussian point source random field $\mathcal{S}$.

4. Deconvolution of Gaussian Random Fields

We have derived in section 3.4 a spectral representation for the least-squares estimate $\mathcal{S}_D$, expressing it as the convolution between the unknown random function $\mathcal{S}$ and a know instrument-specific kernel ζ:

$$S_D(x) = \int_{\mathcal{X}} \zeta(x, y)\,\mathcal{S}(dy), \quad \forall x \in \mathcal{X}. \tag{26}$$

The inverse problem of recovering $\mathcal{S}$ from eq. (26) is called a deconvolution problem. For Gaussian random functions, this deconvolution can equivalently be performed on the covariance function κ_D. The goal is then to recover the spectral measure ν of $\mathcal{S}_D$, verifying the equation:

$$\kappa_D(x, y) = \int_{\mathcal{X}} \zeta(x, u)\zeta^*(y, u)\,\nu(du), \quad \forall (x, y) \in \mathcal{X}^2. \tag{27}$$

Notice that the spectral measure ν is also the control measure of the Gaussian white noise $\mathcal{S}$ and hence its knowledge suffices to fully characterise the white noise (see definition 2.4). This formulation conveniently transfers the problem of deconvolving Gaussian random functions into a deterministic setup.

For point source random functions, eq. (27) simplifies considerably. Indeed, the control measure ν can be written as (see definition 2.5)

$$\nu = \sum_{q=1}^{Q} \sigma_q^2 \delta_{x_q}, \tag{28}$$

with $\sigma_q^2 > 0$ and $x_q \in \mathcal{X}$, for $q = 1, \dots, Q$. Plugging eq. (28) in eq. (26) yields

$$\kappa_D(x,y) = \sum_{q=1}^{Q} \sigma_q^2 \zeta(x, x_q)\zeta^*(y, x_q), \quad \forall (x,y) \in \mathcal{X}^2. \tag{29}$$

Equation (29) seems to suggest a two-stages deconvolution procedure, decoupling the problem of sources localisation from the problem of intensities recovery. Indeed, assuming the sources positions x_q known, the sources' intensities σ_q^2 can easily be estimated by evaluating eq. (29) at Q well-chosen locations in order to form a linear system of Q independent equations. We then face a classic linear regression problem, with minor complications due to the structure of the noise (see section 4.1).

We hence propose in this section a two-stages deconvolution procedure. First, the sources are located using a high-pass filtering on a suitable graph. Then, we estimate the sources intensities by solving a linear regression problem. For simplicity, we postpone the sources localisation procedure to sections sections 4.2 and 4.3.

4.1. Deconvolution by Means of Linear Regression

Assuming the sources locations $\{x_q,\ q = 1, \dots, Q\} \subset \mathcal{X}$ known, we want to estimate the sources intensities $\{\sigma_q^2,\ q = 1, \dots, Q\} \subset \mathbb{R}$ from the equation

$$\kappa_D(x,y) = \sum_{q=1}^{Q} \sigma_q^2 \zeta(x, x_q)\zeta^*(y, x_q), \quad \forall (x,y) \in \mathcal{X}^2. \tag{30}$$

To this end, we take a collection of N points in $\mathcal{X}$

$$\Delta = \{x_k \in \mathcal{X},\ k = 1, \dots, N\},$$

and sample eq. (30) at these selected locations. This yields a form a matrix equation:

$$\mathcal{K} = \mathcal{Z} \Lambda \mathcal{Z}^H, \tag{31}$$

where $\Lambda := \operatorname{diag}(\boldsymbol{\sigma}) \in \mathbb{R}^{Q \times Q}$ is a diagonal matrix with diagonal entries given by $\boldsymbol{\sigma} := (\sigma_1^2, \dots, \sigma_Q^2) \in \mathbb{R}^Q$. The matrices $\mathcal{K} \in \mathbb{K}^{N \times N}$ and $\mathcal{Z} \in \mathbb{K}^{N \times Q}$ are defined as

$$\mathcal{K} := \begin{bmatrix} \kappa_{1,1} & \cdots & \kappa_{1,N} \\ \vdots & \ddots & \vdots \\ \kappa_{N,1} & \cdots & \kappa_{N,N} \end{bmatrix}, \text{ and } \mathcal{Z} := \begin{bmatrix} \zeta_{1,1} & \cdots & \zeta_{1,Q} \\ \vdots & \ddots & \vdots \\ \zeta_{N,1} & \cdots & \zeta_{N,Q} \end{bmatrix},$$

where $\kappa_{i,j} = \kappa(x_j, x_i), \forall i,j = 1, \dots, N$, $\zeta_{k,q} = \zeta(x_i, x_q), \forall k = 1, \dots, N$, $\forall q = 1, \dots, Q$.

The collection of points $\Delta \subset \mathcal{X}$ can either be chosen uniformly over the domain, or according to a prior we have on ν, such as the least-squares estimate of the intensity field $I_D(x) = \operatorname{Var}(S_D(x))$. Indeed, the N points with higher magnitude in the intensity field are likely to be more informative samples about ν. Alternatively, one could also use the eigenfunctions of the covariance operator T_{κ_D} (see eq. (17) on how to compute these eigenfunctions), and select the points with higher magnitudes in the first eigenfunctions containing a significant portion of the total variance. We observe that, to get an overdetermined system, and hence unicity of the solution when invertible, we should choose $N(N+1)/2 \geq K$. In the noiseless case and when the system is invertible, choosing $N(N+1)/2 = K$ is theoretically sufficient to uniquely recover $\{\nu_1, \dots, \nu_K\}$. In the presence of noise, choosing $N(N+1)/2 \geq K$ can bring noise resilience.

