A new SDP relaxation based blind beamforming technique

Erion-Vasilis Pikoulis and Emmanouil Z. Psarakis

Department of Computer Engineering and Informatics, University of Patras, 26500 Rio-Patras, Greece
email:{pikoulis, psarakis}@ceid.upatras.gr
phone: +30 2610 996969, fax: +30 2610 996971

***Abstract*— A unified framework for combining datasets of similar signals in order to increase the SNR of the output is formulated and solutions to the arising problems of signal alignment and weight assignment, are proposed. The signal alignment problem is formulated as a combinatorial optimization problem and an approximate solution is proposed by using the technique of SDP relaxation. On the other hand, the weight assignment problem requires the solution of a quadratic maximization problem which in the vast majority of cases has an analytical solution, while in more challenging noise conditions can be approximately solved via SDP relaxation. The greatest advantage of the proposed technique lies in its ability to obtain very robust and accurate estimations of the pairwise time delays between the signals. This makes it suitable for a variety of applications requiring time delay estimation, signal enhancement, or both. The superior performance of the technique compared to other similar approaches is demonstrated through a number of numerical simulations with several signal and noise models.**

I. Introduction

Datasets of similar signals occur mainly under two scenarios. Either when the signal from a single source is recorded by several similar and closely spaced sensors, or when signals from several similar and closely spaced sources are recorded by a single sensor. Applications involving sensor arrays are the most pronounced examples of the first scenario, while seismic relocation procedures involving the analysis of seismic clusters (that is, groups of similar events), constitute prime representatives of the second scenario. In both cases we are dealing with datasets of similar signals, that are however recorded with unknown delays, and corrupted by noise of unknown statistics.

The process of aligning (by estimating the individual time delays) and combining, (i.e., averaging) the input signals in order to produce a higher quality output with enhanced signal-to-noise ratio (SNR) is generally known as beamforming. The increasing use of sensor arrays in several fields including wireless communication, radar, sonar, speech processing, seismic prospecting, medical imaging, surveillance, and others, has turned beamforming into a topic of intensive and ongoing research.

The specific solution to the problem of beamforming depends on the assumptions that can be made regarding the placement of the sensors, the characteristics of the recorded signals, as well as the statistical properties of the noise. There is an extensive literature treating the problem of beamforming under the scenario of narrowband signals and precisely [1]–[5] or imprecisely [6]–[9] known sensor location and responses.

In case the information regarding the characteristics of the array is (totally or partially) missing, the beamforming problem is referred to as blind beamforming. The majority of techniques that have been proposed for the solution of the blind beamforming problem make specific assumptions regarding the nature of the recorded signals and/or the statistical properties of the noise. The constant modulus algorithms (CMA) [10], [11] and the higher order statistics (HOS) methods [12]–[15] belong to this category of blind beamforming techniques.

In this manuscript, we treat blind beamforming in its most general form, as defined in [16]. We do not make assumption regarding sensor placement, specific signal features, or specific noise models. We only assume a dataset of similar (but not necessarily identical) signals, recorded with unknown delays, and corrupted by uncorrelated noise. Also, it is implicitly assumed that the dataset consists of signals from a single source, or at least a dominant one, i.e. a source whose signal is the strongest in the data set.

The goal of the proposed technique is to obtain an enhanced representative of the dataset, that is, a combined output that retains the signal information of the input and attenuates the noise content. More formally, the problem at hand consists of the estimation of the time shifts (delays) that yield the best possible joint alignment of the signals, and of the non-negative weights that must be applied to the aligned signals, so that the SNR of their weighted sum is maximized.

The proposed beamforming framework is especially applicable in conditions where exact knowledge and control over the input is missing, as is it the case, for example, with seismic arrays and seismological datasets in general [17]–[19]. Particularly in reflection seismology, the problem at hand can be considered as a generalization of what is referred to as optimal stacking [20]–[22]. The generalization occurs because in typical stacking procedures the signals are considered identical (apart from amplitude scaling) and perfectly aligned, with the optimization being targeted exclusively towards the weights.

The remaining of this manuscript is organized as follows. In Section II we present several useful quantities and notation remarks. In Section III we present a general formulation of the problem at hand and subsequently, in Section IV, we give an outline of the proposed blind beamforming technique. In

Sections V and VI the proposed solutions to the ensuing problems of signal alignment and weight assignment, respectively, are presented and in detail. In Section VII we present an outline of the proposed algorithm and hold a brief discussion regarding the estimation of the required parameters. Section VIII contains our simulation results. Section IX contains our conclusions and finally, the mathematical proofs are presented in Appendices A-E.

II. Definitions

Available dataset: Let $x_i(n)$ denote a collection of M time series, defined as follows:

$$\begin{aligned} x_i(n) &= s_i(n) + w_i(n), \\ n = 0, 1, \ldots, N-1, &\ i = 1, \ldots, M, \end{aligned} \tag{1}$$

where $s_i(n)$, $w_i(n)$ denote the i-th signal and i-th noise process, respectively. The involved noise processes are assumed white, and jointly uncorrelated. The variance of the $i-$th noise process is denoted as σ_i^2.

Correlation and signal alignment: The signal correlation function

$$r_{ij}(\kappa) \triangleq \frac{1}{N} \sum_{n=0}^{N-1} s_i(n) s_j(n+\kappa), \tag{2}$$

quantifies the alignment quality between signals $s_i(n)$, $s_j(n)$, at a time lag equal to κ samples. Thus, if we define

$$\kappa_{ij}^{\max} = \underset{\kappa}{\operatorname{argmax}}\ r_{ij}(\kappa), \tag{3}$$

$$r_{ij}^{\max} = \max_{\kappa} r_{ij}(\kappa) = r_{ij}(\kappa_{ij}^{\max}), \tag{4}$$

the (i, j)-th signal pair is optimally aligned when $s_i(n)$ is delayed by $\kappa_{ij}^{\max}$ samples relative to $s_j(n)$.

In this sense, if there exists a set $\tau_1^*, \ldots, \tau_M^*$ of time-delays so that:

$$\tau_i^* - \tau_j^* = \kappa_{ij}^{\max}, \quad \forall\ 1 \leq i, j \leq M, \tag{5}$$

then, delaying $s_i(n)$ by τ_i^* samples, for $i = 1, \ldots, M$, we obtain a delayed version of the dataset, where an optimal alignment has been reached for all possible signal pairs. We will refer to this condition, as a jointly optimal alignment of the signals.

Matrices: The standard inner-product $\langle X, Y \rangle$ of two $m \times n$ matrices X, Y is defined as follows:

$$\langle X, Y \rangle = \sum_{p=1}^{m} \sum_{q=1}^{n} (X)_{pq} (Y)_{pq}, \tag{6}$$

where $(X)_{pq}$ denotes the element at the p-th row and q-th column of X.

We will use $X \geq 0 (> 0)$ in order to denote non-negative (positive) matrices, namely matrices whose elements are non-negative (positive). On the other hand, positive semi-definite matrices will be denoted as $X \succeq 0$.

Finally, we state the following property of non-negative matrices, which is due to the Perron-Frobenius theorem [23]:

$$X \geq 0 (> 0) \Rightarrow \boldsymbol{\xi}_{\max}(X) \geq 0 (> 0), \tag{7}$$

where $\boldsymbol{\xi}_{\max}(X)$ denotes the eigenvector corresponding to the largest eigenvalue of X.

III. Problem formulation

Let $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_M]^t$, $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_M]^T$ denote an array of time-delays and an array of non-negative weights, respectively, and let $\bar{y}(n; \boldsymbol{\gamma}, \boldsymbol{\tau})$ be defined as the weighted average of the delayed version of the available dataset, namely:

$$\bar{y}(n; \boldsymbol{\gamma}, \boldsymbol{\tau}) \triangleq \sum_{i=1}^{M} \gamma_i x_i(n - \tau_i). \tag{8}$$

If we denote as $\mathrm{snr}_{\bar{y}}(\boldsymbol{\gamma}, \boldsymbol{\tau})$ the SNR of $\bar{y}(n; \boldsymbol{\gamma}, \boldsymbol{\tau})$, the problem at hand is finding the optimal $\boldsymbol{\gamma}$,$\boldsymbol{\tau}$, so that $\mathrm{snr}_{\bar{y}}(\boldsymbol{\gamma}, \boldsymbol{\tau})$ is maximized, i.e.:

$$\max_{\boldsymbol{\gamma}, \boldsymbol{\tau}} \mathrm{snr}_{\bar{y}}(\boldsymbol{\gamma}, \boldsymbol{\tau}), \quad \text{s.t. } \boldsymbol{\gamma} \geq 0. \tag{9}$$

In the following sections we first state the two fundamental subproblems that arise from the general problem defined in Eq. (9) and subsequently, we present in detail the methodology used for their solution.

IV. Outline of the proposed solution

After some simple mathematical manipulations and by taking into account the assumed stationarity of the involved processes, the SNR of $\bar{y}(n; \boldsymbol{\gamma}, \boldsymbol{\tau})$, can be expressed compactly as follows:

$$\mathrm{snr}_{\bar{y}}(\boldsymbol{\gamma}, \boldsymbol{\tau}) = \frac{\boldsymbol{\gamma}^T R(\boldsymbol{\tau}) \boldsymbol{\gamma}}{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}, \tag{10}$$

where $R(\boldsymbol{\tau})$, Σ, are the signal correlation and noise covariance matrices, respectively. More specifically, the (i, j)-th element of $R(\boldsymbol{\tau})$ equals $r_{ij}(\tau_i - \tau_j)$, while Σ is a diagonal matrix having σ_i^2 in its i-th diagonal position. $R(\boldsymbol{\tau})$ summarizes the achieved alignment quality when the dataset is delayed according to the entries of $\boldsymbol{\tau}$.

