Sparse acoustical holography from iterated Bayesian focusing

Jérôme Antoni, Thibaut Le Magueresse, Quentin Leclere, Patrice Simard

Version of Record: https://www.sciencedirect.com/science/article/pii/S0022460X1930001X
Manuscript_479676c75b37d071169398ac3aeb25a4

Sparse acoustical holography from iterated Bayesian focusing

Jérôme Antoni

Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, F-69621 Villeurbanne, France

Thibaut Le Magueresse

MicrodB, F-69134, Ecully, France

Quentin Leclère

Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, F-69621 Villeurbanne, France

Patrice Simard

Laboratory Roberval, CNRS UMR 6253, University of Technology of Compiegne, Centre de Recherche de Royallieu, BP 20529-60205 Cédex, Compiègne, France

Abstract

In a previous work, an attempt was made to give a unified view of some acoustic holographic methods within a Bayesian framework. One advantage of the so-called "Bayesian Focusing" approach is to introduce an aperture function that acts like a lens and thus significantly improves the reconstruction results in terms of spatial resolution, but also of quantification over a larger frequency interval than allowed by conventional methods. This is particularly remarkable when the aperture function is allowed to become very narrow as in the case of sparse sources. The aim of the present paper is to demonstrate that the aperture function – which was previously manually tuned by the user – can be automatically estimated, together with the source distribution, in the same inverse problem. The principle is to use the current estimate of the source distribution to update the aperture function in the next iteration. The resulting algorithm is an iterated version of Bayesian Focusing, which can be formalized as an Expectation-Maximization algorithm with proved convergence. The proof of convergence is based on modeling the aperture function as a random quantity, which assigns the source coefficients with prior probability distribution in the form of a "scale mixture of Gaussians" that enforces sparse solutions. Various types of sparsity enforcing priors can thus be constructed, in a much richer setting than the usual ℓ_1 penalized approach, leading to different updating rules of the aperture function. Some immediate byproducts of iterating Bayesian Focusing are 1) to provide a technique for the automatic setting of the regularization parameter, 2) to apply on the cross-spectral matrix of the measurements, and 3) to easily allow the grouping of frequencies for the broadband analysis of sources that are stationary in space. Experimental results confirm that sparse holography improves the reconstruction of sources not only in terms of localization, but also of quantification and of directivity in a frequency range considerably enlarged as compared to classical methods. These improvements can be achieved even with regular arrays, provided that sparser priors than those leading to the standard ℓ_1 penalization are used.

Keywords: Acoustical holography, acoustic imaging, source identification, inverse problem, sparsity

Email address: jerome.antoni@insa-lyon.fr (Jérôme Antoni)

List of acronyms

AF	Aperture Function
BIBF	Broadband Iterated Bayesian Focusing
BF	Bayesian Focusing
GMCG	Generalized multivariate complex Gaussian
IBF	Iterated Bayesian Focusing
LS	Least-Square
MAP	Maximum Aposteriori
MCK	Multivariate complex K-distribution
MCL	Multivariate complex Laplace
MCS	Multivariate complex Student-t
MMCG	Mixture of multivariate complex Gaussians
PDF	Probability Density Function
SMoG	Scaled Mixture of Gaussians
SNR	Signal-to-Noise Ratio
ULA	Uniform Linear Array

1. Introduction

The reconstruction of sound sources from acoustic array measurements is known as a difficult inverse problem. Several methods have been proposed over the years with sustained efforts to reach improved performance. An unavoidable limitation arises from the fact that reconstructing a continuous acoustical field from a limited number of discrete and remote measurements, as returned by an array of microphones, has no unique solution in general. This issue is particularly critical in near-field configurations, where the spectrum of the propagation operator is characterized by a fast exponential decay and is therefore hardly invertible. Consequently, sound sources cannot be reliably reconstructed unless additional information is incorporated into the inverse problem in order to reduce the space of possible solutions. So far, Tikhonov regularization has been the prevalent method to address this issue, by selecting solutions with minimal energy. However, this is often at the expense of accuracy of the reconstruction, and especially in term of spatial resolution. When the sources of interest are known to have a sparse representation, much better strategies than Tikhonov regularization are actually feasible.

Compressed sensing and the related concept of sparsity have recently encountered a huge success for solving ill-posed problems in many scientific fields [1]. Compressed sensing is essentially a new sampling theory which provides a valid alternative to Shannon theory in situations where samples are forced to remain limited. It relies on the premise that there exists a sparse representation of the quantities of interest and furnishes unexpected optimistic bounds on the number of measurements required to reconstruct it without loss. When applied to inverse problems that seek a sparse solution, sparsity intervenes as an alternative to Tikhonov regularization by imposing an apriori on the shape of the expected solution and not only on its energy. This is apt to considerably narrow the space of possible solutions and, in some situations, even shrink it to a unique solution. Technically speaking, the difference is in replacing the ℓ_2-norm penalty used in Tikhonov regularization by a ℓ_p-norm, with p a power strictly less than 2, typically equal to 1. Whereas Tikhonov regularization tries to minimize the energy (reflected by the ℓ_2-norm) of the solution with the side-effect of spreading the source in space, "sparse regularization" controls the source magnitude by forcing it to take only a few high values and shrinking it towards zero everywhere else.

The above benefits have been recognized early for the localizing of sources, a subject which has nourished an abundant literature in array processing (beamforming, Direction Of Arrival) [2, 3]. In these seminal works, sparsity arrives from the assumption that the acoustical field is produced by a few point sources (monopoles) located in the far-field. Remarkable results have been obtained based on minimizing the ℓ_1-norm, which proved superior to those of the state-of-the-art, such as returned by classical high-resolution (parametric) methods (e.g. Capon, MUSIC). Numerous works have followed, which constitute a body of literature known as compressive beamforming [4]. Applications to acoustics are found in Refs. [5, 6, 7, 8, 9, 10].

Incidentally, compressive beamforming shares some similarities with deconvolution algorithms – originally derived as a post-processing of beamforming results – such as DAMAS [11], DAMAS2, SC-DAMAS [12], CLEAN-SC

[13], NNLS, since they all try to return a final beamforming map made of a few point sources [14]. Another connection can be made with covariance matching methods [15], which also enforce sparse solutions as pointed in [16]. This fact has been exploited in Refs. [17, 18] which proposed super-resolution algorithms derived in a Bayesian setting to estimate the power of incoherent point sources in the presence of background noise. In addition to refining spatial resolution, deconvolution and covariance matching methods have been shown to improve the estimation of source powers over a large interval of frequencies; this pushes still further one of the recurrent limits of acoustic imaging techniques. However, this family of techniques critically relies on the assumption of uncorrelated sources, a constraint which is actually relaxed in compressive beamforming [8].

Generalized versions of beamforming with sparsity constraints have also been investigated in acoustics independently from the recent literature on compressive beamforming. The so-called Generalized inverse beamforming (GIBF) is truly an inverse problem that tries to recover a limited number of monopoles and dipoles by imposing a ℓ_1-norm penalty [19, 20]. Strongly inspired by deconvolution algorithms, it has been shown to outperform them in several instances, in particular in the presence of correlated sources. More general configurations have recently been investigated where the acoustical field is parsimoniously represented by a few elements taken in dictionaries different from a set of monopoles or dipoles. Reference [21] proposes to use a dictionary made of spherical harmonics so as to jointly localize the sources and identify their directivity.

Although compressive beamforming (broadly speaking as including all the aforementioned techniques) has probably reached its maturity, it has so far mostly relied on the far-field approximation of the acoustic propagator. This is an explicit assumption in Refs. [5, 6, 7, 8, 9, 10] concerned with DOA and is often implicit in deconvolution approaches based on beamforming. While GIBF is theoretically not subjected to this constraint, it seems that it has been mostly validated in the far-field as well.

There are few reports to date of the application of sparse inversion to (near-field) acoustical holography (NAH). Reference [22] was probably first to address the subject. The objective therein was to demonstrate that a substantial reduction in the number of microphones can be achieved (by more than a factor 10) without affecting the reconstruction performance by exploiting the sparsity of the acoustical field. The two fundamental assumptions are that the acoustical field can be represented by a few plane waves (sparsity in the wavenumber domain) and that the microphones are distributed randomly.

Steered by somewhat different objectives, Ref. [23] recently introduced the concept of wideband acoustical holography (WBH). The motivation in this work is for a method that can bridge the gap between NAH and classical beamforming in order to reconstruct sources over a wide frequency range. On the one hand, NAH can cover the low frequencies but requires the array to be placed at a small distance. On the other hand, beamforming has good performance in a high frequency range but requires the array to be moved at a larger distance. By enforcing sparsity in the Equivalent Source Method (ESM) (i.e. representation by a few point sources), the proposed WBH method is able to reconstruct sources with a remarkable quantification of the sound power over a wide frequency range. A numerical experiment seems to indicate that WBH is also able to consider extended sources such as produced by the vibration of a plate. The algorithm proposed in Ref. [23] is named "fast" because it uses an Iterative Hard Thresholding algorithm instead of a direct ℓ_1-norm minimization, yet it still depends on several user-dependent parameters.

Following a similar same idea, Ref. [24] also proposed a sparse version of ESM, yet based on direct ℓ_1-norm minimization. In accordance with the results of Ref. [22], the authors also stress the importance of using a sampling scheme that leads to low column coherence of the sensing matrix, a condition which is better met when the distance between the measurement plane and the source plane is decreased. As in Ref. [23], the applicability to the reconstruction of extended sources is also demonstrated by placing the equivalent sparse sources behind the actual radiating surface. The investigated frequency range is limited to that of classical NAH, where moderate improvement is observed in the reconstruction of source spatial distribution as compared to ℓ_2 regularization. Reference [25] introduced independently a sparse wave superposition method (with sparsity placed on the so-called "charge points" which actually coincide with the monopoles generating the acoustic field), yet its validation was limited to a numerical simulation. Most recently, the spare ESM has been extended to the case where a limited number of acoustical modes rather than monopoles are recovered in order to explain a source field; in Ref. [26] the "radiation" modes are obtained from the eigenvalue decomposition of the power resistance matrix [27], whereas in Ref. [28] they are computed from a finite element model of the radiating object.

It is noteworthy that Ref. [29] introduced a sparse ESM before the former publications, which was later named Iterative ESM (iESM). It was compared with GIBF in Ref. [30] and also with deconvolution approaches in Ref. [31].

Two remarks are worthwhile in light of the above literature review on compressive beamforming and sparse NAH. First, most of the aforementioned works enforce sparsity through an ℓ_1-norm penalty and essentially differ in the optimization algorithm used to minimize it. Although few methods set up the inverse problem with a general ℓ_p-norm, $p \leq 2$, the inverse problem is often solved with $p = 1$ in order to allow the use of a convex optimization algorithm (e.g Iterated Reweighted Least Squares (IRLS) in [19] and linear programming in [24]). One exception is Ref. [31] which shows the advantage of using $p < 1$ and in particular $p \to 0$ by inserting a thresholding step in IRLS. Another exception is WBH which introduces an optimization algorithm similar to Iterative Hard Thresholding, yet without specifying which norm is actually minimized.

Second, a challenging aspect of all the above approaches is to involve a regularization parameter (sometimes implicitly) that balances the effect of the sparsity enforcing penalty with the data fidelity term. Although setting the regularization parameter is known to be extremely critical, especially in NAH, no systematic rule seems to have found a consensus in the sparse case, contrary to state-of-the-art on ℓ_2-norm Tikhonov regularization [32]. The "cross-validation" criterion is often advocated, yet it is costly to apply since it has no "generalized" equivalent as can be for Tikhonov regularization [33].

The main contribution of the present paper is to demonstrate that sparse NAH is naturally devised by iterating the Bayesian Focusing (BF) method introduced in Ref. [34]. BF solves the inverse acoustical problem by considering the sound sources as random variables. They are first vaguely described by a prior probability density function (PDF) which reflects the user expectation before the experiment is run. Next, the posterior PDF is found as the product of the likelihood function and the prior PDF, which provides the probability of the sound sources after taking the measurements. A typical point estimate is then returned by the maximum value of the posterior PDF – the maximum aposteriori (MAP) – i.e. the source configuration with highest probability of occurrence given the observed data. In BF, the variance of the prior PDF – coined the aperture function (AF) – appears as a key quantity; intuitively, it will take high values where the sound sources are expected to radiate from and nil values elsewhere. Therefore, by shrinking the space of solutions to a confined region, the AF improves the quality of the reconstruction in terms of spatial resolution and of source strength quantification over a larger frequency interval than allowed by conventional methods. This is all the more remarkable as the AF is allowed to become very narrow, as in the case of sparse sources. By analogy with optics, it plays the role of a lens that focuses the light onto a point.

In Ref. [34], the AF was manually tuned by the user. A natural idea is to automate this process by using the current estimate of the source distribution as the AF to apply in the next iteration. It is shown in this paper that iterating BF generally leads to the recovery of sparse solutions. Not only is the common ℓ_1-norm penalty recovered as a particular case, but many other sparsity enforcements – possibly stronger than the ℓ_1-norm – can be devised depending on how the AF is updated. The reason is that when the AF is considered as a random quantity (in the Bayesian setting all unknowns are described by random variables), it then assigns the sound sources with a prior PDF in the form of a "scale mixture of Gaussians" (SMoG), which necessarily promotes sparsity. Based on this finding , the convergence of the iterations can be proved by establishing a formal equivalence with the Expectation-Maximization algorithm. This algorithm will be referred as Iterating Bayesian Focusing (IBF), which is reminiscent of the general concept of "sparse Bayesian learning".

One advantage of IBF is to provide some physical insight into the mechanism of sparsity enforcement. Through the role played by the AF, it explains why promoting sparse solutions not only increases the spatial resolution, but also improves the estimation of the source levels and the source directivity over a larger frequency range.

Another prime advantage of the approach is to inherit the automated regularization of BF [35]. Contrary to Tikhonov regularization for which well established algorithms exist (e.g. the L-curve, GCV), this point remains an issue in compressive beamforming. Here, by taking a Bayesian perspective, the regularization parameter is jointly inferred in IBF with proved convergence.

The Bayesian framework also easily allows the processing of multiple snapshots and multiple frequencies, which both involve the consideration of group sparsity [36]. The multi-snapshot case is addressed here by means of the cross-spectral matrix (CSM), which turns out a sufficient statistic (the approach differs in this aspect from that of Ref. [9]). The advantage of using the CSM is purely practical, since it is often the only quantity recorded by commercial data acquisition systems; indeed, it happens to be sufficient for running most imaging methods while allowing at the same time a substantial reduction of the amount of stored data. The multi-frequency approach is of interest when the sound sources have a constant position in space, independently of their frequency content, thus leading to broadband

processing. Both the multi-snapshot and the multi-frequency considerations are shown to improve the estimation of the sound sources.

The rest of the manuscript is organized as follows. Section 2 starts with a review of BF and of the key concept of an AF. Next, Section 3 explains the focusing mechanism produced by the AF and the resulting improvement in the estimation of the source strength over a wide frequency range. Section 4 then introduces the principle of the IBF method and its formalization by means of prior PDFs in the form of an SMoG. Specific cases are detailed relating to the multivariate complex Student-t, K-distribution, Laplacian, Generalized Gaussian and mixture of Gaussians. Section 5 compares these different possibilities in terms of the number of parameters to set and of systematic and random errors. Section 6 extends IBF to the broadband case. Finally, section 7 illustrates the application of IBF to experimental data. In order to ease the reading of the manuscript, all technical parts are moved to the Appendix.

The following notations are used throughout the document. By convention, vectors are denoted by bold lower-case letters and matrices by bold upper-case letters (e.g. $\mathbf{u}$ and $\mathbf{U}$, respectively). $\mathbf{I}$ stands for the identity matrix. Symbol * stands for the complex conjugate of a complex variable. Symbol ^{H} stands for the transpose conjugate of a vector or a matrix. $\text{Diag}(a_q)$ is the diagonal matrix with element a_q in its q-th diagonal entry. Operator $\mathbb{E}\{X\}$ denotes the expected value of a random variable X and $\mathbb{E}\{X|Y\}$ the expected value of X conditioned to random variable Y. By convention, square brackets are used to denote a PDF (e.g. $[X]$ denotes the PDF of random variable X). Notation $[X|Y]$ stands for the PDF of X conditioned on Y. $\mathcal{N}_{\mathbb{C}}(\mathbf{x}; \mu, \mathbf{\Omega})$ means that vector $\mathbf{x}$ follows a complex Gaussian with mean μ and covariance matrix $\mathbf{\Omega}$. The collection of indexed vectors $\{\mathbf{p}_i\}_{i=1}^N$, $\mathbf{p}_i \in \mathbb{C}^M$ (resp. $\{\mathbf{s}_i\}_{i=1}^N$) is denoted as $\mathbf{P}$ (resp. $\mathbf{S}$), an $M \times N$ matrix whose i-th column is $\mathbf{p}_i$. The ℓ_p-norm of a vector $\mathbf{x}$ in $\mathbb{C}^N$ is defined as $\|\mathbf{x}\|_p \doteq (\sum_{q=1}^N |x_q|^p)^{1/p}$. The position vector in the continuous space is denoted by $\mathbf{r}$. A function of space (e.g. σ) may be noted in three different forms: i) the functional notation, $\sigma(\mathbf{r})$, where it is viewed as a continuous function of the space variable $\mathbf{r}$, ii) the discrete notation, $\sigma_q \doteq \sigma(\mathbf{r}_q)$, where it is viewed as a sample at position $\mathbf{r}_q$, and iii) the matrix notation, $\Sigma \doteq \text{Diag}(\sigma_q)$, which collects all the space samples in a diagonal matrix.

2. Review of Bayesian Focusing (BF)

2.1. Parametrization of the problem

The array is composed of M microphones indexed by the lower-case letter m, located at positions $\mathbf{r}_m$, $m = 1, ..., M$. All measurements are considered at a given frequency f, after application of the Fourier transform on a series of N snapshots indexed by the lower-case letter i ($i = 1, ..., N$). The Fourier coefficients of the sound pressure measured by microphone m and assigned to snapshot i then reads $p_{m,i}$, where explicit dependence on frequency f is dropped to simplify the notation. The M Fourier coefficients $p_{m,i}$, $m = 1, ..., M$ are collected in the column vector $\mathbf{p}_i$. The cross-spectral matrix (CSM) of the measurements averaged over N snapshots is defined as

$$\mathbb{S}_{pp} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{p}_i \mathbf{p}_i^H. \tag{1}$$

In the following, formulations in either continuous or discrete space will be used alternatively, the former because it provides deeper physical insight and the latter because it corresponds to the numerical resolution of the problem. In the continuous formulation, the source distribution is assumed to be a scalar field – e.g. parietal pressure, normal velocity – that adheres on the source surface Γ_s. In a functional notation, it is noted $s_i(\mathbf{r})$ where $\mathbf{r} \in \Gamma_s$ stands for the position vector. In the discrete formulation of the problem, the spatial samples of the source distribution taken at positions $\mathbf{r}_q$, $q = 1, ..., Q$ are noted $s_{q,i} \doteq s_i(\mathbf{r}_q)$. They are referred to as the "source coefficients" and are collected in the column vector $\mathbf{s}_i$, with dimension Q. The power spectrum of the source coefficients at position $\mathbf{r}_q$, averaged over N snapshots, is defined as

$$S_{qq} = \frac{1}{N} \sum_{i=1}^{N} |s_{q,i}|^2. \tag{2}$$

2.2. Direct problem

The relationship between the source distribution and the radiated pressures measured by the microphones is hereafter considered in the form of a Fredholm integral. In a functional notation, the direct problem therefore reads

$$p_{m,i} = \int_{\Gamma_s} G(\mathbf{r}_m|\mathbf{r}) s_i(\mathbf{r}) \mathrm{d}\Gamma(\mathbf{r}) + n_{m,i}, \quad i = 1, ..., N \tag{3}$$

where $G(\mathbf{r}_m|\mathbf{r})$ is the Green function between the source point $\mathbf{r} \in \Gamma_s$ and the microphone location $\mathbf{r}_m$ subjected to Neumann boundary conditions on Γ_s. The quantity $n_{m,i}$ stands for additive noise at microphone m. In anticipation of a numerical resolution of the inverse problem, the discrete version of Eq. (3) is considered,

$$\mathbf{p}_i = \mathbf{G}\mathbf{s}_i + \mathbf{n}_i, \quad i = 1, ..., N, \tag{4}$$

where $\mathbf{G}$ stands for an $M \times Q$ matrix whose entry (i, q) is fed with $G(\mathbf{r}_m|\mathbf{r}_q)\Delta\Gamma(\mathbf{r}_q)$ – with $\Delta\Gamma(\mathbf{r}_q)$ a small surface element at position $\mathbf{r}_q$ – and where vector $\mathbf{n}_i$ stacks the noise terms $n_{m,i}$.

The aim of the inverse problem is to recover an estimate of the source distribution $s_i(\mathbf{r})$ from the observation of the measured pressures $\mathbf{P} \doteq \{\mathbf{p}_i\}_{i=1}^N$ returned by the array of microphone. BF provides a solution to this problem that makes intensive use of prior information about the spatial structure of the source field to be recovered.

2.3. The Bayesian program

The philosophy in the Bayesian approach is to see all unknowns in the inverse problem as random variables and to infer them through their PDFs. Of interest here is the posterior PDF $[\mathbf{s}_i|\mathbf{p}_i]$ of the vector $\mathbf{s}_i$ of coefficients given the measurement $\mathbf{p}_i$. The knowledge of $[\mathbf{s}_i|\mathbf{p}_i]$ completely characterizes the information that can be gained on $\mathbf{s}_i$ once the data have been measured. A popular point estimate used in this paper is the MAP estimate,

$$\hat{\mathbf{s}}_i = \underset{\mathbf{s}_i}{\mathrm{Argmax}}[\mathbf{s}_i|\mathbf{p}_i], \tag{5}$$

to be interpreted as the value of $\mathbf{s}_i$ with maximum probability, yet another plausible estimate is the posterior mean $\mathbb{E}\{\mathbf{s}_i|\mathbf{p}_i\} = \int \mathbf{s}_i[\mathbf{s}_i|\mathbf{p}_i]\mathrm{d}\mathbf{s}_i$.

Bayes's rule then makes it possible to express the "inverse probability" $[\mathbf{s}_i|\mathbf{p}_i]$ in terms of the probabilities assigned to the direct problem, that is

$$[\mathbf{s}_i|\mathbf{p}_i] = \frac{[\mathbf{p}_i|\mathbf{s}_i][\mathbf{s}_i]}{[\mathbf{p}_i]} \tag{6}$$

where $[\mathbf{p}_i|\mathbf{s}_i]$ is the PDF of observing the data given an instance of $\mathbf{s}_i$ (the so-called likelihood function), $[\mathbf{s}_i]$ is the PDF of the values possibly taken by $\mathbf{s}_i$ before the data are observed (the so-called prior) and $[\mathbf{p}_i] = \int [\mathbf{p}_i|\mathbf{s}_i][\mathbf{s}_i]\mathrm{d}\mathbf{s}_i$ is the "evidence". The Bayesian program is to introduce the PDFs $[\mathbf{p}_i|\mathbf{s}_i]$ and $[\mathbf{s}_i]$ according to the specificity of the problem and to deduce the posterior PDF $[\mathbf{s}_i|\mathbf{p}_i]$ by using Eq. (6).

Since the data are processed in the Fourier domain, it holds from the Central Limit Theorem applied to the Fourier transform that the elements of the noise vector $\mathbf{n}_i$ rapidly converge to a complex Gaussian [37]. Assuming for simplicity a zero mean and a covariance matrix proportional to the identity (i.e. a homogeneous field),

$$[\mathbf{p}_i|\mathbf{s}_i] = \frac{\mathrm{e}^{-\beta^{-2}(\mathbf{p}_i - \mathbf{G}\mathbf{s}_i)^H(\mathbf{p}_i - \mathbf{G}\mathbf{s}_i)}}{\beta^{2M}\pi^M} = \mathcal{N}_{\mathbb{C}}(\mathbf{p}_i; \mathbf{G}\mathbf{s}_i, \beta^2\mathbf{I}) \tag{7}$$

where β^2 stands for the value of the noise power spectrum at frequency f (independently of space according to the previous assumption) and $\pi = 3.14159...$ is Archimedes' constant.

