Some uniqueness results in sparse convolutive source separation

Alexis Benichoux, Prasad Sudhakar, Frédéric Bimbot, Rémi Gribonval

Some uniqueness results in sparse convolutive source separation

Alexis Benichoux[1], Prasad Sudhakar[2], Fréderic Bimbot[1] and Rémi Gribonval[1]

[1] METISS Team, INRIA Rennes-Bretagne Atlantique
Campus de Beaulieu, 35042 Rennes CEDEX, France.
[2] ICTEAM/ELEN, Université catholique de Louvain
B-1348, Louvain-la-Neuve, Belgium.
`firstname.secondname@{inria.fr, uclouvain.be}`

Abstract. The fundamental problems in the traditional frequency domain approaches to convolutive blind source separation are 1) arbitrary permutations and 2) arbitrary scaling in each frequency bin of the estimated filters or sources. These ambiguities are corrected by taking into account some specific properties of the filters or sources, or both. This paper focusses on the filter permutation problem, assuming the absence of the scaling ambiguity, investigating the use of temporal sparsity of the filters as a property to aid permutation correction. Theoretical and experimental results bring out the potential as well as the extent to which sparsity can be used as a hypothesis to formulate a well posed permutation problem.

Keywords: sparse filters, convolutive blind source separation, permutation ambiguity, ℓ^p minimization, Hall's Marriage Theorem, bi-stochastic matrices.

1 Introduction

Let $x_i[t]$, $1 \leq i \leq M$ be M mixtures of N source signals $s_j[t]$, resulting from the convolution with filters $a_{ij}[t]$, each of length L such that:

$$x_i[t] = \sum_{j=1}^{N}(a_{ij} \star s_j)[t], \quad 1 \leq i \leq M. \tag{1}$$

where $\star$ denotes convolution. The filter $a_{ij}[t]$ typically models the impulse response between the jth source and the ith sensor. By abuse of notation, $\mathbf{F}a_{ij} = \{a_{ij}[\omega]\}_{0 \leq \omega < L}$ denotes the discrete Fourier transform of the filter seen as a vector $a_{ij} = \{a_{ij}[t]\}_{0 \leq t < L} \in \mathbb{C}^L$. Also, the mixing equation (1) can be rewritten as $\mathbf{X} = \mathbf{A} \star \mathbf{S}$, with $\mathbf{A}$ the matrix of filters $\mathbf{A} := (\{a_{ij}[t]\}_{0 \leq t < L})_{1 \leq i \leq M, \ 1 \leq j \leq N}$, $\mathbf{X}$ the observation matrix and $\mathbf{S}$ the source matrix.

In this context, blind filter estimation refers to the problem of obtaining estimates of the filters $\mathbf{A}$ from the mixtures $\mathbf{X}$, without any explicit knowledge about the sources $\mathbf{S}$. Filter estimation is relevant for tasks such as deconvolution, source localisation, etc. It also has a relationship with the problem of Multiple-Input-Multiple-Output system identification in communications engineering.

2 Permutation and scaling ambiguities in frequency domain filter estimation

A widely used approach for filter estimation relies on the transformation of the mixing model in Eq. (1) into the time-frequency domain, converting a single convolutive filter estimation problem into several complex instantaneous filter estimation problems. Using standard techniques for instantaneous mixing parameter estimation [1], complex mixing filter coefficients $\widetilde{\mathbf{A}}[\omega] = \{\tilde{a}_{ij}[\omega]\}_{1 \leq i \leq M,\ 1 \leq j \leq N}$ are estimated for each frequency bin $0 \leq \omega < L$.

However, without any further assumption on either the filters $a_{ij}[t]$ or the sources $s_j[t]$, one can find filter estimates $\widetilde{\mathbf{A}} = (\tilde{a}_{ij})$, only up to a global permutation and scaling. That is, for every frequency ω we have

$$\tilde{a}_{ij}[\omega] = \lambda_j[\omega] a_{i\sigma_\omega(j)}[\omega], \tag{2}$$

where $\lambda_j[\omega]$ and $\sigma_\omega \in \mathfrak{S}_N$ are the unknown scaling and permutation, with $\mathfrak{S}_N$ being the set of permutations of the integers between 1 and N. Several methods [2] attempt to solve these ambiguities by exploiting properties of either $\mathbf{S}$ or $\mathbf{A}$.