As it can be seen from Eq. (10), if there exists an array $\boldsymbol{\tau}^* = [\tau_1^*, \ldots, \tau_M^*]^T$ that satisfies (5), then $\boldsymbol{\tau}^*$ maximizes $\mathrm{snr}_{\bar{y}}(\boldsymbol{\gamma}, \boldsymbol{\tau})$ with respect to $\boldsymbol{\tau}$, regardless of the value of $\boldsymbol{\gamma} \geq 0$. In this case, a jointly optimal alignment of the signals can be reached and problem (9) becomes separable.

In the general case, a solution of (9) can be obtained via an iterative scheme of alternating optimizations over the two sets of parameters. This approach yields the following two maximization problems:

P_1: *Signal alignment*. For a fixed $\boldsymbol{\gamma}$, the time-delays that maximize the SNR of $\bar{y}(n)$ are the ones that maximize the numerator of Eq. (10), or equivalently, solve the following problem (note that the diagonal elements of $R(\boldsymbol{\tau})$ are independent of $\boldsymbol{\tau}$):

$$\max_{\boldsymbol{\tau}} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \gamma_i \gamma_j r_{ij}(\tau_i - \tau_j), \tag{11}$$

which constitutes the signal alignment problem. It should be noticed that since any global shift of the time series has no effect on the pairwise alignment of the signals, there are (theoretically) infinitely many solutions for problem (11). In other words, if $\boldsymbol{\tau}$ solves (11), then the same holds for any $\boldsymbol{\tau} + c\mathbf{1}$, $c \in \Re$, where $\mathbf{1}$ denotes the all-ones vector.

P_2: *Weight assignment*. For a given $\boldsymbol{\tau}$ (that is, for a given alignment of the signals), the weights that must be used in the averaging process in order to maximize the SNR of $\bar{y}(n)$, are obtained from the solution of the following constrained optimization problem:

$$\max_{\boldsymbol{\gamma}\geq 0} \quad \frac{\boldsymbol{\gamma}^T R(\boldsymbol{\tau})\boldsymbol{\gamma}}{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}. \tag{12}$$

We will refer to this as the weight assignment problem.

As it is obvious from their definitions, P_1 and P_2 are non-convex, NP-hard problems, meaning that in the general case, closed form solutions are not possible. However, by using the available knowledge, which is summarized in the correlation sequences $r_{ij}(\kappa)$ and the noise covariance matrix Σ, we are going to show that approximate solutions for these problems can be found. We then suggest ways to combine these solutions for an ultimate solution of (9).

Finally, we note that the estimation of the required data-related quantities is treated separately in Section VII-A.

V. Solving the signal alignment problem

A. A filtering-based reformulation

Let $\boldsymbol{h}_i = [h_{i,0}, \ldots, h_{i,(L-1)}]^T$, $i = 1, \ldots, M$ denote M filters of length L that are defined as follows:

$$h_{i,n} = \begin{cases} \gamma_i, & n = \tau_i \\ 0, & n \neq \tau_i \end{cases}, \ \gamma_i > 0, \ 0 \leq n \leq L-1, \tag{13}$$

where $0 \leq \tau_i \leq L-1$. In other words, each filter has a single non-zero coefficient. For $\boldsymbol{h}_i$, this coefficient equals $\gamma_i > 0$ [1] and it is located at the τ_i-th position. Finally, we assume that L is greater than the greatest pairwise time-difference of the signals, i.e.:

$$L \geq \max_{1\leq i,j\leq M} |\kappa_{ij}^{\max}|, \tag{14}$$

where $\kappa_{ij}^{\max}$ was defined in Eq. (3).

Let also $\boldsymbol{s}_i(n)$, $i = 1, \ldots, M$ denote a signal array of length L, defined as follows:

$$\boldsymbol{s}_i(n) = [s_i(n), s_i(n-1), \ldots, s_i(n-L+1)]^T. \tag{15}$$

Then, we can write:

$$s_i(n) * h_{i,n} = \boldsymbol{h}_i^T \boldsymbol{s}_i(n) = \gamma_i s(n-\tau_i), \tag{16}$$

meaning that the summation terms in Eq. (11) can be expressed as follows:

$$\gamma_i\gamma_j r_{ij}(\tau_i - \tau_j) = \boldsymbol{h}_i^T R_{ij} \boldsymbol{h}_j, \tag{17}$$

where

$$R_{ij} = \sum_{n=0}^{N-1} \boldsymbol{s}_i(n)\boldsymbol{s}_j^T(n), \tag{18}$$

is an $L \times L$ Toeplitz matrix whose (p,q)-th element equals $r_{ij}(p-q)$. Thus, if we define

$$\mathbf{R} = \begin{bmatrix} 0 & R_{12} & \cdots & R_{1M} \\ R_{21} & 0 & \cdots & R_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ R_{M1} & R_{M2} & \cdots & 0 \end{bmatrix}, \ \mathbf{h} = \begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \\ \vdots \\ \boldsymbol{h}_M \end{bmatrix}, \tag{19}$$

then, problem (11) can be equivalently rewritten as follows:

$$\max_{\mathbf{h}} \quad \mathbf{h}^T \mathbf{R} \mathbf{h} \tag{20}$$

$$\text{s.t.} \quad \boldsymbol{h}_i \text{ satisfies (13)}, \ 1 \leq i \leq M. \tag{21}$$

Thus, for a given set of γ_i's, the signal alignment problem as formulated in (20)-(21), seeks for the optimal placement of the non-zero coefficient in each $\boldsymbol{h}_i$, so that the quadratic form in Eq. (20) is maximized. This is of course a problem of combinatorial complexity.

A very useful property whose importance will become shortly apparent is stated in the following proposition.

Proposition 1: Condition (13) is equivalent to the combination of the following three constraints:

$$\boldsymbol{h}_i^T \boldsymbol{h}_i = \gamma_i^2, \quad (\boldsymbol{h}_i^T \mathbf{1})^2 = \gamma_i^2, \quad \boldsymbol{h}_i \geq 0, \tag{22}$$

where $\mathbf{1}$ denotes the all-ones vector of length L.

Proof: For a proof of Proposition 1, see Appendix A. ■

Using the equivalence of the constraints, we can reformulate the alignment problem as follows:

$$\max_{\mathbf{h}} \quad \mathbf{h}^T \mathbf{R} \mathbf{h}, \tag{23}$$

$$\text{s.t.} \quad \boldsymbol{h}_i^T \boldsymbol{h}_i = \gamma_i^2, \ 1 \leq i \leq M, \tag{24}$$

$$(\boldsymbol{h}_i^T \mathbf{1})^2 = \gamma_i^2, \ 1 \leq i \leq M, \tag{25}$$

$$\mathbf{h} \geq 0. \tag{26}$$

Due to the non-convex equality constraints (24), problem (23)-(26) is still NP-hard. Nevertheless, it presents us with a path towards an approximate solution via the SDP relaxation technique [24], as we are going to see shortly. In order to apply this technique, we first need to transform the quadratic maximization problem (23)-(26) into an equivalent trace maximization problem.

To this end, let us state the following equivalent expressions:

$$\mathbf{h}^T \mathbf{R} \mathbf{h} = \text{trace}(\mathbf{R}\mathbf{h}\mathbf{h}^T) = \text{trace}(\mathbf{R}\mathbf{H}), \tag{27}$$

$$\boldsymbol{h}_i^T \boldsymbol{h}_i = \text{trace}(\boldsymbol{h}_i \boldsymbol{h}_i^T) = \text{trace}(H_{ii}), \tag{28}$$

$$(\boldsymbol{h}_i^T \mathbf{1})^2 = \mathbf{1}^T \boldsymbol{h}_i \boldsymbol{h}_i^T \mathbf{1} = \mathbf{1}^T H_{ii} \mathbf{1}, \tag{29}$$

where $\mathbf{H} = \mathbf{h}\mathbf{h}^T$ is a rank-one, $LM \times LM$ matrix consisting of M^2 $L \times L$ blocks, arranged as follows:

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1M} \\ H_{12}^T & H_{22} & \cdots & H_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ H_{1M}^T & H_{2M}^T & \cdots & H_{MM} \end{bmatrix}, \tag{30}$$

so that $H_{ij} = \boldsymbol{h}_i \boldsymbol{h}_j^T$. Note finally that if $\mathbf{H}$ is non-negative, i.e. $\mathbf{H} \geq 0$, then $\mathbf{h}$ can always be chosen non-negative, i.e. $\mathbf{h} \geq 0$.

By using (27)-(30), the original version of the signal alignment problem can be formulated as the following trace

[1] Since, for $\gamma_i = 0$ the i-th signal will be discarded in the averaging process (thus rendering its alignment to the other signals irrelevant), for simplicity and without loss of generality, we only treat the case where γ_i is strictly positive.

maximization problem:

$$\begin{aligned}\phi_o^* = \max_{\mathbf{H}} \quad & \text{trace}(\mathbf{RH}), && (31)\\ \text{s.t.} \quad & \text{trace}(H_{ii}) = \gamma_i^2,\ 1 \le i \le M, && (32)\\ & \mathbf{1}^T H_{ii} \mathbf{1} = \gamma_i^2,\ 1 \le i \le M, && (33)\\ & \mathbf{H} \ge 0 && (34)\\ & \text{rank}(\mathbf{H}) = 1. && (35)\end{aligned}$$

The only non-convex constraint in (31)-(35) is the rank-one constraint (35), since all other constraints and the objective function are linear in $\mathbf{H}$. By applying the SDP relaxation technique we drop the non-convex constraint (35) and replace it with $\mathbf{H} \succeq 0$, which is convex. Thus, the relaxed version of (31)-(35) takes the following form:

$$\begin{aligned}\phi_r^* = \max_{\mathbf{H}} \quad & \text{trace}(\mathbf{RH}), && (36)\\ \text{s.t.} \quad & \text{trace}(H_{ii}) = \gamma_i^2,\ 1 \le i \le M, && (37)\\ & \mathbf{1}^T H_{ii} \mathbf{1} = \gamma_i^2,\ 1 \le i \le M, && (38)\\ & \mathbf{H} \ge 0, && (39)\\ & \mathbf{H} \succeq 0, && (40)\end{aligned}$$

which constitutes a semi-definite program with variable $\mathbf{H} \in \Re^{LM \times LM}$.