The specification of the prior PDF $[\mathbf{s}_i]$ allows much more flexibility and is actually central in this paper. The assumption initially used in Ref. [34] – to be shortly abandoned from section 4 – is to resort to a complex Gaussian, mainly because it makes analytic calculation tractable and also because it coincides with classical Tikhonov

regularization. Namely,

$$[\mathbf{s}_i] = \frac{\mathrm{e}^{-\alpha^{-2}\mathbf{s}_i^H\Sigma_0^{-2}\mathbf{s}_i}}{\alpha^{2Q}\pi^Q|\Sigma_0^2|} = \mathcal{N}_{\mathbb{C}}(\mathbf{s}_i;\mathbf{0},\alpha^2\Sigma_0^2) \tag{8}$$

where $\alpha^2\Sigma_0^2 = \mathbb{E}\{\mathbf{s}_i^H\mathbf{s}_i\}$ stands for the prior covariance matrix of the coefficients, α^2 for their mean power and Σ_0^2 for a matrix of correlation coefficients conventionally normalized such that $\mathrm{trace}\{\Sigma_0^2\} = Q$. The fact that $\alpha^2\Sigma_0^2$ does not depend on index i reflects the assumption of a stationary acoustic field, i.e. with constant statistical properties during the acquisition of the N snapshots. Using Eq. (6), one then finds the posterior PDF in the form of a complex Gaussian, $[\mathbf{s}_i|\mathbf{p}_i] = \mathcal{N}_{\mathbb{C}}(\mathbf{s}_i;\hat{\mathbf{s}}_i,\mathbf{\Omega})$, with posterior mean and covariance matrix

$$\begin{aligned}\hat{\mathbf{s}}_i &= \Sigma_0^2\mathbf{G}^H(\mathbf{G}\Sigma_0^2\mathbf{G}^H + \eta^2\mathbf{I})^{-1}\mathbf{p}_i, \quad \eta^2 = \frac{\beta^2}{\alpha^2} \\ \mathbf{\Omega} &= \left(\beta^{-2}\mathbf{G}^H\mathbf{G} + \alpha^{-2}\Sigma_0^{-2}\right)^{-1}.\end{aligned} \tag{9}$$

In this case, the posterior mean $\hat{\mathbf{s}}_i$ coincides with the MAP estimate as well as with the Tikhonov regularized least-square solution (LS) returned by

$$\hat{\mathbf{s}}_i = \underset{\mathbf{s}_i}{\mathrm{Argmax}}[\mathbf{p}_i|\mathbf{s}_i][\mathbf{s}_i] = \underset{\mathbf{s}_i}{\mathrm{Argmin}}\left(-\ln[\mathbf{p}_i|\mathbf{s}_i][\mathbf{s}_i]\right) = \underset{\mathbf{s}_i}{\mathrm{Argmin}}\left(\|\mathbf{p}_i - \mathbf{G}\mathbf{s}_i\|_2^2 + \eta^2\|\Sigma_0^{-1}\mathbf{s}_i\|_2^2\right) \tag{10}$$

with $\|\Sigma_0^{-1}\mathbf{s}_i\|_2^2 \doteq \mathbf{s}_i^H\Sigma_0^{-2}\mathbf{s}_i$ and wherein all terms not depending on $\mathbf{s}_i$ have been dropped out. The ratio $\eta^2 = \beta^2/\alpha^2$ in the above equations acts as the Tikhonov regularization parameter which controls the stability of the solution. One benefit of the Bayesian framework is to provide a solution to automatically tune η^2, as described in Ref. [35] and resumed in Appendix B.

It is explicit from Eq. (9) that the MAP estimate directly depends on the setting of the prior covariance Σ_0. This aspect is crucial in the present paper, since it is the mechanism by which sparse solutions will be promoted as explained in the next subsections.

2.4. Definition of a spatial prior through the aperture function

Although the exact source distribution will remain unknown until the inverse problem is solved, prior information is often available before the experiment is run, even in a vague form. Typically, spatial regions may be identified where sources are expected to radiate from – or not – and a certain degree of spatial correlation may be specified on the expected source field. Such elements of information constitute the "spatial prior".

Of specific concern in this paper is the case where a majority of the source coefficients are nil (or nearly so) whereas a few coefficients are allowed to take large values. This is referred to as a "sparse" source distribution. Although this situation will be properly addressed in section 4 by introducing specific non-Gaussian prior PDFs $[\mathbf{s}_i]$, a necessary condition to meet from the onset is to assume *a priori* that the source distribution is incoherent on the surface Γ_s. This will return the finest spatial resolution among all possible spatial priors. It is important to emphasize at this juncture that the assumption of *prior* incoherence does not imply that the reconstructed sources are *a posteriori* incoherent as well: correlation among the estimated source coefficients is truly allowed as seen immediately from the structure of Eq. (9). Indeed, most methods of the literature (e.g. such as NAH or ESM) happen to assume prior incoherence in an implicit manner when interpreted from a Bayesian perspective.

Assuming time stationarity and a finite-power (i.e. with integrable squared-magnitude) spatial spectrum, the spatial correlation of an incoherent source field has the functional form

$$\mathbb{E}\left\{s_i(\mathbf{r}_q)s_i(\mathbf{r}_{q'})^*\right\} = \alpha^2\sigma_0^2(\mathbf{r}_q)\delta_{qq'} \tag{11}$$

where $\delta_{qq'}$ stands for the Kronecker delta and $\sigma_0^2(\mathbf{r}_q)$ reflects the intensity of the field at position $\mathbf{r}_q$. The continuous function $\sigma_0^2(\mathbf{r})$, $\mathbf{r} \in \Gamma_s$, arbitrarily normalized to unit volume, $\int_{\Gamma_s}\sigma_0^2(\mathbf{r})\mathrm{d}\Gamma(\mathbf{r}) = 1$, is called the "aperture function" (AF) [34] for reasons to become clear soon and α^2 the prior "source power". This allows for possible non-homogeneity of the source field with prior local variance $\mathbb{E}\{|s_i(\mathbf{r})|^2\} = \alpha^2\sigma_0^2(\mathbf{r})$. Therefore, spatial regions with high likelihood of

radiation are assigned a relatively high value of $\sigma_0^2(\mathbf{r})$ whereas regions less likely to radiate are given a value of $\sigma_0^2(\mathbf{r})$ close to zero. Based on this model, one has the $Q \times Q$ diagonal prior covariance matrix

$$\Sigma_0^2 = \mathrm{Diag}\left(\sigma_{0,q}^2\right), \tag{12}$$

where $\sigma_{0,q}^2 \doteq \sigma_0^2(\mathbf{r}_q)$. It is proved in Appendix A that this parameterization of the problem actually minimizes the mean-square error of the estimated source. In a discrete setting, this implies the following probabilistic model for the source coefficients,

$$s_{q,i} = \alpha \cdot \sigma_{0,q} \cdot \epsilon_{q,i}, \quad \epsilon_{q,i} \sim \mathcal{N}_{\mathbb{C}}(0,1), \tag{13}$$

where $\epsilon_{i,q}$ denotes a standardized (i.e. zero-mean and unit variance) complex Gaussian random variable. The latter encodes the fluctuations of the source coefficient in time – hence its dependence on the snapshot index i – whereas the (squared) product $\alpha^2\sigma_{0,q}^2$ returns the source power at spatial position $\mathbf{r}_q$ independently of i. Since $\sigma_{0,q}^2$ is normalized, it is convenient to endow the global variable α with the physical dimension of the sources $s_{q,i}$.

Eventually, more insight is provided by the functional notation of Eq. (9) as allowed by the introduction of the AF $\sigma_0^2(\mathbf{r})$. It is seen that the source at position $\mathbf{r}$ is then estimated as

$$\hat{s}_i(\mathbf{r}) = \sigma_0^2(\mathbf{r}) \sum_{m=1}^{M} G(\mathbf{r}|\mathbf{r}_m)^* q_{m,i} \tag{14}$$

where the coefficients $q_{m,i}$ are the elements of vector

$$\mathbf{q}_i = (\mathbb{G} + \eta^2\mathbf{I})^{-1}\mathbf{p}_i \quad \text{with} \quad [\mathbb{G}]_{kl} = \int_{\Gamma_s} G(\mathbf{r}_k|\mathbf{r})G(\mathbf{r}_l|\mathbf{r})^*\sigma_0^2(\mathbf{r})\mathrm{d}\Gamma(\mathbf{r}). \tag{15}$$

Formula (14) explicitly shows that the number of "degrees of freedom" of the estimated source distribution is equal to the number of microphones. Its physical interpretation is that $\hat{s}_i(\mathbf{r})$ is estimated by constructively interfering on surface Γ_s the waves radiated by the M microphones as if they were point sources with "strength" $q_{m,i}$ – from the principle of reciprocity, the conjugated Green function $G(\mathbf{r}|\mathbf{r}_m)^*$ appearing in Eq. (14) means that the direction of propagation is reversed, from the microphone position $\mathbf{r}_m$ to the source coordinate $\mathbf{r}$. At last, the estimated source is focused on the region of interest by multiplication of the sum in Eq. (14) with the AF $\sigma_0^2(\mathbf{r})$.

3. Focusing mechanism and its link to wideband acoustical holography

The AF $\sigma_0^2(\mathbf{r})$ plays a fundamental role in BF. By shrinking the reconstruction of the source distribution on a restrained spatial region where sound is a priori supposed to radiate from, it can considerably improve the estimation performance in terms of spatial resolution, quantification of sound levels, and frequency range. The improvement in spatial resolution is an obvious consequence of using a narrow AF. If S_σ denotes the effective area of the surface covered by $\sigma_0^2(\mathbf{r})$, then the spatial resolution of $\hat{s}(\mathbf{r})$ is on the order of S_σ/M.

The improvement in the quantification of sound levels is perhaps more unexpected. It may be explained as follows: since a narrow AF alleviates the effort of localizing the sources, more "degrees of freedom" are left available to improve the accuracy of the source amplitudes. An extreme example is when a small number – say N_s – of point sources are to be estimated with a priori known positions; the AF is then shrunk into a few Dirac deltas at the source positions so that the inverse problem boils down to estimating N_s amplitudes from M observations. As long as $N_s \leq M$ the problem is potentially over-determined and can be solved more accurately than if the full spatial distribution coded on Q spatial samples had to be inferred.

Eventually, the use of a narrow AF can also enlarge the frequency range of the reconstructed source distribution. This is particularly compelling since traditional approaches in inverse acoustics are known to be fundamentally limited downward and upward in frequency. The reason may be better explained trough an example.

Example 1. Let us consider the case of a monopole source on a linear domain $\Gamma_s = \{r : -1 \leq r \leq 1\}$ parallel to a

uniform linear array (ULA) of $M = 11$ microphones located at a distance $R = 0.35$ m – see Fig. 1. A "high" frequency $f = 6$ kHz is purposely selected that makes the resolution of the inverse problem difficult in this configuration. The signal-to-noise ratio (SNR) is set to 20 dB. The situation without AF is first considered, which is implicitly equivalent to using an uniform AF with unit amplitude over Γ_s. The reconstructed source $\hat{s}(r)$ (whose real and imaginary parts are shown by the black and grey curves, respectively, in Fig. 2(a)) exhibits marked fluctuations over the full domain Γ_s and significantly differs from the theoretical monopole. As a result, the pressure field radiated by $\hat{s}(r)$ (real part shown in Fig. 2(a)) comprises a lot of fictitious interferences. The squared modulus of the radiated pressure field (shown in Fig. 2(d)) – which represents one part of the acoustic energy – evidences a complex radiation of the energy from Γ_s towards the array and some hot spots are seen at the positions of the microphones. This can be understood as an artifact of the classical LS solution with Tikhonov regularization when no AF is used. In this case, the regularization term tries to globally minimize the energy of the source while still matching the measurements at the microphones. Therefore, a source distribution is found with a directivity pattern that primarily points towards the microphone positions. As a result, the average acoustic power produced by the reconstructed source is also largely underestimated: in the present case, it is found equal to only 14.4% of the true value.

The situation is next considered where a narrow AF is inserted in the form of a Hann function (see Fig. 2(c)) in order to encourage the reconstruction of the source in a small neighborhood of $r = 0$. The reconstructed source (real and imaginary parts shown by the black and grey curves, respectively, in Fig. 2(b)) now resembles much more the theoretical monopole and it radiates a pressure field (real part shown in Fig. 2(b)) that correctly reproduces outgoing spherical waves. The squared modulus of the radiated pressure field (shown in Fig. 2(e)) also correctly follows the spherical spreading law of a monopole, without artificial directivity towards the microphones. The average acoustic power is now found equal to 97% of the true value.

The benefit offered by the insertion of a narrow AF can be understood from a different perspective by introducing the concept of a representation basis. From the interpretation following Eq. (14), the source estimate $\hat{s}(\mathbf{r})$ is constructed by interfering waves radiated from the microphones and focused through the AF on the source surface Γ_s. Therefore, these waves provide a spatial basis, $\{\phi_k(\mathbf{r}) = \sigma_0^2(\mathbf{r})G(\mathbf{r}|\mathbf{r}_k)^*, \mathbf{r} \in \Gamma_s\}_{k=1}^M$, for representing $\hat{s}(\mathbf{r})$ (note that the basis is not necessarily complete). Figure 3(a) illustrates how one basis function is produced by the radiation of one microphone. The full family of basis functions relating to Example 1 is illustrated in Fig. 3(b-c). In the absence of an AF – or equivalently when the latter is set uniform – the basis functions reduce to the Green functions, $\{G(\mathbf{r}|\mathbf{r}_k)^*, \mathbf{r} \in \Gamma_s\}_{k=1}^M$, which oscillate in Γ_s with wavelength $\lambda \sim c_0/f$. Because the array is parallel to Γ_s, the basis is composed of the same waveform shifted at different locations. The corresponding spatial spectra are displayed in Fig. 4(a), which all have the same magnitude and only differ by a linear phase. It is seen that the spectra have a marked peak around the spatial frequency f/c_0, which explains why the source distribution in Fig. 2(a) has been reconstructed with so many ripples. On the contrary, the insertion of a narrow AF produces a much richer representation basis, quite close to "Gabor wavelets" as seen in Fig. 3(c). The corresponding spatial spectra (see 4(b)) are close to a filterbank decomposition, where each basis function spans a different frequency band. In addition, the spectral range has been enlarged, which explains why a sharper source can be reconstructed. This is a manifestation of the uncertainty principle stating that the product of the widths of the AF,say Δr, and of its spatial spectrum, say Δk, is upper bounded, i.e. $\Delta r \Delta k \leq C$ for some constant C. There is indeed no limit to the highest spatial frequency that can be reconstructed this way since it is inversely proportional to the AF width which can be tuned arbitrarily narrow – i.e. $\Delta k \sim C/\Delta r$. In particular, a monopole can be reconstructed perfectly when the AF ultimately tends to a Dirac delta. This phenomena – although not explained from this point of view – has been coined "wideband holography" in Ref. [23].

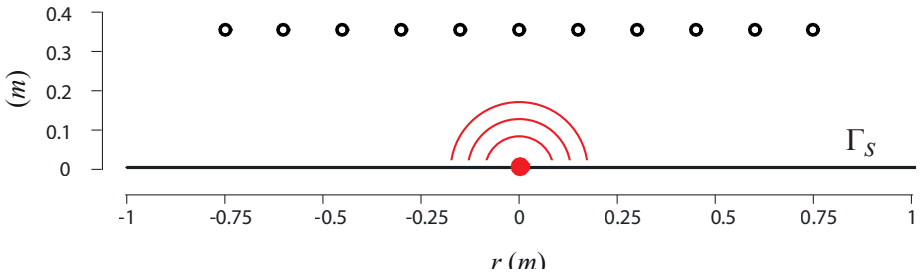

Figure 1: Scheme of a simple inverse problem comprising an ULA of microphones (black circles) and a monopole source (red dot) on a linear source domain Γ_s (black segment).

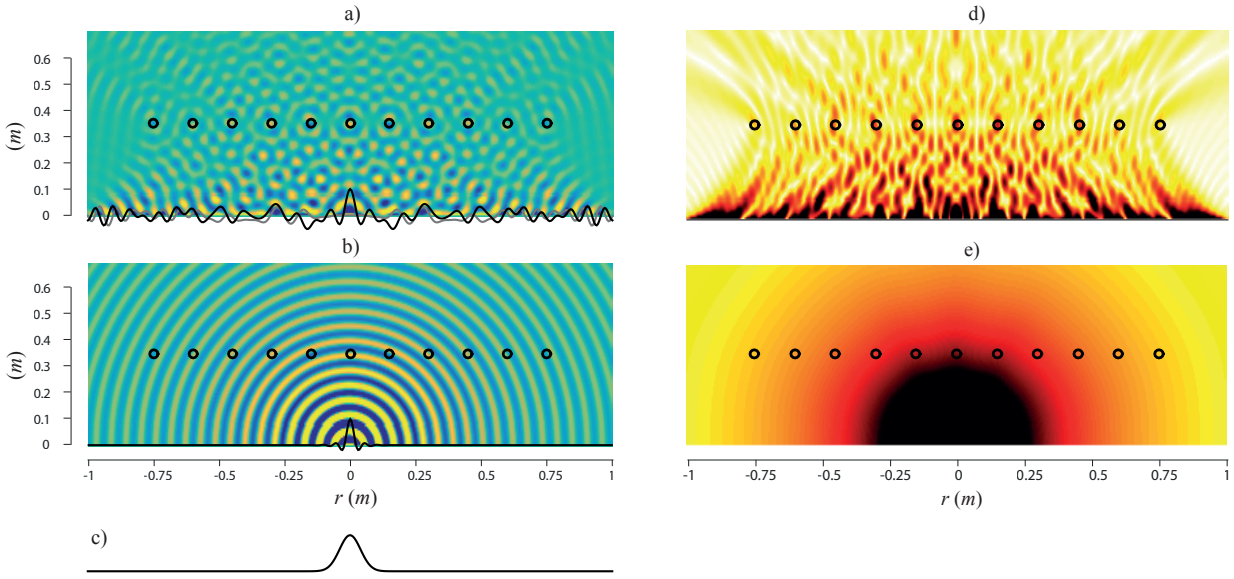

Figure 2: (a-b) Reconstructed source distribution $\hat{s}(r)$ on Γ_s (the real and imaginary parts are shown by the black and grey curves, respectively) together with the corresponding real parts of the radiated pressure field ($f = 6$ kHz). The black circles represent microphones in the array. Configurations (a) and (b) are based on an uniform aperture function and the narrow aperture function displayed in (c), respectively. (d-e) Squared modulus of the pressure fields corresponding to (a) and (b).

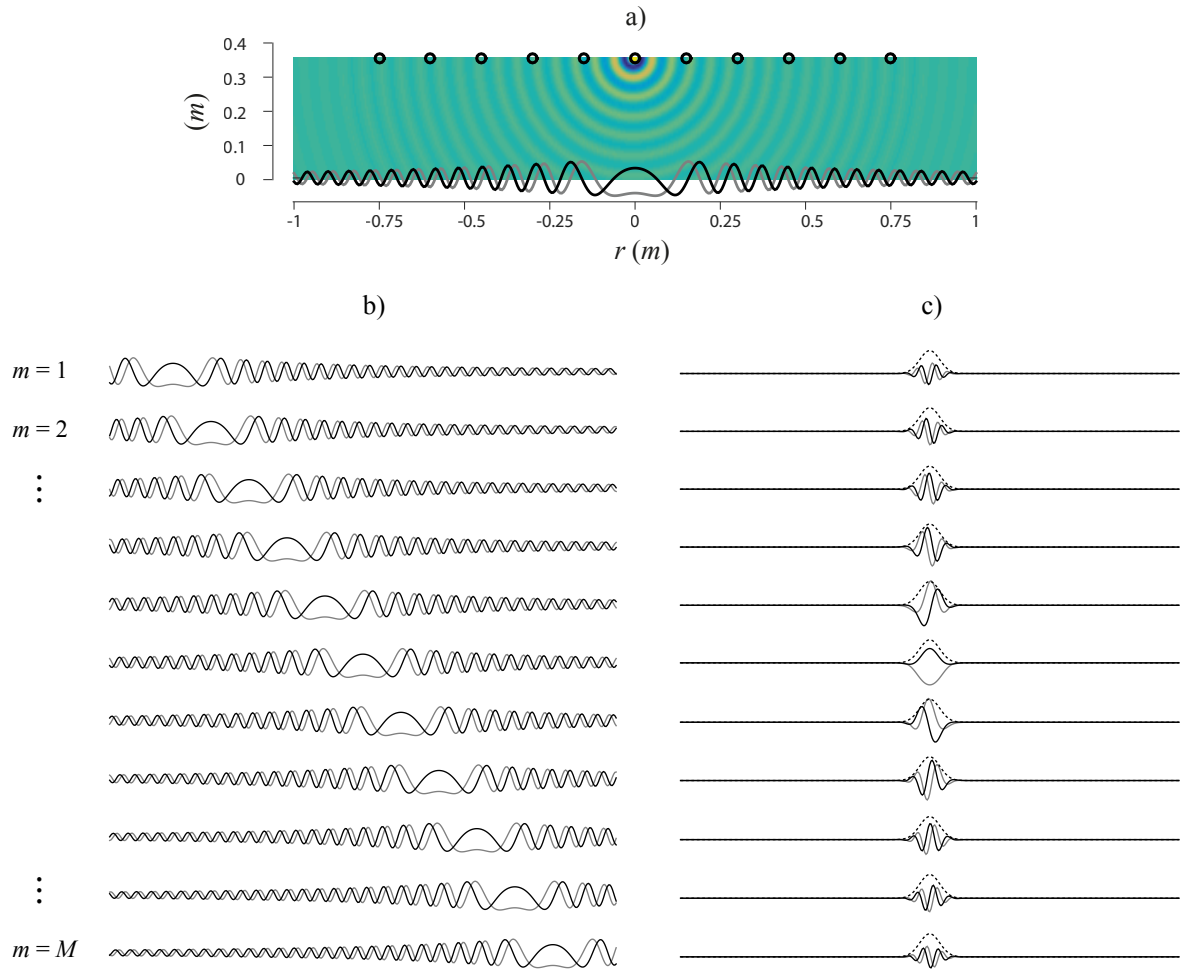

Figure 3: a) b) Real part of the pressure field radiated by one microphone in the array (here the central one) as if it was a source (f = 6 kHz). The black (resp. grey) curve is the real (resp. imaginary) part of the pressure radiated on Γ_s as a function of r. b) Real and imaginary parts (black and grey curves, respectively) of the basis functions (as obtained by the mechanism illustrated in (a)) rooted on an uniform aperture function and c) the narrow aperture function of Fig. 2(c). The aperture function used in (c) is displayed by the black dotted curves.

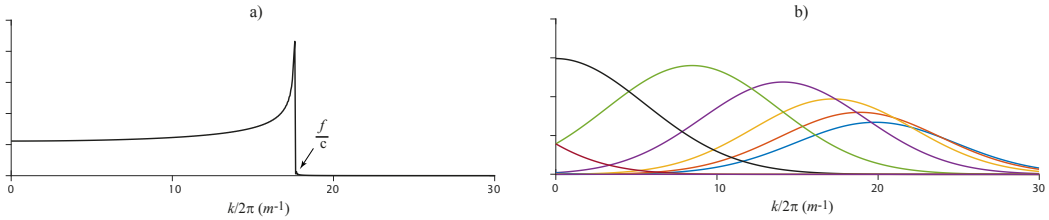

Figure 4: Amplitudes of spatial spectra of the M basis functions of a) Fig. 3(a) and b) Fig. 3(b). Note that all amplitude spectra are superimposed in (a).

4. Sparse Bayesian Focusing

BF, as reviewed in the previous section, involves a manually-tuned AF. If properly designed, the AF can greatly improve the quality of the reconstructed source distribution in terms of localization, quantification and frequency range. The question then arises as whether it is possible to automatically tune the AF, by considering it as another unknown quantity in the inverse problem. An intuitive solution consists in first solving the inverse problem with a large AF and next in using the so-estimated source distribution as a new AF to solve the inverse problem a second time. By iterating the process several times, the initial AF progressively tunes itself so as to focus on the different point sources that constitute the source distribution, as depicted in Fig. 5. This section justifies the strategy from a theoretical point of view in the Bayesian setting. It also demonstrates that it applies with various adaptation rules depending on the choice of the prior PDF of the source coefficients. The principle is to consider the AF as a random quantity that is inferred in an iterative resolution of the inverse problem. How to model the AF as random quantity is first addressed in subsection 4.1. By assigning the AF with a PDF, various source priors can thus be generated, different from the Gaussian one of Eq. (8), which promote sparse sources; they are explored in subsections 4.2. Subsection 4.3 then introduces an iterative algorithm – coined IBF – which jointly infers the source coefficients and updates the AF.

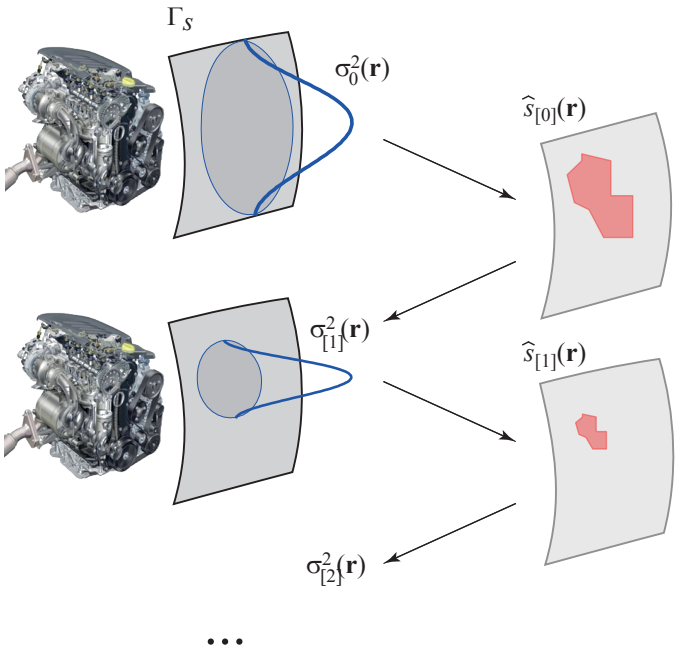

Figure 5: Principle of Iterated Bayesian Focusing: the reconstructed source $\hat{s}_{[k]}(\mathbf{r})$ at the current iteration k is used to refine the aperture function $\sigma^2_{[k+1]}(\mathbf{r})$ used in the next iteration.