2.1 Exploiting sparsity to solve the permutation ambiguity

In this paper, we hypothesize that the filters $\mathbf{A}$ are *sparse* in the time domain and use this property to solve the permutation ambiguity, in the absence of scaling ($\lambda_j[\omega] = 1$). The assumption that $\mathbf{A}$ is sparse means that each filter a_{ij} has few nonzero coefficients, typically measured by the ℓ^0 pseudo-norm

$$\|a_{ij}\|_0 := \sharp\{0 \leq t < L,\ a_{ij}[t] \neq 0\} = \sum_t |a_{ij}[t]|^0.$$

Besides the ℓ^0 pseudo-norm, the following ℓ^p quasi-norms will be used to quantify the sparsity of $\mathbf{A}$:

$$\|\mathbf{A}\|_p^p := \sum_{ij} \|a_{ij}\|_p^p = \sum_{ijt} |a_{ij}[t]|^p, \quad 0 < p \leq 1.$$

The underlying approach in this work is to seek permutations $\widehat{\sigma}_0, \ldots \widehat{\sigma}_{L-1}$ which yield the sparsest estimated time-domain matrix of filters $\widehat{\mathbf{A}} = (\widehat{a}_{ij})$, where $\widehat{a}_{ij}[\omega] := \tilde{a}_{i\widehat{\sigma}_\omega(j)}[\omega]$.

2.2 Main result and structure of the paper

As the main result, we show (Theorem 2) that when the filter length L is prime, and if $\frac{k}{L} \leq \alpha(N)$, with N the number of sources, then k-sparse filters (i.e., $\|a_{ij}\|_0 \leq k$) uniquely minimize the ℓ^0 norm of $\mathbf{A}$ (up to a global permutation).

In Sec. 3, we investigate the interplay of sparsity and frequency permutations of filters and present our main result. We omit the proof of our main result due

to the space constraints, but we describe the main ingredients of the same[3]. In Sec. 4 we propose a combinatorial ℓ^1 minimization algorithm to resolve filter permutations. The effectiveness and limitations of the algorithm is empirically shown, and the observations are related with the theoretical results.

3   Theoretical guarantees

Given an $M \times N$ filter matrix $\mathbf{A}$, made of filters of length L, and an L-tuple $(\sigma_0, \ldots, \sigma_{L-1}) \in \mathfrak{S}_N$ of permutations, we let $\widetilde{\mathbf{A}}$ be the matrix obtained from $\mathbf{A}$ by applying the permutations in the frequency domain, as in (2), without scaling $(\lambda_j[\omega] = 1)$.

The effect of the permutations is said to coincide with that of a global permutation $\pi \in \mathfrak{S}_N$ of the columns of $\mathbf{A}$ if $\widetilde{a}_{ij} = a_{i\pi(j)}$, $\forall i, j$, or equivalently in the frequency domain:

$$\widetilde{a}_{ij}[\omega] := a_{i\sigma_\omega(j)}[\omega] = a_{i\pi(j)}[\omega], \ 0 \leq \omega < L, \ \forall i, j.$$

This is denoted $\mathbf{A} \equiv \widetilde{\mathbf{A}}$. First, we show that for filters with disjoint time-domain supports, permutations cannot decrease the ℓ^p norm, $0 \leq p \leq 1$:

Theorem 1. *Let $\Gamma_{ij} \subset \{0, \ldots, L-1\}$ be the time-domain support of a_{ij}. Suppose that for all i and $j_1 \neq j_2$ we have*

$$\Gamma_{i,j_1} \cap \Gamma_{i,j_2} = \emptyset. \tag{3}$$

Then, for $0 \leq p \leq 1$, we have $\|\widetilde{\mathbf{A}}\|_p \geq \|\mathbf{A}\|_p$.

