GEOMETRICALLY CONSTRAINT ICA FOR CONVOLUTIVE MIXTURES OF SOUND

Mirko Knaak [†‡] *Shoko Araki* [†] *Shoji Makino* [†]

[†] NTT Communication Science Laboratories, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
e-mail: Mirko.Knaak@tu-berlin.de

[‡] Technical University Berlin, Measurement Technology Lab
Einsteinufer 19, 10587 Berlin, Germany

ABSTRACT

The goal of this contribution is a new algorithm using independent component analysis with a geometrical constraint. The new algorithm solves the permutation problem of blind source separation of acoustic mixtures, and it is significantly less sensitive to the precision of the geometrical constraint than an adaptive beamformer. A high degree of robustness is very important since the steering vector is always roughly estimated in the reverberant environment, even when the look direction is precise. The new algorithm is based on FastICA and constrained optimization. It is theoretically and experimentally analyzed with respect to the roughness of the steering vector estimation by using impulse responses of real room. The effectiveness of the algorithms for real-world mixtures is also shown in the case of three sources and three microphones.

1. INTRODUCTION

For many signal processing tasks, such as speech recognition, transmission, or classification of signals, a very good reconstruction of the target signal is essential when the target signal is disturbed by other sources. Adaptive beamformers (ABF) and blind source separation (BSS) are very effective tools for multichannel signal reconstruction.

Although the utility of ABFs is well established [1], they have limited robustness against erroneous parameters. This is very troublesome since the steering vector is always roughly estimated in a reverberant environment, as is shown in this paper. The methods traditionally used to overcome this sensitivity mostly broaden the directivity pattern, resulting in a trade-off between the signal suppression performance and the parameter sensitivity (e.g., [2]).

Independent component analysis ICA is an emerging technique for finding independent components in a multi-channel signal. The main application is BSS which has been shown to be capable of recovering multiple sources from their linear mixture if the sources are independent [3].

In the field of acoustics, convolutive mixtures need to be separated, which involves estimating of many more parameters (see [4] for contributors) than a separation of a scalar mixture. Most approaches simplify the problem to instantaneous separation problems for the frequency components. The scaling and permutation ambiguities left in the recovered frequency components become a serious problem, particularly when the number of sources and microphones becomes larger than two. Different permutations of the frequency components lead to mixed outputs and degraded separation results. There are several approaches to overcome this problem, however, they are restricted to two sources. Hence, the number of real-world applications in the acoustics field is still very limited, and the separation performance is mostly insufficient.

Current publications indicate an equivalence between ABF and BSS, e.g., [5],[6]. BSS is only an intelligent set of ABFs with an adaptive null directivity aimed in the direction of the unnecessary sounds, which have been employed by [7]. This equivalence suggests an application of a geometrical constraint on ICA to solve the permutation and scaling problem.

Geometrically constrained algorithms have been proposed by [8]-[10]. Their contributions do not assess in detail how a rough estimation of the steering vector effects the performance of the algorithms. They only employ the constraint with the assumption that it is estimated correctly. This assumption is very limiting because precise information about the steering vector is very difficult to obtain. The major advantage of using ICA and geometrical information appears when only a rough estimation is possible. Furthermore, their algorithms are rather slow iteration type algorithms.

This paper proposes a new geometrically constrained ICA algorithm that employs the fast convergence properties of the FastICA algorithm [3]. We also analyze the behavior of geometrically constrained ICA algorithms in general with respect to a rough estimation of the constraint. In Sec. 2, we make basic assumptions on the mixture, ABF and BSS algorithms, and analyze the possible reasons for a rough estimation of the steering vector. The new algorithm is introduced and assessed theoretically and experimentally in Sec. 3.

2. BLIND SOURCE SEPARATION AND ADAPTIVE BEAMFORMERS

2.1. Signals and BSS algorithm

In a set $\mathbf{s}^b(t) = \left[s^b_{target}(t), s^b_{i1}(t), \ldots, s^b_{iN-1}(t)\right]^T$ of broadband sources, the first source is the target sound and the others are interfering sources. The sound is measured with an array of M microphones $\mathbf{x}^b(t) = [x^b_1(t), \ldots, x^b_M(t)]^T$. The observed signals are filtered and mixed because the room acoustics impose a different impulse response h^b_{mn} between each source s^b_n and each microphone x^b_m.