It is possible to re-write eq. (31) on the form of a traditional linear system by using the vec( ) operator [14, 9]. We get [1]

$$\begin{aligned} \operatorname{vec}(\mathcal{K}) &= \overline{\mathcal{Z}} \otimes \mathcal{Z} \operatorname{vec}(\Lambda) \\ \boldsymbol{\kappa} &= \overline{\mathcal{Z}} \otimes \mathcal{Z}\ \boldsymbol{\varsigma}, \end{aligned} \tag{32}$$

where $\otimes$ denotes the Kronecker product. Because Λ is a diagonal matrix, eq. (32) simplifies to [1]

$$\begin{aligned} \operatorname{vec}(\mathcal{K}) &= \overline{\mathcal{Z}} \circ \mathcal{Z} \operatorname{diag}(\boldsymbol{\sigma}) \\ \boldsymbol{\kappa} &= \overline{\mathcal{Z}} \circ \mathcal{Z}\ \boldsymbol{\sigma}, \end{aligned} \tag{33}$$

where $\circ$ denotes the Khatri-Rao product, or column-wise Kronecker product. This drastically reduces the dimensionality of the search space, as we go from $\boldsymbol{\varsigma} \in \mathbb{R}^{Q^2}$ to $\boldsymbol{\sigma} \in \mathbb{R}^Q$. It is possible to even further reduce the dimensionality of eq. (33). Indeed, $\mathcal{K}$ being a symmetric matrix, we can restrict our attention to its lower (or equivalently upper) triangular part. This can be done by using the half vectorization operator vech() [9], that vectorizes a symmetric matrix and discards the entries corresponding to its upper triangular part. The vec() and vech() operators are linked through the elimination $\mathcal{E}_n : \mathbb{K}^{N^2} \to \mathbb{K}^{N(N+1)/2}$ and duplication $\mathcal{D}_n : \mathbb{K}^{N(N+1)/2} \to \mathbb{K}^{N^2}$ operators, such that [9]:

$$\operatorname{vec}(A) = \mathcal{D}_N \operatorname{vech}(A), \quad \& \quad \operatorname{vech}(A) = \mathcal{E}_N \operatorname{vec}(A),$$

for every symmetric matrix $A \in \mathbb{R}^{N \times N}$.

With these definitions in hand, we can now re-write eq. (33) as

$$\begin{aligned} \mathcal{E}_{N,L} \operatorname{vec}(\mathcal{K}) &= \mathcal{E}_{N,L} \left(\overline{\mathcal{Z}} \circ \mathcal{Z}\right)\ \boldsymbol{\sigma} \\ \operatorname{vech}(\mathcal{K}) &= \mathcal{E}_{N,L} \left(\overline{\mathcal{Z}} \circ \mathcal{Z}\right)\ \boldsymbol{\sigma} \\ \boldsymbol{\kappa}_h &= \mathcal{E}_{N,L} \left(\overline{\mathcal{Z}} \circ \mathcal{Z}\right)\ \boldsymbol{\sigma} \\ \boldsymbol{\kappa}_h &= \mathcal{M}\ \boldsymbol{\sigma}. \end{aligned} \tag{34}$$

In practice, the matrix $\mathcal{K}$ (and hence $\boldsymbol{\kappa}_h$) has to be estimated from repeated observations of the random fields $\mathcal{S}_D$ at locations $x_1, \dots, x_N \in \mathcal{X}$:

$$\hat{\mathcal{K}} = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} \boldsymbol{S}_D \boldsymbol{S}_D^T,$$

where $\boldsymbol{S}_D = (S_D(x_1), \dots, S_D(x_N)) \sim \mathcal{N}_N(0, \mathcal{K})$. For large enough sample sizes N_s, we can write

$$\operatorname{vec}\left(\hat{\mathcal{K}}\right) = \operatorname{vec}(\mathcal{K}) + \boldsymbol{\epsilon},$$

with $\boldsymbol{\epsilon} \sim \mathcal{N}_{N^2}(\mathbf{0}, R)$ and covariance matrix $R \in \mathbb{R}^{N^2 \times N^2}$ given by [9]

$$R = \frac{1}{N_s} \left(I + I_{(N,N)}\right) (\mathcal{K} \otimes \mathcal{K}),$$

where $I_{(N,N)} \in \mathbb{R}^{N^2 \times N^2}$ is the commutation matrix of order (N, N), defined in general as

$$\operatorname{vec}(A) = I_{(m,n)} \operatorname{vec}(A^T),$$

for any matrix $A \in \mathbb{R}^{m \times n}$. From that we obtain the empirical version of eq. (34)

$$\hat{\boldsymbol{\kappa}}_h = \mathcal{M}\boldsymbol{\sigma}\ +\ \boldsymbol{\epsilon}_h, \tag{35}$$

where $\boldsymbol{\epsilon_h} = \mathcal{E}_{N,L}\boldsymbol{\epsilon} \sim \mathcal{N}(0, R_h)$, with $R_h \in \mathbb{R}^{\frac{N(N+1)}{2} \times \frac{N(N+1)}{2}}$ given by

$$R_h = \mathcal{E}_{N,L}\ R\ \mathcal{E}_{N,L}^T. \tag{36}$$

We observe that eq. (35) is on the form of a classical linear regression problem, with Gaussian additive noise. However, the noise is

correlated from one measurement to another, and this correlation structure can be estimated in practice by plugging $\hat{\mathcal{K}}$ in eq. (36). This suggests a weighted least-squares estimation procedure:

$$\hat{\boldsymbol{\sigma}} = \text{argmin} \left\{ \|W^{\frac{1}{2}}(\boldsymbol{\kappa}_h - \mathcal{M}\boldsymbol{\sigma})\|_2^2 , \quad \boldsymbol{\sigma} \in \mathbb{R}^Q \right\}, \tag{37}$$

where the weighting matrix $W^{\frac{1}{2}}$ is given by $W^{\frac{1}{2}} = \hat{R}_h^{-\frac{1}{2}}$. Efficient computation of this matrix can be carried out by performing a Cholesky factorization on $\hat{R}_h$ and then a backward substitution to obtain the inverse of the Cholesky factor. Provided that $\mathcal{M}$ is full-column rank, a closed form solution for eq. (37) exists and is given by

$$\hat{\boldsymbol{\sigma}} = (\mathcal{M}^T W \mathcal{M})^{-1} \mathcal{M}^T W \hat{\boldsymbol{\kappa}}_h.$$

When $\mathcal{M}$ is ill-conditioned, the above estimate is numerically unstable and can lead to noise magnification. Regularization techniques must then be explored.