4.1. Sparse spatial prior

In the context of this paper, the sources coefficients are assumed to be sparsely distributed in space; this means that only a very elements in vector $\mathbf{s}_i$ (always the same independently of snapshot i) are expected to have significant

values, the other ones being close (or ideally equal) to zero. One way to enforce this property is to assign to $\mathbf{s}_i$ a prior PDF with heavy tails and very peaked at zero as compared to the Gaussian, which is denoted hereafter as leptokurtic (i.e. with a kurtosis greater than the Gaussian). The use of leptokurtic prior PDFs (sometimes referred to as sparse priors, sparsity enforcing priors or shrinkage prior) within the Bayesian framework for enforcing sparse solutions has been extensively studied in the recent statistical literature and is also known as sparse Bayesian learning[1] (see e.g. [38, 39]).

A physical interpretation is given hereafter in terms of SMoGs. Following the same lines as in subsection 2.3, let the source coefficients have the following probabilistic model,

$$s_{q,i} = \alpha \cdot \sigma_{0,q} \cdot \tau_q \cdot \epsilon_{q,i}, \quad \epsilon_{q,i} \sim \mathcal{N}_{\mathbb{C}}(0,1), \tag{16}$$

where the source power α^2, the AF $\sigma^2_{0,q}$ and the random variables $\epsilon_{q,i}$ have been defined in Eq. (13). Equation (16) separates the source coefficients into time-dependent (index i) and space-dependent (index q) variables. All variables are dimensionless, except α which carries the dimension of $s_{q,i}$. All variables are also seen as random, except $\sigma^2_{0,q}$ which is "deterministically" designed by the user to reflect a global (or "initial") AF (e.g. see dotted bell-shaped curves in Fig. 7); it is introduced for the sake of generality, yet it could just as well be ignored to simplify the exposition. As compared to Eq. (13), τ_q is now an additional non-negative random variable, which reflects the "relative intensity" at the spatial position $\mathbf{r}_q$ (see Fig. 6). It may also be interpreted as the variance of $\tau_q \cdot \epsilon_{q,i} \sim \mathcal{N}_{\mathbb{C}}(0, \tau_q^2)$, which is the mechanism by which "local" aperture functions can be produced at position $\mathbf{r}_q$. It is noted at this juncture that α cannot be absorbed into τ_q because the latter will be shortly seen to act as a "sparsifying" variable, with high degree of sparsity achieved for (very) small values of its scale parameter; this means that the scale parameter of τ_q must be left free and the global scaling assigned to α. The stochastic model (16) is actually equivalent to considering right out a random AF

$$\sigma^2(\mathbf{r}) = \sigma_0^2(\mathbf{r})\tau^2(\mathbf{r}), \tag{17}$$

(given here in functional notation with $\tau_q \doteq \tau(\mathbf{r}_q)$) instead of a deterministic one as done in the previous section. The random AF in Eq. (17) yields the conditional prior variance $\mathbb{E}\{|s_i(\mathbf{r})|^2|\tau^2(\mathbf{r})\} = \alpha^2 \cdot \sigma_0^2(\mathbf{r}) \cdot \tau^2(\mathbf{r})$ at position $\mathbf{r}$ on Γ_s, independently of the snapshot index i, which locally and constantly encourages high values of the source coefficients where the relative intensity $\tau^2(\mathbf{r})$ is large and tends to switch them off where the relative intensity is small. As compared to the uniform relative intensity used in BF (see Eq. (11)), the resulting spatial distribution of the source coefficients is thus sparser.

It is noteworthy that the proposed probabilistic model allows enforcing sparsity directly on the modulus of the source coefficients. This is advantageous as compared to the situation imposed by some solvers for ℓ_1 minimization, implemented for real-value data and whose extension to the complex case is not trivial. Sparsity on the modulus – which is equivalent to group sparsity on the real and imaginary parts – easily allows switching off a large number of source coefficients. On the contrary, enforcing separate sparsity on the real and imaginary parts – as would be done by mapping a complex number to a two-dimensional real vector – can only switch off a source coefficient when both the real and imaginary happen to vanish simultaneously, which occurs with a much lower probability. In addition, separate sparsity would produce several instances of exactly nil real (resp. imaginary) parts together with non-zero imaginary (resp. real) parts, which reflects unrealistic physical combinations.

Although it is difficult to specify the values of τ_q^2 before the inverse problem is solved, it is possible to assign them a hyperprior PDF, $\left[\tau_q^2\right]$, that allows local variations in space. Given a spatial position $\mathbf{r}_q$ and considering all snapshots $i = 1, ..., N$ observed at this position, the prior PDF of the source coefficients $\mathbf{S} \doteq \{s_{q,i}\}_{i=1}^N$ then reads

$$[\mathbf{S}] = \int_0^\infty \left[\mathbf{S}|\tau_q^2\right]\left[\tau_q^2\right] \mathrm{d}\tau_q^2 = \mathop{\mathbb{E}}_{\tau_q^2}\left\{\prod_{i=1}^N \mathcal{N}_{\mathbb{C}}\left(s_{q,i}; 0, \alpha^2\sigma^2_{0,q}\tau_q^2\right)\right\} \tag{18}$$

where $\mathbb{E}_{\tau_q^2}$ means that the expected value is taken with respect to the random variable τ_q^2. Inspection of Eq. (18) shows that the prior PDF of the source coefficients is a continuous mixture of Gaussians $\mathcal{N}_{\mathbb{C}}\left(s_{q,i}; 0, \alpha^2\sigma^2_{0,q}\tau_q^2\right)$ weighted by the

[1] Note that sparse Bayesian learning should be undertood as a general terminology that embodies the family of Bayesian methods dedicated to enforcing sparsity, yet it is not an algorithm by itself.

"scale" $\left[\tau_q^2\right]$ – the so-called SMoG. This is in general no longer a Gaussian, but a leptokurtic PDF. Figure 6 illustrates how a leptokurtic PDF is obtained by mixing Gaussians with random variances distributed according to a given law. It is seen that PDFs with very heavy tails can thus be obtained by allowing the variance to be broadly distributed.

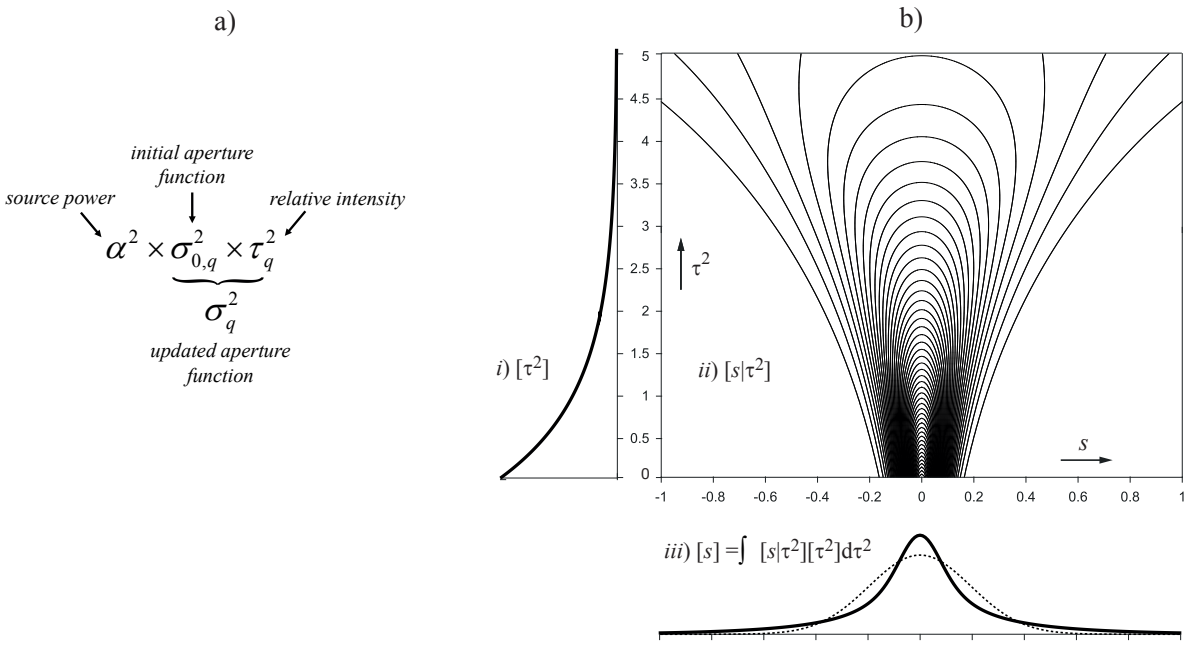

Figure 6: a) Structure of the variance of the sparse spatial prior, where index q relates to spatial position $\mathbf{r}_q$. α^2 is the global source power and the initial aperture function $\sigma_{0,q}^2$ reflects its expected spatial distribution before the experiment is run. The spatial variance $\alpha^2\sigma_{0,q}^2$ is used in the Gaussian prior PDF of BF; its multiplication by the relative intensity τ_q^2 returns the spatial variance $\alpha^2\sigma_{0,q}^2\tau_q^2$ used in the SMoG prior PDF of Iterated Bayesian Focusing. The method updates the aperture function as $\sigma_q^2 \leftarrow \sigma_{0,q}^2\tau_q^2$. b) Principle of construction of an SMoG. *i*) PDF $[\tau^2]$ of the variance τ^2, *ii*) isoprobabiliy contours of the conditional Gaussian PDF, $[s|\tau^2] = \mathcal{N}_{\mathbb{C}}\left(s; 0, \tau^2\right)$, as a function of the random variable s and of the variance parameter τ^2, *iii*) PDF $[s] = \int [s|\tau^2][\tau^2]\mathrm{d}\tau^2$ of the random variable s averaged over all values of the variance τ^2 (continuous curve) together with the Gaussian PDF with same variance (dotted curve). It is seen that a leptokurtic PDF is produced, with heavy tails and a peaked mode.

Examples of prior source fields simulated with the SMoGs investigated in the next subsection are displayed in Fig. 6. In order to allow an easy representation, the source domain is limited to a line segment. The initial AF $\sigma_{0,q}^2$ has been designed in the form of a Hann function that progressively downsizes the values at the extremities of the domain where sound is less likely to radiate from. In each case, the value of α^2 is tuned so as to impose a unit variance (spatially averaged over the source domain). The source coefficients are displayed as a function of the space position for a given snapshot i. It is seen that the real and imaginary parts are jointly sparse, which means that they both have the same variance proportional to the updated AF $\sigma_{0,q}^2\tau_q^2$. In other words, sparsity is enforced on the modulus – and not on the real and imaginary parts, independently – in order to force a majority of source coefficients to take small values and, ideally, to switch off completely.

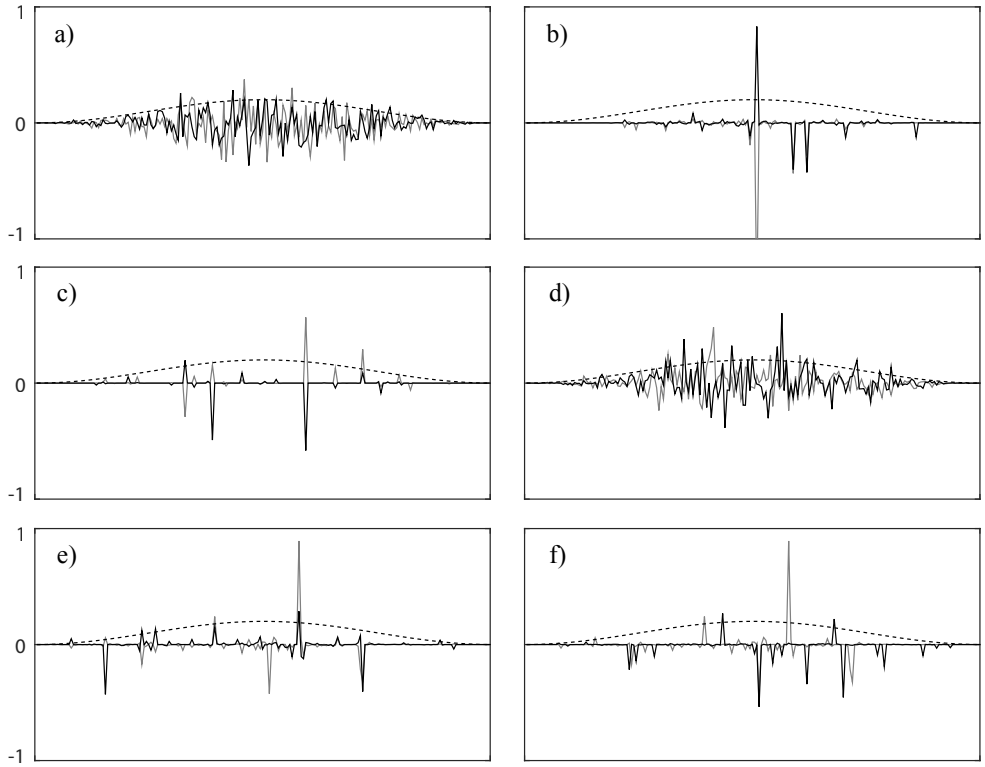

Figure 7: Examples of one-dimensional cuts through complex random fields synthesized with different SMoG PDFs (the real and imaginary parts are displayed by black and gray curves, respectively). a) Complex Gaussian, b) complex Student-t (rooted on the inverse Gamma PDF $[\tau^2]$ with $a = b = 1$), c) complex K-distribution (rooted on the Gamma PDF $[\tau^2]$ with $a = b = 0.05$), d) complex Laplacian (generalized complex Gaussian with $p = 1$), e) generalized complex Gaussian with $p = 0.1$, f) mixture of complex Gaussians with $\bar{\pi} = 0.1$, $\bar{\tau}_1^2 = 0.1$ and $\bar{\tau}_2^2 = 10$. Also shown is the initial AF $\sigma_{0,q}^2$ (dotted curve).

4.2. Prior PDFs in the SMoG family

Different prior PDFs are obtained depending on the choice of the hyperprior $\left[\tau_q^2\right]$ in Eq. (18). Although closed-form results exist in the literature in the case of real-valued random variables, they must be extended here to the scenario with complex-valued source coefficients and with several snapshots – i.e. to the complex multivariate case. Explicit formula are derived in Appendix C for some particular cases of interest and resumed in Tab. 1. As far as the authors know, some of these formula are given here for the first time.

It is noteworthy that the SMoG model allows expressing the prior PDF $[\mathbf{S}]$ by means of the power spectrum of the source coefficients only. Specifically, the quantity

$$\chi_q^2 = N\frac{S_{qq}}{\alpha^2\sigma_{0,q}^2}, \tag{19}$$

which involves the ratio of the source power spectrum S_{qq} (see Eq. (2)) to the prior spatial variance $\alpha^2\sigma_{0,q}^2$ at position q, is a sufficient statistic. It will be used from now on as a shortcut notation.

Prior source distributions of Tab. 1 are illustrated in Fig. 7 and also displayed in Fig. 11 (a-b) for $N = 1$ and $N = 50$ snapshots. It is verified that all PDFs are apt to produce very sparse fields, even though the Laplacian (corresponding to the ℓ_1 penalty) seems slightly inferior than the other candidates.

Table 1: Prior PDFs constructed in the SMoG family

Hyperprior PDF $\left[\tau_q^2\right]$	Hyperparameters	Sparse prior PDF $[\mathbf{S}] = \int \left[\mathbf{S}\middle\vert\tau_q^2\right]\left[\tau_q^2\right]\mathrm{d}\tau_q^2$
Inverse Gamma	shape parameter $a > 0$ scale parameter $b > 0$	Multivariate complex Student-t (MCS)
Gamma	shape parameter $a > 0$ scale parameter $b > 0$	Multivariate complex K-distribution (MCK) particular case: Multivariate complex Laplace (MCL) when $a = N + \frac{1}{2}$ and $b = \frac{1}{4}$
(Not required)	power $0 < p \leq 2$	Generalized multivariate complex Gaussian (GMCG) particular cases: Multivariate complex Gaussian (MCG) when $p = 2$ Multivariate complex Laplace (MCL) when $p = 1$
Discrete: $\bar{\tau}_1^2$ with probability $\bar{\pi}$ $\bar{\tau}_2^2$ with probability $1 - \bar{\pi}$	$\bar{\tau}_1^2, \bar{\tau}_2^2, 0 < \bar{\pi} < 1$ assume $\bar{\tau}_1^2 \ll \bar{\tau}_2^2$	Mixture of multivariate complex Gaussians (MMCG)

4.3. Inference in the SMoG family

The price to pay for the introduction of a non-Gaussian prior PDF is that, generally, a closed-form solution no longer exists for estimating the source coefficients. However, it is shown hereafter that a point estimate can still be expressed with the same structure as in Eq. (9), yet where Σ_0^2 is replaced by a different AF. This is a direct consequence of the fact that an SMoG is a weighted average of Gaussian priors as expressed in Eq. (18). To see this, let us introduce the notation

$$\hat{\mathbf{s}}_i(\Sigma^2) = \Sigma^2\mathbf{G}^H(\mathbf{G}\Sigma^2\mathbf{G}^H + \eta^2\mathbf{I})^{-1}\mathbf{p}_i \tag{20}$$

for the classical penalized LS solution given in Eq. (9) rooted on a given AF, $\Sigma^2 = \mathrm{Diag}(\sigma_q^2)$ (note that, as a particular case, Eq. (9) is exactly recovered with $\Sigma^2 = \Sigma_0^2$). Then, it is established in Appendix D that the posterior mean under an SMoG prior is

$$\mathbb{E}\{\mathbf{s}_i|\mathbf{P}\} = \int \hat{\mathbf{s}}_i(\Sigma^2)[\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2 \tag{21}$$

where $[\Sigma^2|\mathbf{P}]$ is the posterior PDF of the AF given the measurements $\mathbf{P}$. Equation (21) is recognized as a weighted average of $\hat{\mathbf{s}}_i(\Sigma^2)$ over realizations of the AF conditioned on the data. More importance is therefore given to solutions

whose posterior PDF $[\Sigma^2|\mathbf{P}]$ is high. In the extreme case where the latter PDF is very peaked, the posterior mean may be evaluated directly from Eq. (20) wherein the AF is replaced by its MAP estimate $\Sigma^2_{MAP} = \mathrm{Diag}(\sigma^2_{0,q} \cdot \mathrm{Arg\,max}[\tau_q^2|\mathbf{P}])$.

Alternatively, it is proved in Appendix D that the MAP estimate under an SMoG prior – which will be considered from now on – is $\hat{\mathbf{s}}_i(\Sigma^2)$ with AF

$$\Sigma^2 = \mathrm{Diag}\left(\frac{\sigma^2_{0,q}}{\mathbb{E}\left\{\tau_q^{-2}|\hat{\mathbf{S}}\right\}} \right) \tag{22}$$

where $\hat{\mathbf{S}} \doteq \{\hat{\mathbf{s}}_i\}_{i=1}^N$. This means that the MAP is obtained by updating the AF $\sigma^2_{0,q}$ by dividing it with the expected value of τ_q^{-2} conditioned on $\hat{\mathbf{S}}$. This is expected to provide a "sharper" AF than the initial one. It appears in light of these results that the MAP estimate will generally differ from the posterior mean in the SMoG family – contrary to the Gaussian case of section 2 – which reflects the fact that $[\mathbf{s}_i|\mathbf{P}]$ is not a symetric PDF.

Although inserting Eq. (22) into (20) returns an implicit equation with no closed-form solution, it can be solved iteratively as shown in the next subsection.

At this juncture, the inverse problem has been formally reformulated as inference in a Bayesian hierarchical model with a structured prior, as depicted in Fig. 8. In short, the essential difference with BF is that the AF is now an unknown quantity to be inferred simultaneously with the source coefficients.

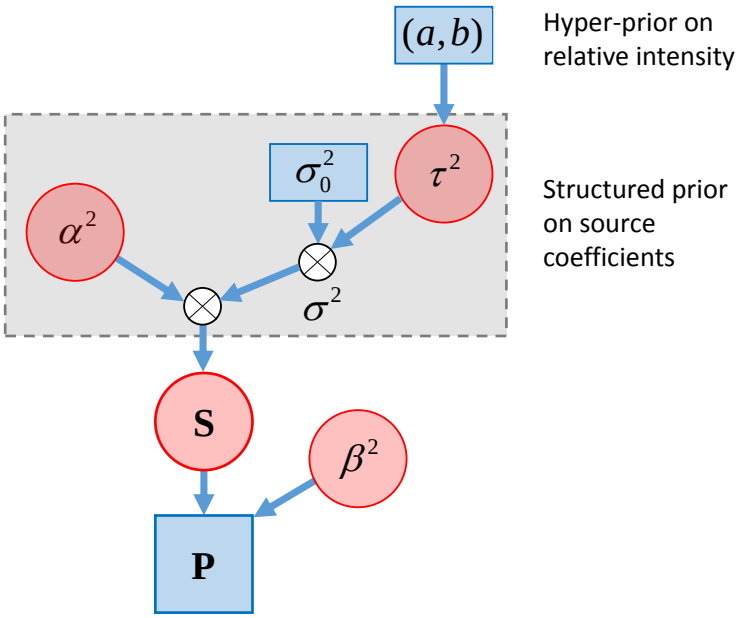

Figure 8: Directed graph of the hierarchical model for inferring **S**. Unknown parameters to infer are shown in circles and known quantities in rectangles. Note that **P** and **S** can be replaced by the CSM $\mathbb{S}_{pp}$ and the source power spectrum S_{qq}.

4.4. Iterated Bayesian Focusing (IBF)

An algorithm is proposed hereafter to reach the MAP estimated given in Eq. (22). The IBF algorithm is described hereafter in terms of the normalized sum of squares

$$\chi^2_{q,[k-1]} = N \frac{S_{qq,[k-1]}}{\alpha^2_{[k-1]}\sigma^2_{0,q}} \tag{23}$$

which involves the power spectrum $S_{qq,[k-1]} = \frac{1}{N}\sum_{i=1}^N |\hat{s}_{q,i,[k-1]}|^2$ of the source coefficients and the source power $\alpha^2_{[k-1]}$ estimated at iteration $k-1$. The reason for introducing this variable is manifold: beyond greatly simplifying

forthcoming notations, it is also a sufficient statistics which plays a central role in the IBF algorithm. As seen from its definition, it also has the physical significance of the "normalized" power of the sources.

Algorithm 1

- **Step 0**: Design an initial AF $\sigma^2_{0,q}$ and construct the diagonal matrix Σ^2_0 whose q-th diagonal entry is $\sigma^2_{0,q}$.
- **Step 1**: Set $k = 0$. Estimate the source power $\alpha^2_{[0]}$ and the regularization parameter $\eta^2_{[0]}$ by using Eqs. (B.3) and (Appendix B) evaluated with Σ^2_0. Initialize the algorithm with the MAP solution $\hat{\mathbf{s}}_{i,[0]}$ of Eq. (9) as returned from a Gaussian prior.

Repeat steps 2 to 5 until convergence.

- **Step 2**: Do $k \longleftarrow k+1$.
  Estimate the relative intensity as

$$\hat{\tau}^2_{q,[k]} = -\left(\frac{\partial}{\partial(\chi^2_q)} \ln\left([\chi^2_q]\right)\right)^{-1}\Bigg|_{\chi^2_q=\chi^2_{q,[k-1]}}, \tag{24}$$

  where $[\chi^2_{q,[k-1]}]$ is the prior PDF expressed as a function of the normalized sum of squares $\chi^2_{q,[k-1]}$ at iteration $k-1$ (closed-form expressions of $\hat{\tau}^2_{q,[k]}$ are given by Eqs. (G.1), (G.2), (G.4), (G.5), (G.6), (G.8), and (G.9)).

- **Step 2(bis)**: Set $\hat{\tau}^2_{q,[k]} = \max\left\{\hat{\tau}^2_{q,[k]}; \epsilon_1 \cdot \max_q\left\{\hat{\tau}^2_{q,[k]}\right\}\right\}$ where $0 < \epsilon_1 \ll 1$.

- **Step 3**: Update the AF as

$$\sigma^2_{q,[k]} = \sigma^2_{q,0}\hat{\tau}^2_{q,[k]} \tag{25}$$

  and construct the diagonal matrix $\Sigma^2_{[k]}$ whose q-th diagonal entry is $\sigma^2_{q,[k]}$.

- **Step 4**: Estimate the source power and the regularization parameter as

$$\left(\alpha^2_{[k]}, \eta^2_{[k]}\right) = \text{Argmax}\left[\alpha^2, \eta^2 | \mathbb{S}_{pp}, \Sigma^2_{[k]}\right] \tag{26}$$

  by using Eqs. (B.3) and (Appendix B) evaluated with the updated AF.

- **Step 5**: Estimate the source coefficients as

$$\hat{\mathbf{s}}_{i,[k]} = \Sigma^2_{[k]}\mathbf{G}^H(\mathbf{G}\Sigma^2_{[k]}\mathbf{G}^H + \eta^2_{[k]}\mathbf{I})^{-1}\mathbf{p}_i. \tag{27}$$

Convergence criterion: Stop iterations when $||\hat{\mathbf{s}}_{i,[k]} - \hat{\mathbf{s}}_{i,[k-1]}||/||\hat{\mathbf{s}}_{i,[k-1]}||$ is smaller than a given threshold, $0 < \epsilon_2 < 1$.

(Step 2(bis) is optional with the MCS and MMCG priors, yet it is recommended with the MCK, MCL and GMCG priors as explained later in subsection 5.4.)