In the frequency domain, the convolutive mixture can

be written as $\mathbf{x}^f = \mathbf{H}^f \cdot \mathbf{s}^f + \mathbf{n}^f$, where $\mathbf{x}^f$ is a narrow-band signal component filtered from $\mathbf{x}^b$ with a band pass centered at f. For simplicity, the index f is omitted hereafter. $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$ consists of the steering vectors $\mathbf{h}_N$, where $\mathbf{h}_1$ is the steering vector of the target sound. Only under anechoic conditions, they can be approximated by the phase shifts caused by the time delays τ_{mn} with $\mathbf{h}_n = [e^{j2\pi f \tau_{1n}}, \ldots, e^{j2\pi f \tau_{Mn}}]^T$. When considering echoes and reverberation, $\mathbf{h}_n$ is the sum of all echo paths.

The goal of the algorithms discussed here is to find an optimal estimation $y_1(t)$ of the target signal s_{target}. To achieve this goal, an unmixing matrix $\mathbf{W}$ or a coefficient vector $\mathbf{w}_1$ is applied to the vector of observations as follows:

$$\mathbf{y} = \mathbf{W} \cdot \mathbf{x} \qquad y_1 = \mathbf{w}_1^H \cdot \mathbf{x}, \tag{1}$$

where $\langle\rangle^H$ is the hermitian (conjugate transposed).

Blind source separation uses ICA to estimate the unmixing matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]^H$ by making the output signals as independent as possible. Essentially, ICA has two steps ($\mathbf{W} = \mathbf{T}^H \cdot \mathbf{V}$). In the first step (sphering), the matrix $\mathbf{V}$ is determined by the principle component analysis (PCA). In the second step (rotation with $\mathbf{T}$), maximization of nongaussianity, nonlinear decorrelation, non-stationary decorrelation, or spatio-temporal decorrelation can be used to determine the rotation matrix $\mathbf{T}$ [3].

2.2. ABF with an imprecise estimation of the steering vector

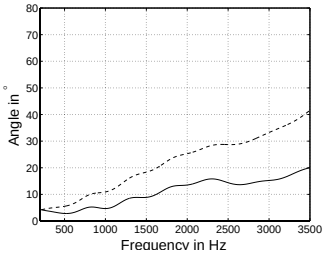
An ABF minimizes the power of the output signal with a constraint: the energy of a signal coming from the direction of the target is passed without changes $\mathbf{w}_1^H \hat{\mathbf{h}}_1 = c_1$. This means that the source position has to be known in advance. It can be estimated by sound localization methods (e.g., MUSIC [1]), determined by image processing, or simply known by geometry.

A major drawback of ABFs is that they rely on the correct estimation of the steering vector. Since the impulse response of a room is normally not available, the steering vector $\hat{\mathbf{h}}_1$ is estimated by time delays of the direct sound only. An estimation error of $\hat{\mathbf{h}}_1$ is caused by two reasons: a wrong position or direction of arrival, or the existence of strong reverberations. The latter reason is due the multiple directions of arrival while only the direct sound is used for the estimation. The estimation becomes rough even when the source position might be well known.

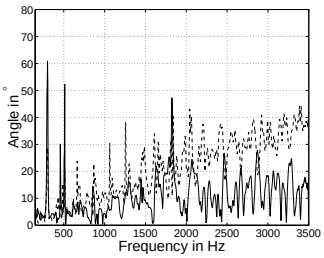
To determine the error made by a rough estimation of $\hat{\mathbf{h}}_1$, we introduce a new measure: the steering vector error angle (SVA) μ. For its definition, we use a generalized cosine $\cos(\mathbf{x}, \mathbf{y}) = ||\mathbf{x}^H \mathbf{y}||/||\mathbf{x}||||\mathbf{y}||$ to define the angle between two complex vectors $\mathbf{x}$ and $\mathbf{y}$. Let $\mu(f)$ be the angle between $\hat{\mathbf{h}}_1(f)$ and $\mathbf{h}_1(f)$ and let $\alpha(f)$ be the angle between the mixing vectors $\mathbf{h}_1(f)$ of the target signal and $\mathbf{h}_2(f)$ of the jammer signal.