Note that filters with disjoint supports need not be very sparse: M filters of length L can have disjoint supports provided that $\max_j \|a_{ij}\|_0 \leq L/M$. Yet, disjointness of supports is a strong assumption, and Theorem 1 only indicates that frequency permutations cannot decrease the ℓ^p norm. Thus, the minimum value of the ℓ^p norm might not be uniquely achieved (up to a global permutation). In our main result, we consider k-sparse filters of prime length, and $p = 0$:

Theorem 2. *Let $\mathbf{A}$ be a $M \times N$ matrix of filters of prime length L. Assume that*

$$\max_{ij} \|a_{ij}\|_0 \leq k. \tag{4}$$

where

$$\frac{k}{L} \leq \alpha(N) := \begin{cases} \frac{2}{N(N+2)} & \text{if } N \text{ is even,} \\ \frac{2}{(N+1)^2} & \text{if } N \text{ is odd.} \end{cases} \tag{5}$$

Then, up to a global permutation, $\mathbf{A}$ uniquely minimises the ℓ^0 pseudo-norm among all possible frequency permutations.

Noticeably, the uniqueness condition does not depend on the number M of mixtures. In order to prove Theorem 2, it is important to quantify the amount of permutations incurred. We use the following definition of the "size" of permutations in the rest of the paper.

---

[3] An extended version of this paper, containing all proofs, has been submitted for possible publication and is available as INRIA Technical Report No 7782.

3.1 Quantification of permutations

Given a reference global permutation π, we define the maximum number of frequencies where each estimated filter actually differs from the (globally permuted) original filters, by:

$$\Delta(\widetilde{\mathbf{A}}, \mathbf{A}|\pi) := \max_{i,j} \|\mathbf{F}(\widetilde{a}_{ij} - a_{i\pi(j)})\|_0 \tag{6}$$

The "size" of permutations is then defined as:

$$\Delta(\widetilde{\mathbf{A}}, \mathbf{A}) := \min_{\pi \in \mathfrak{S}_N} \Delta(\widetilde{\mathbf{A}}, \mathbf{A}|\pi). \tag{7}$$

Note that $\Delta(\widetilde{\mathbf{A}}, \mathbf{A}) = 0$ iff $\widetilde{\mathbf{A}} \equiv \mathbf{A}$. We also use the symbol Δ to denote $\Delta(\widetilde{\mathbf{A}}, \mathbf{A})$.

3.2 Exploitation of an uncertainty principle

Along with the quantification of permutations, the following lemma, which exploits an uncertainty principle, is an intermediate result to prove Theorem 2.

Lemma 1. *Assume that $\widetilde{\mathbf{A}} \not\equiv \mathbf{A}$, that L is a prime integer, and that (4) holds with*

$$2k + \Delta \leq L. \tag{8}$$

Then $\|\widetilde{\mathbf{A}}\|_0 > \|\mathbf{A}\|_0$ and $\|\widetilde{a}_{ij}\|_0 \geq \|a_{ij}\|_0, \forall i, j$. The latter inequality is strict when $\widetilde{a}_{ij} \neq a_{ij}$. For a general L (not necessarily prime), the same conclusions hold when the assumption (8) is replaced with

$$2k \cdot \Delta < L. \tag{9}$$

This lemma states that when the original filters $\mathbf{A}$ are sufficiently sparse and if the size Δ of permutations are controlled, in relation to the filter length L, then the resulting permuted filters $\widehat{\mathbf{A}}$ have a larger ℓ^0 norm than $\mathbf{A}$. Moreover, it also states that each individual filter $\widetilde{a}_{ij}$ will have an ℓ^0 norm that is at least as large as that of the corresponding a_{ij}. The skilled reader will rightly sense the role of uncertainty principles [3, 4], [5, Theorem 1] in the above lemma.

As opposed to Lemma 1, Theorem 2 does not use an explicit quantification of the permutations, Δ. In fact, this quantity is buried inside the constant $\alpha(N)$. It is actually necessary to make some combinatorial arguments concerning the permutations to arrive at the constant $\alpha(N)$, starting from Δ. The objective of these arguments is to bound Δ from above.