Under the assumption of known sources' locations $\{x_1, \dots, x_Q\} \subset \mathcal{X}$ and for large enough sample sizes, marginal and joint confidence intervals for $\hat{\boldsymbol{\sigma}}$ can be obtained. Indeed, we have

$$\boldsymbol{\epsilon_h} \sim \mathcal{N}(0, R_h),$$

and hence

$$\hat{\boldsymbol{\sigma}} \sim \mathcal{N}(0, R_\sigma),$$

where $R_\sigma \in \mathbb{R}^{Q \times Q}$ is given by

$$\begin{aligned} R_\sigma &= (\mathcal{M}^T W \mathcal{M})^{-1} \mathcal{M}^T W \, R_h \, W \mathcal{M} (\mathcal{M}^T W \mathcal{M})^{-1} \\ &= (\mathcal{M}^T W \mathcal{M})^{-1} \mathcal{M}^T W^{T/2} W^{1/2} R_h W^{T/2} W^{1/2} \mathcal{M} (\mathcal{M}^T W \mathcal{M})^{-1} \\ &= (\mathcal{M}^T W \mathcal{M})^{-1} \mathcal{M}^T W^{T/2} \underbrace{R_h^{-1/2} R_h R_h^{-T/2}}_{=I} W^{1/2} \mathcal{M} (\mathcal{M}^T W \mathcal{M})^{-1} \\ &= (\mathcal{M}^T W \mathcal{M})^{-1} \mathcal{M}^T W \mathcal{M} (\mathcal{M}^T W \mathcal{M})^{-1} \\ &= (\mathcal{M}^T W \mathcal{M})^{-1}. \end{aligned}$$

Marginal confidence intervals are then given by

$$\mathcal{IC}(\hat{\sigma}_q, \beta_q) = \left[\sigma_q \pm \Phi_{1-\beta_q} \sqrt{(R_\sigma)_{q,q}} \right], \quad \forall q = 1, \dots, Q.$$

where $0 < \beta_q < 1$ and $\Phi_{1-\beta_q} \in \mathbb{R}$ is the $1 - \beta_q$ quantile of the standard Gaussian distribution. For simultaneous 95% confidence confidence intervals, we can apply the Bonferroni method and choose

$$\sum_{q=1}^{Q} \beta_q = 0.05,$$

which is met for example when $\beta_q = 0.05/Q, \ \forall q = 1, \dots, Q$.

4.2. Sources Localisation by means of Graph Signal Processing

We now address the problem of estimating the positions $\{x_1, \dots, x_Q\}$ of the Q sources within the domain $\mathcal{X}$. For this, we propose to use the least-squares estimate I_D of the intensity function (see fig. 6b) . By looking at the diagonal part of eq. (29), we get

$$I_D(x) = \sum_{q=1}^{Q} \sigma_q^2 |\zeta(x, x_q)|^2, \quad \forall x \in \mathcal{X}.$$

Hence, the intensity function I_D correspond to a weighted summation of squared point spread function $|\zeta(x, x_q)|^2$, each centred at different sources' locations $x_q \in \mathcal{X}$. From fig. 9, we observe that the kernel $\zeta(x, y)$ is *diagonally dominant*, with each of the point spread functions $\zeta(\cdot, y_0)$ admitting a global maximum[1] at y_0. Roughly

[1] This is always the case in practice, by design of the phased array.

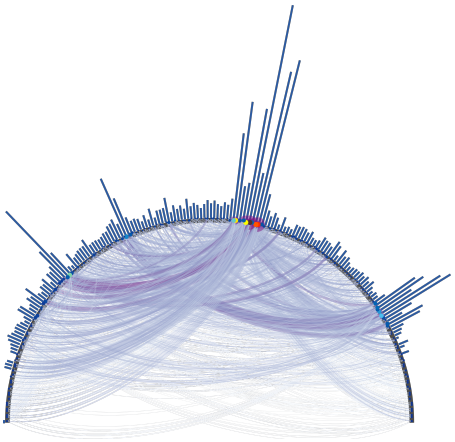


Figure 10 | Sampling and embedding of the least-squares estimate I_D from fig. 6 on a weighted graph, with weights given by the covariance function κ_D.

speaking, this means that each of the functions $|\zeta(x, x_q)|^2$ will exhibit a main strong lobe at their center x_q (see fig. 9). Hence, searching for the positions x_q of the sources, accounts to looking for strong peaks in the intensity function I_D, which would significantly pop out from the rest of the signal. In Fourier analysis terms, we are hence looking for high frequency contents within the signal I_D. An idea could then be to perform a suitable high pass filtering on the signal in the Fourier domain, to retain only significant peaks within it. Although intuitively appealing, this approach fails in our given application, as it neglects potential correlations existing between different parts of the signal (see fig. 11). Indeed, because of the complicated sidelobes structure in ζ (see fig. 9), convolution artefacts from strong sources could be locally interpreted as significantly popping out from the the signal, while they are only "echoes" of an actual source, highly correlated with the latter. Respectively, faint sources could be confused with smooth structures if surrounded by artefacts with comparable magnitudes and hence be discarded during the high-pass filtering operation. In both cases, it seems that accounting for covariance structures within the field could help in ruling out these pitfalls.