B. The effect of relaxation

In this subsection we compare the solutions that are obtained by solving the original and the relaxed versions of the problem. More specifically, our goal is to investigate the properties of matrix $\mathbf{H}$ under the rank-one and the positive semi-definite constraints, respectively. To this end, the following analytic expression of $\text{trace}(\mathbf{RH})$ as a function of the off-diagonal blocks of $\mathbf{H}$:

$$\text{trace}(\mathbf{RH}) = \langle \mathbf{R}, \mathbf{H} \rangle = \sum_{i=1}^{M} \sum_{\substack{j=1\\ j\neq i}}^{M} \langle R_{ij}, H_{ij} \rangle, \tag{41}$$

where $\langle X, Y \rangle$ denotes the inner product of matrices X, Y, will prove itself very helpful, as we are going to see. Let us now discriminate between the two versions of the problem.

Original problem: In the original problem, constraints (32)-(34) impose diagonality on the diagonal blocks H_{ii} of the solution $\mathbf{H}$. Combined with the rank-one constraint, this yields the condition that every diagonal block of $\mathbf{H}$ has a single non-zero element, equal to γ_i^2 and placed in its main diagonal. Thus, if γ_i^2 is located at the (p_i, p_i)-th location of H_{ii}, due to the rank-one constraint, each off-diagonal block H_{ij} has also a single non-zero element, equal to $\gamma_i \gamma_j$ and placed at the (p_i, p_j)-th location, with $1 \le p_i, q_j \le L$, $1 \le i, j \le M$. Thus, by taking into account that the (p_i, p_j)-th element of R_{ij} equals $r_{ij}(p_i - p_j)$, we can write:

$$\langle R_{ij}, H_{ij} \rangle = \gamma_i \gamma_j r_{ij}(p_i - p_j). \tag{42}$$

Consequently, the goal of the original problem is to place the single non-zero element of H_{ij}, in a location that corresponds to the largest possible value of R_{ij} (for all $1 \le i, j \le M$), while ensuring the rank-one requirement of $\mathbf{H}$.

Note also that, since all R_{ij}'s are Toeplitz matrices, by adding the same integer c to all p_i's we obtain a different, equivalent solution for (32)-(34). This is because global translations of the p_i's shift uniformly the locations of the non-zero elements of $\mathbf{H}$, without affecting the diagonals that the non-zero elements are placed at (since, $p_i + c - (p_j + c) = p_i - p_j$).

Relaxed problem: As it was the case with the original problem, constraints (37)-(39) impose diagonality on the diagonal blocks H_{ii} of the solution $\mathbf{H}$. However, since the rank-one constraint is relaxed, there is no limitation on the number of the non-zero elements that can be placed on the main diagonal of each H_{ii}. This, in turn, means that the number of the non-zero elements of the off-diagonal blocks H_{ij} is not restricted either.

Nevertheless, a crucial feature that the relaxed solution inherits from the original one, is the following: while in the original solution the single non-zero element in every H_{ij} equals $\gamma_i \gamma_j$, in the relaxed solution it holds $\Sigma_{H_{ij}} \le \gamma_i \gamma_j$, where $\Sigma_{H_{ij}}$ denotes the sum of the elements of H_{ij}. This is due to the diagonality of the diagonal blocks of $\mathbf{H}$ and results from the application of Proposition 2. Combining this fact with the non-negativity of $\mathbf{H}$, leads to the conclusion that the relaxed problem has a natural tendency towards sparse solutions, as shown by Proposition 3.

Specifically, by taking into account Proposition 3, the objective of the relaxed problem can be expressed as the problem of concentrating the maximum permitted value of $\Sigma_{H_{ij}}$, i.e., $\gamma_i \gamma_j$, to the fewest possible diagonals of H_{ij} (corresponding to the largest possible correlation values), while ensuring the positive definiteness of $\mathbf{H}$.

When this goal is achieved to its fullest, namely when every H_{ij} has a single non-zero diagonal, then the relaxed solution assumes a specific structure that ensures its equivalence to the original solution, as it proven in Lemma 1. Additionally, Lemma 2 informs us that a sufficient (but not necessary) condition for such an equivalence is the ability of the (noisy) signals to become jointly optimally aligned (namely, the existence of delays that satisfy Eq. (5)).

In the general case, the sparsity of $\mathbf{H}$ depends on the structure of $\mathbf{R}$ (and more specifically, on the placement of the correlation values along the diagonals of the R_{ij}'s), which, in turn, depends on the nature of the involved signals. It could be said that, the higher the similarity of the (noisy) signals, the better the joint alignment that can be achieved. Thus, the higher the similarity of the signals, the sparser the solution of the relaxed problem (measured by the non-zero diagonals of H_{ij}) and consequently, the closer it is to the solution of the original problem.

Finally, it must be stressed that in our experience, $\mathbf{H}$ retains its sparsity characteristic (thus ensuring the solution of the original problem) even under very unfavourable noise conditions, provided that the underlying signals share a similar enough structure. The example of Fig. 1 constitutes such a case.

A mathematical analysis follows.

Analysis: Since the solution of the relaxed problem, $\mathbf{H}$, is positive semi-definite, the same must hold for all its principal sub-matrices, including the following ones:

$$\mathbf{H}_{ij} = \begin{bmatrix} H_{ii} & H_{ij} \\ H_{ij}^T & H_{jj} \end{bmatrix}, \quad 1 \leq i, j \leq M, \ i \neq j, \tag{43}$$

which yields the following necessary conditions:

$$(H_{ij})_{pq} \leq \sqrt{(H_{ii})_{pp}(H_{jj})_{qq}} \leq \gamma_i \gamma_j, \tag{44}$$

$$\Sigma_{H_{ij}} \leq \sqrt{\Sigma_{H_{ii}} \Sigma_{H_{jj}}} = \gamma_i \gamma_j, \tag{45}$$

where Σ_X denotes the sum of the elements of matrix X. Eq. (44) represents a well-known property of positive semi-definite matrices, while (45) results from Proposition 2 that follows and by taking into account constraint (38).

Proposition 2: Let Y be a $k \times k$ positive semi-definite matrix presented in the following 2×2 block-symmetric form:

$$Y = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}, \tag{46}$$

where A, C, are square matrices of dimensions $k_1 \times k_1$, $k_2 \times k_2$, respectively, while B is a matrix of size $k_1 \times k_2$, with $k_1 + k_2 = k$. Let also Σ_X denote the sum of the elements of matrix X. Then, the following relations hold:

$$\Sigma_A \geq 0, \quad \Sigma_C \geq 0, \tag{47}$$

$$|\Sigma_B| \leq \sqrt{\Sigma_A \Sigma_C}. \tag{48}$$

Proof: For a proof of Proposition 2, see Appendix B. ■

As we it can be seen from Eqs. (44), (45), under constraints (37)-(40), the sum of all the elements of H_{ij} is bounded by $\gamma_i \gamma_j$, as is every element of H_{ij} separately. Since the elements of H_{ij} are non-negative, this leaves (45) as the only necessary condition.

Thus, the total sum $\Sigma_{H_{ij}}$ can theoretically be distributed to arbitrarily few elements of H_{ij} (even to a single one) without violating Eq. (44). Equivalently, $\Sigma_{H_{ij}}$ can theoretically reach its maximum (namely, $\gamma_i \gamma_j$) regardless of the number of non-zero elements of H_{ij} (provided there is at least one). This is a key feature of the solution of the relaxed problem that results from the diagonality of the diagonal blocks of $\mathbf{H}$, and yields the following important properties:

Proposition 3: Under necessary condition (45),

a) for every H_{ij} with n non-zero diagonals there exists an H'_{ij} with m non-zero diagonals, where $1 \leq m < n$, so that $\langle R_{ij}, H'_{ij} \rangle \geq \langle R_{ij}, H_{ij} \rangle$, with strict inequality holding if the involved correlation values are unique.

b) it holds:

$$\langle R_{ij}, H_{ij} \rangle \leq \Sigma_{H_{ij}} r_{ij}^{\max}, \tag{49}$$

with equality being reached if and only if H_{ij} is non-zero only along its $\kappa_{ij}^{\max}$-th diagonal,

Proof: For a proof of Proposition 3, see Appendix C. ■

Proposition 3 provides the main reason behind the sparsity of the relaxed solution in the general case. On the other hand, Lemma 1 states the specific sparsity requirement that ensures the equivalence between the solutions of the relaxed and the original problem.

Lemma 1: The solution $\mathbf{H}$ of the relaxed problem is equivalent to the solution of the original problem if and only if the following conditions hold:

C_1: Every off diagonal block H_{ij} of $\mathbf{H}$ has exactly one non-zero diagonal,

C_2: Eq. (45) holds with equality.

Proof: For a proof of Lemma 1, see Appendix D. ■

As shown in Appendix D, if conditions C_1, C_2 are satisfied, then $\mathbf{H}$ assumes a specific structure (see the example of Fig.1(b) for a visual reverence) and can be written as a convex combination of rank-one matrices that constitute equivalent optimal solutions for the original problem. Specifically, every eigenvector of $\mathbf{H}$ provides an optimal solution for the original problem.

Finally, Lemma 2 informs us that the equivalence between the solutions of the relaxed and the original problem is always ensured if the involved signals can become jointly optimally aligned.

Lemma 2: The validity of conditions C_1, C_2 stated in Lemma 1 is always ensured if the involved signals can become jointly optimally aligned, that is, if there exists a set of time-delays $\tau_1^*, \ldots, \tau_M^*$ that satisfy Eq. (5).