3.3 Combinatorial arguments

Using Lemma 1 with prime L, a simple combinatorial argument can be used to obtain a weakened version of Theorem 2, with the more conservative constant $\alpha'(N) := 1/2N!$: by the pigeonhole principle, for any L-tuple of frequency permutations among N sources, at least $L/N!$ permutations are identical; as a result,

$\Delta(\widetilde{\mathbf{A}}, \mathbf{A})$ is universally bounded from above by $L - L/N!$; hence if $k \leq L/2N!$ we obtain $2k + \Delta \leq L$ and we can conclude thanks to Lemma 1.

The proof of Theorem 2 with the constant $\alpha(N)$ exploits a stronger universal upper bound $\Delta(\widetilde{\mathbf{A}}, \mathbf{A}) \leq L(1 - 2\alpha(N))$, obtained through an apparently new quantitative application of Hall's Marriage Theorem [6] to bi-stochastic matrices. Bi-stochastic matrices are defined as:

Definition 1 (Bi-stochastic matrix). *An $N \times N$ matrix $\mathbf{B}$ is called* bi-stochastic *if all its entries are non-negative, and the sum of the entries over each row as well as the sum of the entries over each column is one.*

The following lemma connects permutation matrices, which define permutations, and bi-stochastic matrices through Hall's marriage theorem. The subsequent corollary (Corollary 1) provides the bound $\Delta(\widetilde{\mathbf{A}}, \mathbf{A}) \leq L(1 - 2\alpha(N))$, which is crucial to Theorem 2.

Lemma 2. *Let $\mathbf{B}$ be an $N \times N$ bi-stochastic matrix: there exists a permutation matrix $\mathbf{P}$ such that all the entries of $\mathbf{B}$ on the support of $\mathbf{P}$ exceed the threshold*

$$2\alpha(N) = \begin{cases} \frac{4}{N(N+2)} & \text{if } N \text{ is even,} \\ \frac{4}{(N+1)^2} & \text{if } N \text{ is odd.} \end{cases} \tag{10}$$

Corollary 1. *Let $\sigma_0, \ldots, \sigma_{L-1} \in \mathfrak{S}_N$ be L permutations. There exists a global permutation π such that*

$$C_{j\pi(j)} = \sharp\{\ell : \sigma_\ell(j) = \pi(j)\} \geq 2L\alpha(N), \quad \forall 1 \leq j \leq N.$$

The reader may have noticed that Theorem 2, while dropping the *disjoint support* assumption from Theorem 1, introduces new restrictions: the assumption that L is prime, and the restriction to $p = 0$ compared to $0 \leq p \leq 1$ in Theorem 1. How critical are these restrictions? Could they be extended to filters of arbitrary length L and $0 \leq p \leq 1$? This is discussed in the following section.

3.4 Extending Theorem 2 to non-prime filter length L?

As indicated by Lemma 3 below, even for $L \geq 4$, there exists sparse matrices of filters that are the sparsest *but not unique (even up to a global permutation)* solution of the considered problem: certain frequency permutations provide an *equally sparse, but not equivalent, solution.*

Lemma 3. *For any integer k such that $2k$ divides L, there exists a matrix of k-sparse filters $\mathbf{A}$ and a set of $L/2k$ frequency permutations resulting in $\widetilde{\mathbf{A}} \not\equiv \mathbf{A}$ such that for all $0 \leq p \leq \infty$: $\|\widetilde{\mathbf{A}}\|_p = \|\mathbf{A}\|_p$, and*

$$\|\widetilde{a}_{ij}\|_p = \|a_{ij}\|_p, \quad \forall i, j. \tag{11}$$

We have $2k \cdot \Delta(\widetilde{\mathbf{A}}, \mathbf{A}) = L$.

The fact that the filter matrices $\mathbf{A}$ and $\widetilde{\mathbf{A}}$ satisfy $2k \cdot \Delta(\widetilde{\mathbf{A}}, \mathbf{A}) = L$ shows the sharpness of Lemma 1 for the case when L is even: the strict inequality in (9) cannot be improved.

Specializing Lemma 3 to $k = 1$ for even $L \geq 4$ yields ideally 1-sparse filters a_{ij} and a set of $L/2$ frequency permutations such that: $\widetilde{a}_{ij}$ are 1-sparse; $\widetilde{\mathbf{A}}$ is not equivalent to $\mathbf{A}$ and cannot be discriminated from it by any ℓ^p norm.