$$\mu(f) = \cos^{-1}\left(\frac{|\mathbf{h}_1^H \hat{\mathbf{h}}_1|}{|\mathbf{h}_1||\hat{\mathbf{h}}_1|}\right) \qquad \alpha(f) = \cos^{-1}\left(\frac{|\mathbf{h}_1^H \mathbf{h}_2|}{|\mathbf{h}_1||\mathbf{h}_2|}\right) \tag{2}$$

Figure 1 shows the SVA μ for a real room (sampling rate 8 kHz, distance between the microphones $d = 4$ cm, direction of arrival $\theta = 30°$, FFT size: 1024). $\hat{\mathbf{h}}_1$ is approximated by $[1, \ldots, e^{j2\pi \cdot (M-1) d \cdot \cos(\hat{\theta}) \frac{f}{c}}]$ (distance between microphones $d = 4$ cm), and $\mathbf{h}_1$ is the actual impulse measured in real environments with different reverberation times (for the database see Sec. 3.3). As seen in figure 1, in the case of a reverberant environment, a much stronger SVA can occur in some frequency bins than the SVA caused by a rough estimation of the source position.



(a) Incorrect estimation $\Delta\theta = 20°$, $T_R = 0$ ms



(b) Correct estimation $\Delta\theta = 0°$, $T_R = 128$ ms

Figure 1: SVA μ induced by a rough estimation ($\Delta\theta = 20°$) and reverberation. The effect of reverberation exceeds the effect of an imprecise estimation (solid line). The critical angle (see Sec. 3.2) of this mixture is shown as a dashed line.

3. GEOMETRICALLY CONSTRAINED ICA

3.1. Derivation of new algorithm

The algorithm is based on negentropy maximization (3) proposed by [3]. In this approach, the negentropy is approximated by the nonlinear function $G()$ with the derivation $g()$. As usual in ICA approaches, a PCA is applied first. Figure 2 shows a scatter plot of the sphered signals $\mathbf{z} = \mathbf{V}\mathbf{x}$. After sphering, we have the following equations:

$$\arg\min_{\mathbf{t}_1} \mathrm{E}\{G(|\mathbf{t}_1^H \mathbf{z}|^2)\} \tag{3}$$

with a constraint to the target signal:

$$\mathbf{w}_1^H \hat{\mathbf{h}}_1 = \mathbf{t}_1^H \mathbf{V} \hat{\mathbf{h}}_1 = c_1 \tag{4}$$

Although most BSS algorithms claim to be unconstrained (using (3) only), they normally employ the assumption that $\mathbf{T} = [\mathbf{t}_1, \cdots, \mathbf{t}_N]^H$ is unitary and, therefore, $\mathbf{t}_1$ has a unit length. This assumption is necessary to avoid a convergence to the point of origin and to reduce the dimensionality of the optimization problem since it can be done on the unit hyper sphere.

To combine BSS with the constraint, we have to weaken the assumption because a strict assumption collides with the constraint (Fig. 2). The column vectors of $\mathbf{T}$ need to be unitary, but they do not have to have a unit length. A degeneration solution is avoided by the constraint.

According to [3], the Kuhn-Tucker points of (3) are

$$\mathrm{E}\{\mathbf{z} \cdot g(\mathbf{t}^H \mathbf{z})\} = \beta \mathbf{t} \tag{5}$$

when the derivation is done on the unit circle $|\mathbf{t}|^2 = 1$.

We can also use this condition for points of $\mathbf{t}$ that are not on the unit circle. If $\mathbf{t}$ is a Kuhn-Tucker point on the unit hyper sphere, then $\mathbf{t}' = \gamma \mathbf{t}$ (for any γ) is a Kuhn-Tucker point of Lagrangian that constrains the solution to vectors of the same norm. According the theory of FastICA, the maximal nongaussianity only says something about the direction of the unmixing vector while the norm is not decisive. We are looking for the vector that satisfies the constraint (4) and has the highest negentropy of all vectors with the same norm.

