Iterative beamforming for identification of multiple broadband sound sources

Xun Wang\textsuperscript{a,b,*}, Benjamin Quost\textsuperscript{a}, Jean-Daniel Chazot\textsuperscript{b}, Jérôme Antoni\textsuperscript{c}

\textsuperscript{a} Sorbonne University, Université de Technologie de Compiègne, UMR CNRS 7253 Heudiasyc, 60205 Compiègne cedex, France
\textsuperscript{b} Sorbonne University, Université de Technologie de Compiègne, UMR CNRS 7337 Roberval, 60203 Compiègne cedex, France
\textsuperscript{c} University of Lyon, Laboratory of Vibrations and Acoustics, 69621 Villeurbanne cedex, France

\begin{abstract}
The reconstruction of broadband sound sources is an important issue in industrial acoustics. In this paper, a model comprising multiple incoherent Gaussian random sources is considered. The aim is to estimate locations and powers of the sound sources using the pressures measured by an array of microphones. Each measured pressure is interpreted as a mixture of latent signals emitted by different sound sources. Then, an Iterative Beamforming (IB) method is developed to estimate the source parameters. This approach is based on the Expectation-Maximization (EM) algorithm, a well-known iterative procedure for solving maximum likelihood parameter estimation. More specifically, IB iteratively estimates the source contributions and performs beamforming on these estimates. In this work, experiments on real data illustrate the advantage of IB with respect to classical beamforming and Near-field Acoustical Holography (NAH). In particular, the proposed method is shown to work over a wider range of frequencies, to better estimate the source locations, and is able to quantify the powers of the sources. Furthermore, experiments illustrate that IB can not only localize the sources on a given surface, but also accurately estimate their 3D locations.
\end{abstract}

1. Introduction

Noise reduction of complex devices is an important issue in acoustics engineering. For example, in the automobile industry, acoustic quality is now accepted as a decisive criterion to judge the performance of a vehicle, due to requirements in terms of acoustic comfort expressed by the users, as well as the legislation related to noise pollution. For this purpose, the first step consists in modeling the acoustic behavior of the device in order to focus on the main sources of noise and the most annoying frequency bands. However, the noise sources emitted by a machine are multiple and structurally complex.

In order to describe the behaviour of some assemblies of the device, various numerical methods, e.g., finite element method or boundary element method, can be employed. However, it is still difficult to accurately model the complexity of the full structure over a wide frequency range. Some acoustic imaging tools can be used to give a quick overview of the acoustic radiation of complex devices. Beamforming\textsuperscript{[1–4]} and Near-field Acoustical Holography (NAH)\textsuperscript{[5–10]} are probably the most widely used methods dedicated to the aforementioned issue. Both are based on the sound pressures measured by an array of microphones.

\* Corresponding author at: Sorbonne University, Université de Technologie de Compiègne, UMR CNRS 7253 Heudiasyc, 60205 Compiègne cedex, France.
\textit{E-mail addresses:} xun.wang@hds.utc.fr, xunwang00@gmail.com (X. Wang), benjamin.quost@utc.fr (B. Quost), jean-daniel.chazot@utc.fr (J.-D. Chazot), jerome.anti@insa-lyon.fr (J. Antoni).
microphones. NAH back-propagates the sound field over a surface near the sound sources. By taking evanescent waves into account, NAH ensures a high spatial resolution, but it is still limited in frequency due to the distance between the microphones. In its classical formulation, it is also limited to regular arrays of microphones. A variant of NAH, called Statistically Optimized NAH (SONAH) [7–9], can overcome some of these limitations. The pressure field is expanded over the same plane wave basis in both methods, but the analytical formalism of SONAH avoids the errors caused by the discrete spatial Fourier transform used in NAH. However, both classical and statistically optimized NAH reconstruct the sound field on a specific surface and cannot estimate all three coordinates in a 3D space. Beamforming can estimate the direction of arrival of the plane waves in the far-field case, or the point source location in the near-field case, by maximizing the Delay-and-Sum beamformer. However, beamforming is still limited in its frequency range and minimum resolvable source separation. In order to alleviate these limitations, some variants of beamforming [11,12] have been introduced. Besides, classical Tikhonov regularization [13] can be used to deal with the corresponding ill-posed inverse problem. However, beamforming is always based on the single source assumption, and may thus be inappropriate in the case of multiple sources from the statistical point of view. Ref. [14] presented these different acoustic imaging methods in a unified approach. Based on a Bayesian framework with prior information about the source distribution, a super-resolution of the reconstructed sound pressure field is obtained.

In this paper, the multiple sound source estimation problem is solved in a statistical framework. The pressure measured by a microphone is interpreted as a mixture of signals emitted by the various sources. In previous works [15,16], the authors proposed a model estimation method for deterministic signals (i.e., the source amplitudes are assumed as parameters of the model). However, in real world situations, deterministic signals are generally not adequate to model broadband noise. In particular, Gaussian noise, whose amplitude is assumed to be a zero-mean Gaussian random variable, is a realistic assumption in many industrial applications. Therefore, this paper considers the estimation problem of Gaussian random sound sources, which consists in estimating the locations and powers of the sound sources. The power of a zero-mean Gaussian sound source is equivalent to the variance of its random amplitude. In the case of a single source, estimating the parameters by Maximum Likelihood (ML) is proved to be equivalent to beamforming [2]. For multiple sources, however, maximizing the likelihood function of the parameters given the observed pressures is computationally difficult, since the contributions of the various sources to a measured signal are unknown. Clearly, should these contributions be known, computing the Maximum Likelihood Estimates (MLE) would be straightforward as in the single source case. Feder and Weinstein [17–19] investigated the parameter estimation of superimposed signals. More specifically, they proposed to introduce latent variables representing the (unknown) contributions of the sources to the measured pressures. Then, the MLE of the model parameters were estimated via the EM algorithm [20,21]. This procedure alternates iteratively between two steps. The first one consists in computing the expected source contributions given the current fits of the model parameters. Consequently, the model parameters are updated according to the new source contribution estimates. In Ref. [17], two deterministic plane waves with specific angles of arrival were considered and the propagation operator was taken without attenuation. Kabaoglu et al. [22,23] used a similar approach to study the localization of near-field sources. These two contributions solved the problem in 2D and 3D space, respectively. However, the near-field was just taken into account with the Fresnel approximation of the source-microphone distance in the time delay, without the 1/r spatial decrease. Furthermore, in their works, the powers of the sources were not quantified in the estimation process. Refs. [24,25] investigated the multiple deterministic source localization problem via acoustic energy measurement using the EM algorithm. In their models, only the energy of sound wave was considered, but without regard for the phase. In previous works [15,16], the authors used the EM algorithm to estimate the multiple sound source in the case of deterministic signals.