One way of doing that is to redefine our measure of smoothness, and tailor it to the covariance structure of the signal under study. The Fourier basis for example, provides a generic scale for splitting a given signal in components of different degrees of smoothness. However, being an eigenbasis for the classical Laplace-Beltrami operator, it implicitly assumes a local measure of smoothness, where signal's variations are assessed on infinitely small neighbourhoods. It is hence a very ill-suited basis (see fig. 11c) for our current purpose, as large-scale correlations patterns within the signal can only be captured with a global measure of smoothness. One possible cure would be to find an embedding $\mathcal{I}$ from $\mathcal{X}$ to a (possibly infinite dimensional) Riemannian manifold $\mathcal{M}$, such that on this manifold, the notion of *locality* and *connectivity* (as defined by the covariance structure of the signal) would agree [16]. More precisely, if $x_1, x_2 \in \mathcal{X}$ are "more correlated" than $x_3, x_4 \in \mathcal{X}$, then we would like $\mathcal{I}(x_1), \mathcal{I}(x_2) \in \mathcal{M}$ to be closer with respect to the Riemannian metric $d_{\mathcal{M}}$ from one another than $\mathcal{I}(x_3), \mathcal{I}(x_4) \in \mathcal{M}$:

$$\begin{aligned} &\mathbb{E}[S_D(x_1) S_D^*(x_2)] \leq \mathbb{E}[S_D(x_3) S_D^*(x_4)] \\ \Rightarrow \ &d_{\mathcal{M}}(\mathcal{I}(x_1), \mathcal{I}(x_2)) \leq d_{\mathcal{M}}(\mathcal{I}(x_3), \mathcal{I}(x_4)). \end{aligned}$$

Provided the existence of such a manifold, we could then define a new Laplace operator $\Delta_{\mathcal{M}}$, adapted to the new underlying geome-

try. The Fourier basis would then be defined as the eigenvectors of $\Delta_{\mathcal{M}}$, and would be a much better indicated basis to perform the high-pass filtering described above, as tailored to the covariance structure of the signal.

While finding such a manifold $\mathcal{M}$ and deriving the associated Laplace operator $\Delta_{\mathcal{M}}$ and its eigenbasis are challenging tasks at the continuous level, things simplify considerably at the discrete level. Indeed, assuming a very fine grid $\{x_1, \dots, x_{N_p}\}$ with resolution $N_p \in \mathbb{N}$, we can sample the intensity field I_D and embed the signal on a a graph structure [16], skeleton of the underlying geometry of $\mathcal{M}$ (see fig. 10). More precisely, let $\mathcal{G} = (V, E)$ be a fully connected undirected graph, where $V = \{x_1, \dots, x_{N_p}\} \subset \mathcal{X}$ is the vertex set, and $E = V \times V$ is the edge set. We define a signal $\boldsymbol{i}_D \in \mathbb{R}^{N_p}$ on V, given by

$$\boldsymbol{i}_D = [I_D(x_1), \dots, I_D(x_{N_p})],$$

and weight the edges in E so as to reflect potential similarities between nodes,

$$w_{ij} = \kappa_D(x_i, x_j), \quad \forall i, j = 1, \dots, N_p.$$

This weighting effectively encodes in the graph the connectivity patterns described by the covariance function κ_D. We can then define the graph Laplacian $L \in \mathbb{K}^{N_p \times N_p}$ of $\mathcal{G}$, discrete analog of the Laplace-Beltrami operator [16]:

$$L = \widetilde{D} - A,$$

where $A \in \mathbb{K}^{N_p \times N_p}$ is called the affinity matrix of $\mathcal{G}$ (see fig. 11a)

$$A_{ij} = \begin{cases} w_{ij}, & \text{if } (x_i, x_j) \in E, \\ 0, & \text{otherwise,} \end{cases} \qquad i, j = 1, \dots, N_p.$$

and $D \in \mathbb{R}^{N_p \times N_p}$ is a diagonal matrix with the absolute weighted degree of each node on the diagonal $D_{ii} = \sum_{k=1}^{N_p} |w_{ik}|$, $i = 1, \dots, N_p$. Analogously to the classical Fourier transform, the **Graph Fourier Transform (GFT)** [16] can be defined from the eigenvectors $U \in \mathbb{R}^{N_p \times N_p}$ of the Laplacian L:

$$L = U \Upsilon U^H,$$

where $\Upsilon \in \mathbb{R}^{N_p \times N_p}$ is the diagonal matrix of eigenvalues of L sorted in ascending order, and $UU^H = U^H U = I$. The GFT of the signal $\boldsymbol{i}_D \in \mathbb{R}^{N_p}$ defined on V, is then given by

$$\hat{\boldsymbol{i}}_D = U^H \boldsymbol{i}_D.$$

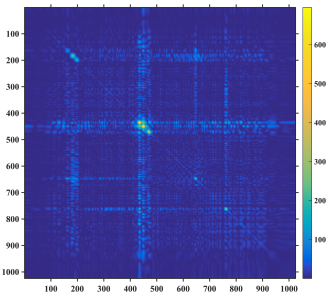
It can be shown that the eigenvectors associated to the largest eigenvalues are more oscillatory behaviours than the ones associated to the smallest eigenvalues [16]. Hence, in the graph setup, performing an ideal high-pass filtering of bandwidth $b \in \mathbb{N}$ accounts to setting to zero the first b coefficients (see fig. 11d). We note the result of this filtering $\tilde{\boldsymbol{i}}_D \in \mathbb{R}^{N_p}$:

$$\tilde{\boldsymbol{i}}_D = \begin{cases} 0, & \text{for } k = 1, \dots, b, \\ \hat{\boldsymbol{i}}_D[k], & \text{for } k = b+1, \dots, N_p. \end{cases}$$

