

# ON MULTIPLE SOLUTIONS OF THE “SEQUENTIALLY DRILLED” JOINT CONGRUENCE TRANSFORMATION (SEDJOCO) PROBLEM FOR SEMI-BLIND SOURCE SEPARATION

Arie Yeredor<sup>1</sup>, Yao Cheng<sup>2</sup>, and Martin Haardt<sup>2</sup>

<sup>1</sup> School of Electrical Engineering  
Tel-Aviv University  
P.O. Box 39040, Tel-Aviv 69978, Israel  
arie@eng.tau.ac.il

<sup>2</sup> Communications Research Laboratory  
Ilmenau University of Technology  
P.O. Box 100565, D-98684 Ilmenau, Germany  
{y.cheng, martin.haardt}@tu-ilmenau.de

## ABSTRACT

In the context of Maximum Likelihood (ML) source separation in a semi-blind scenario, where the spectra of the sources are known and distinct, the likelihood equations amount to a set of matrix decompositions (known as the “Sequentially Drilled” Joint Congruence Transformation (SeDJoCo)). However, quite often multiple solutions of SeDJoCo exist, only one of which is the optimal solution, corresponding to the global maximum. In this paper we characterize the different solutions and propose a procedure for detecting whether a given solution is sub-optimal. Moreover, for such sub-optimal solutions we propose a procedure for re-initializing an iterative solver so as to converge to the optimal solution. Using simulation, we present the empirical probability to encounter a sub-optimal solution (by a given iterative algorithm), as well as the resulting separation improvement when applying our proposed re-initialization approach in such cases.

**Index Terms**— joint matrix transformation, maximum likelihood, semi-blind source separation

## 1. INTRODUCTION

While a myriad of diverse and successful methods for Blind Source Separation (BSS) and Independent Component Analysis (ICA) have been proposed over the past three decades, relatively little effort has been addressed towards Maximum Likelihood (ML) separation. This is mainly due to the fact that in a fully blind scenario, statistical models for the sources are generally unknown (apart from their mutual statistical independence), and ML estimation cannot be applied in such cases. Nevertheless, several Quasi-ML methods have been proposed (e.g., [1]), which involve educated guesses, assumptions or parameterized estimates regarding the missing statistical models. Moreover, in a semi-blind scenario, such statistical models might be known a-priori.

In the semi-blind context of separating Gaussian sources with distinct (known) temporal correlations, it has been

The authors gratefully acknowledge the financial support by the German-Israeli Foundation (GIF), grant number I-1282-406.10/2014.

shown in [2, 3, 4] that for obtaining the ML estimate, a set of matrix decomposition equations (constituting the “Likelihood Equations” in this case) needs to be solved. These equations lead to a special form of hybrid exact-approximate joint diagonalization, termed a “Sequentially Drilled” Joint Congruence transformation (SeDJoCo) in [4]. Interestingly, the same set of equations finds applications also in Coordinated Beamforming (CBF) [4].

While, to the best of our knowledge, no closed-form solution of the SeDJoCo equations is currently known, several iterative solution algorithms have been proposed in recent years [5, 2, 3, 4, 6, 1] (some of these pre-dated the explicit formulation of the general SeDJoCo problem in [5, 4], but provide implicit solutions for the same). In [4] we established sufficient conditions for existence of a solution, but the issue of uniqueness remained unresolved.

Although the number of equations in SeDJoCo equals the number of unknowns, it turns out that the solution is not unique, in general, since the equations are nonlinear. In the context of ML BSS, the iterative algorithms may thus converge to solutions corresponding to local stationary points of the likelihood function, rather than to the global maximum. Although any solution of SeDJoCo usually leads to reasonable separation, sub-optimal solutions result in degraded separation performance and do not share the asymptotic optimality of the ML estimate.

In this work we characterize possible different solutions of SeDJoCo, and propose a procedure for disqualifying sub-optimal solutions which do not correspond to the global maximum. Moreover, we show how such sub-optimal solutions can be used for re-initializing an iterative algorithms, such that following re-initialization the algorithm would converge to the desired solution. The procedure is based on detecting undesired permutations and scaling (if any) in the solution and on correcting these in the re-initialization process. We demonstrate (in simulation) the empirical probability of obtaining sub-optimal solutions, and the resulting improvement in separation performance when applying our proposed procedure in such cases.