This paper solves the multiple sound source identification problem for random Gaussian amplitude sources, in which both sound wave phase and power attenuation are considered. The proposed approach is formalized in a statistical framework: more precisely, it is derived from ML via the EM algorithm. The resulting procedure consists in iteratively estimating the sound source contributions and applying beamforming on these estimates. For this reason, the proposed approach is referred to as iterative beamforming (IB). This approach bears a similarity with the deconvolution methods such as CLEAN [26] and DAMAS [27], in which the estimated sources are iteratively removed to improve the estimation of the remaining sources. However, one major difference is that the proposed IB approach operates on the spectral matrix of the measured pressures (i.e., on the beamformer input), rather than the energetic map returned by the beamformer output. As such, it is expected to be unbiased and benefits from all the optimal properties of the MLE. Compared to the classical methods, IB makes it possible to address a wider frequency range and to separate close sources. Furthermore, the sound sources may not only be localized, but also precisely quantified.

The organization of this paper is as follows. Section 2 begins with a description of the model. Then, IB is introduced as an iterative procedure to compute MLE of the model parameters (locations and powers of the sound source). The link between the proposed method and classical beamforming is also illustrated. Section 3 presents experimental results on real data, by which IB is compared with classical beamforming and NAH. First, experiments with three sources in a given plane demonstrate that the proposed algorithm can work over a wider frequency range than the other two methods and can separate very close sources. Then, an experiment considering coherent sound sources is introduced. Finally, IB is justified that it can accurately estimate the 3D locations of sound sources with different distances to the microphone plane. Based on the studies in this paper, conclusions are drawn in Section 4.
2. Sound source estimation via iterative beamforming

This section first presents the sound propagation model considered in the paper. Beamforming, which makes it possible to estimate the location of a single source, is also briefly recalled. Then, the proposed strategy for estimating multiple sound sources based on the EM algorithm, referred to as iterative beamforming, is presented.

2.1. Model

Fig. 1 displays the sound source estimation problem considered in this paper. Assume $S$ incoherent broadband sound sources with coordinates $\mathbf{r}_s, s = 1, \ldots, S$; the radiated sound is measured by $M$ microphones with coordinates $\mathbf{r}_0, m = 1, \ldots, M$. The measured sound pressures in the time domain can be modeled as stochastic processes, then transformed to the frequency domain to produce a number of snapshots. The $M$-dimensional vector of measured pressures $\mathbf{p}_t = (p_{1t}, \ldots, p_{Mt})^T$ at the frequency $f$ corresponding to the $t$-th snapshot ($t = 1, \ldots, T$) is obtained by:

$$\mathbf{p}_t = \mathbf{G}(\mathbf{r})\mathbf{A}_t + \mathbf{n}_t.$$  

(1)

Here, the $M$-by-$S$ matrix $\mathbf{G}$ of general term $g_{ms}$ describes the free-field sound propagation process from the $s$-th source to the $m$-th microphone:

$$g_{ms} = G(r_s, r_m) = \frac{e^{ik||\mathbf{r}_s - \mathbf{r}_m||}}{4\pi||\mathbf{r}_s - \mathbf{r}_m||}.$$  

(2)

where $k = \frac{2\pi f}{v}$ is the wavenumber and $v$ is the sound speed. The vector $\mathbf{A}_t = (A_{1t}, \ldots, A_{St})^T$ contains the uncorrelated random amplitudes of the (incoherent) sound sources: it is assumed to follow an $S$-dimensional complex-valued Gaussian distribution with $0$-mean and diagonal covariance matrix

$$\text{Cov}(\mathbf{A}_t) = \mathbb{E}(\mathbf{A}_t\mathbf{A}_t^H) = \text{diag}(\mathbf{a}^2).$$  

(3)

where $\mathbf{a}^2 = (a_1^2, \ldots, a_S^2)^T$ is vector of powers (variances of amplitudes) of the sound sources and $\text{diag}(\mathbf{x})$ denotes the diagonal matrix with the elements of $\mathbf{x}$ on the main diagonal. The measurement noise vector $\mathbf{n}_t = (n_{1t}, \ldots, n_{Mt})^T$ is also assumed to be
complex Gaussian distributed with 0-mean and covariance matrix $\Sigma = \sigma^2 I_M$. It is reasonable to further assume that $A_t$ and $n_t$ are independent for any two snapshots $t_1, t_2 \in [1, ..., T]$. Therefore, the sound pressure $p_t$ measured by a microphone array follows a complex-valued multivariate Gaussian distribution with 0-mean and covariance matrix

$$
2 \Sigma_{p}(r, a^2) = E \left( G A_t A_t^H G^H \right) + E(n_t n_t^H) = G \text{ diag}(a^2) G^H + \Sigma.
$$

(4)

The purpose of this work is to use the sound pressures measured by the $M$ microphones to estimate the $S$ sources. Therefore, this is an inverse statistical estimation problem, in which the parameters to estimate are the locations $r_t$ and the powers $a_t^2 (s=1, ..., S)$ of the sound sources.