Lemma 3 actually gives a worst case well-posedness bound for filters with arbitrary lengths, and is pessimistic. But, such a bound is achieved in cases when the filters are associated to Dirac combs, which are highly structured. However, existing probabilistic versions of uncertainty principles (see, e.g., the nice survey [7]) lead us to conjecture that if the sparse filters in $\mathbf{A}$ are drawn at random (e.g. from Bernoulli-Gaussian distribution), the uniqueness guarantee of Theorem 2 will hold except with small probability $O(L^{-\beta})$, provided that $k < c(\beta)L/\log L$, for large L.

4   Numerical experiments

The results achieved so far are theoretical well-posedness guarantee, but do not quite provide algorithms to compute the potentially unique (up to global permutation) solution of the frequency permutation problem. We conclude this paper with the description of a relatively naive optimization algorithm, an empirical assessment of its performance with Monte-Carlo simulations, and a discussion of how this compares with the theoretical uniqueness guarantees achieved above.

4.1   Proposed combinatorial algorithm

Given a "permuted" matrix $\widetilde{\mathbf{A}}$, one wishes to find a set of frequency permutations yielding a new matrix $\widehat{\mathbf{A}}$ with minimum ℓ^p norm. The proposed algorithm starts from $\widehat{\mathbf{A}}_0 = \widetilde{\mathbf{A}}$. Given $\widehat{\mathbf{A}}_n$, a candidate matrix $\widehat{\mathbf{A}}_{n+1,\pi}$ can be obtained by applying a permutation π at frequency $\omega_n \equiv n \ [\mathtt{mod} \ L]$. Testing each possible permutation π and retaining the one π_n which minimizes $\|\widehat{\mathbf{A}}_{n+1,\pi}\|_p$ yields the next iterate $\widehat{\mathbf{A}}_{n+1} := \widehat{\mathbf{A}}_{n+1,\pi_n}$. The procedure is repeated until the ℓ^p norm $\widehat{\mathbf{A}}_n$ ceases to change. Since there is a finite number of permutations to try, the stopping criterion is met after sufficiently many iterations.

In theory, it could happen that the stopping criterion is only met after a combinatorially large number of iterations. However, the algorithm stops much sooner in practice. In fact, if we were to use the ℓ^0 norm, the algorithm would typically stop after just one iteration, because the ℓ^0 norm attains its maximum value $M \times N \times L$ for most frequency permutations except a few very special ones. For this reason, we chose to test the algorithm using ℓ^p norms $p > 0$, which are not as "locally constant" as the ℓ^0 norm. To our surprise, the experiments below will show that the best performance is not achieved for small p, but rather for $p = 2 - \epsilon$ with small $\epsilon > 0$. For $p = 0$ and for $p \geq 2$, the algorithm indeed completely fails.
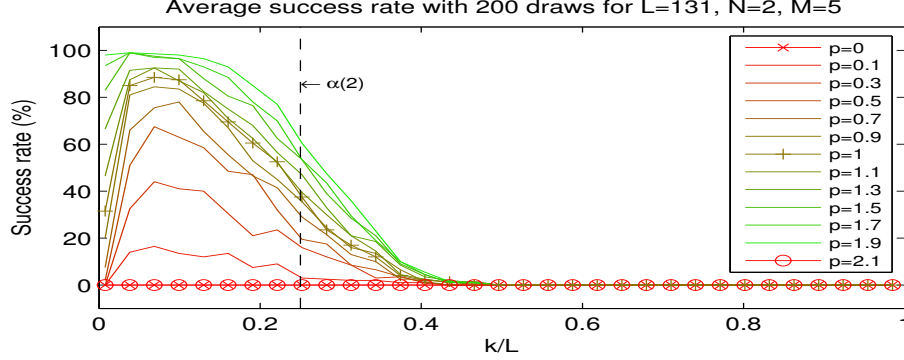
Fig. 1. Filter recovery success as a function of p, $0 \leq p \leq 1.9$