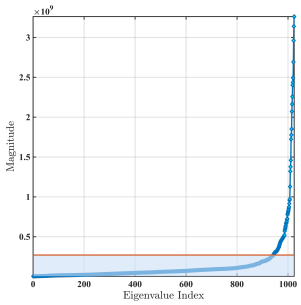
We choose the bandwidth b by finding an elbow in the spectrum Υ of L, allowing us to seperate the high-frequency eigenvectors from the low-frequency ones (see fig. 11b). Finally, we can recover the filtered signal $\boldsymbol{i}_F \in \mathbb{R}^{N_p}$ by applying the inverse GFT:

$$\boldsymbol{i}_F = U \tilde{\boldsymbol{i}}_D.$$

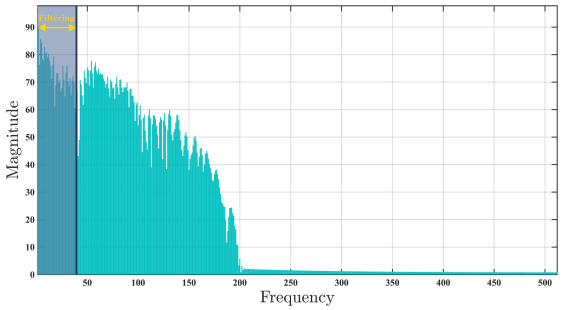
The resulting estimate exhibits much less sidelobes than the original signal $\hat{\boldsymbol{i}}_D$. Candidate sources locations can then simply be inferred from the local maxima of $\boldsymbol{i}_F$, and provided as input to the deconvolution procedure described in section 4.1 to recover the intensities.



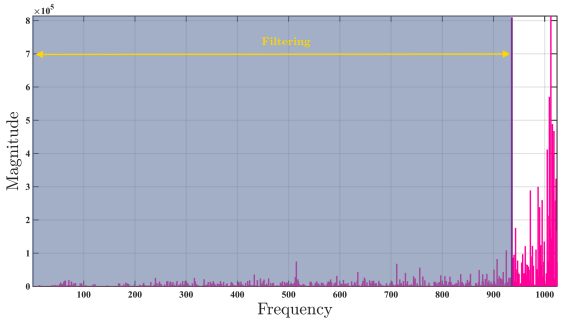
(a) Absolute value of the affinity matrix (covariance function κ_D sampled on a fine grid).



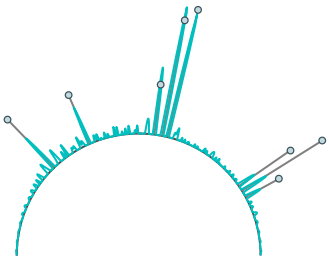
(b) Spectrum of the graph Laplacian. High frequencies correspond to large eigenvalues (above the threshold).



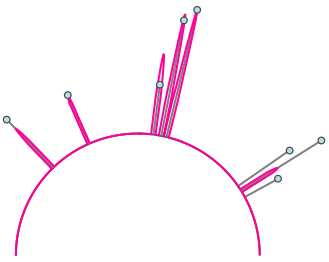
(c) Half spectrum of $\boldsymbol{i}_D$ with the discrete Fourier transform. The spectrum is very diffuse, evidence of the inadequacy of the Fourier basis to represent $\boldsymbol{i}_D$. The high-pass filter bandwidth is chosen according to the first major eigengap in the spectrum (around frequency 20).



(d) Spectrum of $\boldsymbol{i}_D$ with the graph Fourier transform. The spectrum is much more condensed in the high-frequencies.



(e) Intensity function after high-pass filtering in the DFT domain. Many sidelobes remain.



(f) Intensity function after high-pass filtering in the GFT domain. All the sidelobes have disappeared.

Figure 11 | Sources localization by means of a GFT-based high-pass filtering and comparison with the classical DFT. The merits of the GFT are obvious: it is much easier to determine the optimal filter bandwidth, and the convolution artefacts (sidelobes) are fully filtered out.

4.3. A Multi-scale Approach

The graph-based filtering procedure proposed in section 4.2 suffers from two major limitations:

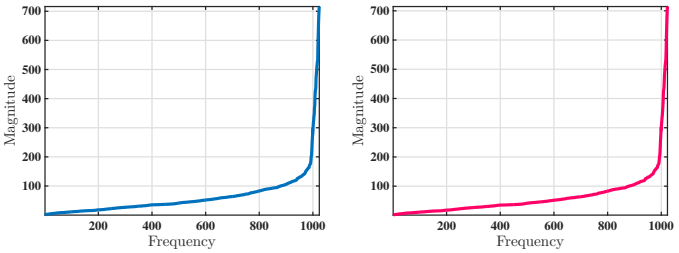
- **Computational Efficiency & Memory Consumption:** the procedure can become very computationally expensive and memory demanding, as it requires to construct and store a potentially dense and large matrix (square of the resolution N_p) as well as to perform its eigendecomposition. For very large N_p, forming the matrix only is already prohibitive. Although sparsification techniques could be explored to store and handle the matrix more efficiently, it remains unsatisfying, as there is a significant computational overhead in the procedure. Indeed, not knowing a priori how the eigenvectors of the graph Laplacian will cluster in high and low frequency components, we have to compute its full eigendecomposition and look for a change of regime in the spectrum, such as an elbow in fig. 11b. Once the bandwith of the high-pass filtering chosen, only a few eigenvectors are selected, the vast majority (corresponding to low frequencies) being discarded. This is a total waste of resources, especially since the discarded eigenvectors are the most expensive to compute! State of the art algorithms such as the Arnoldi method have indeed a convergence rate exponential in the minimal eigengap in the spectrum. Hence, if the spectrum exhibits a flat region, as observed for low-frequencies in fig. 11b, then the corresponding eigenvalues and eigenvectors will be increasingly hard to compute with a prescribed accuracy. It would hence be much more efficient to estimate in advance the elbow in the spectrum, and then compute only the eigenvalues and eigenvectors corresponding to high frequencies. We will show that this is achievable for a specific type of graph Laplacians, that we call **atomic Laplacians**. For this special type of Laplacians, we will further be able to propose a *matrix-free* [17] formulation of the Arnoldi method, hence reducing a lot the memory overhead.
- **Accuracy for Faint Sources:** one the main disadvantages of using the covariance matrix as an affinity matrix, is that it is *scale-dependent*. Hence, in the presence of very bright sources in the field, the covariance matrix can be dominated by the contributions of these sources, making it very hard to recover the faint sources. This is typically what we observe in fig. 11f where the faint sources on the right-hand side were filtered out, considered as smooth components. A multi-scale filtering procedure would hence be welcome, where sources with very different intensity levels would be treated independently from one another.