Proof: For a proof of Lemma 2, see Appendix E. ■

As shown in Appendix E, the existence of time-delays that satisfy Eq. (5) ensures that every H_{ij} has non-zero values only along its $\kappa_{ij}^{\max}$-th diagonal, namely the diagonal that corresponds to the maximum correlation value.

C. Estimating $\boldsymbol{\tau}$ from the solution of the relaxed problem

Let $\mathbf{H}$ be the solution of the relaxed problem and let $\hat{\mathbf{h}} = \boldsymbol{\xi}_{\max}(\mathbf{H})$, where $\hat{\mathbf{h}} = [\hat{\boldsymbol{h}}_1^T \cdots \hat{\boldsymbol{h}}_M^T]^T$, $\hat{\boldsymbol{h}}_i$ is a $L \times 1$ vector and $\boldsymbol{\xi}_{\max}(\mathbf{H})$ denotes the eigenvector that corresponds to the largest eigenvalue of $\mathbf{H}$.

If $\mathbf{H}$ satisfies the conditions of Lemma 1, then $\hat{\mathbf{h}}\hat{\mathbf{h}}^T$ solves the original problem (31)-(35), meaning that every $\hat{\boldsymbol{h}}_i$ has a single non-zero (positive) element. In this case, the loci of the positive elements produce the set of time-delays that solve the signal alignment problem.

If this is not the case, then $\hat{\mathbf{h}}\hat{\mathbf{h}}^T$ represents the best, in the Frobenius sense, rank-one approximation of $\mathbf{H}$. Moreover, $\hat{\mathbf{h}}$ is non-negative since $\mathbf{H}$ is non-negative (see Eq. (7)). Although there is no restriction on the number of the non-zero elements in each filter, typically $\hat{\boldsymbol{h}}_i$ is a sparse vector with a dominant positive element. The reasons behind the sparsity of the relaxed solution are discussed in Section V-B.

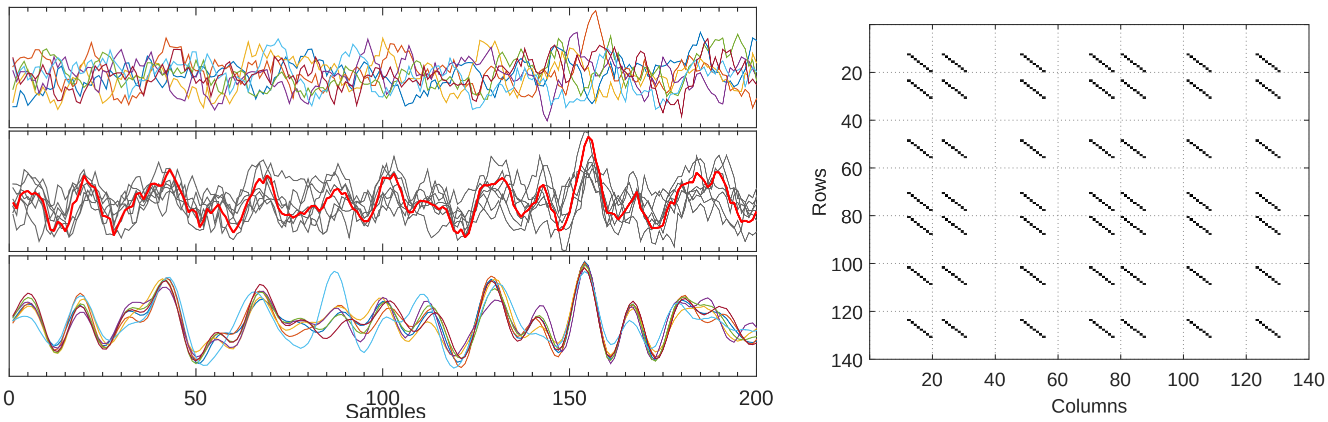
In any case, an estimate of the time-delays that lead to the optimal joint alignment of the signals, can be obtained as follows:

$$\hat{\tau}_i = \underset{0 \leq n \leq L-1}{\operatorname{argmax}} \ \hat{h}_{i,n}, \tag{50}$$

where $\hat{\tau}_i$, $\hat{h}_{i,n}$ denote the elements of $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{h}}_i$, respectively.

D. Computational complexity and over-relaxation

Although it is polynomially solvable, problem (36)-(40) is still computationally intensive. The main factor contributing to its complexity lies in Eq. (39) which introduces $LM(LM -$



a) b)

Fig. 1: (a) A synthetic dataset of $M = 7$ signals in their pure and aligned form (bottom) and after the introduction of random delays in the range $[-10, 10]$ samples, plus the addition of AR-modelled noise for an individual SNR of -5 dB (top). The results obtained from the application of the proposed blind beamforming technique with a filter length of $L = 20$, are shown in the middle plot. The aligned signals are shown in grey, while the obtained average, in red. (b) The solution of the relaxed problem is a $ML \times ML = 140 \times 140$ matrix containing $M^2 = 49$ blocks of size 20×20 $(= L \times L)$. Non-zero elements are shown in shades of grey, while zero elements, in white (the grid lines have been added as visual guidelines). In this case, every block of $\mathbf{H}$ has exactly one non-zero diagonal, without this diagonal being the $\kappa_{ij}^{\max}$-th one in every H_{ij}. This means that the solution of the original alignment problem has been achieved, although a jointly optimal alignment of the (noisy) signals is not possible.

1)/2 inequality constraints to the problem. As a result, (36)-(40) can become impractical for large filter lengths, namely, for large values of L (e.g. for $L > 25$). This depends of course on the number of signals as well (namely, M), which, however, is not a parameter of the problem.

Since L needs to be greater than the greatest pairwise time-difference of the signals, in cases where the signals can be separated by large time intervals, we propose solving the following version of the problem:

$$\phi_{rr}^* = \max_{\mathbf{H}} \quad \mathrm{trace}(\mathbf{R}_{+}\mathbf{H}), \tag{51}$$

$$\text{s.t.} \quad \mathrm{trace}(H_{ii}) = \gamma_i^2,\ 1 \le i \le M, \tag{52}$$

$$\mathbf{1}^T H_{ii} \mathbf{1} = \gamma_i^2,\ 1 \le i \le M, \tag{53}$$

$$H_{ii} \ge 0,\ 1 \le i \le M, \tag{54}$$

$$\mathbf{H} \succeq 0, \tag{55}$$

where, $\mathbf{R}_{+}$ retains only the non-negative values of $\mathbf{R}$ (the rest are set to 0).

As it can be seen in Eq. (54), the non-negativity constraint is applied to the diagonal blocks of $\mathbf{H}$ (thus preserving their diagonality), but not to the off-diagonal ones (which is compensated to a certain extent by the non-negativity of $\mathbf{R}_{+}$). This reduces the non-negativity constraints by a factor of M and ultimately speeds-up the solution by several orders of magnitude.

In our experience, (51)-(55) retains the robustness of (36)-(40), but trades off a slight degradation in accuracy for a substantial gain in execution time. Thus, although (51)-(55) has a very satisfactory performance in its own, the best possible results are obtained by using it in conjunction with (36)-(40), in order to improve the accuracy of the solution. Under this scenario, we would first obtain an alignment of the signals using (51)-(55) with a large L (namely, greater than the maximum expected pairwise time-difference of the signals), and then we would fine-tune the results by using (36)-(40) with a small L (e.g., no more than 10 samples).

VI. The Optimal Weights

As it is obvious from Eq. (10), for a given $\boldsymbol{\tau}$, $\mathrm{snr}_{\bar{y}}(\boldsymbol{\gamma})$ constitutes a generalized Rayleigh quotient, meaning that the following holds:

$$\mathrm{snr}_{\bar{y}}(\boldsymbol{\gamma}) \le \lambda_{\max}(Q), \tag{56}$$

with equality being reached for

$$\boldsymbol{\gamma}_{\max} = \Sigma^{-1/2} \boldsymbol{\xi}_{\max}(Q), \tag{57}$$

where $Q = \Sigma^{-1/2} R(\boldsymbol{\tau}) \Sigma^{-1/2}$. Thus, if $\boldsymbol{\gamma}_{\max}$ has non-negative elements, that is, if $\boldsymbol{\gamma}_{\max} \ge 0$, then $\boldsymbol{\gamma}_{\max}$ solves problem (12). If this is not the case, then, as we did with the alignment problem, we can relax (12) into the following semi-definite program with variable $\Gamma \in \Re^{M \times M}$:

$$\max_{\Gamma} \quad \mathrm{trace}(Q\Gamma), \tag{58}$$

$$\text{s.t.} \quad \mathrm{trace}(\Gamma) = 1, \tag{59}$$

$$\Gamma \ge 0, \tag{60}$$

$$\Gamma \succeq 0. \tag{61}$$

The optimal weights are then estimated as

$$\hat{\boldsymbol{\gamma}} = \Sigma^{-1/2} \boldsymbol{\xi}_{\max}(\Gamma). \tag{62}$$

Two crucial points have to be made. Firstly, since the estimation of $\boldsymbol{\gamma}$ succeeds the alignment of the signals, it is reasonable to expect that in the vast majority of cases, the correlation matrix $R(\boldsymbol{\tau})$ will be non-negative, which is a sufficient condition for $\boldsymbol{\gamma}_{\max} \ge 0$ to hold. In fact, in the numerous experiments that we conducted, there were only a handful of very problematic cases which required the relaxed formulation (58)-(61) for the estimation of the weights. Secondly, it seems that (58)-(61) constitutes an equivalent formulation of (12) rather than its relaxation (namely, (58)-(61) has always a rank-one solution), although we have not been able to prove such a claim, nor have we found an analogous proof in the literature. We can only state that throughout our experiments involving (58)-(61) we were not able to find an example that contradicts this claim. The outline of the proposed algorithm follows.