4.2 Monte-Carlo simulations

For various sparsity levels k and dimensions M, N, random sparse filter matrices $\mathbf{A}$ made of independent random k-sparse filters of length $L = 31$ were generated. Each filter was drawn by choosing: a) a support of size k uniformly at random; b) i.i.d. Gaussian coefficients on this support. For each configuration (k, M, N), 200 random sparse filter matrices $\mathbf{A}$ were drawn. For each $\mathbf{A}$, independent random frequency permutations were applied to obtain $\widetilde{\mathbf{A}}$. The algorithm was then applied to obtain $\widehat{\mathbf{A}}$. The rate of recovery was then computed for each configuration (k, M, N), with an SNR threshold of 100 dB to consider the estimation as a success. We observed that in case of success the SNR was actually more than 300 dB, while in case of failure it was essentially 0 dB.

Figure 1 displays the success rate as a function of the relative sparsity k/L, for various choices of the ℓ^p criterion, with filters of prime length $L = 131$, $N = 2$ sources and $M = 5$ channels. The vertical dashed line indicates the threshold $k/L \leq \alpha(2)$ associated with the well-posedness guarantee (using an ℓ^0 criterion) of Theorem 2. Surprisingly, one can observe that the success rate increases when $0 < p < 2$ is increased. The maximum success rate is achieved when $p = 2 - \epsilon$ with small $\epsilon > 0$.

Beyond the well-posedness regime suggested by the theory (i.e., to the right of the vertical dashed line), the algorithm can succeed, but at a rate that rapidly decreases when the relative sparsity k/L increases. In the regime of well-posed problems, the proposed algorithm is often successful but can still fail to perfectly recover the filters, especially –and surprisingly– for small values of k. This phenomenon is strongly marked for $p < 1$ and essentially disappears for $p > 1$. It remains an open question to determine the respective roles of the ℓ^p criterion and of the naive greedy optimization algorithm in this limited performance for $k/L \ll 1$ when the problem is well-posed with respect to the ℓ^0 norm.

5 Conclusions

It is well known that a sufficient sparsity assumption can be used to make underdetermined linear inverse problems well-posed: without the sparsity assumption, the problem admits an affine set of solutions, which intersects at only one point with the set of sparse vectors. Besides this well-posedness property, a key factor that has lead to the large deployment of sparse models and methods in various fields of science is the fact that a convex relaxation of the NP-hard ℓ^0 minimization problem can be guaranteed to find this unique solution under certain sparsity assumptions. The availability of efficient convex solvers then really makes the problem tractable.

The problem considered in this paper is not a linear inverse problem. Even though it is a simplification of the original permutation *and scaling* problem arising from signal processing, it remains a priori a much harder problem than linear inverse problems in terms of the structure of the solution set: each solution comes with a herd of solutions that are equivalent up to a global permutation.

It is encouraging that we have obtained well-posedness results in this context, but this is at best the beginning of the story: even if the solution is unique, how do we efficiently compute it? Can these results be extended to the original permutation and scaling problem? Why does the proposed naive algorithm perform better for $p > 1$? Answers to these questions can have an impact in fields like blind source separation with sparse multipath channels.

Acknowledgment

References

1. Comon, P. and Jutten, C., Eds., Handbook of Blind Source Separation, Independent Component Analysis and Applications, Academic Press, 2010.
2. Pedersen, M. S., Larsen, J.,Kjems, U. and Parra, L. C., A survey of convolutive blind source separation methods, Multichannel Speech Processing Handbook, Citeseer.
3. Donoho, D. L. and Stark, P. B., Uncertainty Principles and Signal Recovery, SIAM Journal on Applied Mathematics, 49(3), pp. 906-931, 1989.
4. Elad, M. and Bruckstein, A. M., A generalized uncertainty principle and sparse representation in pairs of bases, IEEE Trans. on Information Theory, 48(9), pp. 2558-2567, 2002.
5. Tao, T., An Uncertainty Principle for Cyclic Groups of Prime Order, Mathematical Research Letters, 12, 121-127, 2005.
6. Hall, P., On Representatives of Subsets, J. London Math. Soc., 10(1), 26-30, 1935.
7. Tropp, J., On the Linear Independence of Spikes and Sines, Journal of Fourier Analysis and Applications, 14(5), 838-858, 2008.