To overcome both these limiations, we propose a multi-scale graph-based filtering procedure based on the spectral approximation eqs. (22) and (23) of κ_D and I_D:

$$\kappa_D(x,y) \simeq \sum_{i=1}^{K_0} \lambda_i\ \epsilon_i(x)\epsilon_i^*(y), \quad \forall (x,y) \in \mathcal{X}^2,$$

$$I_D(x) \simeq \sum_{i=1}^{K_0} \lambda_i\ |\epsilon_i(x)|^2, \quad \forall x \in \mathcal{X},$$

where K_0 is the truncation level for the chosen accuracy level τ_0, and $\epsilon_i = \Phi\alpha_i / \|\Phi\alpha_i\|$, $i = 1, \dots, K_0$. From the above equation and the orthogonality of the ϵ_i's, we note that we can decompose the intensity function I_D in independent contributions $|\epsilon_i|^2$ (see fig. 6c). To each of these contributions, we can associate a covariance structure $\epsilon_i\epsilon_i^*$, and hence define a graph structure as in section 4.2. The hope is that the sources will be clustered in groups of similar scales across these different subgraphs, providing a much better chance of recovery for faint sources, even in the presence of strong sources.



(a) True spectrum. (b) Spectrum estimated with eq. (40).

Figure 12 | Estimation of the spectrum of an atomic Laplacian, with the approximation eq. (40). We observe that the approximation is very good, with the two sepctrums almost indistinguishable from one another.

Sampling the above quantities on a grid $\{x_1, \dots, x_{N_p}\} \subset \mathcal{X}$ we get K_0 signals

$$\boldsymbol{i}_D^{(k)} = \left[|\epsilon_k(x_1)|^2, \dots, |\epsilon_k(x_{N_p})|^2\right] \in \mathbb{K}^{N_p}, \quad k = 1, \dots, K_0,$$

defined over K_0 graphs, with affinity matrices given by

$$A_k = \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H \in \mathbb{K}^{N_p \times N_p}, \quad k = 1, \dots, K_0,$$

where $\boldsymbol{\epsilon}_k = \left[\epsilon_k(x_1), \dots, \epsilon_k(x_{N_p})\right] \in \mathbb{K}^{N_p}$, $k = 1, \dots, K_0$. Being outer products of a vector by itself, these affinity matrices admit a Laplacian of the form

$$L_k = \|\boldsymbol{\epsilon}_k\|_1 \text{diag}(|\boldsymbol{\epsilon}_k|) - \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H, \quad k = 1, \dots, K_0. \tag{38}$$

We call graph Laplacians of the form (38) **atomic Laplacians**. The eigenvalues of atomic Laplacians can be shown to be given by the roots of a certain function:

Proposition 4.1 — ***Eigenvalues of Atomic Laplacians.*** *Let $L \in \mathbb{K}^{N\times N}$ be a square matrix, given by*

$$L = \|\boldsymbol{\epsilon}\|_1 diag(|\boldsymbol{\epsilon}|) - \boldsymbol{\epsilon}\boldsymbol{\epsilon}^H,$$

with $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}[1], \dots, \boldsymbol{\epsilon}[N]]^T \in \mathbb{K}^N$, and $\boldsymbol{\epsilon}[n] \neq 0,\ \forall\, k = 1, \dots, N$. Then, the eigenvalues of L are the roots of the function

$$f(\tau) = 1 - \sum_{n=1}^{N} \frac{|\boldsymbol{\epsilon}[n]|^2}{\|\boldsymbol{\epsilon}\|_1 |\boldsymbol{\epsilon}[n]| - \tau}. \tag{39}$$

■ **Proof 4.1** Let $\lambda_n = \|\boldsymbol{\epsilon}\|_1 |\boldsymbol{\epsilon}[n]|$, for $n = 1, \dots, N$. We will first show that the λ_n's are not eigenvalues of L. By contradiction, assume the existence of a nonnull vector $\boldsymbol{v} \in \mathbb{K}^N$, such that, for a given $n \leq N$ we have

$$(\|\boldsymbol{\epsilon}\|_1 \text{diag}(|\boldsymbol{\epsilon}|) - \lambda_n I)\boldsymbol{v} = \boldsymbol{\epsilon}\boldsymbol{\epsilon}^H \boldsymbol{v}.$$

Looking at the nth coordinate we get

$$0 = \langle \boldsymbol{\epsilon}, \boldsymbol{v} \rangle \boldsymbol{\epsilon}[n],$$

and because $\boldsymbol{\epsilon}[n] \neq 0, \forall n$, we must have $\langle \boldsymbol{\epsilon}, \boldsymbol{v} \rangle = 0$. But then, all the coordinates on the left hand-side are null, yielding

$$(\|\boldsymbol{\epsilon}\|_1 |\boldsymbol{\epsilon}[i]| - \lambda_n)\boldsymbol{v}[i] = \|\boldsymbol{\epsilon}\|_1 (|\boldsymbol{\epsilon}[i]| - |\boldsymbol{\epsilon}[n]|)\boldsymbol{v}[i].$$

Again, because $\boldsymbol{\epsilon}[i] \neq 0, \forall i$, we find that $\boldsymbol{v} = 0$ and hence λ_n is not an eigenvalue of L as the kernel of $L - \lambda_n I$ is trivial, for every $n = 1, \dots, N$.