2.2. Beamforming and maximum likelihood

The classical (or Bartlett) beamforming technique is based on the assumption that the signals are emitted by a single source. The beamformer power output of measurement $p_t$ is defined as

$$
|B_t|^2 = |w^H p_t|^2 = w^H p_t p_t^H w.
$$

(5)

The idea behind each beamformer is to “steer the microphone array” to search for the source position. The steering vector $w$ is obtained by maximizing the beamformer (5) which gives the source location estimate [2]:

$$
\hat{r} = \arg \max_{r} \prod_{t=1}^{T} \frac{G_t^H p_t p_t^H G_t}{\Sigma_t^H}.
$$

(6)

Note that Eq. (6) is also a MLE of the source position.

In the case of multiple sources, classical beamforming solves the problem in the same way as previously; a single source is assumed and the beamformer output is computed for all the points in the considered region. In the two-source case, for example, a secondary source is related to a secondary local maximum of the beamformer. However, in this case classical beamforming is not a parametric approach and thus has a limited spatial resolution, i.e., the sources cannot be separated when they are close to each other. In the rest of Section 2, a generalized beamforming method addressing the case of multiple sources is proposed, based on the multiple-source (parametric) assumption and ML principle.

2.3. Latent source contributions

Since the sound pressures $p = (p_1, ..., p_T)$ follows a complex-valued Gaussian distribution [28], its log-likelihood function can be expressed according to the source locations $r$ and the powers $a^2 = (a_1^2, ..., a_T^2)^T$ as follows:

$$
L(r, a^2|p) = -\log(\det(\Sigma_p)) - \frac{1}{T} \sum_{t=1}^{T} p_t^H \Sigma_p^{-1} p_t
$$

$$
= -\log(\det(\Sigma_p)) - \text{tr} \left( \Sigma_p^{-1} \hat{\Sigma}_p \right),
$$

(7)

where $\hat{\Sigma}_p$ is the sample covariance matrix defined by

$$
\hat{\Sigma}_p = \frac{1}{T} \sum_{t=1}^{T} p_t p_t^H,
$$

(8)

and where det($Q$) and tr($Q$) respectively stand for the determinant and trace of the matrix $Q$. The MLE of $r$ and $a^2$ can be obtained by maximizing Eq. (7). Note that maximizing the log-likelihood Eq. (7) yields the same estimates of the sound source location than beamforming Eq. (6) in the case of a single source [2].

When multiple sources are in presence ($S \geq 2$), computing the MLE requires to solve a computationally expensive optimization problem. Indeed, $4S$ parameters must be estimated at the same time (for each source, the 3D coordinates of $r_t$ and the power $a_t^2$). However, by introducing latent variables, this optimization problem can be largely simplified. A latent variable can be defined as an unknown information that would make the estimation process straightforward, should it be available. In this problem, the latent variables are assumed to be the contributions $c_t$ of the various sound sources to the measured pressures. Note that the sound propagation model detailed in Eq. (1) can be written in the form of superimposed signals:

$$
p_t = \sum_{s=1}^{S} G_s(r_t) A_{st} + n_t,
$$

(9)

in which $G_s$ stands for the $s$-th column of the matrix $G$ and $G_s A_{st}$ represents the contributions from the $s$-th source to the measured pressures. Then, source contributions $c_t = (c_{t1}, ..., c_{tS})$ to the $t$-th snapshot are introduced, related to the model parameters by

$$
c_{st} = G_s A_{st} + n_{st}, s = 1, ..., S.
$$

(10)

Here the noise component $n_{st}$ can be obtained by arbitrarily decomposing the total noise $n_t$ into $S$ components, i.e.,
\[ \sum_{i=1}^{S} n_{it} = n_t. \] Thus, the measurement \( p_i \) is related to the latent source contributions \( c_{it} \) by

\[ p_i = \sum_{s=1}^{S} c_{it}. \quad (11) \]

By assuming further that \( n_{it} \) are mutually independent Gaussian random vectors with 0-mean and covariance matrix \( \Sigma_s = \frac{1}{2} I_t \), the source contribution vector \( c_{it} \) is Gaussian with 0-mean and covariance matrix

\[ \Sigma_{c_{it}}(n_{it}, \sigma^2_t) = E(c_{it}c_{it}^T) = E\left(\left|A_{it}\right|^2 G_{hi}^T\right) + E(n_{it}n_{it}^T) = \alpha^2_t G_{hi}^T + \sigma^2 I_t. \quad (12) \]

Then, the log-likelihood function of the model parameters given the latent source contributions \( c = (c_t, t = 1, \ldots, T) \) can be written as

\[ L(r, a^2 | c) = -\sum_{s=1}^{S} \log(\text{det}(\Sigma_c)) \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{S} c_{it}^T \Sigma_{c_{it}}^{-1} c_{it} \]

\[ = -\sum_{s=1}^{S} \log(\text{det}(\Sigma_c)) - \sum_{s=1}^{S} \text{tr}\left(\Sigma_{c_{it}}^{-1} \Sigma_{c_{it}}\right). \quad (13) \]

where \( \Sigma_{c_{it}} \) is the sample covariance matrix of the source contribution \( c_{it} \):

\[ \Sigma_{c_{it}} = \frac{1}{T} \sum_{t=1}^{T} c_{it} c_{it}^T. \quad (14) \]

Obviously, if the contributions of the sound sources to the measured pressures were available, they could be used to estimate each source separately from the others by solving a single source localization problem.