IBF belongs to the family Iterative Re-weighted Least Squares (IRLS) algorithms [40], since at each iteration it solves a weighted least square problem whose solution is given by Eq. (9) with Σ^2_0 replaced by it updated version $\Sigma^2_{[k]}$ as obtained from Eq. (25). Yet, it is more rigorously derived in Appendix E as an Expected-Maximization (EM)

algorithm, which makes possible to prove its convergence to the MAP solution – insightful connections between IRLS and EM have been investigated in Ref. [41]. The proposed IBF algorithm has the following features:

1. It applies generally in the SMoG family, as long a closed-form expression of the prior PDF is available. Consequently, different structures and degrees of sparsity can be promoted by choosing among the various possible prior PDFs of subsection 4.2.
2. The regularization parameter η^2 is iteratively updated together with the source coefficients with proved convergence (see Appendix B). This is to be contrasted with other approaches found in the literature where the regularization parameter is either arbitrarily imposed or tuned afterwards, typically by using cross-validation or the L-curve. The latter approaches require running the optimization as many times as values of the regularization parameter are to be tested, and are therefore computationally more demanding – if not prohibitive in some applications such as that illustrated in subsection 6. In addition, the estimation of the regularization parameter by means of the Bayesian criterion used in Step 4 has been found more efficient than the cross-validation or the L-curve in many inverse acoustic scenarios [35].
3. The algorithm applies to complex-valued source coefficients and sparsity is enforced to their modulus (and not to the real an imaginary coefficients independently). Consequently, a majority of sources are forced to switch off.
4. A key step in the algorithm is the update of the AF in Eq. (25) rooted on the current estimate of the relative intensity given by Eq. (24). Closed-form expressions of the updating rules corresponding to the prior PDFs of Tab. 1 are given in Appendix G.
5. The updating rule involves the consideration of possibly several snapshots and therefore involves the power spectra of the source coefficients – see Eq. (23). Accordingly, the IBF algorithm can be formulated in terms of the CSM of the measurements, $\mathbb{S}_{pp}$, only (see Eq. (1)). As explained in the introduction, this may be determinant in practice, since some commercial data acquisition system used in acoustic imaging only record the CSM by default. The algorithm is as follows.

Algorithm 2
Step 1 to 4: Unchanged.
Step 5: Estimate the power spectrum of the source coefficients at position $\mathbf{r}_q$ as

$$S_{qq,[k-1]} = \sigma^4_{q,[k]} \mathbf{g}_q^H (\mathbf{G}\boldsymbol{\Sigma}^2_{[k]}\mathbf{G}^H + \eta^2_{[k]}\mathbf{I})^{-1} \mathbb{S}_{pp} (\mathbf{G}\boldsymbol{\Sigma}^2_{[k]}\mathbf{G}^H + \eta^2_{[k]}\mathbf{I})^{-1} \mathbf{g}_q \tag{28}$$

where $\mathbf{g}_q$ is the q-th column of matrix $\mathbf{G}$.

Example 2. The IBF algorithm is illustrated on Example 1 of section 3 which involved the reconstruction of a monopole source at $f = 6$ kHz with an ULA placed at a distance $R = 0.35$ m (see Fig. 2) under an SNR of 20 dB. The MCS prior is used with hyperparameters $a = b = 0.01$ (comparison between the several priors introduced in subsection 4.2 will be addressed in the next section). To start with, only $N = 1$ snapshot is recorded. The IBF algorithm is initialized with the BF solution of section 2 with an AF $\sigma^2_{q,0}$ shaped like a Hann window (raised cosine) over the domain $\Gamma_s = \{r : -1 \leq r \leq 1\}$ (with discretizaton step of 0.01 m). The algorithm is stopped after 10 iterations. Figure 9(a-b) displays the evolution of the AF $\sigma^2_{q,[k]}$ on linear and dB scales with respect to the iteration number k. It is seen that the initial AF quickly tuned itself into a narrow peak around the actual source position shown by a red dot. At the same time, the dynamic range of the AF improves from 5 dB (for $\sigma^2_{q,0}$) to more than 70 dB (for $\sigma^2_{q,[9]}$). As explained in section 3, the shrinkage of the AF introduces a focusing mechanism that improves the accuracy of the estimated acoustic power. Whereas it was only 17.6% of the true value at the first iteration (note a slight improvement as compared to 14.4% for the uniform window used in section 3), it is 99.9% at the ninth iteration.

Figure 10 displays the spectra of the estimated acoustic power (relatively to the true value) for several working frequencies in the range $[0.02; 50]$ kHz. The monopole is driven by a white Gaussian signal and results are averaged over 100 independent draws of complex Gaussian noise added to the measurements with an SNR of 20 dB. As expected, the BF solution has an upper frequency limit on the order of $f_{max} \sim c_0/d \simeq 2.3$ kHz with $d = 0.15$ m the spacing between the microphones. On the other hand, the IBF solution does not show such a limit. In the present case

with a monopole source, it is virtually unlimited in frequency since the AF is allowed to become arbitrarily narrow. The IBF solution is remarkably accurate over the full frequency range as compared to the BF one. It is also found more precise with a 3-sigma confidence interval about half as large as that of the BF solution. In addition, the coefficient of variation (CV) is constant (CV = 0.44%) as compared to that of the BF solution which increases when approaching the upper frequency limit f_{max}.

The same experiment is now run with two incoherent monopole sources driven by white Gaussian signals in order the check the ability of the IBF algorithm to self-tune the AF around several distinct source points. The first monopole is located at $r = 0$ m and the second one at $r = -0.05$ m. They are both fed with mutually independent complex Gaussian random noises. The level of the second monopole is set to 4% the level of the first one. In this case, $N = 100$ snapshots are recorded in order to better estimate the partial power due to each monopole (i.e. to average out incoherent cross-terms). The IBF is run with the same settings as before. Figure 9(c-d) displays the evolution of the AF on linear and dB scales; it is seen to quickly split in two narrow peaks anchored to the two monopole positions. As before, the dynamic range after 10 iterations is greater than 70 dB. The two peaks have different magnitudes – which actually makes it difficult to see the smallest one on the linear scale although its presence is obvious on the dB scale – yet their respective areas converges to exactly the same ratio as the levels of the two monopoles, i.e. 0.04. This might be an issue only for detecting very small sources masked by a dominant source (here below 70 dB of the largest one), a situation which places the performance of IBF much higher than CV or BF in this respect. Despite the two peaks of the updated AF having different magnitudes, the estimated source coefficients are perfectly calibrated to recover the correct strengths of the two monopoles. After convergence, the acoustic power radiated by the first monopole is found equal to 100.1% of theoretical power and that of the second one to 100.7%. This is to be compared to 8.6% and 304.6%, respectively, for the initial estimation at $k = 0$, which on the one hand still underestimates the largest source but on the other hand largely overestimates the smallest one due to the presence of sidelobes from the dominant source.

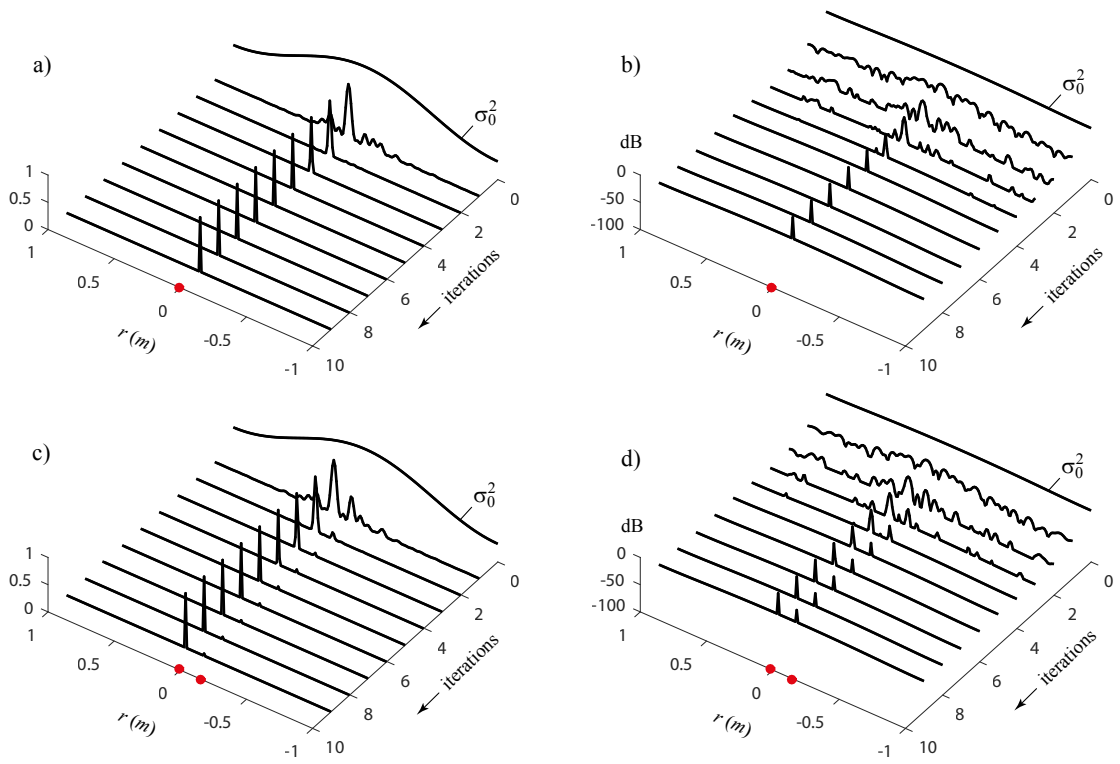

Figure 9: Update of the aperture function $\sigma^2_{q,[k]}$ with respect to iteration numbers $k = 0, ..., 9$ for the MCS prior. The zeroth iteration corresponds to the initial aperture function $\sigma^2_{q,0}$. Configurations with a-b) one monopole at $r = 0$ m and c-d) two incoherent monopoles at $r = 0$ m and $r = -0.05$ m (shown by red dots) with relative levels in the ratio 0.04. The two columns relate to linear and dB scales, respectively.

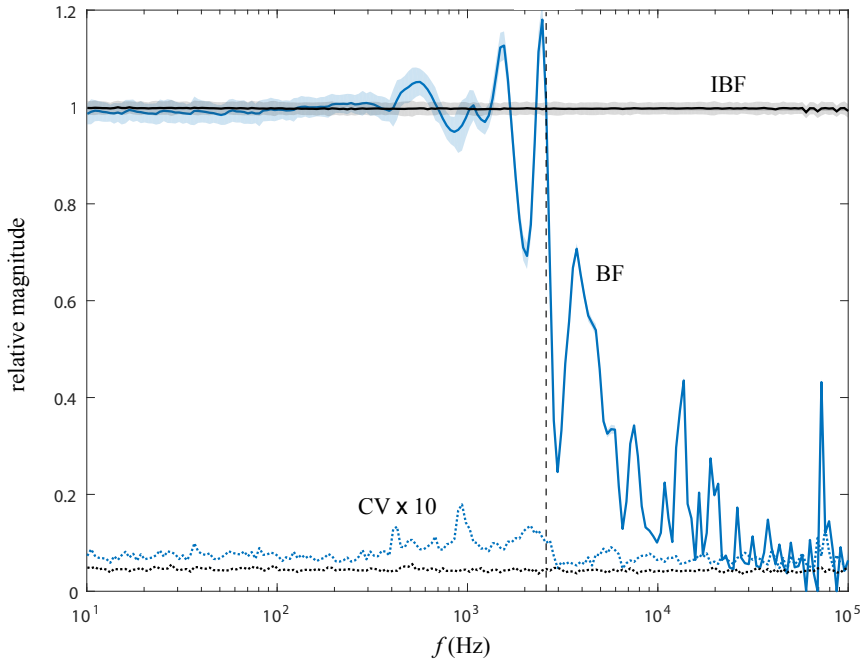

Figure 10: Spectra of the acoustic power (relatively to the true value) radiated by a monopole source estimated with Bayesian Focusing (BF) (blue curve) based on the large aperture function σ_0^2 shown in Fig. 9 and with Iterated Bayesian Focusing (IBF) (black curve). The shaded areas indicate 3-sigma confidence intervals and the dotted curves the coefficients of variation of BF (blue) and IBF (black) amplified by a factor 10. The vertical dotted line corresponds to the cutting frequency $f_{max} \sim c_0/d$. Results are averaged over 100 draws of the independent complex Gaussian noise with 20 dB SNR. (CV = coefficient of variation.)

5. Comparison of priors

The use of an SMoG model has been shown to yield several possible candidates for the prior PDF of the source coefficients. The question naturally arises as which one(s) should be selected in practice. Although the exact answer to this question is surely problem dependent, some considerations are possible based on the following arguments.

5.1. Number of hyperparameters

Unless the user has a very accurate knowledge of the prior PDF, it is often convenient to adopt a choice that involves the fewest hyperparameters. In this respect, The MCL arrives in first position without any hyperparameter to set. The GMCG arrives in second position with only one hyperparameter, p, which directly reflects the amount of expected sparsity. The MCS and the MCK come next with two hyperparameters, a and b. However, as seen in Eq. (G.1), the effect of hyperparameter a in the MCS tends to vanish when the number of snapshots becomes large. Therefore, a convenient choice is to set $a = 0$ and keep b as the only hyperparameter to tune; another possibility is to set $a = b$ to a small value that promotes sparsity. No similar simplification seems to hold for the MCK, for which the presence of two hyperparameters is the price to make it very flexible. The MMCG comes last with three hyperparameters, $\bar{\tau}_1$, $\bar{\tau}_2$ and $\bar{\pi}$. It has been checked in simulations that their ratio is really determinant, so that one can set $\bar{\tau}_2 = 1$ for instance from the onset. These findings are resumed in Tab. 2.

Table 2: Hyperparameters in the SMoG family

Sparse prior PDF	Number of relevant hyperparameters	Setting
Multivariate complex Laplace (MCL)	0	
Generalized multivariate complex Gaussian (GMCG)	1	$0 < p \leq 2$
Multivariate complex Student-t (MCS)	2 reduced to 1	$a = 0$ and b small or $a = b$ small
Multivariate complex K-distribution (MCK)	2	a and b
Mixture of multivariate complex Gaussians (MMCG)	3 reduced to 2	$\bar{\pi}$ and $\bar{\tau}_1^2/\bar{\tau}_2^2$

5.2. Sensitivity to hyperparameters

The hyperparameters of the SMoG model arrive in the last level of the Bayesian hierarchical model (see Fig. 8) and are therefore, in theory, much less influential than hyperparameters in lower levels such as typically the regularization parameter η^2 to which the inverse problem is known to be extremely sensitive (see e.g. Ref. [42] for an excellent discussion on the influence of hyperparameters in high-level priors). As explained in the previous subsection, the general recommendation is to tailor the prior PDF to make it sparse. Extensive experiments have shown that, on the one hand, the MCS prior is very stable when a and b are set small as indicated in Tab. 2. On the other hand, the MCK, the MMCG and the GMCG are quite sensitive to the setting of their hyperparameters. This is indeed the price to pay for the PDF to gain flexibility in its shape. The GMCG has yet an advantage in that it is controlled by only one parameter, the power p, with perhaps a clearer meaning than for the MCS and MCK hyperparameters. In this respect, the GMCG prior is a relevant choice if the prior PDF has to be carefully tuned in a simple way, as will be evidenced in the experiments of section 7. Automatic tuning of the power p remains challenging; although theoretically possible in the Bayesian setting [43], this issue is outside the scope of the present paper.

5.3. Bias

The choice of a particular SMoG has a direct impact on the estimation error of the source coefficients. This may be investigated by considering independently the bias and the variance of the MAP estimate. Based on the results of subsection 4.3, the bias can be expressed as (see Appendix I)

$$\mathbf{s}_i - \mathbb{E}\left\{\hat{\mathbf{s}}_i\right\} = \left(\mathbf{I} + \eta^{-2}\mathbf{G}^H\mathbf{G} \cdot \mathrm{Diag}\left(\sigma_{0,q}^2 \mathbb{E}\left\{\tau_q^{-2} | \hat{\mathbf{S}}\right\}^{-1}\right)\right)^{-1} \mathbf{s}_i. \tag{29}$$

Therefore, it is seen that a reduced bias is generally achieved by SMoG models which can return high values of the relative intensity $\hat{\tau}_q^2$. As seen in Fig. 11(c-d), this puts the MCS, the MCK with $N > a$, and the GMCG with $p \gtrsim 0$ (symbol $\gtrsim$ means close to but greater than), in first position. On the opposite, the MMCG is expected to return the strongest bias as its relative intensity in Eq. (G.9) is bounded upward by $\bar{\tau}_2^2$.

Incidentally, expression (29) for the bias also indicates that matrix $\mathbf{G}$ should be designed such that $\mathbf{G}^H\mathbf{G}$ is as close as possible to a diagonal matrix in order for the bias at position $\mathbf{r}_q$ not to be influenced by the non-zero values of the source coefficients at other positions $\mathbf{r}_{q'} \neq \mathbf{r}_q$. This condition is actually reminiscent of the "restricted isometry property" [6] and "mutual coherence criterion" [8, 24] of compressed sensing.

Eventually, it is seen from Eq. (29) that the bias is inflated by the regularization parameter η^2, which is the price to pay for stabilizing an ill-posed inverse problem. Given a non-zero AF, the bias completely vanishes only in the particular case when η^2 is nil.

5.4. Variance and rate of convergence

The exact expression of the covariance matrix of the estimated source coefficients is not easy to derive. However, it can be shown within the framework of the EM algorithm that it directly depends on the so-called "missing information matrix" $\mathcal{I}_m$ [44], which measures the discrepancy of information due to the introduction of the *unknown* relative intensity τ_q^2 in the EM algorithm as compared to the case where it would be known for all indices q: the larger the norm of $\mathcal{I}_m$, the larger that of the covariance matrix. Besides, the speed of convergence of the EM algorithm is similarly controlled by the norm of $\mathcal{I}_m$: the larger the "missing information" and the slower the convergence.

It is proved in Appendix J that the missing information matrix averaged over N snapshots takes the following expression

$$\mathcal{I}_m = \frac{1}{\alpha^2 N}\mathrm{Diag}\left(\frac{\chi_q^2}{\sigma_{0,q}^2}\cdot \mathrm{Var}\left\{\tau_q^{-2}|\hat{\mathbf{S}}\right\}\right) \quad \text{where} \quad \mathrm{Var}\left\{\tau_q^{-2}|\hat{\mathbf{S}}\right\} = \frac{\partial^2}{(\partial\chi_q^2)^2}\ln\left[\chi_q^2\right]. \tag{30}$$

This result shows that the measure of missing information essentially depends on the conditional variance of the inverse of the relative intensity, $\mathrm{Var}\{\tau_q^{-2}|\hat{\mathbf{S}}\}$, which is the price paid for sparsity. This unveils a connection between the missing information matrix and sparsity: missing information increases when more sparsity is introduced; intuitively, this corresponds to requiring more information to tune the unknown random variables τ_q^2 when moving from a smooth LS solution (no sparsity) to a highly structured (highly sparse) solution. This equivalently reflects a balance between high sparsity and reduced estimation variance. Equation (30) also evidences that the variance decreases proportionally with the number of snapshots; therefore taking N large should be encouraged in general. The conditional variance of the PDFs introduced in subsection 4.2 are displayed in Fig. 11(e-f) for $N = 1$ and $N = 50$ snapshots.

An important requirement for the convergence of the IBF is to verify that the diagonal elements of $\mathcal{I}_m$ remain bounded for all $0 \leq \chi_q^2 < \infty$. It is established in Appendix J that this is the case for the MCS and the MMCG, yet not for the GMCG with $p < 2$ neither for the MCK in general. This has a practical implication for the IBF algorithm based on the GMCG and MCK priors since the AF should be controlled in some way in order to keep it stable when χ_q^2 takes very small values. One such strategy is to systematically set to zero the relative intensities $\hat{\tau}_{q,[k]}^2$ that converge below a small threshold and to remove them from the subsequent iterations. Another simple strategy is to replace them by the value of the threshold. This is the rationale behind Step 2(bis) in Algorithm 1.

In conclusion, the considerations of subsections 5.1, 5.3 and 5.4 place the MCS in first position since it 1) allows very sparse solutions while 2) requiring the setting of only one significant hyperparameter and 3) having a reduced bias and 4) an unconditionally bounded variance.

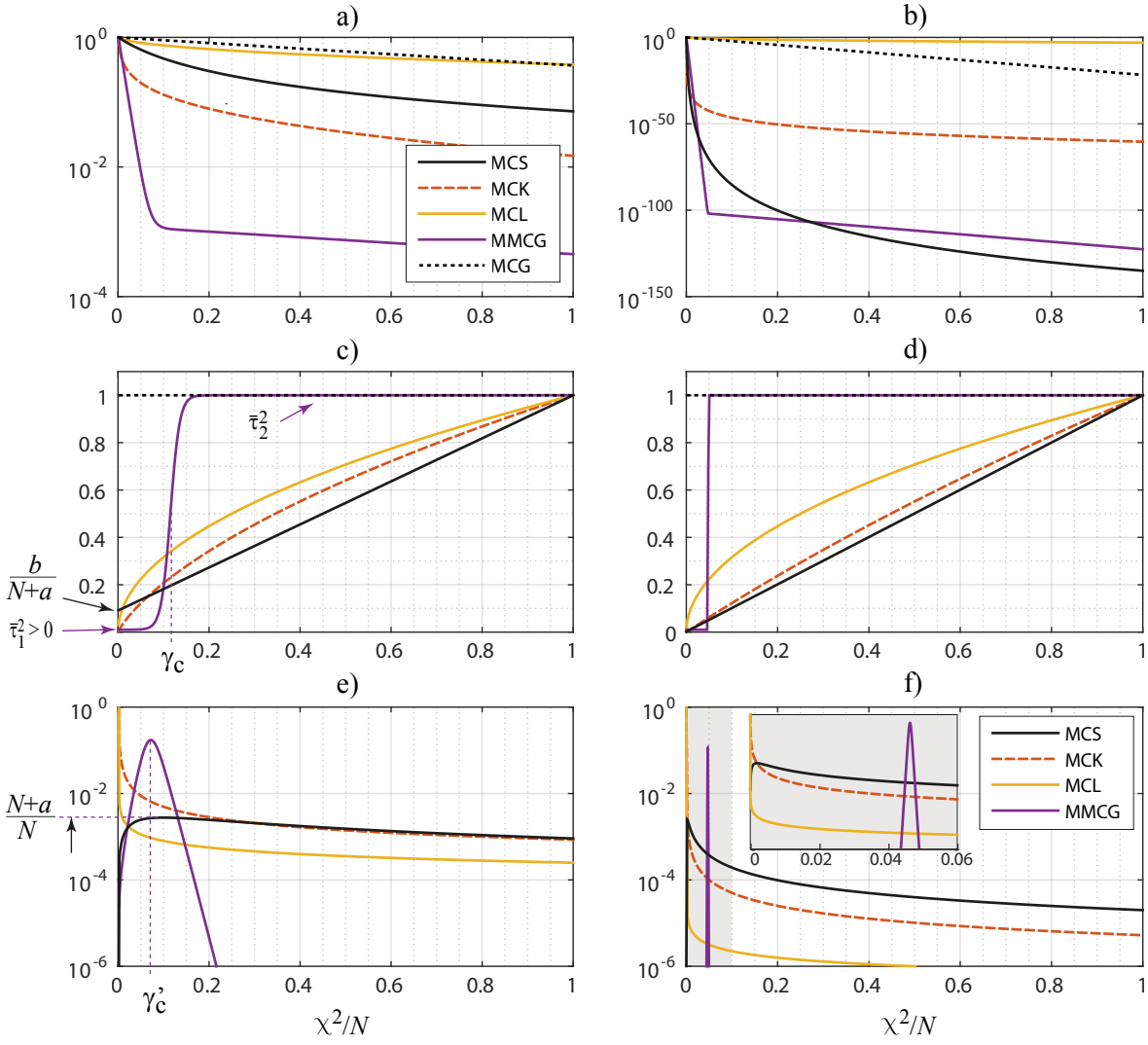

Figure 11: a) Prior PDF $[\chi_q^2]$ for $N = 1$ (in logarithmic scale) and b) for $N = 50$ in the MCS case with $(a = 0.1, b = 0.2)$ (black line), the MCK case with $(a = 3N/4, b = 1)$ (red dotted line), the MCL case (yellow line) and the MMCG case with $(\bar{\pi} = 0.9, \bar{\tau}_1^2 = 0.0001, \bar{\tau}_2^2 = 1)$ (purple line) compared with the multivariate complex Gaussian (MCG) case (dotted black line). Sparse behaviors are essentially recognized by the distance to the Gaussian at the origin ($\chi_q^2 = 0$) and towards infinity ($\chi_q^2 \to \infty$). c) Corresponding relative intensity $\bar{\tau}_q^2 = \mathbb{E}\{\tau_q^{-2}|\hat{\mathbf{S}}\}^{-1} = -(\partial \ln([\chi_q^2]/\partial(\chi_q^2))^{-1}$ for $N = 1$ and d) for $N = 50$, all normalized to 1 at $\chi_q^2/N = 1$. e) Corresponding variance $\mathrm{Var}\{\tau_q^{-2}|\hat{\mathbf{S}}\} = \partial^2 \ln[\chi_q^2]/(\partial\chi_q^2)^2$ for $N = 1$ and f) for $N = 50$ with excerpt in the range $0 \le \chi_q^2/N \le 0.06$.