## 2. THE “SEQUENTIALLY DRILLED” JOINT CONGRUENCE TRANSFORMATION (SEDJOCO)

For completeness of the exposition, we briefly review the SeDJoCo problem formulation in this section.

Assume the classical static mixture model  $\mathbf{X} = \mathbf{A}\mathbf{S}$ , where  $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \cdots \ \mathbf{s}_K]^T \in \mathbb{R}^{K \times T}$  denotes a matrix of  $K$  statistically independent source signals ( $\mathbf{s}_1, \dots, \mathbf{s}_K \in \mathbb{R}^T$ ),  $\mathbf{A} \in \mathbb{R}^{K \times K}$  denotes an unknown (but invertible) mixing matrix, and  $\mathbf{X} \in \mathbb{R}^{K \times T}$  denotes the matrix of observed mixture signals. From  $\mathbf{X}$ , it is desired to estimate the mixing matrix and, subsequently, the source signals. When the source signals are zero-mean Gaussian and each has its own temporal (and known) covariance matrix  $\mathbf{C}_k = \mathbb{E}\{\mathbf{s}_k \mathbf{s}_k^T\} \in \mathbb{R}^{T \times T}$  (for  $k = 1, \dots, K$ ) it can be shown (see [2], [3] (Chapter 7) or [1] for more details) that the ML estimate  $\hat{\mathbf{A}}$  of  $\mathbf{A}$  can be obtained (up to a sign ambiguity) as follows. First, construct  $K$  symmetric “target-matrices” as

$$\mathbf{Q}_k = \frac{1}{T} \mathbf{X} \mathbf{C}_k^{-1} \mathbf{X}^T \in \mathbb{R}^{K \times K}, \quad k = 1, \dots, K \quad (1)$$

(note that this is an *ordered* set). Then, look for a matrix  $\hat{\mathbf{A}} \in \mathbb{R}^{K \times K}$  which decomposes these target matrices as

$$\mathbf{Q}_k = \hat{\mathbf{A}} \mathbf{D}_k \hat{\mathbf{A}}^T, \quad k = 1, \dots, K, \quad (2)$$

such that the matrices  $\mathbf{D}_1, \dots, \mathbf{D}_K \in \mathbb{R}^{K \times K}$  are not necessarily diagonal, but satisfy the property

$$\mathbf{D}_k \mathbf{e}_k = \mathbf{e}_k, \quad k = 1, \dots, K, \quad (3)$$

where the pinning vector  $\mathbf{e}_k$  denotes the  $k$ -th column of the  $K \times K$  identity matrix. In other words, the  $k$ -th column of the  $k$ -th matrix  $\mathbf{D}_k$  must equal  $\mathbf{e}_k$ , namely, must be all-zeros except for a “1” in its  $k$ -th element (and, since each  $\mathbf{D}_k$  must be symmetric by construction, this also applies to its  $k$ -th row), or, from another perspective,  $\mathbf{e}_k$  should be an eigenvector of  $\mathbf{D}_k$  with eigenvalue 1. We term this structure a “drilled” structure, as illustrated in Fig. 1. A substitution with  $\mathbf{B} \triangleq \hat{\mathbf{A}}^{-1}$  (the ML estimate of the unmixing matrices) is also possible by reformulating (2) as

$$\hat{\mathbf{B}} \mathbf{Q}_k \hat{\mathbf{B}}^T = \mathbf{D}_k, \quad k = 1, \dots, K. \quad (4)$$

Interestingly, in [4, 7] we have shown that the very same fundamental transformation is also useful in the context of CBF for the multi-user MIMO downlink, where a base station with  $K$  antennas transmits to  $N \leq K$  users each having  $K$  antennas as well. The elimination of the multi-user interference is achieved via CBF, employing a solution of SeDJoCo for a special set of target matrices, constructed from the (flat fading) channels’ coefficients matrices.