### 2.4. Maximum likelihood estimation of multiple sound sources

Unfortunately, the complete log-likelihood (13) cannot be exploited directly, since the latent contributions \( c_{it} \) are generally unknown. However, the EM algorithm [20] makes it possible to proceed with Eq. (13) by treating the missing data as random variables. Hereafter, this algorithm is applied to the multiple source estimation problem.

Let \( c = (c_1, \ldots, c_T) \) be the complete data (here, the latent contributions of the sources) and \( p \) be the realization of the incomplete data (here, the observed sound pressures). The proposed method can maximize the observed data log-likelihood with respect to the parameters \( \phi = (r, a^2) \) by proceeding iteratively with the expected complete log-likelihood given the incomplete data. This latter is defined as the expectation of the complete data log-likelihood with respect to the latent variables, computed using a current fit for the parameter vector \( \phi \). More precisely, the algorithm starts from an initial parameter vector \( \phi^0 \) and iterates back and forth between two steps:

- the Expectation step (or E-step) consists in computing the conditional expectation \( Q(\phi^t | \phi) = \mathbb{E}(L(r, a^2 | c) | p, \phi^t) \) of the log-likelihood with respect to the latent variables \( c \), given the observed incomplete data \( p \) and a current fit \( \phi^t \) for the model parameters;
- in the Maximization step (or M-step), new fits for the parameters are determined by maximizing \( Q(\phi^t | \phi) \) with respect to \( \phi \).

The algorithm guarantees that the observed data likelihood \( L(r, a^2 | p) \) increases at each iteration [20]. However, whether it converges to the global maximum depends on the choice of the initial parameter \( \phi^0 \) [21]. Therefore, it is well-advised to initialize the parameters with different values that cover a large part of the parameter space, to compute the corresponding estimates, and to finally retain the result associated with the highest likelihood.

**E-step.** The E-step amounts to computing

\[ Q(\phi | \phi^t) = \mathbb{E}\left(L(r, a^2 | c) | p, \phi^t, (a^2)^t\right) \]

\[ = -\sum_{s=1}^{S} \log(\text{det}(\Sigma_c)) - \sum_{s=1}^{S} \text{tr}\left(\Sigma_{c_{it}}^{-1} e_{it}^t\right). \quad (15) \]

where \( e_{it} = \mathbb{E}(\hat{\Sigma}_c \dot{c}_{it} | p, r^t, (a^2)^t) \) is the conditional expectation of the sample covariance matrix of the latent contribution \( c_{it} \).

Note that the joint vector \( (\dot{c}_{it}) \) is Gaussian with 0-mean and covariance matrix \( \left(\Sigma_p, \Sigma_p \Sigma_{c_{it}}^{-1} \Sigma_{c_{it}} \right) \). Therefore, the conditional distribution of \( c_{it} \) given \( p_t \) is Gaussian [29]:

\[ c_{it} | p_t \sim \mathcal{N}\left(\dot{c}_{it} | \Sigma_p, \Sigma_{c_{it}} \right). \quad (16) \]

Hence,

\[ e_{it}^t = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(c_{it}c_{it}^T | p_t, r^t, (a^2)^t) = \Sigma_{c_{it}} - \Sigma_{c_{it}} \left(\Sigma_p^{-1} \Sigma_{c_{it}} \right)^{-1} \Sigma_p \left(\Sigma_{c_{it}} \left(\Sigma_p^{-1} \Sigma_{c_{it}} \right)^{-1}\right)^H, \quad (17) \]
in which \( \Sigma_c \) and \( \Sigma_p \) are computed using the current fits for the model parameters:

\[
\begin{align*}
\Sigma_c &= (a_c^2)^{\frac{1}{2}} G_c(r_c) G_c(r_c)^\dagger + \sigma^2 I_M, \\
\Sigma_p &= G(r) \text{diag}((a_1^2)^{\frac{1}{2}}, \ldots, (a_N^2)^{\frac{1}{2}}) G(r)^\dagger + \sigma^2 I_M.
\end{align*}
\] (18)

**M-step.** The M-step amounts to computing new estimates of the model parameters by solving, for \( s = 1, \ldots, S \),

\[
(r_{s}^{l+1}, (a_s^2)^{l+1}) = \arg\min_{r, \sigma^2} \left( \log(\det(\Sigma_c)) + \text{tr}(\Sigma_c^{-1} e_s^t) \right).
\] (19)

Eq. (19) can be simplified such that the estimates of the source location and amplitude variance are computed separately:

\[
r_{s}^{l+1} = \arg\max_{r_s} \frac{G_s^t e_s^t G_s}{\|G_s\|^2} 
\] (20)

and

\[
(a_s^2)^{l+1} = \frac{(G_s^{l+1})^t e_s^t G_s^{l+1}}{\|G_s^{l+1}\|^4} - \frac{\sigma^2}{S} \frac{1}{\|G_s^{l+1}\|^2},
\] (21)

where \( G_s^{l+1} = G_s(r_s^{l+1}) \). The derivation of Eqs. (20) and (21) can be found in Appendix A. Note that Eq. (20) is a 3-parameter optimization problem, which is much easier to solve than directly maximizing the log-likelihood function of the measured pressures Eq. (7) (a 4S-parameter optimization problem).

By rewriting Eq. (20) as

\[
r_{s}^{l+1} = \arg\max_{r_s} 1 \sum_{t=1}^{T} \frac{G_s^t e_s^t p_t^t, (a_s^2)^{l+1} G_s}{G_s^t G_s},
\] (22)

and comparing it with Eq. (6), it is clear that the source location estimate computed at each iteration is equivalent to performing a beamforming projection for each latent source \( c_s \). Beside, by multiplying both sides of Eq. (21) with \( \|G_s^{l+1}\|^2 \), it is noticeable that each source power estimate is obtained from the fact that the sound power propagated from the source to the microphone array \((a_s^2)^{l+1} \|G_s^{l+1}\|^2\) is equal to the beamformer power output of the corresponding contribution of measurement minus the power of the measurement error \((e_s)^2\). At each iteration, each source is first estimated via beamforming by using the corresponding source contribution, and then the contributions are revised according to the estimated source parameters. For this reason, the proposed method is referred to as iterative beamforming (IB). The parametric mechanism of IB achieves a high spatial resolution: the separation between two sources is not limited by the main-lobe width and the presence of side-lobes of the beamformer. Furthermore, it should be stressed that the IB strategy makes it possible not only to localize the multiple sound sources but also to quantify their powers (via Eq. (21)). The IB strategy for estimating the model parameters is summarized in **Algorithm 1**.