Example 3. Experimental comparison of the above priors is carried on in a configuration similar to Examples 1 and 2. The range distance is $R = 0.5$ m and an ULA with 20 microphones and spacing 0.1 m is used. The SNR is set to 20 dB. The source domain $\Gamma_s = \{r : -1 \leq r \leq 1\}$ (in meters) is discretized with a spatial step of 0.01 m. Initialization of the IBF algorithm is done as in Example 2 with a Hann window for the AF. A threshold set to $\epsilon_2 = 10^{-3}$ times the highest amplitude is imposed with the GMCG and MCK priors in order to avoid possible divergence of the EM algorithm (see discussion of subsection 5.4). The threshold used in the stopping criterion of the iteration is $\epsilon_1 = 10^{-3}$. The MCS prior is used with $a = b = 0.01$, the MCK prior with $a = b = 0.1$ and the MMCG prior with $\bar{\pi} = 0.99$ and $\bar{\tau}_2^2/\bar{\tau}_1^2$ either set to 10 or to 100. These parameters have been kept constant for all experiments in order to fairly test the robustness of the method in this respect. With these settings, the EM algorithm has been found to converge within a few iterations (about 10 on the average), with the fastest speed attained with the MMCG prior and the slowest one with the MCL prior.

The first experiment investigates the reconstruction of two incoherent monopoles with equal strengths, located at $r_1 = -0.05$ m and $r_2 = 0.1$ m, at frequency $f = 1$ kHz and with $N = 1$ snapshot. Reconstruction results are superimposed for 4 independent draws of the snapshot. Figure 12(a) displays the results of conventional beamforming and BF, which both suffer from a poor spatial resolution in this case. Apart from an offset due to a different scaling (the amplitude of beamforming returns the source strength whereas BF returns a source density), the two curves are quite similar. Figure 12(b) displays the result of the MCL prior (corresponding to the ℓ_1-norm penalty), which improves the spatial resolution but still has a limited dynamic range. Numerous sidelobes are however present, probably due to the use of a uniform array, and a marked dispersion is observed with respect to snapshot draws. The estimation using the MCK prior is found very similar to the MCL prior in this case (Figure 12(e)). The GMCG (ℓ_0-norm penalty) and MCS priors in Figure 12(c-d) show the sparsest results, by refining even more the spatial resolution and greatly increasing the dynamic range. At the same time, they are found rather stable with respect to snapshot draws. The MMCG reconstruction shown in Fig. 12(f) shows intermediate sparsity, which have been checked to strongly depend on the value of the ratio $\bar{\tau}_2^2/\bar{\tau}_1^2$. In some cases, only one of the two sources is identified.

The second experiment repeats the first one, yet using the CSM computed with $N = 100$ snapshots. The main finding is that results have been considerably stabilized in accordance with the conclusions of subsection 5.4 (they are now almost invariant with respect to draws of the CSM). While the use of the CSM has clearly upgraded the MCK estimation by placing it to close to the GMCG ($p = 0$) and MCS results, it has reduced the spatial resolution of MMCG. This is explained by the fact that the updating rule (G.9) tends to a hard threshold when N increases, which basically clips the source coefficients estimated in the first iteration of the IBF algorithm, whose magnitudes are smaller than γ_c as given by Eq. (G.10).

The third experiment tackles a more difficult configuration with a very low frequency $f = 100$ Hz and with the first monopole strength 20% smaller than the second one (i.e. 14 dB difference in power level). Only results obtained with the CMS computed on 100 snapshots are displayed (see Fig. 14). It is seen that both beamforming and BF fail to detect the smallest source due to a poor spatial resolution. Although the MCL can correctly identify the two sources, even better results are achieved by sparser priors such as the GMCG ($p = 0$) and the MCS (note a small bias on the localization of the small source). Both the MCK and the MMCG priors failed to reconstruct the small source with the given parameter settings. The spatial resolution of MMCG is only slightly better than that of beamforming.

The fourth experiment investigates another extreme case with a very high frequency $f = 10$ kHz and the same two unequal source strengths (see Fig. 15). Beamforming and BF both fail in this case. The MCL prior clearly improves the situation, yet GMCG ($p = 0$) and MCS still return the best estimates with the deepest dynamic range achieved by the latter. Although sparse, the MMCG fail to correctly identify the two sources.

Similar conclusions have been obtained in several other experimental configurations not reported in the paper. In general, it turned out that sparser results than allowed by the ℓ_1-norm penalty could be obtained by using the other priors investigated in the paper. However, only the GMCG (ℓ_0-norm penalty) and the MCS priors showed a constant stability of the estimates. On the contrary, the MMCG prior was found extremely dependent on the setting of its parameters $\bar{\tau}_1$ and $\bar{\tau}_2$. Overall, results have been found remarkably independent of whether the point sources are incoherent or partially coherent (with the above settings, perfectly coherent sources of equal strengths could be resolved, while sources with 14 dB difference in power could be correctly identified up to 80% coherence).

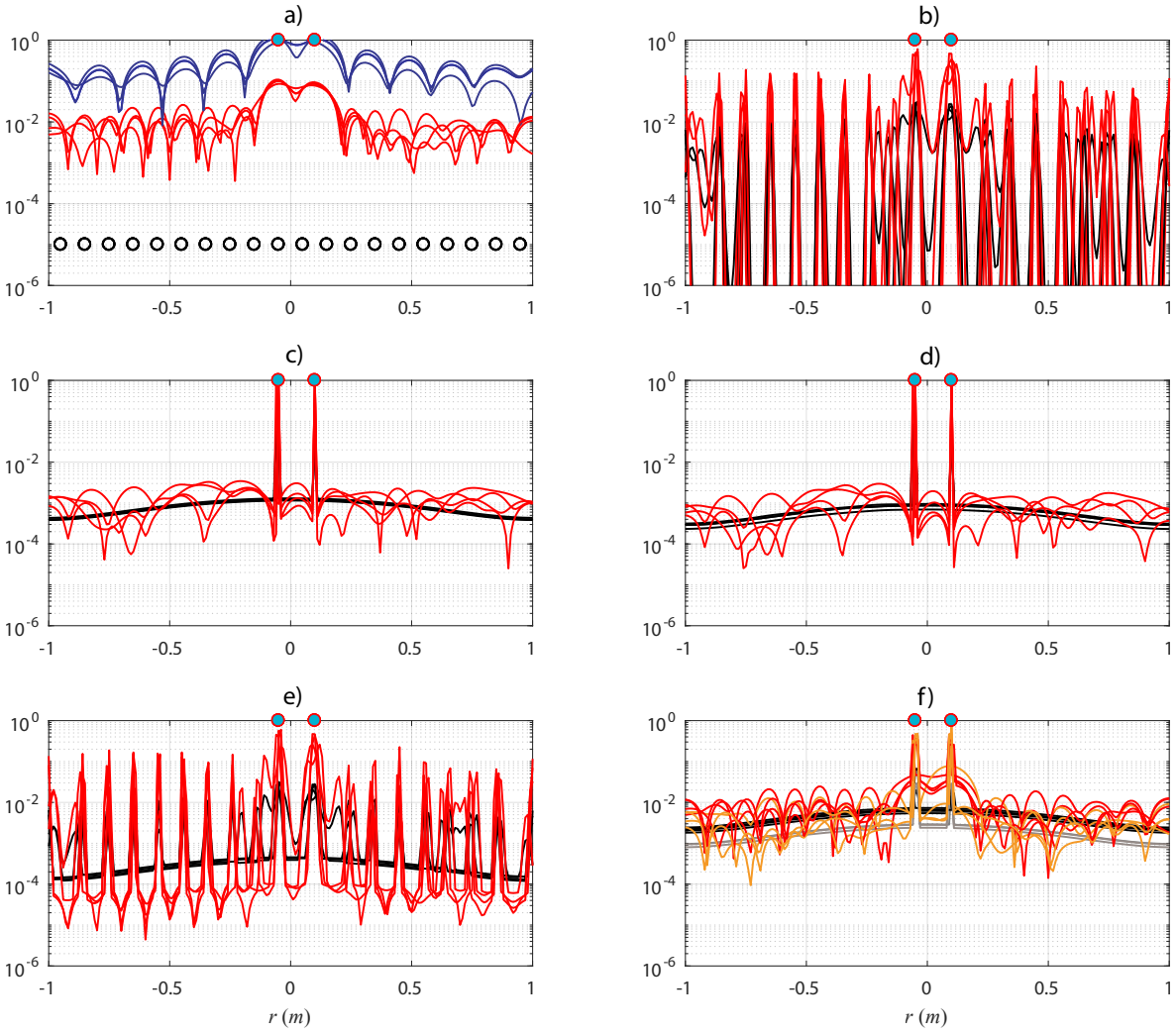

Figure 12: Estimated source power spectrum S_{qq} at $f = 1kHz$ with $N = 1$ snapshot in the case of two incoherent monopoles with equal strengths (indicated by blue circles). Results from four independent snapshot realizations are superimposed. The positions of the microphones is marked by black circles in (a). a) Results returned by conventional beamforming (blue) and Bayesian Focusing (red). Results returned by Iterated Bayesian Focusing with b) the MCL prior (GMCG with $p = 1$), c) the GMCG prior with $p = 0$, d) the MCS prior ($a = b = 0.01$), e) the MCK prior ($a = b = 0.1$) and f) the MMCG prior with $\bar{\pi} = 0.99$, $\bar{\tau}_2^2 = 1$, $\bar{\tau}_1^2 = 0.01$ (red) and $\bar{\tau}_1^2 = 0.001$ (orange). The red/orange curves in (b-f) relate to the source power spectrum S_{qq} and the black/gray curves to the updated aperture function σ_q^2.

6. Broadband Iterated Bayesian Focusing (BIBF)

So far, a different AF has been assumed at each frequency f. Nevertheless, it is meaningful to consider instances where the AF is independent of frequency, as would typically occur when the sound source is attached to a specific device (e.g. an obstruction, an opening, etc.). This situation has been considered for example in acoustic imaging in Refs. [45, 46] and is reminiscent of the concept of "multi-frequency" group sparsity, where the same sparse structure is enforced to the source distribution for a group of frequencies.

Forcing a constant AF in a frequency band is straightforward within the proposed IBF framework. It is also found to yield improved results. The algorithm proceeds as follows (see Appendix K).

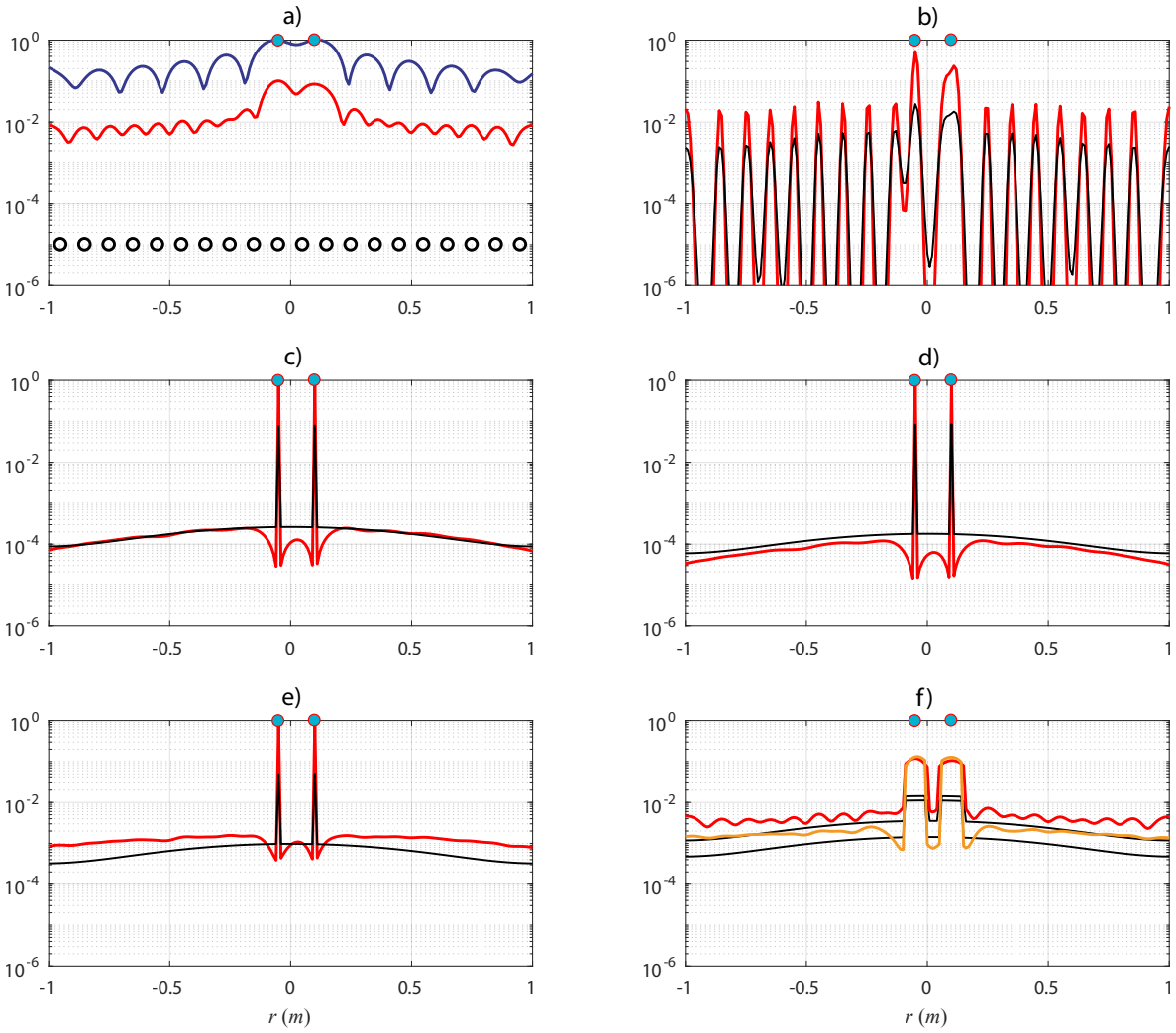

Figure 13: Same configuration as in Fig. 12 with $N = 100$ snapshots.

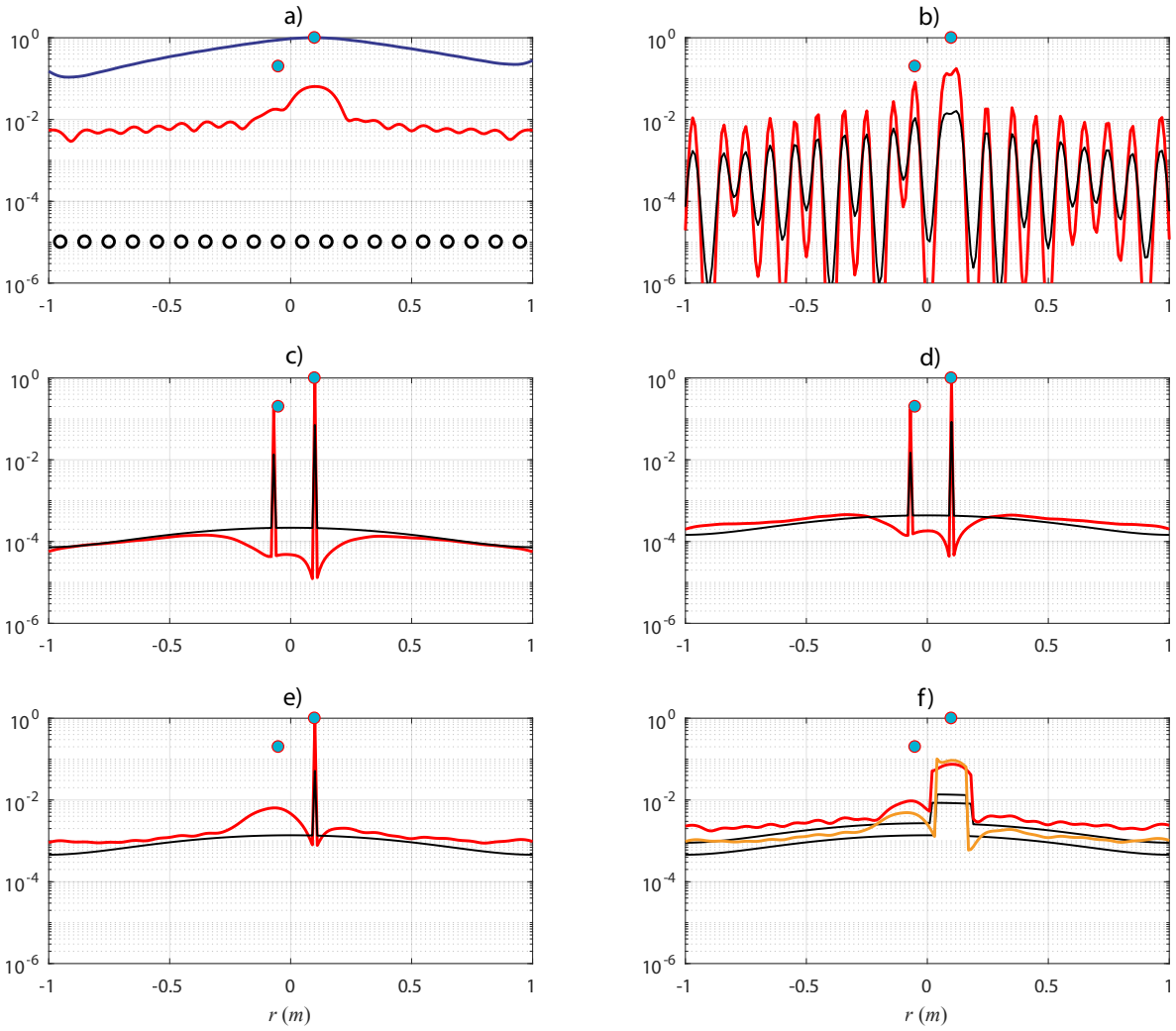

Figure 14: Same configuration as in Fig. 12 at $f = 0.10$ kHz, $N = 100$ snapshots and two incoherent monopoles with 14 dB difference in levels.

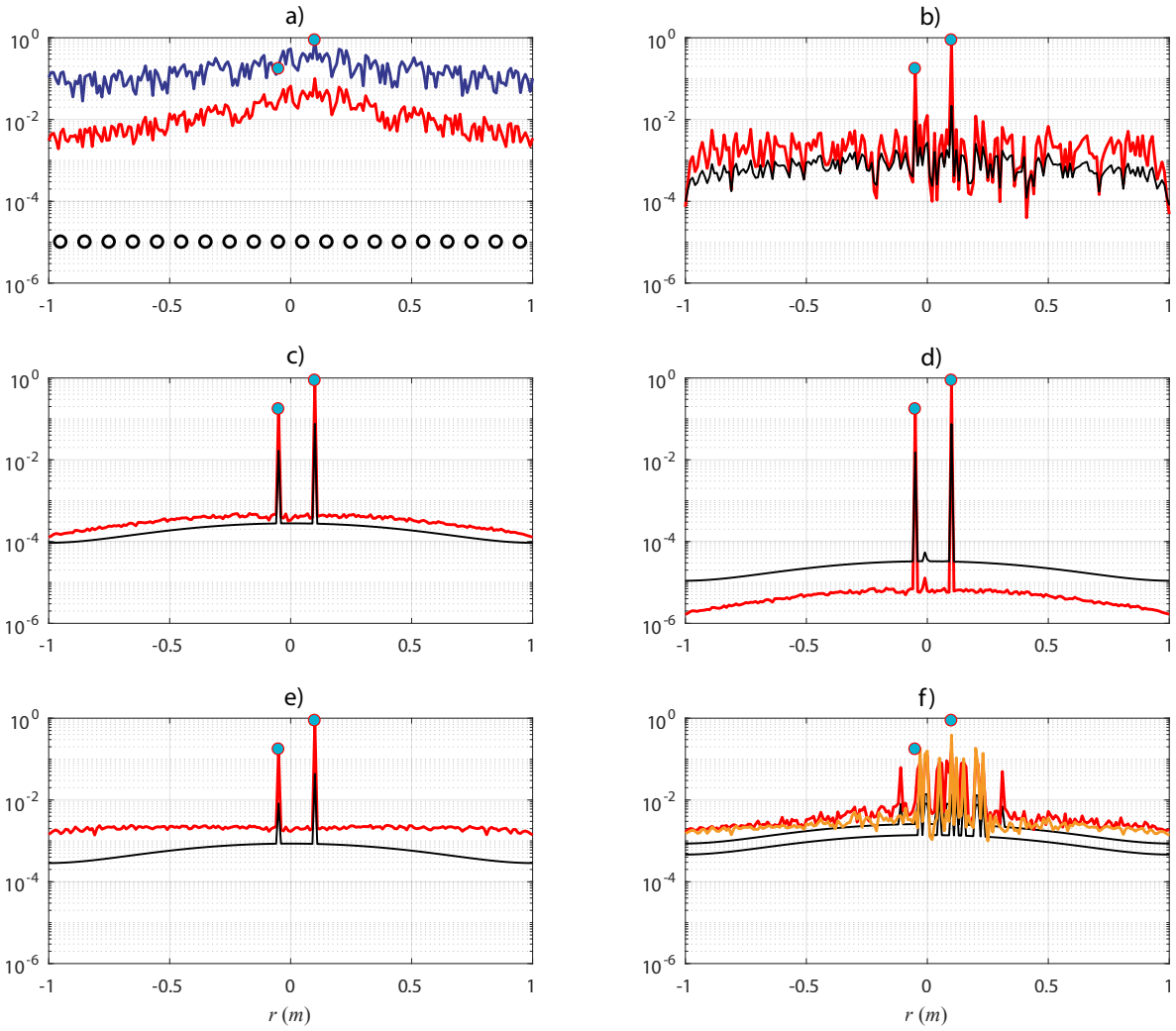

Figure 15: Same configuration as in Fig. 14 at $f = 10$ kHz.

Algorithm 3

- Define a frequency band $\mathcal{B} = \{f_1, ..., f_B\}$ as a collection of frequency bins (as typically returned by the discrete Fourier transform) where the self-tunning AF – as introduced in Eq. (17) – is invariant.
- Apply Algorithm 1 of section 4.4 where
  - steps 0 and 3 are unchanged,
  - step 2 is modified with

$$\overline{\chi^2}_{q,[k-1]} = \frac{\sum_{l=1}^{B}\sum_{i=1}^{N}|\hat{s}_{q,i,[k-1]}(f_l)|^2}{\alpha^2_{[k-1]}\sigma^2_{0,q}} = N\frac{\sum_{l=1}^{B} S_{qq,[k-1]}(f_l)}{\alpha^2_{[k-1]}\sigma^2_{0,q}} \tag{31}$$

    substituted for $\chi^2_{q,[k-1]}$, where $\hat{s}_{q,i,[k-1]}(f_l)$ stands for the evaluation of $\hat{s}_{q,i,[k-1]}$ at frequency f_l and snapshot i,
  - steps 1, 4 and 5 are unchanged but performed for *each* frequency f_k in band $\mathcal{B}$.

Equation (31) is the main modification to address the broadband case. The quantity $\overline{\chi^2}_{q,[k-1]}$ may be interpreted as the normalized energy of the source coefficients – as measured by their power spectra – summed in band $\mathcal{B}$. A resolution in terms of the CSM is also trivially by proceeding as in Algorithm 3. Eventually, it is noteworthy that a different regularization parameter is reasonably assumed at each frequency.

Since BIBF solves the inverse problem at once for *all* frequencies in band $\mathcal{B}$ (say a total B frequencies), its computational cost is obviously increased as compared to one call of BF (for *one* frequency). However, it is lower than repeating Algorithm 1 successively for each frequency in the band, since only one AF is inferred instead of B. In addition, according to subsection 5.4, BIBF is expected to converge more rapidly than IBF solved at a single frequency since the corresponding missing information matrix (see Eq. (30)) is divided by an extra factor B. Overall, this endows BIBF with less computational effort than the repeated call of BF for scanning all frequencies in the band.

Example 4. BIBF is illustrated here for the reconstruction of four uncorrelated monopole sources driven by white noises in a wide frequency interval $\mathcal{B}$ ranging from 100 Hz to 20 kHz. The sources are located at $r_1 = -0.1$, $r_2 = 0$, $r_3 = 0.1$ and $r_4 = 0.2$ m on the source domain $\Gamma_s = \{r : -1 \leq r \leq 1\}$ (in meters) and have decreasing levels in the proportion 1, 0.25, 0.04 and 0.01 (i.e. 0dB, -6dB, -14dB, -20dB). The array is the same as in Example 3 and the range distance is $R = 0.1$ m. The SNR is set to 20 dB. BIBF is run with the CSM computed on 100 snapshots and a spatial resolution of 10 Hz is used (thus leading to $B = 1991$ frequency bins). The results of conventional beamforming, BF, IBF and BIBF are compared in Fig. 16. It is again checked that beamforming and BF both suffer from 1) a poor spatial resolution which makes the localization of the sources difficult below 2 kHz and 2) a limited dynamic range which prevents the two smallest sources from being identified. On the one hand, whereas IBF greatly upgrades the results, it is limited downward to 1 kHz for the reconstruction of the strongest source and to 2 kHz for the smallest source. The identification of the two small sources is actually intermittent and the smallest one is missed at most frequencies. On the other hand, BIBF completely fixes these errors. The accuracy in terms of localization and quantification is excellent over the full frequency range and for all sources. This is because a frequency-independent AF is better estimated from an "average" of the map in Fig. 16(c), which in turns improves the estimation of the sources at low frequencies and low SNRs.

Figure 17 displays the estimated hyperparameters α^2 and η^2 as a function of frequency. The prior source power spectrum $\alpha^2(f)$ is constant on the average, as expected from sources driven by white noises. The spectrum of the regularization parameter, $\eta^2(f)$, shows a hyperbolic decrease because the conditioning of the inverse problem actually improves with frequency. The fact that theses hyperparameters have to estimated repeatedly for all frequencies in the band $\mathcal{B}$ justifies the need for the automatic process based on Bayesian regularization given in Appendix B.