Moreover, the characteristic polynomial of L is given by

$$\begin{aligned} \det(L - \tau I) &= \det(\|\boldsymbol{\epsilon}\|_1 \text{diag}(|\boldsymbol{\epsilon}|) - \tau I - \boldsymbol{\epsilon}\boldsymbol{\epsilon}^H) \\ &= (\lambda_1 - \tau) \cdots (\lambda_N - \tau) \left[1 - \sum_{n=1}^{N} \frac{|\boldsymbol{\epsilon}[n]|^2}{\|\boldsymbol{\epsilon}\|_1 |\boldsymbol{\epsilon}[n]| - \tau}\right], \end{aligned}$$

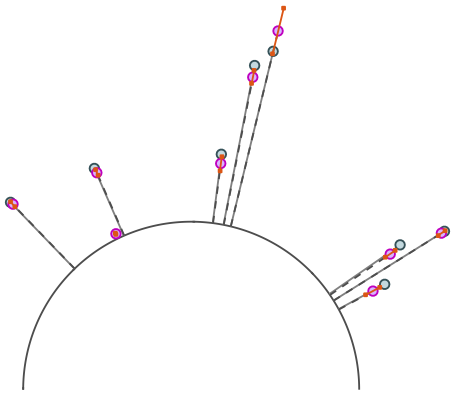


Figure 13 | Deconvolved intensity field, after sources localization by means of multi-scale graph-based high pass-filtering (see fig. 14) and intensity recovery by means of linear regression (see section 4.1). Recovered sources are in pink and actual sources in green. Confidence intervals for the intensities are superimposed in orange.

where we have used the identity:

$$\det(B + \boldsymbol{x}\boldsymbol{y}^H) = \left(1 + \boldsymbol{y}^H B^{-1}\boldsymbol{x}\right)\det(B),$$

valid for any invertible matrix $B \in \mathbb{K}^{N\times N}$. As proven above, the factors $(\lambda_n - tau)$ are all nonzero and hence the eigenvalues of L are the roots of the function:

$$f(\tau) = 1 - \sum_{n=1}^{N} \frac{|\boldsymbol{\epsilon}[n]|^2}{\|\boldsymbol{\epsilon}\|_1|\boldsymbol{\epsilon}[n]| - \tau}.$$

■

Because we are only interested in estimating roughly the location of the elbow in the specturm, we do not compute exactly the roots of eq. (39), but rather approximate them very efficiently. Indeed, a quick study of the function f reveals that there is exactly one root between each forbidden value of f, yielding for $k = 1, \ldots, K_0$

$$\tau_i^{(k)} \simeq \frac{\tilde{\boldsymbol{\varepsilon}}_k[i] + \tilde{\boldsymbol{\varepsilon}}_k[i-1]}{2}, \quad i = 2, \ldots, N, \tag{40}$$

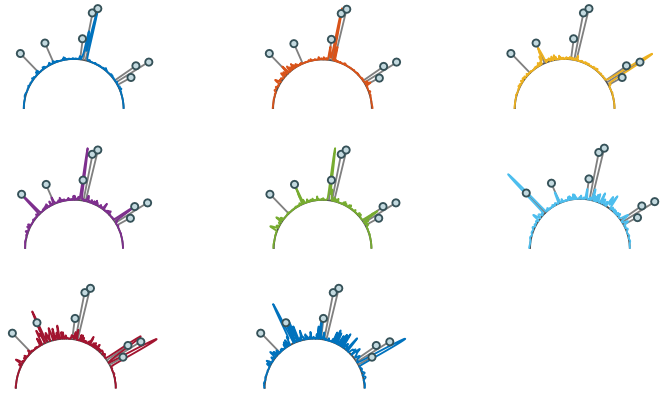
where $\tau_1^{(k)} < \ldots < \tau_{N-1}^{(k)}$ and $\tilde{\boldsymbol{\varepsilon}}_k$ are the vectors $\|\boldsymbol{\epsilon}_k\|_1|\boldsymbol{\epsilon}_k|$ sorted in ascending order.

From eq. (40), we can hence very efficiently approximate the spectrum of the atomic Laplacians L_k as well as the location of their respective elbows (see fig. 12). Then, the selected high-frequency eigenvectors and eigenvalues are very efficiently computed with a matrix-free formulation of the Arnoldi method. Indeed, the Arnoldi method does not require the explicit knowledge of the matrices L_k but only the result of their multiplication with any vector $\boldsymbol{v} \in \mathbb{K}^{N_p}$. In our case, given the form eq. (38) of the atomic Laplacians L_k, we can very efficiently compute such evaluations with:

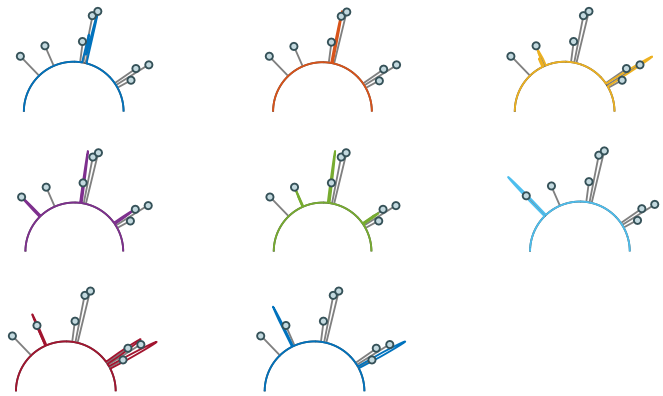
$$L_k\boldsymbol{v} = \|\boldsymbol{\epsilon}_k\|_1|\boldsymbol{\epsilon}_k| \odot \boldsymbol{v} - \langle \boldsymbol{v}, \boldsymbol{\epsilon}_k\rangle\boldsymbol{\epsilon}_k, \quad \forall \boldsymbol{v} \in \mathbb{K}^{N_p},\ k = 1, \ldots, K_0. \tag{41}$$

where $\odot$ represents the point-wise or Hadamard product. Equation eq. (41) is particularly memory efficient, as it permits to run the Arnoldi method on very large Laplacians $L_k \in \mathbb{K}^{N_p\times N_p}$ without having to even form these matrices.