In [4] two iterative solution algorithms have been presented. One approach is based on Newton’s method, employing a conjugate gradient solution if desired, for enhanced

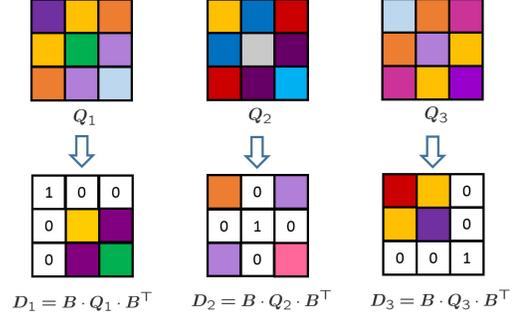


Fig. 1. Illustration of SeDJoCo for  $K = 3$

computational efficiency (termed NCG). The other method (termed STJOCO) is based on a modification of an existing AJD algorithm that uses LU decompositions. Another possible iterative algorithm can be based on an approach which was originally proposed in the context of the joint multi-user MIMO system (JMMS) [8]. All these algorithms may converge to a solution which does not correspond to the global maximum of the likelihood function. However, the former two can depend on the initialization and, as we shall show, NCG can be “directed” to the desired solution if initialized in its vicinity - which can be deduced from a sub-optimal solution using the procedure proposed in the next section.

## 3. MULTIPLE SOLUTIONS

To characterize the multiple possible solutions of SeDJoCo, let us denote by  $\mathbf{B}$  a given solution satisfying

$$\mathbf{B} \mathbf{Q}_k \mathbf{B}^T \mathbf{e}_k = \mathbf{e}_k \quad k = 1, 2, \dots, K. \quad (5)$$

Now assume that the ordered set of target matrices  $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_K$  is permuted into a new set, assuming, for simplicity of the exposition, that only  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  are swapped. Namely, define a new set of target matrices, denoted  $\tilde{\mathbf{Q}}_1, \tilde{\mathbf{Q}}_2, \dots, \tilde{\mathbf{Q}}_K$ , such that  $\tilde{\mathbf{Q}}_1 = \mathbf{Q}_2, \tilde{\mathbf{Q}}_2 = \mathbf{Q}_1$ , and  $\tilde{\mathbf{Q}}_k = \mathbf{Q}_k$  for all  $3 \leq k \leq K$ . Now denote a SeDJoCo solution for the newly ordered set as  $\tilde{\mathbf{B}}$ , satisfying

$$\tilde{\mathbf{B}} \tilde{\mathbf{Q}}_k \tilde{\mathbf{B}}^T \mathbf{e}_k = \mathbf{e}_k \quad k = 1, 2, \dots, K. \quad (6)$$

Next, define  $\mathbf{\Pi}_{1,2}$  as the (symmetric) permutation matrix that swaps the first and second elements of a vector, namely  $\mathbf{\Pi}_{1,2} \mathbf{e}_1 = \mathbf{e}_2, \mathbf{\Pi}_{1,2} \mathbf{e}_2 = \mathbf{e}_1$  and  $\mathbf{\Pi}_{1,2} \mathbf{e}_k = \mathbf{e}_k$  for all other  $3 \leq k \leq K$ . Now consider the matrix  $\mathbf{B}' = \mathbf{\Pi}_{1,2} \tilde{\mathbf{B}}$ . We assert that this matrix solves the SeDJoCo equations for the *original* set  $\{\mathbf{Q}_k\}_{k=1}^K$ , since

$$\begin{aligned} \mathbf{B}' \mathbf{Q}_1 \mathbf{B}'^T \mathbf{e}_1 &= \mathbf{\Pi}_{1,2} \tilde{\mathbf{B}} \tilde{\mathbf{Q}}_2 \tilde{\mathbf{B}}^T \mathbf{\Pi}_{1,2}^T \mathbf{e}_1 \\ &= \mathbf{\Pi}_{1,2} \tilde{\mathbf{B}} \tilde{\mathbf{Q}}_2 \tilde{\mathbf{B}}^T \mathbf{e}_2 = \mathbf{\Pi}_{1,2} \mathbf{e}_2 = \mathbf{e}_1, \end{aligned} \quad (7)$$

and, similarly,  $B'Q_2B'^T e_2 = e_2$ , with  $B'Q_kB'^T = e_k$  for all other  $3 \leq k \leq K$ . This means that we have found an additional solution to the *original* SeDJoCo problem, via a permutation of a solution to a *permuted* SeDJoCo problem.