Hence, we do not change the solution of (5) by projecting it to the constraint. We obtain the following algorithm, with the convergence illustrated in Fig. 2.

$$\mathbf{t}^{k+1} = \mathbf{t}_k - \frac{\mathrm{E}\{\mathbf{z} g(\mathbf{t}_k^H \mathbf{z})\} + \beta \mathbf{t}_k}{\mathrm{E}\{\mathbf{z} g(\mathbf{t}_k^H \mathbf{z})\} + \beta} \tag{6}$$

$$\mathbf{t}^{k+1_{new}} = \frac{\mathbf{t}_{k+1}}{|\mathbf{t}_{k+1}^H \mathbf{V} \hat{\mathbf{h}}_1|} \tag{7}$$

$$\beta = \frac{\mathrm{E}\{\mathbf{t}_k^H \mathbf{z} \cdot g(\mathbf{t}_k^H \mathbf{z})\}}{|\mathbf{t}_k|^2} \tag{8}$$

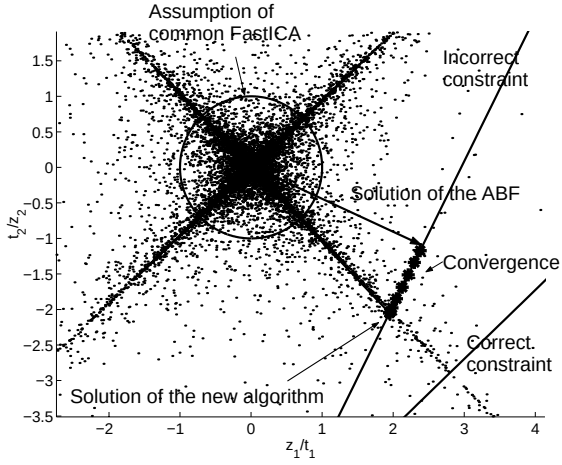
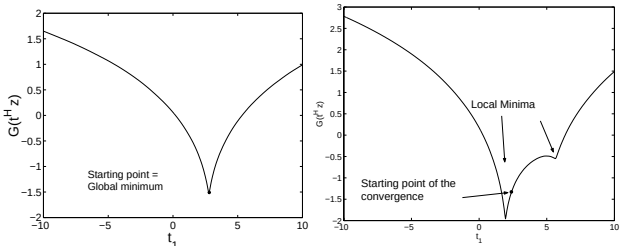


Figure 2: Convergence of the constrained FastICA algorithm. The constrained ICA algorithm starts at the solution of the ABF and converges on the constraint line to the correct solution.

The new algorithm starts with $\mathbf{t}^0 = \hat{\mathbf{t}}_1 = \mathbf{V}\hat{\mathbf{h}}_1$. If the estimation of $\hat{\mathbf{h}}_1$ is correct, the estimation itself is already the correct solution. Then, the algorithm converges according to (6)-(8) to a saddle point of Lagrangian. Since it does not indicate whether it is a minimum or maximum of the cost function, an additional maximality check is introduced into the algorithm.



(a) Correct estimation, $\mu = 0°$, $\alpha = 27°$

(b) Incorrect estimation, $\mu = 12°$, $\alpha = 27°$

Figure 3: Cost function on the constraint

3.2. Theoretical assessment

The Newton method guarantees the algorithms to converge against a saddle point of the Lagrangian. Figure 3(a) shows that there is only one global minimum in the case of a precise estimation of the steering vector $\hat{\mathbf{h}}_1$.

When the estimation becomes rough, two local minima and a local maximum appear (Fig. 3(b)). The two minima belong to the two signals. This means that since convergence is ensured, the algorithm converges either towards the target or the jammer. A convergence to the jammer signal enforces the jammer signal and suppresses the target signal. This is equivalent to the permutation problem of the unconstrained BSS.