**Algorithm 1.** Estimation of random sound sources using IB.

For \( l = 0 \), pick starting values \( r^0, (a^2)^0 \) for the model parameters. For \( l \geq 1 \):

repeat

estimate the covariance matrix \( e_s \) of the latent source contributions by Eq. (17), for \( s = 1, \ldots, S \);

estimate the new source location \( r_s^{l+1} \), for \( s = 1, \ldots, S \), by Eq. (20);

estimate the new source power \( (a_s^2)^{l+1} \), for \( s = 1, \ldots, S \), by Eq. (21);

until the relative increase of the observed data log-likelihood (7) is less than a given threshold \( \kappa \):

\[
\frac{\log L_{IB}^{(l+1)} - \log L_{IB}^{(l)}}{\log L_{IB}^{(l)}} \leq \kappa.
\]

Let \( T_{IB} \) and \( T_{BF} \) be the computation times of IB and classical beamforming; the former can be roughly estimated by \( T_{IB} \approx S \times I^* \times T_{BF} \), where \( I^* \) represents the number of iterations. Here, only the optimization step Eq. (20) is considered: other steps may be reasonably ignored in terms of computation time. Since classical beamforming can return a result almost instantly, the computational cost of IB is also not expensive.

3. **Experiments**

This section illustrates the behavior of the proposed method on experimental data. First, an experiment with three incoherent sound sources located in the same plane is conducted and the proposed method is systematically compared with classical beamforming and NAH. Then, the performances of these three methods are presented for coherent sound sources. Finally, two incoherent sources with different distances to the microphone plane are considered and three \((x, y, z)\) coordinates of each source are estimated.
3.1. Planar source identification

First, an experiment with three incoherent sources in the plane \( z = 0.37 \text{ m} \) is proposed. The experimental setup and the distribution of the microphones are shown in Fig. 2. A microphone array is placed in the plane \( z = 0 \) and centered at the origin. Three sound sources are placed at \( r_1 = (-0.16, 0.03, 0.37) \), \( r_2 = (-0.16, -0.16, 0.37) \) and \( r_3 = (0.26, -0.11, 0.37) \). In this experiment, independent random signals with zero-mean complex-valued Gaussian distributed amplitudes are played by the loudspeakers during 60 seconds. The sound pressures are recorded by the 60 microphones and the sampling frequency is 16384 Hz. Each measured signal is divided into 239 segments with a duration of 0.5 second and an overlap of 0.25 second (0.25 seconds of the signal are shared by two adjacent segments). Then, each segment is transformed to the frequency domain using a Discrete Fourier Transform with a Hanning window, so that 239 snapshots in the frequency domain are obtained. Here, the estimation results at three different frequencies \( f = 700 \text{ Hz}, 2000 \text{ Hz} \) and \( 5500 \text{ Hz} \) are shown. Actually, the far-field is defined by a microphone-source distance greater than the Fraunhofer distance \( d_F = \frac{\pi}{2D} \), where \( D \) is the dimension of the array and \( \lambda \) is the wavelength. In this experimental setup \( D \approx 1 \), so that \( d_F \) is respectively 6.4 m, 18.3 m, and 50 m for the above three frequencies. Therefore, all the experiments are run in the near-field (more exactly in the Fresnel region).

The estimated sound sources obtained via IB are compared with those obtained via classical beamforming and NAH. The computation results are displayed in Figs. 3–5 for the three different frequencies. In all the three methods, the \( z \)-coordinates of the sources are assumed to be known. Since the loudspeakers are not point sources but vibrating membranes, estimating a model with three point sources may not yield accurate estimates of the radiated pressures. A higher number of sources may be more efficient to model these radiating surfaces, for instance with a layer of monopoles or dipoles. Therefore, different numbers of sources are tested in IB (\( S = 3, 6 \) and \( 12 \)).

Fig. 3 (a) and (b) show the reconstructed sound pressure fields obtained via classical beamforming and NAH at \( f = 700 \text{ Hz} \). The estimated sound pressure levels in the source plane are shown in the figures (the reference sound pressure is \( 2 \times 10^{-5} \text{ Pa} \)). The grid spacing of the both methods is 0.02 m, which is much smaller than the wavelength. At this low frequency, both methods can separate the source from the two others, but cannot separate the two sources close to each other. Fig. 3 (c-e) present the results obtained with the proposed IB method on the same configuration. In this case, the initial sources are obtained via classical beamforming estimates as follows. The two local maxima of Fig. 3 (a) are taken as two initial sources, located at \( \hat{r}_1 = (-0.18, -0.05, 0.37) \) and \( \hat{r}_2 = (0.25, -0.1, 0.37) \). The estimated powers of these sources

---

1 In this section, the spatial coordinates are in meter (m), except where noted.
2 Actually, all the frequencies mentioned in this section represent the centers of narrow bands with a 2 Hz bandwidth.
3 In the experiments of this section, the NAH results are obtained from the statistically optimized approach (SONAH) [7] with a Bayesian regularization [14].
are then obtained with the least square (LS) solution of Eq. (1):

$$\langle \hat{a}_1^2, \hat{a}_2^2 \rangle = \frac{1}{T} \sum_{t=1}^{T} \left| (\hat{G}^T \hat{G})^{-1} G(\hat{r}) p_t \right|^2,$$