Note further, that the nonlinear set of “permuted” SeDJoCo equations applied to the elements of  $\tilde{B}$  in (6) is essentially different from the original equations applied to the elements of  $B$  in (5), namely - different elements of the target matrices multiply different products of elements of the  $B$  matrix in these two sets. Consequently, the resulting elements of  $B'$  would generally be essentially different from the elements of  $B$  (i.e., the difference between these solutions would generally not be confined merely to permutation of their elements).

Now, since any permutation matrix can be expressed as the product of two-elements-permutation matrices, we may extend the above result to claim that, in general, the number of “essentially different” solutions of a  $K$ -dimensional SeDJoCo problem may be at least as large as the number of possible permutations of the set, which is  $K!$ . Each of these provides an exact solution of the likelihood equations (in the context of ML BSS), and therefore corresponds to a stationary point of the likelihood function (of  $B$ ), but only one of these corresponds to the true (global) ML estimate of  $B$ . It is therefore of significant interest to be able to detect a sub-optimal solution, and, moreover, to try to reach the “correct” solution. Fortunately, this is possible (at least under asymptotic conditions) in the context of BSS when the source signals are stationary - as we shall explain in the following section.

#### 4. IDENTIFYING AND RECTIFYING SUB-OPTIMAL SOLUTIONS

Under asymptotic conditions in BSS, even sub-optimal solutions of SeDJoCo provide reasonable (albeit sub-optimal) separation of the sources, because the solution of the SeDJoCo equations always amounts to hybrid exact-approximate joint diagonalization of the sources [5], where the exact diagonalization applies to the “drilled” row (and column) in each target matrix, and the approximate diagonalization follows by applying to all other rows (and columns) in these matrices. Such a solution, even when not related to the global maximum of the likelihood, provides reasonable separation.

Quite commonly in BSS, the solution is only obtained up to possible permutation and scaling of the sources, which are, in a fully blind scenario, inevitable ambiguities. Fortunately, however, in the context of ML BSS, when the statistical models of the sources are available, the permutation and scaling ambiguities can be eliminated, since each of the reconstructed sources can be tested against its available statistical model and the different sources can be identified and ordered according to their empirical match to the respective models. This feature enables us to detect cases where the reconstructed (separated) sources are permuted (and possibly scaled) relative to their

expected order - which would in turn mean that the solution at hand is a “permuted” solution, and as such is too far from the “true” (ML) solution (the likelihood’s global maximum).

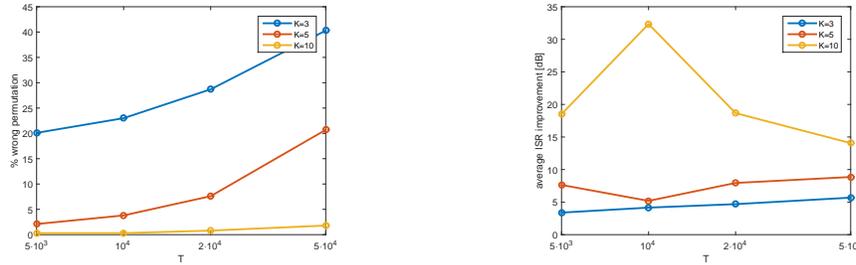
Moreover, after identifying the permutation (and possible scaling) of the reconstructed sources, we may de-permute (and rescale) the estimated separation matrix, thereby bringing it closer to the “true” ML solution. We can then use the de-permuted  $B$  matrix to initialize an iterative SeDJoCo solution, increasing the chances of convergence to the true solution, since the starting-point would be much “closer” to that solution.