The shape of the Lagrangian and the starting point of the iteration control the convergence. A convergence can be guaranteed when $\hat{\mathbf{t}}_1 = \mathbf{V}\hat{\mathbf{h}}_1$ is closer to the correct solution $\mathbf{t}_1$ than to the permutated solution $\mathbf{t}_2$ because it is in convergence range of $\mathbf{t}_1$.

In following, we derive a condition for the SVA μ and the mixing matrix that ensures that $\hat{\mathbf{t}}_1$ is close enough to $\mathbf{t}_1$. Thereto, we define angles μ_{t_1} and μ_{t_2} between $\hat{\mathbf{t}}_1$ and $\mathbf{t}_1$, and $\hat{\mathbf{t}}_1$ and $\mathbf{t}_2$, respectively, and we restrict ourselves to $N = M = 2$ and equivalent contributions from both sources.

Let $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$ be the spatial covariance matrix of the observed signals. Since it is a Hermitian matrix, it has M non zero real eigenvalues $\mathbf{\Lambda} = diag[\lambda_1, \ldots, \lambda_M]$ and M orthogonal eigenvectors $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$ belonging to them.

In the above defined case, $\mathbf{R}$ has the following eigenvectors: $\mathbf{u}_1 = \frac{\mathbf{h}_1 + \mathbf{h}_2}{|\mathbf{h}_1 + \mathbf{h}_2|}$, $\lambda_1 = \sigma_s^2 \mathbf{h}_1^H (\mathbf{h}_1 + \mathbf{h}_2)$, $\mathbf{u}_2 = \frac{\mathbf{h}_1 - \mathbf{h}_2}{|\mathbf{h}_1 - \mathbf{h}_2|}$ and $\lambda_2 = \sigma_s^2 \mathbf{h}_1^H (\mathbf{h}_1 - \mathbf{h}_2)$, and the sphering matrix can be written as $\mathbf{V} = \mathbf{\Lambda}^{1/2}\mathbf{U}$. Using the eigenvalue decomposition of $\mathbf{R}$, the definition of the scalar product in (2), and the mixing vectors angle α and the SVA μ defined in Sec. 2.2,

$$\begin{aligned} \cos(\mu_{t_1}) &= \frac{\hat{\mathbf{t}}_1^H \mathbf{t}_1}{|\hat{\mathbf{t}}_1||\mathbf{t}_1|} = \frac{\hat{\mathbf{h}}_1^H \mathbf{V}^H \mathbf{V} \mathbf{h}_1}{|\hat{\mathbf{t}}_1||\mathbf{t}_1|} = \frac{\hat{\mathbf{h}}_1^H \mathbf{R}^{-1} \mathbf{h}_1}{|\hat{\mathbf{t}}_1||\mathbf{t}_1|} \\ &= \frac{\hat{\mathbf{h}}_1^H \mathbf{h}_1 + \hat{\mathbf{h}}_1^H \mathbf{h}_2}{|\hat{\mathbf{t}}_1||\mathbf{t}_1|} + \frac{\hat{\mathbf{h}}_1^H \mathbf{h}_1 - \hat{\mathbf{h}}_1^H \mathbf{h}_2}{|\hat{\mathbf{t}}_1||\mathbf{t}_1|} \end{aligned}$$

$$= \frac{\cos\mu - \cot\alpha \sin\mu}{|\hat{\mathbf{t}}_1||\mathbf{t}_1|}$$

$$\cos(\mu_{t_2}) = \frac{\hat{\mathbf{t}}_1^H \mathbf{t}_2}{|\hat{\mathbf{t}}_1||\mathbf{t}_2|} = \frac{\frac{\sin\mu}{\sin\alpha}}{|\hat{\mathbf{t}}_1||\mathbf{t}_2|} \quad (9)$$

Combining the results in (9), we can define the convergence range: The algorithm will converge to the target (at least) when the miss-estimation is smaller than a critical angle $\mu_{critical}$ in (10).

$$\cot\mu < \cot\mu_{critical} = \frac{1+\cos\alpha}{\sin\alpha} \quad (10)$$

For a real-world acoustical application, it is important to analyze (10) with realistic numbers for the SVA μ and the mixing vectors angle α in each frequency bin. Figure 1 also shows the critical angle. In both cases, the SVA μ is in almost all frequency bins under the critical angle.