(24)
Fig. 5. Sound source estimates using classical beamforming, NAH and IB at $f = 5500$ Hz. Black crosses represent the loudspeaker centers. Subfigures (a), (b), (d), (f) and (h): reconstructed sound pressure levels (the reference pressure is $2 \times 10^{-5}$ Pa) in the source plane using classical beamforming, NAH and IB estimates with $S = 3, 6, 12$ respectively. Subfigures (c), (e) and (g): estimated source positions by IB with $S = 3, 6$ and 12 respectively. (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)
where \( \mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2) \). Then, when the number of sources \( S \) chosen in the IB method is different from the number of initial sources estimated via classical beamforming, it is necessary to divide the \( S \) sources between the estimated sources according to their intensity or to their span. In the present case, the first source seems more spread out, so \( \frac{S}{2} \) of the initial sources are taken at the first location \( \mathbf{r}_1 \) with a power \( a^1 \), while the other \( \frac{S}{2} \) initial sources are placed at the second position \( \mathbf{r}_2 \) with a power \( a^2 \). The final estimated sources with the IB approach are the estimates of their positions \( \mathbf{r}_i^* \) and powers \( a_i^* \), \( s = 1, \ldots, S \), and the sound pressures are then reconstructed from these estimates at some discrete points \( \mathbf{r} \neq \mathbf{r}_i^* \) in the source plane with

\[
p^*(\mathbf{r}) = \left( \sum_{i=1}^{S} a_i^* |G(\mathbf{r}; \mathbf{r}_i^*)|^2 \right)^{\frac{1}{2}}.
\]

This reconstructed sound field presented in Fig. 3 (c-e) for different source numbers \( S = 3, 6, 12 \) shows that, contrary to classical beamforming and NAH, the proposed method can clearly separate the three sources at the low frequency.

Fig. 4 shows the corresponding results at 2000 Hz. In this case, the initialization of IB is similarly achieved the three local maxima returned by classical beamforming (in Fig. 4 (a)) are used as initial values. The results demonstrate that at this medium frequency all three methods can accurately identify the three sources.

Finally, an experiment at a high frequency \( (f = 5500 \text{ Hz}) \) is performed. Fig. 5 (a) and (b) show that both classical beamforming and NAH perform poorly. In this case, it is not possible to obtain the initial sources from the beamforming estimates. The initial values of IB are therefore randomly generated. For each source, its initial location is generated according to a 2D uniform distribution having a 60 cm \( \times \) 60 cm support centered on each loudspeaker. The initial powers are taken at random according to a uniform distribution, whose support centers at the LS solution (obtained as in Eq. (24)) and has a range of 6 dB in sound power level. As mentioned in Section 2.4, in order to guarantee that the likelihood function converges to the global maximum, the estimation procedure detailed in Algorithm 1 is run 100 times with initial parameter values selected at random Fig. 5 (c) displays the 100 IB estimates in the given z-plane using blue points and the retained estimate (having the highest likelihood value) using red crosses for \( S = 3 \). In this case, different initial values generate different IB estimates. The source location estimates characterized by the highest likelihood value are accurate approximations to the actual ones. Subfigures (e) and (g) display the retained estimated sources positions (with a maximum likelihood among the 100 candidates) and their estimated random amplitude standard deviations for the cases with \( S = 6 \) and \( S = 12 \) sources, respectively. Subfigures (d), (f) and (h) show the sound field reconstructed in the source plane with the IB estimates in the cases \( S = 3, S = 6 \) and \( S = 12 \) respectively. In the two latter cases \((S = 6 \text{ and } S = 12)\), satisfactory results are obviously obtained even when the number of assumed point sources is greater than the actual number of sources (IB might need several point sources to properly represent each actual loudspeaker membrane). On the other hand, since the loudspeaker diameter (around 10 cm) is not dramatically larger than the wavelength in this particular case, three point sources already model the three loudspeakers satisfactorily; for example, the nine other sources in Fig. 5 (g) can be neglected compared to the three most important sources with high power. This demonstrates the stability of the method with respect to the number of sources.

The super-resolution obtained with IB is quite noticeable compared to classical beamforming. Under the assumption of far-field and uniform array (with constant inter-element spacing \( d \)), the upper frequency limit of beamforming [1] is returned by \( f_{\text{max}} = \frac{c}{2d} \). In the experimental setup of this work, it is around 2000 Hz when \( d \) is taken as the average value of inter-element spacing (8.6 cm). Since the array is not uniform but irregular, the upper frequency limit can be pushed up. Meanwhile, the grating lobes of the beamformer are expected to produce local maxima in the likelihood function which are

---

4 Obviously, a large number of initial values can increase the probability of reaching the global maximum of the likelihood, but the computational cost has to be considered as well. Our experiments showed that choosing 100 initial values makes it possible to achieve good results at a reasonable computational cost.
slightly lower than the global maximum associated with the main lobe, as illustrated in Fig. 6. This figure displays the point source function for the three considered frequencies \( f = 700 \text{ Hz}, 2000 \text{ Hz} \) and \( 5500 \text{ Hz} \), which are obtained from the beamformer output for a single source at \((0.0, 0.37)\). At a high frequency, the strategy based on several random initializations of the parameters has been devised to help finding the global maximum. Furthermore, since the IB approach is parametric, the grating lobes never explicitly appear in the final results (Table 1).