A most convenient case in which the reconstructed sources can be compared against their statistical models is the case of stationary source signals with known spectra [2]. Assuming such stationarity, our proposed identification of sub-optimal solutions and the associated remedy take the following steps:

- i. Given a solution  $\hat{B}$ , estimate the separated sources as  $\hat{S} = \hat{B}X$ ;
- ii. Use any consistent spectrum estimation method to estimate the spectra of the  $K$  reconstructed sources;
- iii. To each of the  $K$  estimated spectra, find the “best match” among the  $K$  known spectra. More specifically, denoting the known spectra as  $\{P_k(e^{j\omega})\}_{k=1}^K$  and the estimated spectra as  $\{\hat{P}_k(e^{j\omega})\}_{k=1}^K$ , the goal is to find for each  $k$  its respective index  $\ell_k \in \{1, \dots, K\}$  and scaling parameter  $\gamma_k \in \mathbb{R}$ , via

$$\{\ell_k, \gamma_k\} = \arg \min_{\ell, \gamma} \int_0^{2\pi} \left( P_\ell(e^{j\omega}) - \gamma^2 \cdot \hat{P}_k(e^{j\omega}) \right)^2 d\omega. \quad (8)$$

However, we would need to find the optimal association of each of the known spectra to each of the estimated spectra. Rather than trying all possible associations (permutations), we can take a sub-optimal greedy approach, e.g., by first associating the best-matched pair, then eliminating the already matched spectra from both groups and proceeding to the next best match, and so forth. At the end of this association process we end up with a permutation matrix  $\Pi$ , such that its  $k$ -th row has a “1” at the corresponding  $\ell_k$ -th column. We also construct a diagonal scaling matrix  $\Gamma$ , which has the respective optimal scaling coefficients,  $\gamma_k$ , as its  $(k, k)$ -th elements. Note that normally, if  $\hat{B}$  is an optimal solution, we should get  $\Pi = I_K$  (the identity matrix) and  $\Gamma \approx I_K$ , because the optimal solution should be permutation-free, and the respective scaling is taken care of in the SeDJoCo solution (the only deviations of  $\Gamma$  from  $I_K$  would be due to errors in the estimated spectra). However, sometimes  $\Pi$  would be different from  $I_K$  - implying a sub-optimal SeDJoCo solution, for which we should take the remedy steps outlined in the sequel.



**Fig. 2.** (left) Empirical percentage of obtaining a sub-optimal (permuted) solution; (right) Average improvement in ISR attained by the re-initialization strategy (for the cases of sub-optimal solutions only).

- iv. If  $\Pi \neq \mathbf{I}_K$ , a new initial guess for  $\mathbf{B}$  can be constructed as  $\hat{\mathbf{B}} \triangleq \Gamma \Pi \hat{\mathbf{B}}$ . Note that, by construction, if  $\hat{\mathbf{B}}$  is applied to the mixtures, the new separated sources  $\hat{\mathbf{S}} = \hat{\mathbf{B}}\mathbf{X}$  would each be associated with its respective spectrum (with a scaling factor close to 1). Note, however, that  $\hat{\mathbf{B}}$  (as opposed to  $\mathbf{B}$ ) is *not* a solution of the SeDJoCo equations, and therefore it is not the ML estimate of  $\mathbf{B}$ . Nevertheless, we can now create a new set of target matrices as  $\tilde{\mathbf{Q}}_k \triangleq \hat{\mathbf{B}}\mathbf{Q}_k\hat{\mathbf{B}}^T$  (for  $k = 1, \dots, K$ ), and feed this new set to an appropriate iterative SeDJoCo-solving algorithm. In effect, the transformation of the set constitutes a new “initial guess” (or “starting point”) for the algorithm, and if that algorithm has a local convergence property, it would be likely to converge to a new  $\hat{\mathbf{B}}'$ , which would be close to an identity matrix, yet would provide an exact SeDJoCo solution for the new set.
- v. The ultimate solution of the original SeDJoCo problem would then be given by  $\mathbf{B}_o \triangleq \hat{\mathbf{B}}'\hat{\mathbf{B}}$ , which, by construction, provides an exact solution to the original SeDJoCo problem (with the original  $\mathbf{Q}_k$  matrices), and yet (if the iterative algorithm did not deviate too far from the initial state) has the correct permutation and scaling.