Concerning computational time, the BIBF algorithm was actually found faster than repeating the IBF for all frequencies in the scrutinized band. As previoulsy explained, this is because accounting for all frequencies at once makes the convergence of the iterations faster (convergence was achieved after 7 iterations for BIBF while each run

of the IBF required about twice as many iterations on the average).

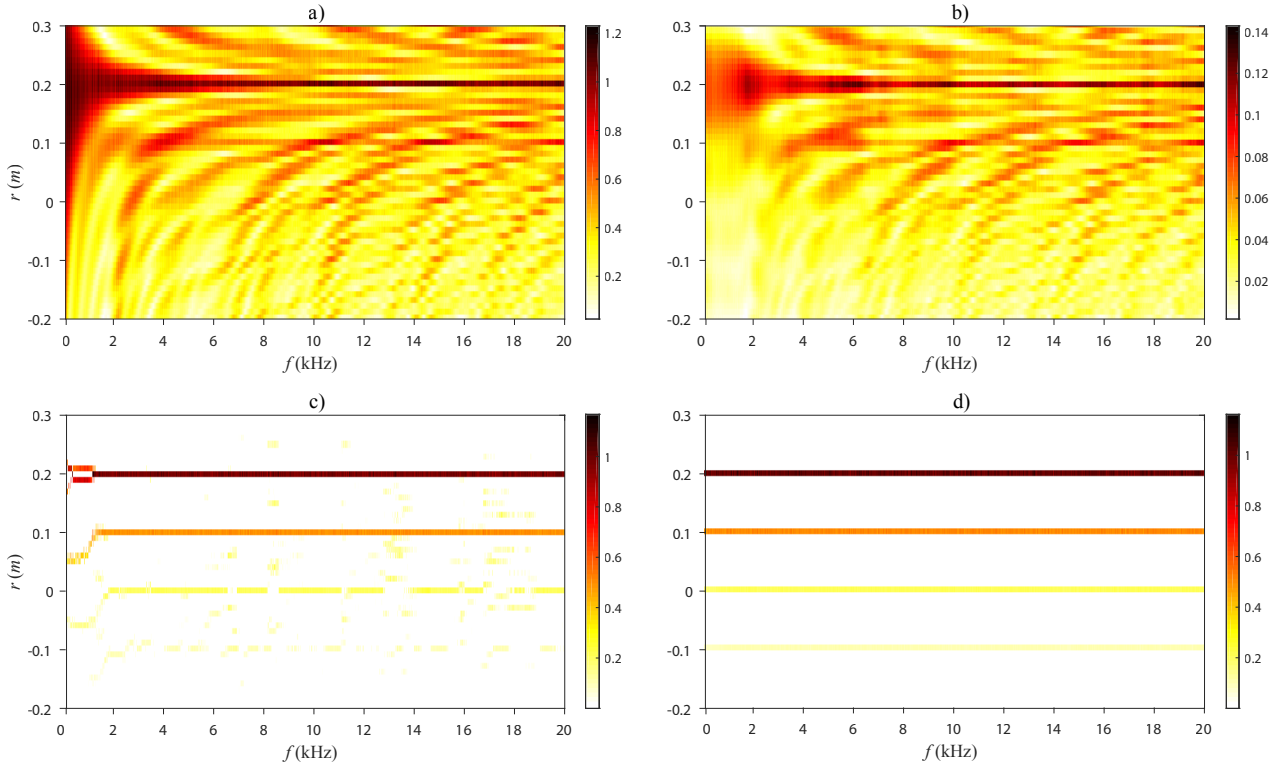

Figure 16: Estimated source distribution on the 1D domain $\Gamma_s = \{r : -0.2 \leq r \leq 0.3\}$, as a function of frequency, when the theoretical distribution comprises four monopoles placed at $r = 0.2$ m, $r = 0.1$ m, $r = 0$ m and $r = -0.1$ m, driven by uncorrelated white Gaussian signals with levels in the proportion 1, 0.25, 0.04 and 0.01, respectively, and SNR of 20 dB. Estimates returned by a) BF with a fixed uniform aperture function, b) conventional beamforming, c) IBF with the MCS prior, and d) broadband IBF with the MCS prior.

7. Experimental application

Experiments have been conducted to validate the theoretical results of the previous sections. The experimental setup is displayed in Fig. 18. It consists of an 117 microphone LMS Sound Camera Digital Array placed in a semi-anechoic room, 30 cm away from "point" sources flush-mounted on a wood plate and spaced apart by a distance of 14 cm. The sources are fed by mutually uncorrelated white-noise signals with a global level difference of about 12 dB. The radiation of each source with the other one switched off was first recorded by an intensity probe in order to provide a reference based on a normative sound power measurement technique.

Since the radiation of the acoustic energy is in the half space, the baffled-source Green functions were used in the inverse problem. In addition, since the position of the sources is expected to be invariant in frequency, the experimental setup provides an excellent example to test the BIBF algorithm. Two priors are chosen to enforce sparsity of the reconstructed sources: *i*) the MCS prior (with $a = b = 0.01$ =) because it has been previously found to enforce high sparsity while being simple to use and *ii*) the GMCG prior with $p = 1.3$ because it best matches the actual sources which, in this experience, do not reduce to perfect points. The results are displayed by means of the sound power,

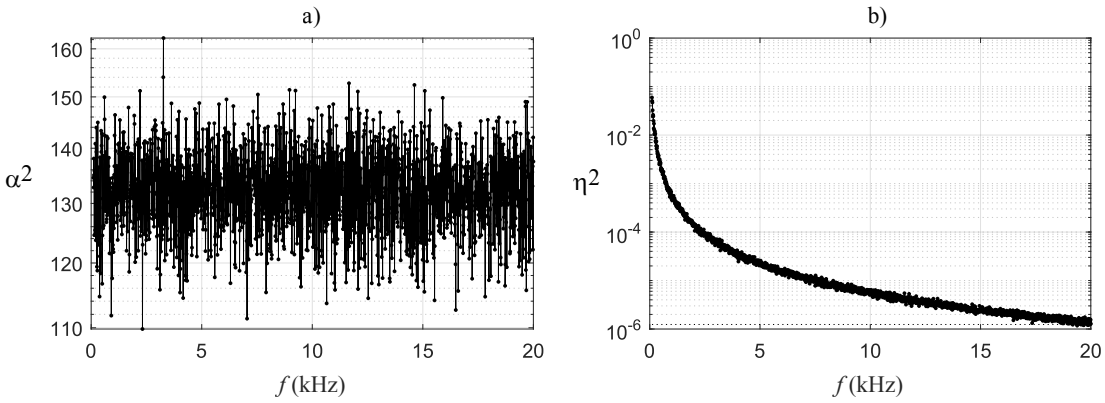

Figure 17: Estimated hyperparameters a) α^2 and b) η^2, as a function of frequency, returned by broadband IBF with the MCS prior.

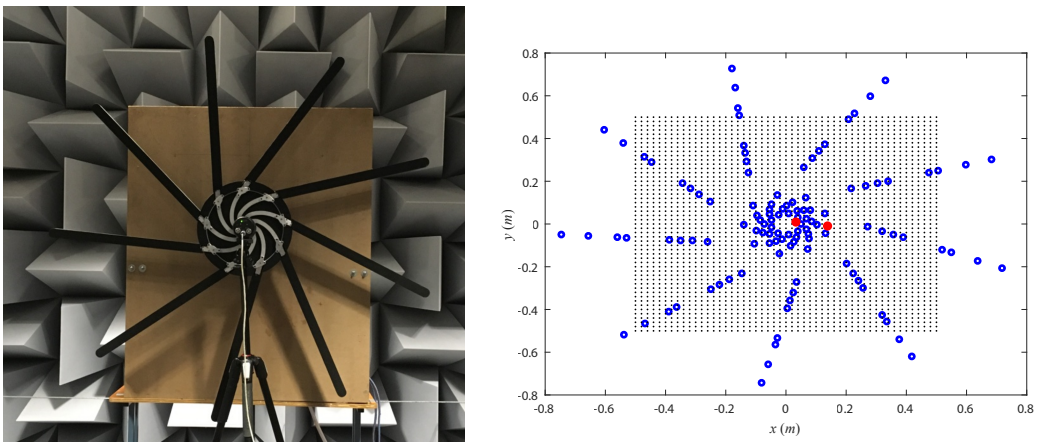

Figure 18: Layout of the experimental set-up (left) and its geometrical configuration (right) showing the locations of the two sources (red dots), the microphone positions (blue circles) and the nodes of the calculation grid (black dots).

$W(\mathbf{r}_q)$, calculated at each node q of the calculation grid with the formula

$$W(\mathbf{r}_q) = \frac{\rho c_0 k^2}{8\pi} S_{qq} - \frac{\rho c_0 k}{2} \sum_{l \neq q} \mathrm{Im}\left\{G(\mathbf{r}_q|\mathbf{r}_l) S_{ql}\right\}, \tag{32}$$

where operator $\mathrm{Im}\{\cdots\}$ extracts the imagy part, k stands for the wavenumber, ρ for the air density, c_0 for the sound speed, and S_{ql} is the estimated cross-power spectrum between the complex volume velocities at positions $\mathbf{r}_q$ and $\mathbf{r}_l$ [47].

Examples of sound power maps reconstructed in the third-octave band at 1 kHz by IBF and BIBF and using the two aforementoned priors are shown in Fig. 19(a). It is seen that the benefit of using BIBF instead of IBF is to improve the spatial resolution and thus to better reveal the presence of the smallest source. This is further evidenced by the one dimensional cut of the sound power distribution at $y = 0$ shown in Fig. 19(b), where the very low frequency 300 Hz is now considered. Besides, it is observed in Fig. 19 that the MCS prior returns a sparser reconstruction than the GMCG prior with $p = 1.3$. Although this may be advantageous for localization, it does not necessarily reflects to the actual source characteristics as shown hereafter.

Figure 20(a) displays the sound power spectra estimated from BF and IBF with the two different priors. The sound

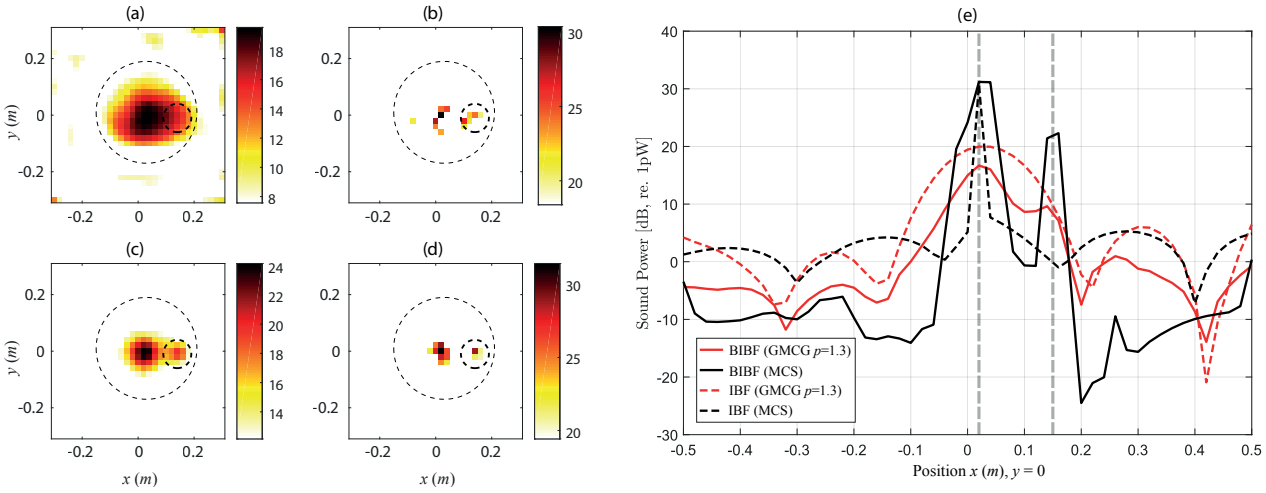

Figure 19: Sound power maps reconstructed in the 1 kHz third-octave band from iterated Bayesian Focusing using a) the GMCG prior (with $p = 1.3$) and b) the MCS prior (with $a = b = 0.01$). Reconstruction from Broadband Iterated Bayesian Focusing MCS using c) the GMCG prior and d) the MCS prior. e) Comparison of the sound power distribution ($|x| \leq 0.5$m, $y = 0$) at 300 Hz estimated from Iterated Bayesian Focusing and Broadband Iterated Bayesian Focusing with GMCG ($p = 1.3$) and MCS ($a = b = 0.01$) priors.

power was evaluated from Eq. (32) by integrating over the large circular region depicted in Fig. 19(a), which encloses the two sources. The spectra are also compared with the reference returned by the intensity probe, which imposes the upper limit of 6.5 kHz of the frequency interval. On the one hand, it is seen that BF underestimates the actual sound power, especially in the higher frequency range. This complies with previous numerical simulations, even though the under-estimation is not as important as before since the array now covers a large solid angle from the source point of view. On the other hand, the MCS prior is found to overestimate the sound power. This is not contradictory with the results of the paper since the actual sources used in the experiment are not true monopoles: their supports are not exactly point-wise on surface Γ and their directivities are not exactly omnidirectional. Consequently, the GMCG prior with a controlled degree of sparsity ($p = 1.3$) seems to better reflect their actual nature. This observation points out the importance of correctly choosing the prior type and, therefore, of having a general algorithm that allows testing several of them.

Finally, Fig. 20(b) displays the estimated sound power of each source compared with the intensity reference obtained with one of the two soures alternatively switched off. For the secondary source this is achieved by integrating over the small circular region depicted in Fig. 19(a) and, for the primary source, by taking the difference with the spectra of Fig. 20(a). Only the results obtained with the GMCG prior are shown, yet compared for IBF and BIBF. It is again checked that BIBF allows a better resolution of the secondary source in the very low frequency range, down to 300 Hz, which is compliant with the result shown in Fig. 19(b).

8. Conclusion

This paper has demonstrated how sparse acoustical holography is naturally achieved by simply iterating the Bayesian Focusing method previously proposed in Ref. [34]. The principle consists in using the estimated sources in the current iteration to update the aperture function in the next iteration. Formally, this is equivalent to considering the aperture function as a random quantity (endowed with a prior probability density function) that is estimated conjointly with the sources. Not only does this point of view provide physical insight into the mechanism of sparse acoustical holography – and in particular why it can considerably improve the estimation of source quantification and directivity high in frequency – but it also allows direct extension to different scenarios. For instance, the method is easily formulated in terms of the cross-spectral matrix and adapted to the broadband case (i.e. frequency group sparsity) where source positions are assumed stationary in space. The method also allows complete flexibility in the choice

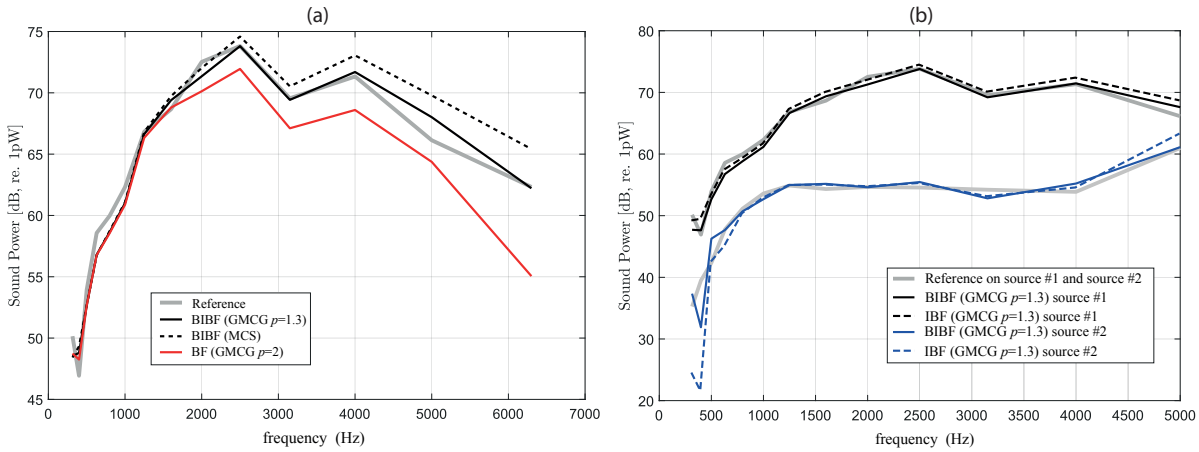

Figure 20: a) Overall sound power spectra estimated with Iterated Bayesian Focusing, with the MCS ($a = b = 0.01$) and GMCG ($p = 1.3$) priors, compared to Bayesian Focusing and to the intensity probe reference. b) Sound power spectra of each source estimated with Iterated Bayesian Focusing and Broadband Iterated Bayesian Focusing, with the GMCG ($p = 1.3$) prior.

of the sparsity penalty. Several updating rules for the aperture function have been investigated, rooted on different source priors. The multivariate complex Student$-t$ prior appears to be a good candidate, which returns sparser results than the typical ℓ_1-norm approach while at the time offering stable results. However, the Generalized multivariate complex Gaussian may be advantageous in cases where the degree of sparisty needs to be controlled, for instance when the actual sources are not truly pointwise. A last but not least advantage is that Iterated Bayesian Focusing inherits the possibility of automatic setting of the regularization parameter, an issue which has remained critical in sparse methods.

Overall, it is believed that Iterated Bayesian Focusing provides a comprehensive view of sparse acoustical holography that might not be shared by other approaches.

Beyond methodological aspects, this paper has also confirmed the experimental results of Ref. [23] establishing that sparse holography improves the reconstruction of sources not only in terms of localization – as does compressive beamforming in the far-field – but also of quantification and of directivity, in a frequency range that is considerably enlarged as compared to classical methods. It has been found that these improvements can be achieved even with regular arrays, provided that sparser priors than those leading to the standard ℓ_1 penalty are used; therefore, using random arrays is not a necessary condition as is in compressed beamforming or ESM.

Acknowledgments

This work was performed within the framework of the Labex CeLyA of Université de Lyon, operated by the French National Research Agency (ANR-10-LABX-0060/ANR-11-IDEX-0007), and supported by the Labcom P3A (ANR-13-LAB2-0011-01).

Appendix A. Optimal basis for sparse sources

In a general setting and in functional notation, the source distribution is expanded onto a spatial basis $\{\phi_k(\mathbf{r})\}_{k=1}^{Q}$, $M \leq Q$, i.e.

$$s_i(\mathbf{r}) = \sum_{k=1}^{Q} \phi_k(\mathbf{r}) c_{k,i} \tag{A.1}$$

where coefficients $c_{k,i}$, $k = 1, ..., Q$, $i = 1, ..., N$, are the unknowns of the problem. In discrete form, this reads $\mathbf{s}_i = \mathbf{\Phi}\mathbf{c}_i$ where $[\mathbf{\Phi}]_{kq} \doteq \phi_k(\mathbf{r}_q)$ and vector $\mathbf{c}_i$ collects coefficients $c_{k,i}$. This parametrization imposes a priori the spatial

correlation

$$\mathbb{E}\left\{\mathbf{s}_i\mathbf{s}_i^H\right\} = \mathbf{\Phi}\mathbb{E}\left\{\mathbf{c}_i\mathbf{c}_i^H\right\}\mathbf{\Phi}^H = \mathbf{\Phi}\mathbf{\Omega}_c\mathbf{\Phi}^H \tag{A.2}$$

where $\mathbf{\Omega}_c$ stands as the prior covariance matrix of $\mathbf{c}_i$. Therefore, by assigning to $\mathbf{c}_i$ a prior PDF in the form of the complex Gaussian, $[\mathbf{c}_i] = \mathcal{N}_{\mathbb{C}}(\mathbf{c}_i; \mathbf{0}, \mathbf{\Omega}_c)$, one finds the MAP estimate

$$\hat{\mathbf{c}}_i = \mathbf{\Omega}_c\mathbf{\Phi}^H\mathbf{G}^H\left(\mathbf{G}\mathbf{\Phi}\mathbf{\Omega}_c\mathbf{\Phi}^H\mathbf{G}^H + \beta^2\mathbf{I}\right)^{-1}\mathbf{p}_i \tag{A.3}$$

with η^2 as defined in Eq. (9). Equation (A.3) generally applies for any choice of $\mathbf{\Phi}$ and $\mathbf{\Omega}_c$. In the case of sparse sources, it make sense to expect a priori an incoherent source field, yet possibly non-homogeneous, as given by Eq. (11). It is proved in Ref. [35] that the prior covariance matrix $\mathbf{\Omega}_c$ which makes $\mathbb{E}\left\{\mathbf{s}_i\mathbf{s}_i^H\right\}$ in Eq. (A.2) as close as possible to $\alpha^2\Sigma_0$ – with Σ_0^2 diagonal as defined in Eq. (12) – in the mean-square sense is

$$\mathbf{\Omega}_c = \alpha^2\left(\mathbf{\Phi}^H\Sigma_0^{-2}\mathbf{\Phi}\right)^{-1}. \tag{A.4}$$

With such a choice, it is further proved in Ref. [34] that the optimal $\mathbf{\Phi}$ that minimizes the mean-square error (or returns the highest MAP) of the estimated source distribution is

$$\mathbf{\Phi} = \Sigma_0^2\mathbf{G}^H \in \mathbb{C}^{Q\times M}. \tag{A.5}$$

This shows in passing that for $\mathbf{\Phi}$ to be a basis, one must have $M = Q$. Substituting expressions (A.4) and (A.5) back into Eqs. (A.1) and (A.3) then returns Eq. (9). Comparison of Eqs. (A.3) and (9) shows that, with a complex Gaussian prior PDF $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \alpha^2\Sigma_0^2)$, the optimal parameterization that enforces an incoherent source field is actually equivalent to using the canonical basis $\mathbf{I}_{Q\times Q}$ together with the prior covariance matrix $\alpha^2\Sigma_0^2$ on the spatial samples $\mathbf{s}_i$ of the source seen as the unknowns of the problem. The same parameterization can actually be adopted afterwards with $Q > M$, provided that Eq. (9) remains invertible by properly setting the value of the regularization parameter η^2.

Although this parameterization turns out somewhat similar to the Equivalent Source Method (ESM) [48, 24], it bears several differences which are worth being highlighted.

- First, the present proof is deductive and it establishes that the parameterization of the problem is optimal under the assumption of prior incoherence of the source distribution. Other assumptions on the prior spatial correlation would lead to other parameterizations.
- Second, it emphasizes that the actual dimension of the problem (i.e. the dimension of the spatial basis $\mathbf{\Phi}$ as given in Eq. (A.5)) is $K = M$ (the number of microphones) and not Q (the dimension of the discretized source surface), which is advantagous since one often has $M << Q$. For this reason, the optimal basis introduced in Eq. (A.5) leads to a minimal parameterization. This is not explicit in the ESM method.
- Third, it strictly involves Green functions satisfying Neumann boundary conditions on Γ_s and not necessarily free-field Green functions (of monopoles) as is commonplace with the ESM.

Appendix B. Bayesian regularization

Following the Bayesian program, the regularization parameter appearing in Eqs. (9) together with the source power needed in Eq. (23) are estimated as the MAP of the joint PDF $[\alpha^2, \eta^2|\mathbb{S}_{pp}, \Sigma^2]$ with $\mathbb{S}_{pp}$ the CSM (see Eq. (1)) and Σ^2 a given AF. This subsection resumes the results of Ref. [35] where the solution to this problem is addressed in details. The algorithm is as follows.

- **Step 1**: Compute the eigenvalue decomposition

$$\mathbf{G}\Sigma^2\mathbf{G}^H = \sum_{m=1}^{M} s_m^2\mathbf{u}_m\mathbf{u}_m^H \tag{B.1}$$

where $\mathbf{u}_k$ are and s_k^2, $k = 1, ..., M$, are the eigenvectors and eigenvalues of $\mathbf{G}\Sigma^2\mathbf{G}^H$, respectively.

- **Step 2**: Project the data in the space spanned by the eigenvectors:

$$< |y_k|^2 > \doteq \mathbf{u}_k^H \mathbb{S}_{pp} \mathbf{u}_k, \quad k = 1, ..., M. \tag{B.2}$$

- **Step 3**: Estimate the source power as

$$\hat{\alpha}^2(\eta^2) = \frac{1}{M} \sum_{k=1}^{M} \frac{< |y_k|^2 >}{s_k^2 + \eta^2}. \tag{B.3}$$

- **Step 4**: Estimate the regularization parameter as the minimum of the cost function

$$J(\eta^2) = \frac{1}{M} \sum_{k=1}^{M} \ln(s_k^2 + \eta^2) + \left(M - \frac{2}{N}\right) \ln \hat{\alpha}^2(\eta^2), \tag{B.4}$$

which may be easily found by a simple grid search and eventually substitute back this result into Eq. (B.3).

Appendix C. Prior PDFs in the SMoG family

This appendix provides expressions of some prior PDFs in the SMoG family obtained from different choices of the hyperprior $\left[\tau_q^2\right]$ in Eq. (18). Expressions are given in terms of the sufficient statistic $\chi_q^2 \doteq NS_{qq}/\alpha^2\sigma_{0,q}^2$.