VII. Algorithm

Algorithm 1 SDP-based blind beamforming

1: **procedure** SDPBB
2: Estimate correlation functions, noise variances and L.
3: Set $\boldsymbol{\gamma} = \mathbf{1}$.
4: Solve (51)-(55) and estimate initial $\boldsymbol{\tau}$.
5: **repeat**
6: Update L.
7: Solve (36)-(40) and estimate $\boldsymbol{\tau}$.
8: Calculate $R(\boldsymbol{\tau})$ and update $\boldsymbol{\gamma}$.
9: **until** The change in $\boldsymbol{\tau}$ becomes insignificant.
10: Return $\boldsymbol{\tau}$, $\boldsymbol{\gamma}$.
11: **end procedure**

Parameter estimation is discussed in Secrion VII-A, while $\boldsymbol{\tau}$, $\boldsymbol{\gamma}$, are estimated by using Eqs. (50), (57) (or (58)-(62)), respectively. In step 6, the filter length is updated so that it reflects the current alignment (that, is the current maximum pairwise time-delay) of the signals. We note also that, whether the estimation of $\boldsymbol{\tau}$ in steps 4, 7, is going to be based on the solution of the relaxed or the over-relaxed versions of the alignment problem, depends on the nature of the signals, the requirements of the application, as well as on the available (computational) resources. We present what in our opinion seems the most computationally effective approach that does not sacrifice estimation accuracy.

As it is clear, initially the signals are aligned having been assigned equal weights. If an optimal (joint) alignment is reached, then the algorithm stops after the first iteration (in this case problem (9) is separable). Otherwise, it re-iterates by giving greater importance to signals with lager weights. Typically, no more than 2-3 iterations are needed.

A. Parameter Estimation

Three data-related quantities are required by the proposed technique, namely, the signal correlation functions $r_{ij}(\kappa)$, the filter length L and the noise variances σ_i^2.

Firstly, we assume that the noise variance are given or can be estimated from "quiet" intervals of the data. Since the noise processes are assumed white and jointly uncorrelated, the signal correlation functions $r_{ij}(\kappa)$ can be estimated from the available samples as follows: $\hat{r}_{ij}(\kappa) = r_{ij}^x(\kappa)$, $i \neq j$, $\hat{r}_{ii}(\kappa) = r_{ii}^x(\kappa) - \sigma_i^2$, $1 \leq i, j \leq M$, where $r_{ij}^x(\kappa)$ denotes the pairwise correlation functions of the available noisy data:

$$r_{ij}^x(\kappa) = \frac{1}{N}\sum_{n=0}^{N-1} x_i(n)x_j(n+\kappa) \approx r_{ij}(\kappa) + E[w_i(n)w_j(n)].$$

Finally, since the filter length must be larger that the largest time separation of the signals, an indication for the value of L can be obtained by the maximum of the lags that maximize the pairwise correlation functions $\hat{r}_{ij}(\kappa)$, namely the quantity $\max_{ij} |\operatorname{argmax}_\kappa \hat{r}_{ij}(\kappa)|$.

VIII. Simulation Results

This section contains the most representative cases of the numerous simulation experiments that were conducted, with various signal/noise model combinations. Regarding noise, the results for two "extreme" cases are shown. The first one concerns additive white Gaussian noise (AWGN), which represents the most favourable noise scenario and it is included as a "control sample". The second noise scenario concerns an first-order AR process with a pole of magnitude 0.8, that is used to test the performance of the proposed method in more realistic cases of highly correlated noise.

On the other hand, the synthetic signals (namely $s_i(n)$ in Eq. (1)) were modelled as low-pass filtered white Gaussian noise. The corner frequency of the filter was set at 10Hz, and the sampling frequency at 100Hz. In order to control the degree of similarity among the signals of the dataset (in their pure noiseless form), the M signals were produced by M different filters. Each filter occurred by adding a random perturbation of a controlled magnitude to a common low-pass filter prototype. Finally, in order to construct the synthetic dataset, every signal was delayed by an arbitrary amount (within specified limits) and multiplied by a constant gain in order to achieve the desired SNR.

The performance of the proposed method is compared against two other techniques that share the same assumptions with the proposed one (see also Section I). More specifically, the first one is the technique proposed in [17] for jointly aligning waveforms of closely spaced seismic events with the purpose of improving phase arrival estimates. The author of [17] estimates the pairwise lags that maximize the respective cross-correlation functions (namely, the quantities $\kappa_{ij}^{\max}$ defined in (3)) and seeks for a set of time-delays $\tau_1^*, \ldots, \tau_M^*$ that minimizes the error between the theoretical optimal lag, namely $\tau_i^* - \tau_j^*$ and the observed one, namely $\hat{\kappa}_{ij}^{\max}$, for all possible signal pairs. This leads to an overdetermined system of $M(M-1)/2$ equations with M unknowns. An approximate solution is obtained by minimizing the L_1-norm of the residual vector. The second technique included in our comparisons is the blind beamforming technique proposed in [16]. In this case the desired time-delays are obtained via the maximal eigenvector of a correlation matrix which is identical to the matrix $\mathbf{R}$ defined in (19), with the exception that the diagonal blocks of the matrix used in [16] contain the autocorrelation matrices of the M signals, rather than being empty.

Finally, it must be stressed that all included results are focused on the signal alignment aspect of the problem rather than on the weight assignment one. This is because on the one hand, the solution of the signal alignment problem constitutes the main contribution of the manuscript and on the other, the solution to the weighting problem is obtained in a closed form and it is in total agreement with almost all other beamforming techniques.

Experiment I

The goal of the first experiment is to asses the sensitivity of the proposed alignment technique with respect to the quality (i.e., the SNR) of the available dataset. To this end, we used signals with a very high degree of similarity in their pure form (having pairwise correlation coefficients in the neighbourhood of 0.9), by limiting the amount of perturbations introduced to the signal-producing filters. In doing so, we ensured that in the vast majority of cases, an optimal joint alignment of the pure noiseless signals could be achieved. In other words, in the absence of noise the proposed technique would solve the original alignment problem (11) and we would obtain error-free estimates. An example of the used signals is shown in the top diagram of Fig. 3.(c).

We then tested the performance of the technique against the techniques of [17], [16], under AWGN and AR-modelled noise scenarios, for several SNR values (in this case, all input time series shared the same SNR). For every SNR value 100 synthetic datasets were constructed, each containing 15 signals with arbitrary delays of up to 10 samples from a reference point (pairwise delays of up to 21 samples) and corrupted by noise.

We estimated the pairwise delays by using the techniques of [17] (L1), [16] (MaxEig), the over-relaxed problem (Proposed-OR) (51)-(55) and the relaxed problem (36)-(40) (Proposed-R). We then calculated the mean histogram of the estimation error by averaging the histograms obtained from each dataset. Finally, from the mean histogram, we obtained the (empirical) cumulative distribution functions (CDFs) of the error, which are shown in Fig. 2. The errors are in samples and a point (x, y) on the curve signifies that the probability of having an estimation error of at most x samples, is equal to y.

The filter length was set to $L = 25$ for the first step of the proposed technique (i.e. for the solution of the proposed over-relaxed problem), as well as for the technique of [16]. The relaxed problem was solved with a filter length of $L = 6$ (for a correction of up to ± 5 samples), having the solution of the over-relaxed problem as a starting point.

Although there is not much to differentiate the performance of the compared techniques under the ideal AWGN scenario, the advantages of the proposed technique become readily apparent in the realistic case of the AR-modelled noise, where it clearly outperforms its rivals. Furthermore, as it can be seen in the bottom row of Fig. 2, the results of the first experiment confirm our belief that by solving the efficient over-relaxed problem (51)-(55) we obtain a very robust initial solution to the alignment problem, which can then be refined by using (36)-(40) in order to improve the accuracy of the obtained estimates.

Experiment II

In the second experiment we evaluated the performance of the proposed technique in cases where the hypothesis of signals that can become jointly optimally aligned collapses. Two scenarios rejecting this hypothesis were tested. In the first one, we included outliers (that is, irrelevant signals) in every dataset and assessed the estimation errors regarding the useful (similar) signals (which were constructed as in Experiment I). Out of the 15 signals contained in every dataset, 7 were outliers.

In the second scenario, we lowered the pairwise similarity of the dataset by increasing the perturbations introduced to the signal-producing filters. This led to signals having pairwise correlation coefficients in the neighbourhood of 0.6 in their pure form (see the bottom diagram of Fig. 3.(c) for an example). It must be stressed that in the second scenario, due to the low similarity of the signals, even the total absence of noise does not ensure error-free estimates (that is, jointly optimally aligned signals), regardless of the used technique.

The parameters used were identical to the first experiment and the results obtained for the AR noise case with SNR$=$ -6 dB are shown in Fig. 3. Although there is an apparent drop in the performance of all used techniques (see also Fig. 2.(f) in comparison), especially in the case of low-similarity signals (as it was expected), the proposed technique manages to outperform its rivals by a safe margin.

IX. Conclusions

Blind beamfroming is formulated as a combination of the signal alignment and weight assignment problems. The signal alignment problem is a problem of combinatorial complexity which is relaxed and approximately solved by using the technique of SDP relaxation. An efficient over-relaxed version of the problem is also proposed. On the other hand, the weight assignment problem leads to a quadratic maximization problem which in the vast majority of cases has an exact solution, while in more challenging conditions is solved in an approximate fashion. The main contribution of the manuscript is the use of SDP for the solution of the signal alignment problem, which results in very robust and accurate estimations of the pairwise time delays between the signals. This makes it suitable for a great variety of applications that rely in time delay estimation in order to solve enhancement or localization problems. The superiority of the proposed technique was demonstrated through a number of numerical simulations with various signal/noise combinations.