## 5. SIMULATION RESULTS

To demonstrate the proposed identification and correction approach, we simulated the following BSS scenario:  $K$  sources are generated by passing white Gaussian noise through  $K$  FIR filters of length 12, with randomly chosen impulse response coefficients (independent, Gaussian distributed with zero mean and unit variance). These sources are then mixed using a random mixing-matrix  $\mathbf{A}$  (similarly drawn from a Gaussian distribution). We use  $T$  samples to construct the  $K$  target matrices  $\mathbf{Q}_1, \dots, \mathbf{Q}_K$  as per (1), where the known covariance matrices  $\mathbf{C}_k$  are obtained from the FIR filters (the stationarity and the ensuing Toeplitz structure of all  $\mathbf{C}_k$  are

used for efficient computation in the spectral domain, using the sources’ known spectra, see [2] for details).

We ran 1000 independent trials, showing (for several values of  $K$  and  $T$ ) the empirical percentage of sub-optimal (permuted) solutions, as well as the average improvement in separation performance (in terms of the residual Interference to Source Ratio (ISR)) resulting from the correction strategy proposed above. The presented improvement applies only to the cases of permuted solutions - when the optimal (non-permuted) solution is obtained upfront, it is detected as such (by the proposed approach), and no correction is needed or applied. All of the presented simulation results were obtained using the NCG algorithm presented in [4], which has a local convergence property, and is therefore sensitive to initialization and can be guided to the optimal solution by the proposed method. We note that not all iterative SeDJoCo algorithms share that property - for example, according to our experience, a JMMS-based algorithm (based on [8]) almost always converges to a permuted solution, and is almost insensitive to initialization.

Evidently, the probability of obtaining a permuted solution decreases with  $K$  but increases with  $T$ . This can be attributed to the fact that for smaller values of  $K$  and for larger values of  $T$ , the target matrices are almost jointly-diagonalizable, which means that the different SeDJoCo solutions are in fact “close” to each other (in some sense), and the probability of moving from the optimal to a sub-optimal solution increases. However, sub-optimal solutions are significantly more adverse when  $K$  is large, hence the corrective action attains a more significant improvement in such cases.

## 6. CONCLUSION

We identified and characterized the existence of multiple solutions to the SeDJoCo problem. In the context of ML Semi-Blind Source Separation, solutions which do not correspond to the global maximum of the likelihood function are sub-optimal and are characterized by a detectable permutation of the resulting separated sources. The detection can be used to re-initialize an iterative algorithm so as to have it redirected to the optimal solution.

## 7. REFERENCES

- [1] D.-T. Pham, "Blind Separation of Instantaneous Mixture of Sources via the Gaussian Mutual Information Criterion," *Signal Processing*, vol. 81, pp. 855 – 870, 2001.
- [2] A. Yeredor, "Blind Separation of Gaussian Sources with General Covariance Structures: Bounds and Optimal Estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 10, pp. 5057 – 5068, 2010.
- [3] P. Comon and C. Jutten, *Handbook of Blind Source Separation*, Academic Press, 2010.
- [4] A. Yeredor, B. Song, F. Roemer, and M. Haardt, "A "Sequentially Drilled" Joint Congruence (SeDJoCo) Transformation with Applications in Blind Source Separation and Multi-User MIMO Communication," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2744 – 2757, 2012.
- [5] A. Yeredor, "On Hybrid Exact-Approximate Joint Diagonalization," *Proc. 3rd IEEE workshop on Computational Advances in Multi-Channel Sensor Array Processing, (CAMSAP'2009)*, 2009.
- [6] S. Dgerine and A. Zadi, "Separation of an instantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach," *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1499 – 1512, 2004.
- [7] B. Song, F. Roemer, and M. Haardt, "Using a New Structured Joint Congruence (STJOCO) Transformation of Hermitian Matrices for Precoding in Multi-User MIMO Systems," in *Proc. IEEE ICASSP 10*, 2010.
- [8] K. Wong, R. D. Murch, and K. B. Letaief, "A Joint-Channel Diagonalization for Multiuser MIMO Antenna Systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, July 2003.