The source powers have also been characterized separately according to the norm ISO 9614-2 (Intensity Scanning⁵). Table 2 presents the sound power levels measured for each loudspeaker with this Intensity Scanning method and the same sound power levels estimated via the IB method (power levels of \( a_{fi}^{2} \) in dB) at the three previous frequencies. Note that the dB levels in Table 2 and Figs. 3–5 refer to different quantities: the former is the sound power level of each source and the latter corresponds to the sound pressure levels reconstructed in the source plane (obtained from Eq. (25)). Table 2 shows that the proposed IB method is also efficient to quantify the source power except at very high frequencies where the accuracy seems to decrease. In fact, the Intensity Scanning measures the total source power from all the directions, while the estimated source powers are computed according to the measurements only made in the front plane. The assumption behind the proposed model and IB method is that the source is omnidirectional. At a high frequency, however, this property vanishes [30], such that the IB method overestimates the source power by assuming that the monopole source radiates the same power through other faces as through the front face. By the above discussion, the estimation efficiency of source power can be summed up as follows: i) IB is able to accurately estimate the power of each source radiated over the measurement surface (the front face); ii) it is able to accurately estimate the total power when the sources are omnidirectional (the cases of 700 Hz and 2000 Hz in this experiment); iii) it would be possible to accurately evaluate the power of a directive source (the case of 5500 Hz in this experiment) if its directivity could be known or the measurements in other directions were also available.

Different microphone-source distances are now tested. Three sound sources are placed at \( r_1 = (-0.16, -0.03, z) \), \( r_2 = (-0.16, 0.16, z) \) and \( r_3 = (0.26, -0.11, z) \), with a microphone-source distance \( z \) equal to 0.14 m or 1.07 m. Fig. 7 displays the estimation results at 2000 Hz using classical beamforming, NAH and IB with an assumed source number \( S = 12 \), respectively. Subfigures (a)–(c) present the results when \( z = 0.14 \) m, which show that classical beamforming cannot work in this case. By contrast, subfigures (d)–(f) display the case \( z = 1.07 \) m, in which NAH fails to separate the two close sources. However, the proposed IB method gives good results in both cases.

The model (1) assumes that the sound sources are incoherent. Two coherent sources are now tested to see the performance of the proposed method in this case. In this experiment, two loudspeakers emit the same signal at 1000 Hz. The loudspeakers are placed at \( r_1 = (-0.16, 0.03, z) \) and \( r_2 = (0.26, -0.11, z) \). Fig. 8 displays the experimental results for two different microphone-source distance \( z = 0.1 \) m and \( z = 0.7 \) m. Each figure shows the reconstructed sound field in the source plane using classical beamforming, NAH or IB. Although the proposed method is developed under the assumption of incoherent sources, these figures show that it still works with coherent sources contrary to classical beamforming that fails to separate two coherent sources. This result therefore increases the application range of the proposed IB method.

### 3.2. 3D source localization

In this section, an experiment with two sources having different \( z \)-coordinates is considered. The two sources are located at \( r_1 = (-0.16, 0.03, 0.58) \) and \( r_2 = (0.26, -0.11, 0.32) \); both radiate at a frequency \( f = 1000 \text{ Hz} \). In this experiment, all the three coordinates of both sources are unknown.

Classical beamforming and NAH are based on a predefined \( z \)-plane, so that it is impossible to localize the sources in a 3D space. However, the proposed IB method still works: all the three coordinates can be estimated. Fig. 9 (a) displays on the left side the 3D estimation results using IB, in which the actual (black) and estimated (red) \( z \)-coordinate of the sources are labeled on the \( z \)-axis. The right side shows the projection of the actual and the estimated source locations in the microphone

---

⁵ The sound power of each source (the other sources being switched off) is measured over five faces of the loudspeaker (except the bottom face which is considered as totally reflecting). Then, the power is integrated over the five planes to obtain the total sound power level of the source.
plane ($z=0$). Both figures illustrate that the proposed method can accurately localize the sources in different planes with unknown $z$-coordinates. On the other hand, classical beamforming and NAH estimates based on five different planes with $z_1=0.2$ m, $z_2=0.32$ m (the $z$-plane of $r_2$), $z_3=0.45$ m, $z_4=0.58$ m (the $z$-plane of $r_1$) and $z_5=0.7$ m are shown. The results are
presented in Fig. 9 (b) and (c) which clearly show that both methods are able to estimate the x- and y-coordinates of the sources in some planes, but cannot estimate the distances to the microphone plane.

4. Conclusions and perspectives

This paper addresses the problem of estimating multiple broadband sound sources from sound pressures measured by an array of microphones. The source locations and powers are estimated via maximum likelihood. The parameter estimation problem can be simplified by introducing latent variables that capture the contributions of the various sources to the
measured pressures. Then, applying the EM algorithm results in an iterative procedure in which the source contributions and the model parameters are alternatively estimated.

At each step, the proposed approach gives an estimate of each source location from the current estimated source contributions. This step is equivalent to performing a classical beamforming. For this reason, the procedure is referred to as iterative beamforming. Furthermore, besides the well-known properties guaranteed by maximum likelihood estimation, the procedure exhibits two major advantages when it is compared to classical beamforming: it incorporates the estimation of the source contributions, and also makes it possible to accurately quantify the powers of the sound sources.

The results obtained on experimental data show the interest of the proposed method. It works indeed over a wider range of frequency than classical beamforming and NAH. The accuracy of the sound source quantification and the ability to separate very close sources are also demonstrated. Moreover, an experiment with two sources at different $z$-coordinates illustrates that the proposed method is able to accurately localize sources in the 3D space, while classical beamforming and NAH can hardly estimate the distances between the sources and the microphone array. Finally, the IB method still works even when the assumption of incoherent sources is not satisfied.

Future works may be conducted in several directions. The modification of the method to take into account coherent sources with a non-diagonal covariance matrix in Eq. (3) would be a natural extension of the current IB, but this would be to the detriment of the computational cost. A deeper study of coherent sources would therefore be interesting to evaluate the limits of the current method. It would be also interesting to automatically determine the optimal number of sources in real life applications.