3.3. Real-world assessment

Several tests with realistic mixtures were executed. Japanese language sounds were mixed with room impulse responses measured in real rooms. The reverberation times were 0, 150 and 300 ms. The impulse responses where $N = M = 3$ were from the RWCP database, and a complete description is available at http://tosa.mri.co.jp/ sounddb/index.htm. Furthermore, the algorithm has been tested with real-world mixtures for the case of $N = M = 2$.

Figure 4 shows the signal-to-interference-ratio (SIR) in each frequency bin. In almost all frequencies the algorithm converged against the correct solution and yielded a high SIR. Negative values indicated an incorrect permutation that happened mostly in the low frequency range, but it also occasionally occurred in the higher frequency ranges. This is attributed to a large estimation error due to the multi-path mixture and amplification by the PCA.

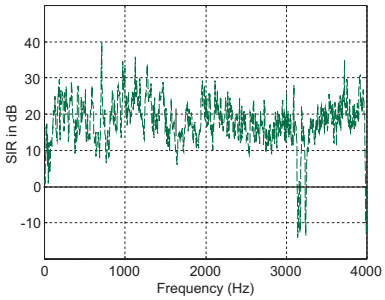
Incorrect steering vectors caused by an incorrect look direction (a) and by reverberation (b) were used in all plots of Fig. 4. A high improvements of the frequency SIR over 10dB were achieved, even in the 3×3 ($N = M = 3$) case. This demonstrates the effectiveness of the new algorithm in its major domain when a rough estimation of the steering vector is available. Although the computational cost has not been analyzed yet, the convergence is very fast due to the Newton method of the underlying FastICA algorithm.

4. CONCLUSION

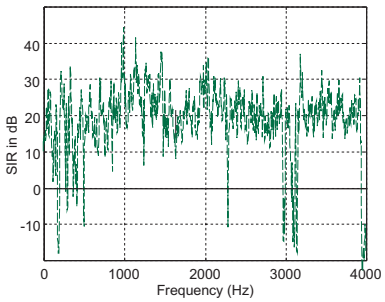
We introduced a new ICA algorithm with a geometrical constraint and showed its effectiveness both theoretically by defining a convergence range and experimentally by using impulse responses from a real room. The new algorithm solves the permutation problem of BSS of acoustic mixtures, particularly when the number of sources and microphones becomes larger than two.

5. ACKNOWLEDGMENT

We thank Stefan Winter for revising the algorithms, Hiroshi Sawada and Ryo Mukai for daily cooperation and valuable discussions, and Dr. Dieter Filbert and Dr. Shigeru Katagiri for their continuous encouragement.



(a) $T_R = 0$ ms, $\Delta\theta = 20°$



(b) $T_R = 150$ ms, $\Delta\theta = 0°$

Figure 4: Real-world simulation results in $N = M = 3$

REFERENCES

[1] D. Johnson and D. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] I. Thng, A. Cantoni, and Y. H. Leung, "Constraints for maximally flat optimum broadband antenna arrays," *IEEE Trans. on Signal Processing*, vol. 43, no. 6, pp. 1334–1347, June 1995.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley, New York, 2001.

[4] K. Torkkola, "Blind separation of delayed and convolved mixtures," in *Unsupervised adaptive filtering*, S. Haykin, Ed., New York, 2000, vol. 1, John Wiley.

[5] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," in *Eurospeech*, 2001, pp. 2595–2598.

[6] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Processing* (to appear).

[7] H. Saruwatari, S. Kurita, and T. Takeda, "Blind source separation combining frequency-domain ICA and beamforming," in *ICASSP*, 2001 pp. 2733–2736.

[8] M. Knaak and D. Filbert, "Acoustical semi-blind source separation for machine monitoring," in *3rd. Int. Conf. on BSS and ICA*, 2001, pp. 361–366.

[9] M. Knaak, M. Fausten, D. Filbert, "Acoustical machine monitoring using blind source separation," in *4th Int. Conf. on acoust. and vibratory surveillance*, 2001, pp. 401–412.

[10] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," in *NNSP*, 2001, pp. 273–282.