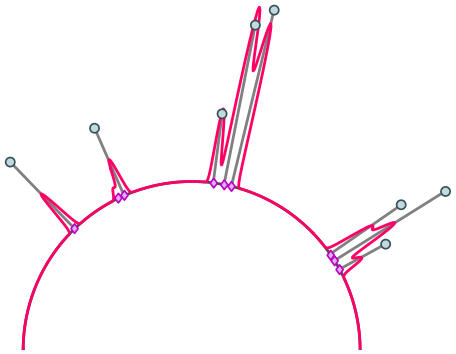
We can hence very efficiently perform a GFT-based high-pass filtering on each of the K_0 independent contributions $\boldsymbol{i}_D^{(k)}$, embedded on their respective graphs with affinity matrix A_k (see fig. 14). Each of these filtering operations being independent from one another, they can be performed in parallel. The K_0 filtered signals $\boldsymbol{i}_F^{(k)}$ can then be combined together as:



(a) Squared magnitude of the sampled eigenfunctions $\boldsymbol{i}_D^{(k)}$ associated with the $K_0 = 8$ largest eigenvalues of the covariance operator T_{κ_D}, in decreasing order.



(b) Squared magnitude of the high-passed filtered eigenfunctions $\boldsymbol{i}_F^{(k)}$.



(c) Intensity function $\tilde{\boldsymbol{i}}_F$ recovered from the high-pass filtered eigenfunctions $\boldsymbol{i}_F^{(k)}$. Candidate locations for the sources are inferred from local maxima, and marked of pink diamonds.

Figure 14 | Sources localization by means of a multi-scale GFT-based high-pass filtering. Thanks to the multiscale approach, the faint sources missed in fig. 11f are recovered.

$$\tilde{\boldsymbol{i}}_F = \sum_{k=1}^{K_0} \lambda_k \boldsymbol{i}_F^{(k)}.$$

The resulting estimate exhibits much less sidelobes than the original signal $\hat{\boldsymbol{i}}_D$. Moreover, simulations reveal that it is more likely to detect faint sources even in the presence of very bright sources, because of its multi-scale approach (see fig. 14). Candidate sources locations can then simply be inferred from the local maxima of $\tilde{\boldsymbol{i}}_F$,

and provided as input to the deconvolution procedure described in section 4.1 to recover the intensities (see fig. 13).

5. Conclusion

In conclusion, we have proposed a generic and efficient deconvolution algorithm for point source Gaussian random fields as sensed by phased arrays. Based on the procedure described in [5], the algorithm first performs very efficiently a functional principal component analysis on the least-squares estimate of the random field. This permits to conveniently represent the signal in independent contributions, allowing explicit tradeoff between accuracy and efficiency, denoising and parallel post-processing.

We then sample each of these contributions at a high enough resolution and use their associated covariance kernel to embed them on weighted graphs. A Graph Fourier Transform (GFT) is then leveraged in order to filter out the convolution artifacts within each of the independent contributions. We show that this step can be performed in a computationally and memory efficient manner for atomic Laplacians. The filtered eigenfunctions are then summed together to form an artefact-free intensity function. Candidate locations of the sources are then identified with local maxima of the resulting intensity function. From these locations a deconvolution problem is finally solved by means of weighted linear regression and the intensities of the sources within the field recovered.

The algorithm was tested on simulations for a variety of different randomly generated Gaussian point source random functions, defined over the sphere. For reasonable numbers of sources at least (less than 40), the algorithm seemed to perform very well, with very few false negative and a good accuracy in the recovered intensities and locations. However, a more formal performance analysis should be performed, with a much larger number of sources and on the sphere. Different noise conditions should also be tested. Finally, future work may include extending the filtering technique proposed in this paper at the continuous level, without having to sample the quantities at a given resolution.

References

[1] A.-J. van der Veen and S. J. Wijnholds, "Signal processing tools for radio astronomy," in "Handbook of Signal Processing Systems," (Springer, 2013), pp. 421–463.

[2] O. Ocal, "Data reduction algorithms for radio astronomy antenna stations," Master's thesis, EPFL (Spring 2014).

[3] M. Simeoni, "Towards more accurate and efficient beamformed radio interferometry imaging," Master's thesis, EPFL (Spring 2015).

[4] M. Simeoni, "Statistical inference in positron emission tomography," Tech. rep., EPFL (2014).

[5] M. Simeoni, "Attributing aliasing artifacts in dirty intensity fields to their parent source," Tech. rep., EPFL (2016).

[6] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of signal processing* (Cambridge University Press, 2014).

[7] M. Lifshits, *Lectures on Gaussian processes* (Springer, 2012).

[8] M. A. Lifshits, *Gaussian random functions*, vol. 322 (Springer Science & Business Media, 2013).

[9] K. Jinadasa, "Applications of the matrix operators vech and vec," Linear Algebra and its Applications **101**, 73–79 (1988).

[10] P. A. Dirac, "A new notation for quantum mechanics," in "Proceedings of the Cambridge Philosophical Society," , vol. 35 (Cambridge Univ Press, 1939), vol. 35, pp. 416–418.

[11] P. Hurley and M. Simeoni, "Flexibeam: analytic spatial filtering by beamforming," in "International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE," (March 2016 (to appear)).

[12] O. Ocal, P. Hurley, G. Cherubini, and S. Kazemi, "Collaborative randomized beamforming for phased array radio interferometers," in "International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE," (2015).

[13] M. Van Haarlem, M. Wise, A. Gunst, G. Heald, J. McKean, J. Hessels, A. De Bruyn, R. Nijboer, J. Swinbank, R. Fallows *et al.*, "Lofar: The low-frequency array," Astronomy & astrophysics **556**, A2 (2013).

[14] C. F. Van Loan, "The ubiquitous kronecker product," Journal of computational and applied mathematics **123**, 85–100 (2000).

[15] B. N. Parlett, *The symmetric eigenvalue problem*, vol. 7 (SIAM, 1980).

[16] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," IEEE Signal Processing Magazine **30**, 83–98 (2013).

[17] S. Diamond and S. Boyd, "Matrix-free convex optimization modeling," arXiv preprint arXiv:1506.00760 (2015).