Appendix A. Proof of Proposition 1

The equivalence of the constraints can be established as follows:

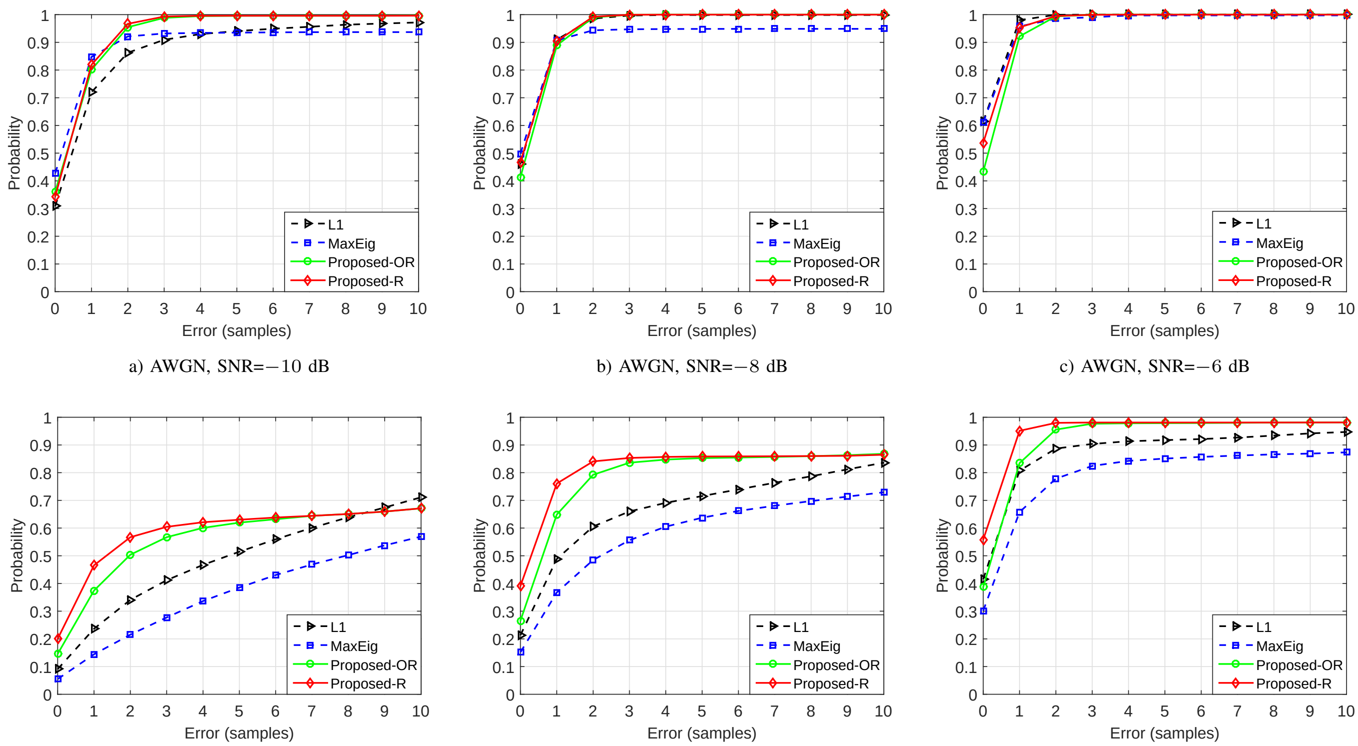
$$
\begin{aligned}
(\boldsymbol{h}_i^T \mathbf{1})^2 &= \sum_{n=0}^{L-1} \sum_{m=0}^{L-1} h_{i,n} h_{i,m} && (63)\\
&= \sum_{n=0}^{L-1} h_{i,n}^2 + \sum_{n=0}^{L-1} \sum_{\substack{m=0\\ m\neq n}}^{L-1} h_{i,n} h_{i,m}. && (64)
\end{aligned}
$$

By combining (22), (64), we have

$$\gamma_i^2 = \gamma_i^2 + \sum_{n=0}^{L-1} \sum_{\substack{m=0\\ m\neq n}}^{L-1} h_{i,n} h_{i,m}, \tag{65}$$

which yields

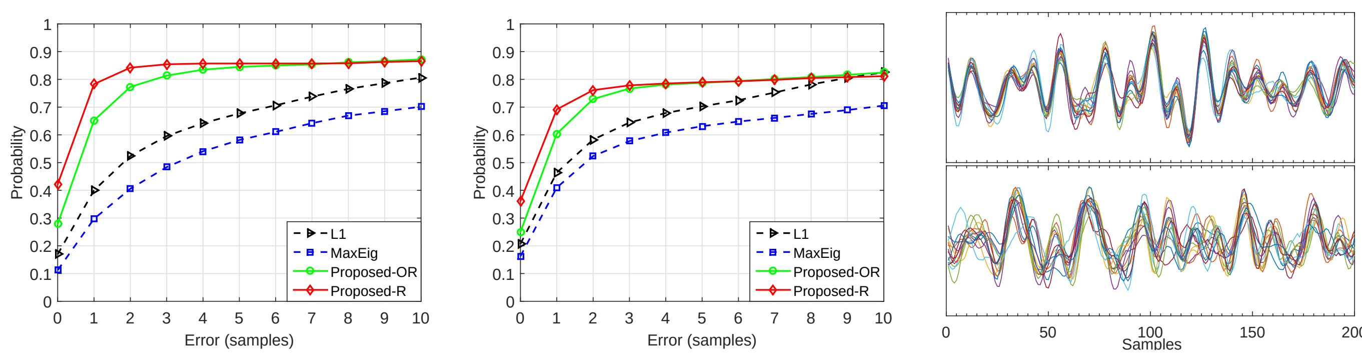
$$\sum_{n=0}^{L-1} \sum_{\substack{m=0\\ m\neq n}}^{L-1} h_{i,n} h_{i,m} = 0. \tag{66}$$



a) AWGN, SNR=-10 dB b) AWGN, SNR=-8 dB c) AWGN, SNR=-6 dB

d) AR, SNR=-10 dB e) AR, SNR=-8 dB f) AR, SNR=-6 dB

Fig. 2: Empirical CDFs of the pairwise time-delay estimation errors obtained by the techniques of [17] (L1), [16] (MaxEig), the over-relaxed problem (51)-(55) (Proposed-OR) and the relaxed problem (36)-(40) (Proposed-R), respectively, for the scenarios of white Gaussian noise (top row) and AR-modelled noise (bottom row), in Experiment I.



a) AR, SNR=-6 dB b) AR, SNR=-6 dB c) Signal examples

Fig. 3: (a)-(b) Empirical CDFs of the pairwise time-delay estimation errors obtained in Experiment II for the scenarios of (a) outliers being present and (b) low-similarity datasets, in Experiment II. In both cases the signals were corrupted by AR-modelled noise with an SNR of -6 dB. (c) Examples of signals with average pairwise correlation coefficient in the neighbourhood of 0.9 (top) and 0.6 (bottom). The presented signals are noiseless and aligned.

It is obvious that under the condition $\boldsymbol{h}_i \geq 0$, (66) can be satisfied if and only if $\boldsymbol{h}_i$ has at most one nonzero element. Thus, the combination of the three constraints in Eq. (22) ensures that $\boldsymbol{h}_i$ has exactly one nonzero element, equal to γ_i, thereby concluding the proof. ■

Appendix B. Proof of Proposition 2

Since Y is positive semi-definite, then there exists a factorization $Y = VV^T$, for some $k \times r$ matrix V, with $1 \leq r \leq k$. If V is partitioned as

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \tag{67}$$

where the dimensions of V_1, V_2 are $k_1 \times r$, $k_2 \times r$, respectively, then, $A = V_1 V_1^T$, $C = V_2 V_2^T$, $B = V_1 V_2^T$, thereby establishing the following relations:

$$\Sigma_A = \mathbf{1}^T A \mathbf{1} = \mathbf{1}^T V_1 V_1^T \mathbf{1} = \mathbf{v}_1^T \mathbf{v}_1 = \|\mathbf{v}_1\|_2^2, \tag{68}$$

$$\Sigma_C = \mathbf{1}^T C \mathbf{1} = \mathbf{1}^T V_2 V_2^T \mathbf{1} = \mathbf{v}_2^T \mathbf{v}_2 = \|\mathbf{v}_2\|_2^2, \tag{69}$$

$$\Sigma_B = \mathbf{1}^T B \mathbf{1} = \mathbf{1}^T V_1 V_2^T \mathbf{1} = \mathbf{v}_1^T \mathbf{v}_2, \tag{70}$$

where $\mathbf{v}_i = V_i^T \mathbf{1}$, $i = 1, 2$, are $r \times 1$ vectors, $\| \cdot \|_2$ denotes the Euclidean norm, while $\mathbf{1}$ denotes the all-ones vector of compatible length. Then, (68), (69), impose non-negativity on Σ_A, Σ_C, respectively, while (48) results from the application of the Cauchy-Schwarz inequality to the inner product of vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and by using (68)-(69). ■