**Acknowledgements**

This work was partially funded by the European Union committed in Picardy with the FEDER. It was carried out in the framework of the Labex MS2T, which was funded by the French Government, through the program "Investments for the future" by the National Agency for Research (Reference ANR-11-IDEX-0004-02).

The authors would also like to thank Félix Focial and Jean-Marc Gherbezza for their helps in the experimental part.

**Appendix A. Derivation of Eqs. (20) and (21)**

In order to compute $\det(\Sigma_c)$ in Eq. (19), the eigenvalue decomposition of $\Sigma_c$ is obtained by solving the equation $\det(A_M - \Sigma_c) = 0$. One eigenvalue of $\Sigma_c$ is $\left(\frac{\sigma^2}{S} + a_s^2 \|G_s\|^2\right)$ with eigenvector $G_s$. The remaining $M - 1$ eigenvalues are all $\frac{\sigma^2}{S}$ and the corresponding $M - 1$ eigenvectors can be chosen as a basis in the orthogonal complement of the space $G_s$ and obtained by solving the $(M - 1)$-dimensional linear equation system $G_s x = 0$. Therefore, there exists an $M$-by-$M$ invertible matrix $Q$ of eigenvectors such that

$$
\Sigma_c = Q \text{diag}\left(\frac{\sigma^2}{S} + a_s^2 \|G_s\|^2, \frac{\sigma^2}{S}, ..., \frac{\sigma^2}{S}\right) Q^{-1},
$$

which gives

$$
\det(\Sigma_c) = \left(\frac{\sigma^2}{S} + a_s^2 \|G_s\|^2\right)^{M-1} \left(\frac{\sigma^2}{S}\right)^{M-1}.
$$

Furthermore, by the Matrix Inversion Lemma,

$$
\Sigma_c^{-1} = \frac{S}{\sigma^2 A_M} - \frac{S}{\sigma^2 \frac{\sigma^2}{S} + a_s^2 \|G_s\|^2} a_s G_s G_s^H.
$$

Therefore, by Eqs. (27) and (28), (19) can be written as

$$
\begin{align*}
\begin{bmatrix} r_s \end{bmatrix}^{t+1} & = \begin{bmatrix} a_s \end{bmatrix}^{t+1} \\
& = \arg \min_{r_s, a_s} \left\{ \log\left(\frac{\sigma^2}{S} + a_s^2 \|G_s\|^2\right) - \frac{S}{\sigma^2 \frac{\sigma^2}{S} + a_s^2 \|G_s\|^2} \right\}.
\end{align*}
$$

In order to further simplify Eq. (29), the following function of $a_s^2$ is defined:

$$
g(a_s^2) = \log\left(\frac{\sigma^2}{S} + a_s^2 \|G_s\|^2\right) - \frac{S}{\sigma^2 \frac{\sigma^2}{S} + a_s^2 \|G_s\|^2}.
$$

Its first and second derivatives can be computed:

$$
g'(a_s^2) = \frac{\|G_s\|^2 \left(\frac{\sigma^2}{S} + a_s^2 \|G_s\|^2\right) - G_s^H e_i e_i G_s}{\left(\frac{\sigma^2}{S} + a_s^2 \|G_s\|^2\right)^2}
$$
and
\[ g' (a^2_s) = \frac{\| G_s \|^4 \left( \frac{a^2_s + a^2_i}{\| G_s \|^2} \right)^2 - 2 \left( \frac{a^2_s + a^2_i}{\| G_s \|^2} \right) \| G_s \|^2 \left( \frac{a^2_s + a^2_i}{\| G_s \|^2} \right) - G_s^l e_i G_i}{(\frac{a^2_s + a^2_i}{\| G_s \|^2})^4}. \] (32)

The unique stationary point of \( g(a^2_s) \), denoted as \( a^2_{s*} \), is obtained by solving the equation \( g'(a^2_s) = 0 \):
\[ a^2_{s*} = \frac{G_s^l e_i G_i}{\| G_s \|^4} - \frac{\| G_s \|^2}{ \| G_s \|^2} \frac{1}{S}. \] (33)

Since
\[ g' (a^2_s) = \frac{\| G_s \|^4 \left( \frac{a^2_s + a^2_i}{\| G_s \|^2} \right)^2 - 2 \left( \frac{a^2_s + a^2_i}{\| G_s \|^2} \right) \| G_s \|^2 \left( \frac{a^2_s + a^2_i}{\| G_s \|^2} \right) - G_s^l e_i G_i}{(\frac{a^2_s + a^2_i}{\| G_s \|^2})^4} > 0, \] (34)

the stationary point \( a^2_{s*} \) is proved to be the minimum of \( g(a^2_s) \). Therefore, given any source location estimate \( r_s \), the estimate of the source amplitude variance is computed as
\[ (a^2_{s*})^{-1} = \frac{G_s^l e_i G_i}{\| G_s \|^4} - \frac{\| G_s \|^2}{ \| G_s \|^2} \frac{1}{S}. \] (35)

By substituting Eq. (35) back into Eq. (29), the source location estimate is then obtained by
\[ r_{s*} = \arg \min \left\{ \log \left( \frac{G_s^l e_i G_i}{\| G_s \|^2} \right) - \frac{S}{\| G_s \|^2} \frac{G_i^l e_i G_i}{\| G_s \|^2} \right\}. \] (36)

Since \( h(x) = \log x - \frac{S}{\| G_s \|^2} x \) is a monotonically decreasing function in \( \{x: x > \frac{S}{\| G_s \|^2} \} \) and \( \frac{G_i^l e_i G_i}{\| G_s \|^2} > \frac{S}{\| G_s \|^2} \) (Eq. (33) is positive), Eq. (36) finally reduces to
\[ r_{s*} = \arg \max \left\{ \frac{G_i^l e_i G_i}{\| G_s \|^2} \right\}. \] (37)

References