- **Multivariate complex Student-*t* (MCS).** Let us assume that τ_q^2 is distributed like an *inverse Gamma*, with shape and scale parameters $a > 0$ and $b > 0$, respectively:

$$\left[\tau_q^2\right] = \frac{b^a}{\Gamma(a)} \frac{\mathrm{e}^{-b\tau_q^{-2}}}{\tau_q^{2(a+1)}}. \tag{C.1}$$

It then holds that

$$\left[\chi_q^2\right] = \frac{\Gamma(N+a)b^a}{(\pi\alpha^2\sigma_{0,q}^2)^N \Gamma(a) \left(\chi_q^2 + b\right)^{(N+a)}}, \tag{C.2}$$

which is recognized as a multivariate complex version of the Student-*t* distribution. This PDF is made very leptokurtic by setting the value of a small. By substituting the MCS for the prior PDF into the MAP (see Eq. (E.1) in appendix Appendix E), one arrives at an LS formulation with a logarithmic penalty, viz

$$\begin{aligned} \hat{\mathbf{s}}_i &= \underset{\mathbf{s}_i}{\mathrm{Argmax}}[\mathbf{P}|\mathbf{S}][\mathbf{S}] = \underset{\mathbf{s}_i}{\mathrm{Argmin}}\,(-\ln[\mathbf{p}_i|\mathbf{s}_i][\mathbf{S}]) \\ &= \underset{\mathbf{s}_i}{\mathrm{Argmin}} \left(\|\mathbf{p}_i - \mathbf{G}\mathbf{s}_i\|_2^2 + \beta^2(N+a) \ln \left(\alpha^{-2} \sum_{j=1}^{N} \mathbf{s}_j^H \Sigma_0^{-2} \mathbf{s}_j + b \right) \right). \end{aligned} \tag{C.3}$$

The logarithmic penalty is here naturally endowed with a regularization term $b > 0$ that prevents it from diverging.

- **Multivariate complex K-distribution (MCK).** let us assume that τ_q^2 is distributed like a *Gamma*, with shape and rate parameters $a > 0$ and $b > 0$, respectively:

$$\left[\tau_q^2\right] = \frac{b^a}{\Gamma(a)} \mathrm{e}^{-b\tau_q^2} \tau_q^{2(a-1)}. \tag{C.4}$$

It then holds that

$$\left[\chi_q^2\right] = \frac{2b^{\frac{(N+a)}{2}}}{(\pi\alpha^2\sigma_{0,q}^2)^N \Gamma(a)} \left(\chi_q^2\right)^{\frac{(a-N)}{2}} K_{N-a}\left(2\sqrt{b\chi_q^2}\right), \tag{C.5}$$

where $K_\nu(z)$ is the modified Bessel function of the second kind with order ν (note that $K_\nu(z) = K_{-\nu}(z)$). This is recognized as a multivariate complex version of the K-distribution [49]. This PDF is made very leptokurtic by setting the value of a small. The MCK also leads to an equivalent penalized LS problem with a flexible penalty function that can be given various shapes depending on the values of a, b and N.

- **Multivariate complex Laplacian (MCL).** As a special case when $a = N + 1/2$ and $b = 1/4$, the MCK becomes

$$\left[\chi_q^2\right] = \frac{1/2}{(\pi\alpha^2\sigma_{0,q}^2)^N}\frac{\Gamma(N)}{\Gamma(2N)}\mathrm{e}^{-\sqrt{\chi_q^2}} \tag{C.6}$$

  which is a multivariate complex version of the Laplacian distribution [49] as well as a particular case of the generalized multivariate complex Gaussian introduced in the next paragraph. The Laplacian PDF is a popular choice to enforce sparsity of real-value coefficients since it corresponds to a cost function with a penalty term based on the ℓ_1-norm. This is exploited in the Bayesian LASSO [50, 51]. It is noteworthy that in the complex case, the Laplacian originates from an SMoG rooted on a different prior than in the real case (i.e. an exponential PDF). The equivalent penalized LS problem reads

$$\hat{\mathbf{s}}_i \quad = \quad \underset{\mathbf{s}_i}{\mathrm{Argmin}}\left(\|\mathbf{p}_i - \mathbf{G}\mathbf{s}_i\|_2^2 + \frac{\beta^2}{\alpha}\left(\sum_{j=1}^{N}\mathbf{s}_j^H\Sigma_0^{-2}\mathbf{s}_j\right)^{\frac{1}{2}}\right) \tag{C.7}$$

  where the penalty function happens to mix the ℓ_1 and the ℓ_2 norms in a way that may not be obvious without using the Bayesian approach.

- **Generalized multivariate complex Gaussian (GMCG).** The multivariate complex version of a generalized Gaussian PDF reads

$$\left[\chi_q^2\right] = \frac{p/2}{(\pi\alpha^2\sigma_{0,q}^2)^N}\frac{\Gamma(N)}{\Gamma(2N/p)}\mathrm{e}^{-(\chi_q^2)^{\frac{p}{2}}}, \quad 0 < p \leq 2. \tag{C.8}$$

  Since the generalized Gaussian (also known as the exponential power distribution in the statistical literature) is an SMoG, the GMCG is also an SMoG (actually for $p \leq 2$). The expression of $\left[\tau_q^2\right]$ happens to be quite intricate, yet only the expression of the marginal PDF (C.8) will be needed in the following. Formula (C.8) extends the result of Ref. [52] to the multivariate case.

  Sparsity is enforced by setting small values of p, strictly smaller than 2 and typically smaller than or equal to 1. As special cases, $p = 2$ corresponds to the multivariate complex Gaussian (MCG) and $p = 1$ to the MCL.

  The GMCG prior gives rise to the penalized LS problem,

$$\hat{\mathbf{s}}_i = \underset{\mathbf{s}_i}{\mathrm{Argmin}}\left(\|\mathbf{p}_i - \mathbf{G}\mathbf{s}_i\|_2^2 + \frac{\beta^2}{\alpha^p}\left(\sum_{j=1}^{N}\mathbf{s}_j^H\Sigma_0^{-2}\mathbf{s}_j\right)^{\frac{p}{2}}\right), \tag{C.9}$$

  where the penalty function mixes the ℓ_2 and the ℓ_p norms.

- **Mixture of multivariate complex Gaussians (MMCG).** Mixture of multivariate complex Gaussians are obtained from the SMoG model when τ_q^2 is a discrete random variable. The simplest situation is when it takes two values, say $\bar{\tau}_1^2$ and $\bar{\tau}_2^2$, with probabilities $\bar{\pi}$ and $(1 - \bar{\pi})$, respectively, which means that a fraction $\bar{\pi}$ of the coefficients are distributed like a complex Gaussian with variance $\bar{\tau}_1^2$ and a fraction $1 - \bar{\pi}$ like a complex Gaussian with variance $\bar{\tau}_2^2$. The PDF then reads $\left[\tau_q^2\right] = \bar{\pi}\delta(\tau_q^2 - \bar{\tau}_1^2) + (1 - \bar{\pi})\delta(\tau_q^2 - \bar{\tau}_2^2)$ (with δ the Dirac delta) and the corresponding SMoG is

$$\left[\chi_q^2\right] = \bar{\pi} \cdot \mathcal{N}_{\mathbb{C},N}(\chi_{q,[k-1]}^2; 0, \pi\alpha^2\sigma_{0,q}^2\bar{\tau}_1^2) + (1 - \bar{\pi}) \cdot \mathcal{N}_{\mathbb{C},N}(\chi_{q,[k-1]}^2; 0, \pi\alpha^2\sigma_{0,q}^2\bar{\tau}_2^2) \tag{C.10}$$

  where $\mathcal{N}_{\mathbb{C},N}(\chi_{q,[k-1]}^2; 0, \sigma^2) \doteq \sigma^{-2N}\mathrm{e}^{-\chi_q^2/\sigma^2}$. Very sparse PDF can thus be constructed when $\bar{\tau}_1 << \bar{\tau}_2$ and $\bar{\pi} \to 1$. The extreme scenario when $\bar{\tau}_1 \to 0$ returns the Bernouilli-Gauss model where a fraction $\bar{\pi}$ of the coefficients

are exactly zero on the average, the other coefficients being distributed like a complex Gaussian with variance $\bar{\tau}_2^2$. The equivalence with a penalized LS problem is here less obvious.

Appendix D. Point estimates in the SMoG family of priors

It is proved in this Appendix that, in the SMoG family of priors, the MAP and the posterior mean estimates of $\mathbf{s}_i$ both take the structure of the penalized LS solution given in Eq. (20).

Appendix D.1. Posterior mean

The posterior mean is defined as $\mathbb{E}\{\mathbf{s}_i|\mathbf{P}\} = \int \mathbf{s}_i[\mathbf{s}_i|\mathbf{P}]\mathrm{d}\mathbf{s}_i$. Since

$$[\mathbf{s}_i|\mathbf{P}] = \int [\mathbf{s}_i|\mathbf{P}, \Sigma^2][\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2 = \int [\mathbf{s}_i|\mathbf{p}_i, \Sigma^2][\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2 \tag{D.1}$$

where the notation $\Sigma^2 \doteq \mathrm{Diag}(\sigma_q^2)$ has been used, $\mathbb{E}\{\mathbf{s}_i|\mathbf{P}\} = \int\int \mathbf{s}_i[\mathbf{s}_i|\mathbf{p}_i, \Sigma^2][\Sigma^2|\mathbf{P}]\mathrm{d}\mathbf{s}_i\mathrm{d}\Sigma^2$. Recognizing that $\int \mathbf{s}_i[\mathbf{s}_i|\mathbf{p}_i, \Sigma^2]\mathrm{d}\mathbf{s}_i$ is the posterior mean $\hat{\mathbf{s}}_i(\Sigma^2)$ conditioned on Σ^2 as given in Eq. (9), one has

$$\begin{aligned}\mathbb{E}\{\mathbf{s}_i|\mathbf{P}\} &= \int \hat{\mathbf{s}}_i(\Sigma^2)[\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2 \\ &= \int \left(\Sigma^2\mathbf{G}^H(\mathbf{G}\Sigma^2\mathbf{G}^H + \eta^2\mathbf{I})^{-1}\mathbf{p}_i\right)[\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2. \end{aligned}\tag{D.2}$$

Appendix D.2. MAP estimate

The MAP estimate satisfies the optimality condition $\frac{\partial}{\partial \mathbf{s}_i^H}[\mathbf{S}|\mathbf{P}] = 0$. Using Eq. (D.1), this is equivalent to $\int \frac{\partial}{\partial \mathbf{s}_i^H}[\mathbf{S}|\mathbf{P}, \Sigma^2][\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2 = 0$. Since $[\mathbf{s}_i|\mathbf{p}_i, \Sigma^2] = \mathcal{N}_{\mathbb{C}}(\mathbf{s}_i; \hat{\mathbf{s}}_i, \mathbf{\Omega})$ with $\mathbf{\Omega} = (\beta^{-2}\mathbf{G}^H\mathbf{G} + \alpha^{-2}\Sigma^{-2})^{-1}$ and $\hat{\mathbf{s}}_i = \beta^{-2}\mathbf{\Omega}\mathbf{G}^H\mathbf{p}_i$, the optimality condition becomes

$$\begin{aligned}&\int \mathbf{\Omega}^{-1}(\mathbf{s}_i - \hat{\mathbf{s}}_i)[\mathbf{S}|\mathbf{P}, \Sigma^2][\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2 = 0, \\ &\Leftrightarrow \left(\beta^{-2}\mathbf{G}^H\mathbf{G}[\mathbf{S}|\mathbf{P}] + \alpha^{-2}\int \Sigma^{-2}[\mathbf{S}|\mathbf{P}, \Sigma^2][\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2\right)\mathbf{s}_i - \beta^{-2}\mathbf{G}^H\mathbf{p}_i[\mathbf{S}|\mathbf{P}] = 0\end{aligned}\tag{D.3}$$

where $\int [\mathbf{S}|\mathbf{P}, \Sigma^2][\Sigma^2|\mathbf{P}]\mathrm{d}\Sigma^2 = [\mathbf{S}|\mathbf{P}]$ has been used. Dividing all terms by $\beta^{-2}[\mathbf{S}|\mathbf{P}]$ and recognizing that

$$\frac{[\mathbf{S}|\mathbf{P}, \Sigma^2][\Sigma^2|\mathbf{P}]}{[\mathbf{S}|\mathbf{P}]} = [\Sigma^2|\mathbf{S}, \mathbf{P}] = [\Sigma^2|\mathbf{S}] \tag{D.4}$$

then yields

$$\left(\beta^{-2}\mathbf{G}^H\mathbf{G} + \alpha^{-2}\underbrace{\int \Sigma^{-2}[\Sigma^2|\mathbf{S}]\mathrm{d}\Sigma^2}_{\mathbb{E}\{\Sigma^{-2}|\mathbf{S}\}}\right)\mathbf{s}_i = \beta^{-2}\mathbf{G}^H\mathbf{p}_i. \tag{D.5}$$

This proves Eq. (20).

Appendix E. Convergence of the IBF algorithm

Let us first assume that the regularization parameter η^2 and the prior source power α^2 are known. Then, considering all snapshots together, the Bayesian program is to find the MAP estimate

$$\begin{aligned}\hat{\mathbf{s}}_i &= \underset{\mathbf{s}_i}{\text{Argmax}}\,[\mathbf{S}|\mathbf{P}] \\ &= \underset{\mathbf{s}_i}{\text{Argmax}}\,\ln[\mathbf{S}|\mathbf{P}] \\ &= \underset{\mathbf{s}_i}{\text{Argmax}}\,(\ln[\mathbf{P}|\mathbf{S}] + \ln[\mathbf{S}])\,.\end{aligned}$$

The second line results from the application of Bayes rule (6) where the evidence $[\mathbf{P}]$ has been ignored since it does not depend on the unknown quantity $\hat{\mathbf{s}}_i$. Next, replacing $[\mathbf{P}|\mathbf{S}]$ by its multivariate complex Gaussian expression and using Eq. (18) to express $[\mathbf{S}]$ as an SMoG, it comes

$$\hat{\mathbf{s}}_i = \underset{\mathbf{s}_i}{\text{Argmax}}\left(-\beta^{-2}\|\mathbf{p}_i - \mathbf{G}\mathbf{s}_i\|^2 + \ln \text{Diag}\left(\mathop{\mathbb{E}}_{\tau_q^2}\left\{\prod_{j=1}^{N}\left[s_{q,j}|\tau_q^2\right][\tau_q^2]\right\}\right)\right) \tag{E.1}$$

where only functions of the unknown quantity $\hat{\mathbf{s}}_i$ have been kept in the expression to be maximized, the diagonal operator applies with respect to index q and $\mathbb{E}_{\tau_q^2}$ means that the expected value is taken with respect to the random variable τ_q^2. At this stage, it is noticed that if the order of the expectation operator and of the logarithm were reversed in the last term of Eq. (E.1), the cost function would be quadratic in $\hat{\mathbf{s}}_i$ and its optimization would considerably simplify. This is the essence of the EM algorithm, which solves problem (E.1) by iterating between the following two steps [44].

E step: Calculate $Q(\mathbf{S};\mathbf{S}_{[k-1]}) = \mathbb{E}_{\tau_q^2}\left\{\ln\left(\prod_{j=1}^{N}\left[s_{q,j}|\tau_q^2\right]\right)|\mathbf{P},\mathbf{S}_{[k-1]}\right\}$ where the expectation is now conditioned to the data $\mathbf{p}_j$ and the previous estimate of the source coefficients $\mathbf{s}_{j,[k-1]}$ for all $j = 1, ..., N$.

M step: Calculate $\mathbf{s}_{i,[k]} = \text{Argmax}_{\mathbf{s}_i}\left(-\beta^{-2}\|\mathbf{p}_i - \mathbf{G}\mathbf{s}_i\|^2 + Q(\mathbf{S};\mathbf{S}_{[k-1]})\right)$.

The E step is further worked out as

$$\begin{aligned}Q(\mathbf{S};\mathbf{S}_{[k-1]}) &= \sum_{j=1}^{N}\mathop{\mathbb{E}}_{\tau_q^2}\left\{\ln\left[s_{q,j}|\tau_q^2\right]|\mathbf{P},\mathbf{S}_{[k-1]}\right\} \\ &= -\mathbb{E}\{\tau_q^{-2}|\mathbf{P},\mathbf{S}_{[k-1]}\}\sum_{j=1}^{N}\frac{|s_{q,j}|^2}{\alpha^2\sigma_{q,0}^2} - \mathbb{E}\{\ln(\tau_q^2)|\mathbf{P},\mathbf{S}_{[k-1]}\} - N\ln(\pi\alpha^2\sigma_{q,0}^2)\end{aligned}$$

where $[s_{q,j}|\tau_q^2] = \mathcal{N}_{\mathbb{C}}(s_{q,j};0,\alpha^2\sigma_{q,0}^2,\tau_q^2)$ has been used (see Eq. (18)). By ignoring all quantities which do not depend on $\mathbf{s}_{j,[k-1]}$ and inverting the sign of the cost function, the EM algorithm takes the simplified form:

E step: Calculate

$$\tau_{q,[k]}^{-2} = \mathbb{E}\{\tau_q^{-2}|\mathbf{P},\mathbf{S}_{[k-1]}\} \tag{E.2}$$

which, in the SMoG, family takes the simplified form $\mathbb{E}\{\tau_q^{-2}|\chi_{q,[k-1]}^2\}$ where conditioning is constricted to the normalized sum of squares $\chi_{q,[k-1]}^2$ defined in Eq. (19) and evaluated at the previous iteration $k-1$.

M step: Calculate

$$\begin{aligned}\mathbf{s}_{i,[k]} &= \text{Argmin}\left(\beta^{-2}\|\mathbf{p}_i - \mathbf{G}\mathbf{s}_i\|^2 + \alpha^{-2}\mathbf{s}_i^H\text{Diag}(\sigma_{q,0}^{-2}\tau_q^2,[k]^{-2})\mathbf{s}_i\right) \\ &= \left(\mathbf{G}^H\mathbf{G} + \eta^2\text{Diag}(\sigma_{q,0}^{-2}\tau_{q,[k]}^{-2})\right)^{-1}\mathbf{G}^H\mathbf{p}_i \qquad \text{(E.3a)} \\ &= \text{Diag}(\sigma_{q,0}^2\tau_{q,[k]}^2)\mathbf{G}^H(\mathbf{G}\text{Diag}(\sigma_{q,0}^2\tau_{q,[k]}^2)\mathbf{G}^H + \eta^2\mathbf{I})^{-1}\mathbf{p}_i. \qquad \text{(E.3b)}\end{aligned}$$

This last equation is recognized as being equivalent to Eq. (9). Eventually, it is found that source coefficients can

be estimated for each snapshot i, independently, yet by using estimates of η^2 and τ_q^2 which are evaluated by using all available snapshots $j = 1, ..., N$.

Let us now consider the situation where the regularization parameter and the source power are unknown. Then, the goal is to estimate them jointly with the source coefficients, viz

$$\left(\hat{\mathbf{s}}_i, \hat{\alpha}^2, \hat{\eta}^2\right) \quad = \quad \underset{\mathbf{s}_i, \alpha^2, \eta^2}{\text{Arg max}} \left[\mathbf{S}, \alpha^2, \eta^2 | \mathbf{P}\right].$$

The maximization of the above cost function can be significantly simplified by alternating between the maximization of $[\mathbf{S}|\alpha^2, \eta^2, \mathbf{P}]$ and $[\alpha^2, \eta^2|\{\mathbf{s}_j, \mathbf{p}_j\}_{j=1}^N]$, which is the principle of the Expectation-Conditional Maximization (ECM) algorithm [44]. The convergence of the ECM is further accelerated by replacing the maximization of $[\alpha^2, \eta^2|\{\mathbf{s}_j, \mathbf{p}_j\}_{j=1}^N]$ by that of its marginal over the source coefficients, that is

$$[\alpha^2, \eta^2|\mathbf{P}] = \int [\alpha^2, \eta^2|\mathbf{P}, \mathbf{S}]\, [\mathbf{S}]\, \mathrm{d}\mathbf{S} \tag{E.4}$$

(where conditioning on $\mathbf{P}$ is eventually recognized to be equivalent to conditioning on $\mathbb{S}_{pp}$). This is known as "collapsing" [53]. This justifies interlacing the Bayesian regularization algorithm described in Appendix Appendix B with the EM steps.

Appendix F. Estimation of the relative intensity

The conditional expectation (E.2) is evaluated as follows

$$\tau_{q,[k]}^{-2} = \int_0^\infty \tau_q^{-2} \left[\tau_q^2 | \mathbf{P}, \mathbf{S}_{[k-1]}\right] \left[\tau_q^2\right] \mathrm{d}\tau_q^2 = \int_0^\infty \tau_q^{-2} \left[\tau_q^2 | \{s_{q,j,[k-1]}\}_{j=1}^N\right] \left[\tau_q^2\right] \mathrm{d}\tau_q^2 \tag{F.1}$$

where it has been recognized that τ_q^2 depends directly and only on $s_{q,j,[k-1]}$. Using Bayes' rule,

$$\tau_{q,[k]}^{-2} = \frac{\int_0^\infty \tau_q^{-2} \left[\{s_{q,j,[k-1]}\}_{j=1}^N | \tau_q^2\right] \left[\tau_q^2\right] \mathrm{d}\tau_q^2}{\int_0^\infty \left[\{s_{q,j,[k-1]}\}_{j=1}^N | \tau_q^2\right] \left[\tau_q^2\right] \mathrm{d}\tau_q^2} = \frac{\int_0^\infty \tau_q^{-2} \prod_{j=1}^N \left[s_{q,j,[k-1]} | \tau_q^2\right] \left[\tau_q^2\right] \mathrm{d}\tau_q^2}{\int_0^\infty \prod_{j=1}^N \left[s_{q,j,[k-1]} | \tau_q^2\right] \left[\tau_q^2\right] \mathrm{d}\tau_q^2}. \tag{F.2}$$

Now, exploiting the fact that

$$\prod_{j=1}^N \left[s_{q,j,[k-1]} | \tau_q^2\right] = \prod_{j=1}^N \mathcal{N}_{\mathbb{C}}\left(s_{q,j,[k-1]}; 0, \alpha^2 \sigma_{0,q}^2 \tau_q^2\right) = \frac{1}{(\alpha^2 \sigma_{0,q}^2 \tau_q^2)^N} \mathrm{e}^{-\frac{\sum_{j=1}^N |s_{q,j,[k-1]}|^2}{\tau_q^2 \alpha^2 \sigma_{0,q}^2}} \tag{F.3}$$

in the SMoG model and setting $\chi_{q,[k-1]}^2 = \frac{\sum_{j=1}^N |s_{q,j,[k-1]}|^2}{\alpha^2 \sigma_{0,q}^2}$, one finds

$$\tau_{q,[k]}^{-2} = -\frac{\partial}{\partial(\chi_{q,[k-1]}^2)} \ln\left(\int_0^\infty \prod_{j=1}^N \left[s_{q,j,[k-1]} | \tau_q^2\right] \left[\tau_q^2\right] \mathrm{d}\tau_q^2\right) = -\frac{\partial}{\partial(\chi_{q,[k-1]}^2)} \ln\left(\left[\chi_{q,[k-1]}^2\right]\right). \tag{F.4}$$

Application of the latter formula to the prior PDFs (C.2), (C.5), (C.8) and (C.10) returns Eqs. (G.1), (G.2), (G.6) and (G.9), respectively. Incidentally, the above result also proves that $\tau_{q,[k]}^{-2}$ depends on $\chi_{q,[k-1]}^2$ only, which justifies the contracted notation $\tau_{q,[k]}^{-2} = \mathbb{E}\{\tau_q^{-2}|\chi_{q,[k-1]}^2\}$ instead of the full notation of (E.2).

Appendix G. Updating rules for the aperture function

This appendix provides closed-form expressions for $\hat{\tau}_{q,[k]}^2$ in the case of the prior PDFs introduced in section 4.2 (see Eqs. (C.2), (C.5), (C.8) and (C.10)).

- **Multivariate complex Student-*t* (MCS).** The expected relative intensity related to the MCS prior PDF is easily found by application of Eq. (24) to Eq. (C.2):

$$\hat{\tau}^2_{q,[k]} = \frac{\chi^2_{q,[k-1]} + b}{N + a}. \tag{G.1}$$

As seen in Fig. 11(c-d), the estimated relative intensity $\hat{\tau}^2_{q,[k]}$ is a linearly increasing function of the normalized sum of squares $\chi^2_{q,[k-1]}$, which means that source coefficients with high values are progressively emphasized by the updating rule (25) in the IBF algorithm. At convergence, only the coefficients with largest moduli will survive. Very sparse solutions are expected when small values are quickly attenuated, which requires b to be small, and large values are amplified, which requires a to be small as well.

- **Multivariate complex K-distribution (MCK).** Using the properties of the modified Bessel functions, the expected relative intensity related to the MCK is found as

$$\hat{\tau}^2_{q,[k]} = \frac{\chi^2_{q,[k-1]}}{N-a}\left(1 + \frac{\sqrt{b\chi^2_{q,[k-1]}}}{N-a}\frac{K_{N-a-1}\left(2\sqrt{b\chi^2_{q,[k-1]}}\right)}{K_{N-a}\left(2\sqrt{b\chi^2_{q,[k-1]}}\right)}\right)^{-1}, \quad N > a, \tag{G.2}$$

with asymptotic behavior

$$\hat{\tau}^2_{q,[k]} \sim \frac{\chi^2_{q,[k-1]}}{N-a}, \quad N > a, \quad \text{for small} \quad \chi^2_{q,[k-1]} \tag{G.3}$$

and

$$\hat{\tau}^2_{q,[k]} \sim \frac{\chi^2_{q,[k-1]}}{N-a}\left(1 + \frac{\sqrt{b\chi^2_{q,[k-1]}}}{N-a}\right)^{-1}, \quad N > a, \quad \text{for large} \quad \chi^2_{q,[k-1]}. \tag{G.4}$$

This is a crescent function of the normalized sum of squares $\chi^2_{q,[k-1]}$, yet with a nonlinear behavior intermediate between the linear and the square root case. It is noteworthy that for large $\chi^2_{q,[k-1]}$ the MCK prior behaves like the MCL given hereafter (see Eq. (G.6)).