APPENDIX C. PROOF OF PROPOSITION 3

a) By taking into account the structure of R_{ij} (see also Eq. (18)), under the condition that H_{ij} has exactly n non-zero diagonals, the inner product $\langle R_{ij}, H_{ij} \rangle$ can be written as,

$$\langle R_{ij}, H_{ij} \rangle = \sum_{m=1}^{n} \beta_m r_{ij}(\kappa_m), \tag{71}$$

where $\kappa_1, \ldots, \kappa_n$, denote the non-zero diagonals of H_{ij}, with $-(2L-1) \leq \kappa_m \leq 2L-1$, $m = 1, \ldots, n$, while β_m denotes the sum of the elements along the κ_m-th diagonal, with $\beta_m > 0$, $m = 1, \ldots, n$ and $\sum_m \beta_m = \Sigma_{H_{ij}} \leq \gamma_i \gamma_j$ (since H_{ij} satisfies condition (45)). Let us now assume without loss of generality that the correlation values involved in Eq. (71) are ordered in a descending fashion so that for every $1 \leq k < l \leq n$ we have $r_{ij}(\kappa_k) \geq r_{ij}(\kappa_l)$. Then, we can write:

$$\begin{aligned}\langle R_{ij}, H_{ij} \rangle &= \sum_{m=1}^{n-1} \beta_m r_{ij}(\kappa_m) + \beta_n r_{ij}(\kappa_n) &(72)\\ &\leq \sum_{m=1}^{n-1} \beta_m r_{ij}(\kappa_m) + \beta_n r_{ij}(\kappa_{n-1}) &(73)\\ &= \sum_{m=1}^{n-1} \beta'_m r_{ij}(\kappa_m), &(74)\\ &= \langle R_{ij}, H'_{ij} \rangle &(75)\end{aligned}$$

where $\beta'_m = \beta_m$, $1 \leq m < n-1$, $\beta'_{n-1} = \beta_{n-1} + \beta_n$. Thus, if the elements along the κ_n-th diagonal of H_{ij} are set to zero and β_n is allocated to the elements of the κ_{n-1}-th diagonal, we obtain a new matrix H'_{ij} with $n-1$ non-zero diagonals that satisfies (45) (since $\sum_m \beta'_m = \sum_m \beta_m \leq \gamma_i \gamma_j$), yielding $\langle R_{ij}, H'_{ij} \rangle \geq \langle R_{ij}, H_{ij} \rangle$. Note also that if the involved correlation values are unique, Eq. (73) holds with strict inequality. This completes the proof of the first statement of Proposition 3.

b) Following the notation we used in the first part, we can write in the general case:

$$\begin{aligned}\langle R_{ij}, H_{ij} \rangle &= \sum_{m=1}^{2L-1} \beta_m r_{ij}(\kappa_m) &(76)\\ &= \beta_1 r_{ij}(\kappa_1) + \sum_{m=2}^{2L-1} \beta_m r_{ij}(\kappa_m), &(77)\end{aligned}$$

where by definition we have $\kappa_1 = \kappa_{ij}^{\max}$, $r_{ij}(\kappa_1) = r_{ij}(\kappa_{ij}^{\max}) = r_{ij}^{\max}$. Since the sum of all β_m's equals the sum of the elements of H_{ij}, namely $\Sigma_{H_{ij}}$, we can write:

$$\beta_1 = \Sigma_{H_{ij}} - \sum_{m=2}^{2L-1} \beta_m r_{ij}(\kappa_m). \tag{78}$$

Substituting β_1 in (77), gives us:

$$\langle R_{ij}, H_{ij} \rangle = \Sigma_{H_{ij}} r_{ij}^{\max} + \sum_{m=2}^{2L-1} \beta_m (r_{ij}(\kappa_m) - r_{ij}^{\max}). \tag{79}$$

Assuming r_{ij} has a unique maximum, the difference inside the parenthesis on the right-hand side of (79) is always negative. Thus, since all β_m's are by definition non-negative, $\langle R_{ij}, H_{ij} \rangle$ attains its maximum when $\beta_m = 0$ holds for all $2 \leq m \leq 2L-1$. This means that for every valid value of $\Sigma_{H_{ij}}$, we can write

$$\langle R_{ij}, H_{ij} \rangle \leq \Sigma_{H_{ij}} r_{ij}^{\max}, \tag{80}$$

with equality holding if and only if H_{ij} is non-zero only along its $\kappa_1 (= \kappa_{ij}^{\max})$-th diagonal. This completes the proof of Proposition 3. ■

APPENDIX D. PROOF OF LEMMA 1

Since, as already explained in SectionV-B, any feasible solution for the original problem satisfies conditions C_1, C_2, if the solutions of the original and the relaxed problem are equivalent, then, by definition, the relaxed problem is solved by a matrix that satisfies C_1, C_2.

We will now prove that the opposite is also true. To this end, let $\mathbf{H}$ denote the solution of the original problem and let H_{ij} have non-zero (positive) elements only along its κ_{ij}-th diagonal, with $-(L-1) \leq \kappa_{ij} \leq L-1$. Let also the sum of the elements of H_{ij} be equal to $\gamma_i \gamma_j$.

Without overburdening the notation, let us denote as $(H_{ij})_{(p+k)(q+k)}$, $k = 0, \ldots, K-1$, with $p - q = \kappa_{ij}$, the elements along the κ_{ij}-th diagonal of H_{ij}. Since the κ-th diagonal of an $L \times L$ matrix consists of $L - |\kappa|$ elements, we have $K = L - |\kappa_{ij}|$. (As an example, if $\kappa_{ij} = -2$ and $L = 5$, the κ_{ij}-th diagonal of H_{ij} has $K = 5 - 2 = 3$ elements, namely, $(H_{ij})_{13}$, $(H_{ij})_{24}$, and $(H_{ij})_{35}$. In this case, we would have $p = 1$, $q = 3$.)

Then, due to the positive semi-definiteness of $\mathbf{H}$ (and especially of $\mathbf{H}_{ij}$ defined in (43)), the following must hold for all $0 \leq k \leq K-1$:

$$(H_{ij})_{(p+k)(q+k)} \leq \sqrt{(H_{ii})_{(p+k)(p+k)} (H_{jj})_{(q+k)(q+k)}}. \tag{81}$$

Additionally, since H_{ij}, H_{jj}, H_{ii}, contain non-negative elements and the following are true,

$$\sum_{k=0}^{K-1} \left[(H_{ij})_{(p+k)(q+k)} \right] = \Sigma_{H_{ij}} = \gamma_i \gamma_j, \tag{82}$$

$$\sum_{k=0}^{K-1} \left[(H_{ii})_{(p+k)(p+k)} \right] \leq \operatorname{trace}(H_{ii}) = \gamma_i^2, \tag{83}$$

$$\sum_{k=0}^{K-1} \left[(H_{jj})_{(q+k)(q+k)} \right] \leq \operatorname{trace}(H_{jj}) = \gamma_j^2, \tag{84}$$

where (83), (84), are due to constraint (37), we can write:

$$(H_{ij})_{(p+k)(q+k)} = \alpha_k \gamma_i \gamma_j, \tag{85}$$
$$(H_{ii})_{(p+k)(p+k)} = u_k \gamma_i^2, \tag{86}$$
$$(H_{jj})_{(q+k)(q+k)} = v_k \gamma_j^2, \tag{87}$$

$k = 0, \ldots, K-1$, where the coefficients α_k, u_k, v_k, satisfy the following conditions, respectively:

$$\alpha_k \geq 0, \quad \sum_{k=0}^{K-1} \alpha_k = 1, \tag{88}$$
$$u_k \geq 0, \quad \sum_{k=0}^{K-1} u_k \leq 1, \tag{89}$$
$$v_k \geq 0, \quad \sum_{k=0}^{K-1} v_k \leq 1. \tag{90}$$

Thus, substituting (85), (86), (87) into (81), we obtain:

$$\alpha_k \gamma_i \gamma_j \leq \sqrt{u_k v_k \gamma_i^2 \gamma_j^2} \Rightarrow \alpha_k \leq \sqrt{v_k u_k}. \tag{91}$$

We are now going to prove that Eqs. (88)-(91), yield the following condition:

$$\alpha_k = u_k = v_k, \; k = 0, \ldots, K-1. \tag{92}$$

To this end, let $\mathbf{u}$, $\mathbf{v}$, be the vectors whose elements are $\sqrt{u_k}$, $\sqrt{v_k}$, respectively, for $k = 0, \ldots, K-1$. Then, due to (89), (90), we can write $\|\mathbf{u}\|_2 \leq 1$, $\|\mathbf{v}\|_2 \leq 1$, respectively. Moreover, we have

$$\sum_{k=0}^{K-1} \sqrt{u_k v_k} = \mathbf{u}^T \mathbf{v} \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \leq 1, \tag{93}$$

with the last inequality holding due to (89)-(90). On the other hand, using (88), (91), the following inequality is also true:

$$\sum_{k=0}^{K-1} \sqrt{u_k v_k} \geq \sum_{k=0}^{K-1} \alpha_k = 1, \tag{94}$$

It is obvious that (93), (94), can both be valid if and only if all the involved inequalities hold with equality, which occurs only when $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$, $\mathbf{u} = \mathbf{v}$, and (91) holds with equality for all $k = 0, \ldots, K-1$. This proves (92).

Additionally, since (92) must hold for all $1 \leq i, j \leq M$, the following statements can be proved by induction:

1) The sequence $\alpha_0, \ldots, \alpha_{K-1}$ of non-negative coefficients is identical for all H_{ij}, $1 \leq i, j \leq M$.
2) α_k is located at the same row for all the blocks of the same block-row (i.e., for all H_{ij}, $1 \leq j \leq M$).
3) α_k is located at the same column for all the blocks of the same block-column (i.e., for all H_{ij}, $1 \leq i \leq M$).
4) $K \leq L - \max_{ij} |\kappa_{ij}|$.

Statements 1-3 are obtained from the repeated application of Eq. (92) for all H_{ij}'s, while statement 4 reflects the fact that, since (due to statement 1) all H_{ij}'s must hold exactly the same number of non-zero elements along their respective κ_{ij}-th diagonal, K cannot exceed the length of the shortest non-zero diagonal, namely $L - \max_{ij} |\kappa_{ij}|$.

Thus, so far we have established that if $\mathbf{H}$ solves the relaxed problem and conditions C_1, C_2, are satisfied, then for all $1 \leq i, j \leq M$, H_{ij} holds $\alpha_0 \gamma_i \gamma_j, \ldots, \alpha_{K-1} \gamma_i \gamma_j$ (possibly accompanied by additional zero elements) in its κ_{ij}-th diagonal. All other elements of H_{ij} are equal to zero. See also Fig.1(b) for a visual reference.

Let us now elaborate on the significance of statements 2,3. To this end, let $p_i + k$ be the row number of α_k for all the blocks of the i-th block-row, with $1 \leq i \leq M$, $k = 0, \ldots, K-1$. Similarly, let $q_j + k$ be the column number of α_k for all the blocks of the j-th block-column, with $1 \leq j \leq M$, $k = 0, \ldots, K-1$. Then, for all $1 \leq i, j \leq M$, $k = 0, \ldots, K-1$, we have $(H_{ij})_{(p_i+k)(q_j+k)} = \alpha_k \gamma_i \gamma_j$. Moreover, the following must hold:

$$p_i = q_i, \tag{95}$$
$$p_i - q_j = \kappa_{ij}, \tag{96}$$

where (95) is due to the diagonality of the diagonal blocks.

Let us now define $\boldsymbol{h}_i(k) = [h_{i,1}(k), \ldots, h_{i,L}(k)]^T$, as an $L \times 1$ vector with a single non-zero element at its $p_i + k$-th position, namely:

$$h_{i,l}(k) = \begin{cases} \gamma_i, & l = p_i + k \\ 0, & l \neq p_i + k \end{cases}, \; 1 \leq l \leq L, \tag{97}$$

Let also $\mathbf{h}(k)$ denote the $LM \times 1$ vector that is obtained by stacking the M vectors $\boldsymbol{h}_i(k)$, for $i = 1, \ldots, M$. Then, the following statements can be verified without great difficulty:

5) $\mathbf{h}(k)$ is an eigenvector of $\mathbf{H}$, namely:
$$\mathbf{H}\mathbf{h}(k) = \alpha_k \|\mathbf{h}(k)\|^2 \mathbf{h}(k), \; k = 1, \ldots, K-1. \tag{98}$$
6) $\mathbf{h}(k)^T \mathbf{h}(l) = 0$ for all $0 \leq k, l \leq K-1$ with $k \neq l$ and the following decomposition of $\mathbf{H}$ holds:
$$\mathbf{H} = \sum_{k=0}^{K-1} \alpha_k \mathbf{h}(k) \mathbf{h}(k)^T. \tag{99}$$
7) $\mathbf{h}(k)\mathbf{h}(k)^T$ constitutes a feasible solution of the original problem (31)-(35).

If $\phi_r^* = \mathrm{trace}(\mathbf{RH})$ denotes the maximum of the objective function under the relaxed constraints, then based on Eq. (99) and the linearity of the trace, we can write:

$$\phi_r^* = \sum_{k=0}^{K-1} \alpha_k \, \mathrm{trace}\left(\mathbf{R}\mathbf{h}(k)\mathbf{h}(k)^T\right), \tag{100}$$
$$= \sum_{k=0}^{K-1} \alpha_k \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \langle R_{ij} \boldsymbol{h}_i(k) \boldsymbol{h}_j(k)^T \rangle, \tag{101}$$

with,

$$\langle R_{ij} \boldsymbol{h}_i(k) \boldsymbol{h}_j(k)^T \rangle = \gamma_i \gamma_j (R_{ij})_{(p_i+k)(p_j+k)} \tag{102}$$
$$= \gamma_i \gamma_j r_{ij}(p_i - p_j) \tag{103}$$
$$= \gamma_i \gamma_j r_{ij}(p_i - q_j) \tag{104}$$
$$= \gamma_i \gamma_j r_{ij}(\kappa_{ij}), \tag{105}$$

where (102) is due to the definition of $\boldsymbol{h}_i(k)$ in Eq. (97), while (103)-(105) result from Eqs. (95), (96). Thus, since

$\langle R_{ij}\boldsymbol{h}_i(k)\boldsymbol{h}_j(k)^T\rangle$ is independent of k, combining (100), (101), with (88), we can write

$$\phi_r^* = \mathrm{trace}(\mathbf{R}\mathbf{H}) = \mathrm{trace}\left(\mathbf{R}\mathbf{h}(k)\mathbf{h}(k)^T\right). \tag{106}$$

Finally, since $\mathbf{h}(k)\mathbf{h}(k)^T$ constitutes a feasible solution for the original problem, then we must have $\phi_o^* = \phi_r^* = \mathrm{trace}\left(\mathbf{R}\mathbf{h}(k)\mathbf{h}(k)^T\right)$, where ϕ_o^* denotes the maximum of the objective function under the original constraints. This is because by definition it holds $\phi_o^* \leq \phi_r^*$. Thus, $\mathbf{h}(k)\mathbf{h}(k)^T$ constitutes an optimal solution for the original problem.

Consequently, we have shown that if conditions C_1, C_2 hold, then the solution of the relaxed problem can be written as a convex combination of rank-one matrices that constitute equivalent optimal solutions for the original problem. This concludes the proof. ∎

Appendix E. Proof of Lemma 2

By taking into account Eqs. (4), (41), (45), (49), we can see that for every correlation matrix $\mathbf{R}$, the maximum of the objective function under the relaxed constraints has the following theoretical bound:

$$\phi_r^* \leq \phi^*(\mathbf{R}) = \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} \gamma_i\gamma_j r_{ij}^{\max}. \tag{107}$$

As it is also obvious from (49), ϕ_r^* reaches its theoretical bound if and only if every off-diagonal block of $\mathbf{H}$ is non-zero only along its $\kappa_{ij}^{\max}$-th diagonal and (45) holds with equality. In other words, if $\phi_r^* = \phi^*(\mathbf{R})$, then the optimal solution of the relaxed problem satisfies always conditions C_1, C_2, or equivalently, $\phi_r^* = \phi^*(\mathbf{R})$ constitutes a sufficient condition for the equivalence between the solutions of the relaxed and the original problem. Thus, in order to prove Lemma 2, it suffices to show that if the signals of the dataset can become jointly optimally aligned, then a solution that ensures $\phi_r^* = \phi^*(\mathbf{R})$ can always be found.

To this end, let us assume that the time-delays $\tau_1^*, \ldots, \tau_M^*$ satisfy (5). Then, by using Eq. (99) and by selecting $p_i = \tau_i^*$ in (97), for any valid K (please, see statement 4 in the proof of Lemma 1) and for any set of coefficients α_k that satisfy Eq. (88), we obtain an optimal solution $\mathbf{H}$. This is because in this case, we have $\langle R_{ij}H_{ij}\rangle = \gamma_i\gamma_j r_{ij}^{\max}$, which leads to $\mathrm{trace}(\mathbf{R}\mathbf{H}) = \phi^*(\mathbf{R})$. Thus, $\mathbf{H}$ is an optimal solution that ensures $\phi_r^* = \phi^*(\mathbf{R})$. This concludes the proof. ∎

References

[1] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York: Wiley, 2002.

[2] R. Roy, A. Paulraj, and T. Kailath, "Esprit - a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1340 – 1342, 1986.

[3] R. Schmidt, "Multiple emitter location and signal estimation," *IEEE Trans. Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[4] P. Stoica and A. Nehorai, "Music, maximum likelihood, and the cramer-rao bound," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.

[5] N. Yilmazer, J. Koh, and T. K. Sarkar, "Utilization of a unitary transform for efficient computation in the matrix pencil method to find the direction of arrival," *IEEE Trans. Antennas and Propagation*, vol. 54, no. 1, pp. 175–181, 2006.

[6] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust capon beamformer," *IEEE Trans. Signal Processing*, vol. 52, pp. 2407–2423, 2004.

[7] J. Li and P. Stoica, *Robust Adaptive Beamforming*. Wiley, 2005.

[8] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Processing*, vol. 53, pp. 1684–1696, 2005.

[9] A. Khabbazibasmenj and A. Vorobyov, Sergiy A.and Hassanien, "Robust adaptive beamforming based on steering vector estimation with as little as possible prior information," *IEEE Trans. Signal Processing*, vol. 60, no. 6, pp. 2974–2987, 2012.

[10] R. Gooch and J. Lundell, "The cm array: An adaptive beamformer for constant modulus signals," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 11, pp. 2523–2526, 1986.

[11] A. Van Der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1136–1155, 1996.

[12] J. F. Cardoso, "Source separation using higher order moments," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2109–2112, 1989.

[13] D. T. Pham, "Blind separation of instantaneous mixture of sources based on order statistics," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 363–375, 2000.

[14] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1545–1553, 2002.

[15] X. Huang, H.-C. Wu, and J. C. P. Cardoso, "Robust blind beamforming algorithm using joint multiple matrix diagonalization," *IEEE Sensors Journal*, vol. 7, no. 1, pp. 130–136, 2007.

[16] K. Yao, R. E. Hudson, C. W. Reed, D. Chen, and F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system," *IEEE J.Sel. A. Commun.*, vol. 16, no. 8, pp. 1555–1567, 1998.

[17] P. Shearer, "Improving local earthquake locations using the l1 norm and waveform cross correlation: application to the whittier narrows, california, aftershock sequence," *J. Geophys. Res.*, vol. 102, pp. 8269–8283, 1997.

[18] F. Waldhauser and W. L. Ellsworth, "A double difference earthquake location algorithm: method and application to north hayward fault, california," *Bull. Seism. Soc. Am.*, vol. 90, pp. 1353–1368, 2000.

[19] S. Rost and C. Thomas, "Array seismology: Methods and applications," *Reviews of Geophysics*, vol. 40, no. 3, pp. 2–1–2–27, 2002.

[20] J. C. Robinson, "Statistically optimal stacking of seismic data," *Geophysics*, vol. 35, no. 3, pp. 436–446, 1970.

[21] M. Schoenberger, "Optimum weighted stack for multiple suppression," *Geophysics*, vol. 61, no. 3, pp. 891–901, 1996.

[22] G. Liu, S. Fomel, L. Jin, and X. Chen, "Stacking seismic data using local correlation," *Geophysics*, vol. 74, no. 3, pp. V43–V48, 2009.

[23] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. SIAM, 1994.

[24] Y. S. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optim. Methods Softw.*, vol. 9, no. 1-3, pp. 141–160, 1998.