In practice, the numerical evaluation of the modified Bessel functions may be troublesome at high orders and for small values of $\chi^2_{q,[k-1]}$. A good approximation is proposed for large orders $N-a$ is (see Appendix Appendix H)

$$\hat{\tau}^2_{q,[k]} \approx \frac{\chi^2_{q,[k-1]}}{N-a}\left(1 + \frac{\sqrt{b\chi^2_{q,[k-1]}}}{N-a}\left(1 - \mathrm{e}^{-\frac{\sqrt{b\chi^2_{q,[k-1]}}}{N-a-1}}\right)\right)^{-1}, \quad N > a+1. \tag{G.5}$$

- **Generalized multivariate complex Gaussian (GMCG).** The expected relative intensity related to the GMCG is readily found as

$$\hat{\tau}^2_{q,[k]} = \frac{2}{p}\left(\chi^2_{q,[k-1]}\right)^{1-\frac{p}{2}}, \quad 0 < p \leq 2. \tag{G.6}$$

This is again a monotonic power function of $\chi^2_{q,[k-1]}$, where high sparsity can be reached by setting p small. The special case $p = 1$ returns a square root behavior similar to the asymptote of the MCK. One definite advantage of the GMCG is to involve only one parameter p.

The updating rule of the GMCG is particularly intuitive since it reads, in logarithmic scale,

$$\begin{aligned} \ln \sigma^2_{q,[k]} &= \ln\left(\sigma^2_{q,0}\right) + \ln\left(\hat{\tau}^2_{q,[k]}\right) \\ &= \frac{p}{2}\ln\left(\sigma^2_{q,0}\right) + \left(1 - \frac{p}{2}\right)\ln\left(S_{qq,[k-1]}\right) + C \end{aligned} \tag{G.7}$$

where $C \doteq \ln(2/p) + (1-p/2)\ln\left(N/\alpha^2\right)$ is a global scaling that does not depend on the local position $\mathbf{r}_q$. Therefore, the updating is interpreted as the weighted geometric mean between the initial AF $\sigma^2_{q,0}$ and the previous estimation of the source power spectrum $S_{qq,[k-1]}$. The value of p specifies the balance between the two terms. For $p = 1$, uniform weights of value $1/2$ are assigned to the two terms, whereas for $0 < p < 1$, more weight is given to the data-dependent term.

- **Multivariate complex Laplacian (MCL).** Since the multivariate complex Laplacian is a particular case of the GMCG with $p = 1$, the expected relative intensity is

$$\hat{\tau}^2_{q,[k]} = 2\sqrt{\chi^2_{q,[k-1]}} \tag{G.8}$$

with updating rule $\ln\sigma^2_{q,[k]} = \frac{1}{2}\ln\left(\sigma^2_{q,0}\right) + \frac{1}{2}\ln\left(S_{qq,[k-1]}\right) + C$. This is found to generalize the Bayesian LASSO [50] to the multivariate complex case.

- **Mixture of multivariate complex Gaussians (MMCG).** According to Eq. (C.10),

$$\hat{\tau}_q^{-2} = \frac{p_1 \cdot \bar{\tau}_1^{-2} + p_2 \cdot \bar{\tau}_2^{-2}}{p_1 + p_2} \tag{G.9}$$

where $p_1 \doteq \bar{\pi} \cdot \mathcal{N}_{\mathbb{C},N}(\chi^2_{q,[k-1]}; 0, \bar{\tau}_1^2)$ and $p_2 \doteq (1-\bar{\pi}) \cdot \mathcal{N}_{\mathbb{C},N}(\chi^2_{q,[k-1]}; 0, \bar{\tau}_2^2)$. Equation (G.9) has an insightful interpretation: $\hat{\tau}_q^{-2}$ is obtained as the weighted averaged of $\bar{\tau}_1^{-2}$ and $\bar{\tau}_2^{-2}$ where weights p_1 and p_2 reflect the likelihood of the data $\chi^2_{q,[k-1]}$ under a Gaussian model with relative intensities $\bar{\tau}_1$ and $\bar{\tau}_2$, respectively.

The MMCG case stands out from the other cases by producing a soft threshold, as shown in Fig. 11(c-d). Values of $\chi^2_{q,[k-1]}$ below a certain quantity γ_c are attenuated by a factor nearly equal to $\bar{\tau}_1^2$ whereas values above γ_c are amplified by a factor nearly equal to $\bar{\tau}_2^2$. The value of the threshold may be assessed from the inflection point of $\hat{\tau}^2_{q,[k]}$ as a function of $\chi^2_{q,[k-1]}$, viz

$$\gamma_c = \bar{\tau}_1^2 \cdot \frac{\ln\left(\frac{\bar{\pi}}{1-\bar{\pi}}\right) + N\ln\left(\bar{\tau}_2^2/\bar{\tau}_1^2\right)}{1 - \bar{\tau}_1^2/\bar{\tau}_2^2}, \quad \bar{\tau}_2 > \bar{\tau}_1. \tag{G.10}$$

For large values of N, this tends to a hard threshold where all values below it would be exactly zeroed.

Appendix H. Proof of approximation (G.5)

For $N > a + 1$, the ratio of the modified Bessel functions in Eq. (G.2) is very well fitted by a function of the form $1 - \exp(-c\sqrt{b\chi^2_{q,[k-1]}})$, where coefficient $c = 1/(N-a-1)$ is found so as to match the derivative at $\chi^2_{q,[k-1]} = 0$.

Appendix I. Derivation of the bias – Eq. (29)

Setting $\mathbf{p}_i = \mathbf{G}\mathbf{s}_i$ in Eq. (20) and using the matrix inversion lemma, one has

$$\begin{aligned} \mathbf{s}_i - \mathbb{E}\{\hat{\mathbf{s}}_i\} &= \left(\mathbf{I} - \Sigma^2\mathbf{G}^H(\mathbf{G}\Sigma^2\mathbf{G}^H + \eta^2\mathbf{I})^{-1}\mathbf{G}\right)\mathbf{s}_i \\ &= \left(\mathbf{I} - (\mathbf{G}^H\mathbf{G} + \eta^2\Sigma^{-2})^{-1}\mathbf{G}^H\mathbf{G}\right)\mathbf{s}_i \end{aligned} \tag{I.1}$$

with Σ^2 as given in Eq. (22). Now, setting $\mathbf{I} = (\mathbf{G}^H\mathbf{G} + \eta^2\Sigma^{-2})^{-1}(\mathbf{G}^H\mathbf{G} + \eta^2\Sigma^{-2})$ in the former equality yields Eq. (29)

Appendix J. Missing information matrix

In accordance with Ref. [44], let define the *observed information matrix* of the incomplete data (i.e. without introducing τ_q^2),

$$\mathbf{I}_i = -\frac{\partial^2}{\partial \mathbf{s}_i \partial \mathbf{s}_i^H} \ln\left[\mathbf{S}|\mathbf{P}\right], \tag{J.1}$$

and the expected information matrix of the complete data (i.e. marginalized over $\tau = \{\tau_q^2, q = 1, .., Q\}$)

$$\mathcal{I}_{c,i} = -\mathop{\mathbb{E}}_{\tau^2}\left\{\frac{\partial^2}{\partial \mathbf{s}_i \partial \mathbf{s}_i^H} \ln\left[\mathbf{S}|\mathbf{P}, \tau^2\right]|\mathbf{P}, \mathbf{S}\right\}, \tag{J.2}$$

both conditioned to the observations $\mathbf{P}$ and evaluated for snapshot i. The missing information matrix is then defined as

$$\mathcal{I}_{m,i} = \mathcal{I}_{c,i} - \mathbf{I}_i, \tag{J.3}$$

which measures information deficiency when the relative intensity τ_q^2 is not directly observed.

As explained in Ref. [44], an estimate of the covariance matrix of the estimated source coefficients is returned by the inverse of the observed information matrix $\mathbf{I}_i = \mathcal{I}_{c,i} - \mathcal{I}_{m,i}$ averaged over all snapshots, whereas the speed of convergence of the EM algorithm is controlled by the largest eigenvalue of $\mathbf{J}_i = \mathcal{I}_{c,i}^{-1}\mathcal{I}_{m,i}$, both evaluated at the MAP solution $\hat{\mathbf{s}}_i$. In both cases, the missing information matrix is seen to act as the main leverage and its norm should therefore be minimized in order to achieve good convergence of the estimates.

The missing information matrix is calculated hereafter from Eq. (J.2) after evaluating the expressions for $\mathcal{I}_{c,i}$ and $\mathbf{I}_i$. Since $\left[\mathbf{S}|\mathbf{P}, \tau_q^2\right] \propto [\mathbf{P}|\mathbf{S}]\left[\mathbf{S}|\tau_q^2\right]$ where $[\mathbf{P}|\mathbf{S}]$ and $\left[\mathbf{S}|\tau_q^2\right]$ are complex Gaussians PDFs as given by Eqs. (7) and (8), it readily comes that

$$\mathcal{I}_{c,i} = \beta^{-2}\mathbf{G}^H\mathbf{G} + \alpha^{-2}\mathrm{Diag}\left(\sigma_{0,q}^{-2}\mathbb{E}\{\tau_q^{-2}|\mathbf{S}\}\right) \tag{J.4}$$

where $\mathbb{E}\{\tau_q^{-2}|\mathbf{S}\} = -\frac{\partial}{\partial(\chi_q^2)}\ln\left[\chi_q^2\right]$ from Eq. (K.4). Similarly, since $[\mathbf{S}|\mathbf{P}] \propto [\mathbf{P}|\mathbf{S}]\,[\mathbf{S}]$,

$$\mathbf{I}_i = \beta^{-2}\mathbf{G}^H\mathbf{G} - \frac{\partial^2}{\partial \mathbf{s}_i \partial \mathbf{s}_i^H}\ln\left[\mathbf{S}\right] = \beta^{-2}\mathbf{G}^H\mathbf{G} - \alpha^{-2}\mathrm{Diag}\left(\sigma_{0,q}^{-2}\left(\frac{\partial}{\partial(\chi_q^2)} + \frac{|s_{q,i}|^2}{\alpha^2\sigma_{0,q}^2}\frac{\partial^2}{(\partial\chi_q^2)^2}\right)\ln\left[\chi_q^2\right]\right) \tag{J.5}$$

where $[\mathbf{S}]$ is given by Eq. (18) and the chain rule, $\frac{\partial}{\partial s_{q,i}} = \frac{s_{q,i}^*}{\alpha^2\sigma_{0,q}^2}\frac{\partial}{\partial(\chi_q^2)}$ with χ_q^2 as defined in Eq. (23), has been used. Therefore,

$$\mathcal{I}_{m,i} = \mathrm{Diag}\left(\frac{|s_{q,i}|^2}{(\alpha^2\sigma_{0,q}^2)^2}\frac{\partial^2}{(\partial\chi_q^2)^2}\ln\left[\chi_q^2\right]\right), \tag{J.6}$$

whose average over snapshot $i = 1, ..., N$ and evaluation at the MAP solution returns the first equality in (30). Now, considering that $[\chi_q^2] \propto \int \mathrm{e}^{-\chi_q^2/\tau_q^2}[\tau_q^2]\mathrm{d}\tau_q^2$, it can be checked that

$$\frac{\partial^2}{(\partial\chi_q^2)^2}\ln\left[\chi_q^2\right] = \int_0^\infty \tau_q^{-4}\left[\tau_q^2|\{s_{q,j,[k-1]}\}_{j=1}^N\right]\left[\tau_q^2\right]\mathrm{d}\tau_q^2 - \left(\int_0^\infty \tau_q^{-4}\left[\tau_q^2|\{s_{q,j,[k-1]}\}_{j=1}^N\right]\left[\tau_q^2\right]\mathrm{d}\tau_q^2\right)^2 = \mathrm{Var}\left\{\tau_q^{-2}|\mathbf{S}\right\}, \tag{J.7}$$

which proves the second equality of Eq. (30).

The above results lead to the following particular cases:

- The diagonal elements of the information matrix of the MCS prior are upper bounded provided that $b > 0$:

$$[\mathcal{I}_m]_{qq}^{\mathrm{MCS}} = \frac{1}{\alpha^2\sigma_{0,q}^2}\frac{(N+a)}{N}\frac{\chi_q^2}{(\chi_q^2+b)^2} \leq \frac{1}{\alpha^2\sigma_{0,q}^2}\frac{(N+a)}{N}. \tag{J.8}$$

- The information matrix of the MMCG prior, $[\mathcal{I}_m]_{qq}^{\mathrm{MMCG}}$, is found upper bounded by $\frac{1}{\alpha^2\sigma_{0,q}^2}\frac{\gamma_c'}{4\bar{\tau}_1^2}\left(\frac{\bar{\tau}_1^2}{\bar{\tau}_2^2}-1\right)^2$, a maximum located at $\chi_q^2=\gamma_c'=\bar{\tau}_1^2\ln(\frac{1-\bar{\pi}}{\bar{\pi}}(\frac{\bar{\tau}_1^2}{\bar{\tau}_2^2})^N)$.

- The information matrix of the GMCG prior with $p<2$ (which includes the MCL) is seen to diverge at $\chi_q^2=0$ since

$$[\mathcal{I}_m]_{qq}^{\mathrm{GMCG}}=\frac{1}{\alpha^2\sigma_{0,q}^2}=\frac{p(1-p/2)}{2N}\frac{1}{(\chi_q^2)^{1-p/2}}. \tag{J.9}$$

- The MCK has unbounded variance,

$$[\mathcal{I}_m]_{qq}^{\mathrm{MCK}}\sim\frac{1}{\alpha^2\sigma_{0,q}^2}\frac{N-a}{N}\frac{1}{\chi_q^2},\quad N>a,\quad \text{for small}\quad \chi_{q,[k-1]}^2, \tag{J.10}$$

  which diverges at $\chi_q^2=0$.

Appendix K. Broadband Iterated Bayesian focusing (BIBF)

The essential difference in BIBF is to assume that the AF is not a function of frequency. Therefore, the relative intensity τ_q^2 should be estimated once for all frequencies. The equivalent of Eq. (F.1) then becomes

$$\tau_{q,[k]}^{-2}=\int_0^\infty\tau_q^{-2}\left[\tau_q^2|\{s_{q,i,[k-1]}(f_l);i=1,...,N,l=1,...,B\}\right]\left[\tau_q^2\right]\mathrm{d}\tau_q^2 \tag{K.1}$$

(with explicit dependence of the source coefficients $s_{q,i,[k-1]}(f_l)$ on frequency f_l). Using Bayes' rule, this gives

$$\tau_{q,[k]}^{-2}=\frac{\int_0^\infty\tau_q^{-2}\prod_{i=1}^N\prod_{l=1}^B\left[s_{q,i,[k-1](f_l)}|\tau_q^2\right]\left[\tau_q^2\right]\mathrm{d}\tau_q^2}{\int_0^\infty\prod_{i=1}^N\prod_{l=1}^B\left[s_{q,i,[k-1](f_l)}|\tau_q^2\right]\left[\tau_q^2\right]\mathrm{d}\tau_q^2} \tag{K.2}$$

where independence of the source coefficients have been assumed with respect to frequencies (which holds true under the assumption of a (time) stationary sound field). Since

$$\prod_{i=1}^N\prod_{l=1}^B\left[s_{q,i,[k-1](f_l)}|\tau_q^2\right]=\frac{1}{(\alpha^2\sigma_{0,q}^2\tau_q^2)^{NB}}\mathrm{e}^{-\frac{\sum_{l=1}^B\sum_{i=1}^N|s_{q,i,[k-1](f_l)}|^2}{\tau_q^2\alpha^2\sigma_{0,q}^2}} \tag{K.3}$$

or, equivalently, $\left[\overline{\chi^2}_{q,[k-1]}\right]\propto\mathrm{e}^{-\overline{\chi^2}_{q,[k-1]}}$ with $\overline{\chi^2}_{q,[k-1]}$ as defined in Eq. (31), one eventually finds

$$\tau_{q,[k]}^{-2}=-\frac{\partial}{\partial(\overline{\chi^2}_{q,[k-1]})}\ln\left(\left[\overline{\chi^2}_{q,[k-1]}\right]\right). \tag{K.4}$$

References

[1] J. L. Starck, M. J. Fadili, An overview of inverse problem regularization using sparsity, Proceedings - International Conference on Image Processing, ICIP (2009) 1453–1456.

[2] J. J. Fuchs, Linear programming in spectral estimation. application to array processing, in: 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Vol. 6, 1996, pp. 3161–3164 vol. 6.

[3] D. Malioutov, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. on Signal Processing 53 (2005) 3010–3022.

[4] A. C. Gurbuz, J. H. McClellan, V. Cevher, A compressive beamforming method, in: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp. 2617–2620.

[5] G. F. Edelmann, C. F. Gaumond, Beamforming using compressive sensing, The Journal of the Acoustical Society of America 130 (4) (2011) EL232–EL237. doi:10.1121/1.3632046.
[6] P. Simard, J. Antoni, Acoustic source identification: Experimenting the ℓ_1 minimization approach, Applied Acoustics 74 (7) (2013) 974 – 986.
[7] S. Zhong, Q. Wei, X. Huang, Compressive sensing beamforming based on covariance for acoustic imaging with noisy measurements, The Journal of the Acoustical Society of America 134 (5) (2013) EL445–EL451.
[8] A. Xenaki, P. Gerstoft, K. Mosegaard, Compressive beamforming, The Journal of the Acoustical Society of America 136 (1) (2014) 260–271.
[9] P. Gerstoft, A. Xenaki, C. F. Mecklenbruker, Multiple and single snapshot compressive beamforming, The Journal of the Acoustical Society of America 138 (4) (2015) 2003–2014.
[10] A. Xenaki, P. Gerstoft, Grid-free compressive beamforming, The Journal of the Acoustical Society of America 137 (4) (2015) 1923–1935.
[11] T. F. Brooks, W. M. Humphreys, A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays, Journal of Sound and Vibration 294 (4) (2006) 856–879. doi:10.1016/j.jsv.2005.12.046.
[12] T. Yardibi, J. Li, P. Stoica, L. N. C. III, Sparsity constrained deconvolution approaches for acoustic source mapping, The Journal of the Acoustical Society of America 123 (5) (2008) 2631–2642. doi:10.1121/1.2896754.
[13] P. Sijtsma, CLEAN Based on Spatial Source Coherence, International Journal of Aeroacoustics 6 (4) (2007) 357–374. doi:10.1260/147547207783359459.
URL http://journals.sagepub.com/doi/10.1260/147547207783359459
[14] Z. Chu, Y. Yang, Comparison of deconvolution methods for the visualization of acoustic sources based on cross-spectral imaging function beamforming, Mechanical Systems and Signal Processing 48 (1) (2014) 404 – 422.
[15] T. Yardibi, J. Li, P. Stoica, N. S. Zawodny, L. N. C. III, A covariance fitting approach for correlated acoustic source mapping, The Journal of the Acoustical Society of America 127 (5) (2010) 2920–2931.
[16] Z. Yang, L. Xie, C. Zhang, A discretization-free sparse and parametric approach for linear array signal processing, IEEE Transactions on Signal Processing 62 (19) (2014) 4959–4973.
[17] N. Chu, A. Mohammad-Djafari, J. Picheral, Robust Bayesian super-resolution approach via sparsity enforcing a priori for near-field aeroacoustic source imaging, Journal of Sound and Vibration 332 (18) (2013) 4369–4389.
[18] N. Chu, J. Picheral, A. Mohammad-Djafari, N. Gac, A robust super-resolution approach with sparsity constraint in acoustic imaging, Applied Acoustics 76 (2014) 197–208.
[19] T. Suzuki, L1 generalized inverse beam-forming algorithm resolving coherent/incoherent, distributed and multipole sources, Journal of Sound and Vibration 330 (24) (2011) 5835 – 5851.
[20] P. Zavala, W. D. Roeck, K. Janssens, J. Arruda, P. Sas, W. Desmet, Generalized inverse beamforming with optimized regularization strategy, Mechanical Systems and Signal Processing 25 (3) (2011) 928 – 939.
[21] A. Peillot, G. Chardon, L. Daudet, Acoustic sources joint localization and characterization using compressive sampling to cite this version : Acoustic sources joint localization and characterization using compressive sampling, Proceedings of Acoustics2012, Nantes.
[22] G. Chardon, L. Daudet, A. Peillot, F. Ollivier, N. Bertin, R. Gribonval, Near-field acoustic holography using sparse regularization and compressive sampling principles, The Journal of the Acoustical Society of America 132 (3) (2012) 1521–1534.
[23] J. Hald, Fast wideband acoustical holography, The Journal of the Acoustical Society of America 139 (4) (2016) 1508–1517.
[24] E. Fernandez-Grande, A. Xenaki, P. Gerstoft, A sparse equivalent source method for near-field acoustic holography, The Journal of the Acoustical Society of America 141 (1) (2017) 532–542.
[25] N. M. Abusag, D. J. Chappell, Near-field acoustic holography using the method of superposition, Journal of Computational Acoustics 24 (2016) 1–22.
[26] C.-X. Bi, Y. Liu, L. Xu, Y.-B. Zhang, Sound field reconstruction using compressed modal equivalent point source method, The Journal of the Acoustical Society of America 141 (1) (2017) 73–79.
[27] Z. H.-C. Nie Yong-Fa, Acoustic field reconstruction using source strength density acoustic radiation modes, Acta Physica Sinica 63 (10) (2014) 104303.
[28] A. Canclini, M. Varini, F. Antonacci, A. Sarti, Dictionary-based equivalent source method for near-field acoustic holography, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 166–170.
[29] A. Pereira, Acoustic imaging in closed spaces, Ph.D. thesis, University of Lyon, 2013.
[30] B. Oudompheng, A. Pereira, C. Picard, B. Nicolas, A theoretical and experimental comparison of the iterative equivalent source method and the generalized inverse beamforming, Proceedings of the 5th Berlin Beamforming Conference (2014) 1–15.
[31] Q. Leclère, A. Pereira, C. Bailly, J. Antoni, C. Picard, A unified formalism for acoustic imaging based on microphone array measurements, International Journal of Aeroacoustics 16 (4-5) (2017) 431–456.
[32] P. C. Hansen, Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998.
[33] G. H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (2) (1979) 215–223.
[34] J. Antoni, A bayesian approach to sound source reconstruction: Optimal basis, regularization, and focusing, The Journal of the Acoustical Society of America 131 (4) (2012) 2873–2890.
[35] A. Pereira, J. Antoni, Q. Leclère, Empirical bayesian regularization of the inverse acoustic problem, Applied Acoustics 97 (Supplement C) (2015) 11 – 29.
[36] J. Huang, T. Zhang, The benefit of group sparsity, Ann. Statist. 38 (4) (2010) 1978–2004. doi:10.1214/09-AOS778.
[37] D. R. Brillinger, Time Series: Data Analysis and Theory, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.
[38] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Transactions on Signal Processing 56 (6) (2008) 2346–2356.
[39] S. D. Babacan, R. Molina, A. K. Katsaggelos, Bayesian compressive sensing using laplace priors, IEEE Transactions on Image Processing 19 (1) (2010) 53–63.
[40] I. F. Gorodnitsky, B. D. Rao, Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm, IEEE

Transactions on Signal Processing 45 (3) (1997) 600–616.
[41] F. Champagnat, J. Idier, A connection between half-quadratic criteria and EM algorithms, IEEE Signal Processing Letters 11 (9) (2004) 709–712.
[42] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin, Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science), Chapman and Hall/CRC, London, 2014.
[43] M. Aucejo, O. De Smet, On a full Bayesian inference for force reconstruction problems, Mechanical Systems and Signal Processing 104 (2018) 36–59.
[44] G. McLachlan, T. Krishnan, The EM algorithm and extensions, 2nd Edition, Wiley series in probability and statistics, Wiley, Hoboken, NJ, 2008.
[45] Y. Wang, J. Li, P. Stoica, M. Sheplak, T. Nishida, Wideband relax and wideband clean for aeroacoustic imaging, The Journal of the Acoustical Society of America 115 (2) (2004) 757–767.
[46] K. L. Gemba, S. Nannuru, P. Gerstoft, W. S. Hodgkiss, Multi-frequency sparse bayesian learning for robust matched field processing, The Journal of the Acoustical Society of America 141 (5) (2017) 3411–3420.
[47] A. Pereira, Acoustic imaging in enclosed spaces, Ph.D. thesis, University of Lyon, France (2013).
[48] A. Pereira, Q. Leclère, J. Antoni, A theoretical and experimental comparison of the equivalent source method and a bayesian approach to noise source identification, Proceedings of Berlin Beamforming Conference, Berlin 2012 (2012) 1–12.
[49] T. Eltoft, T. Kim, T.-W. Lee, Multivariate Scale Mixture of Gaussians Modeling, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 799–806.
[50] T. Park, G. Casella, The Bayesian Lasso, Journal of the American Statistical Association 103 (482) (2008) 681–686.
[51] C. Hans, Bayesian lasso regression, Biometrika 96 (4) (2009) 835–845.
[52] M. Novey, T. Adali, A. Roy, A complex generalized gaussian distribution – characterization, generation, and estimation, IEEE Transactions on Signal Processing 58 (3) (2010) 1427–1433.
[53] D. A. van Dyk, T. Park, Partially collapsed gibbs samplers, Journal of the American Statistical Association 103 (482) (2008) 790–796.