



HAL
open science

Alternating group lasso for block-term tensor decomposition with application to ECG source separation

José Henrique de Morais Goulart, Pedro Marinho R. de Oliveira, Rodrigo Cabral Farias, Vicente Zarzoso, Pierre Comon

► To cite this version:

José Henrique de Morais Goulart, Pedro Marinho R. de Oliveira, Rodrigo Cabral Farias, Vicente Zarzoso, Pierre Comon. Alternating group lasso for block-term tensor decomposition with application to ECG source separation. *IEEE Transactions on Signal Processing*, In press. hal-01899469v2

HAL Id: hal-01899469

<https://hal.science/hal-01899469v2>

Submitted on 14 Aug 2019 (v2), last revised 1 Apr 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Alternating Group Lasso for Block-Term Tensor Decomposition and Application to ECG Source Separation

J. H. de M. Goulart, P. M. R. de Oliveira, R. C. Farias, V. Zarzoso, *Senior Member, IEEE*,
and P. Comon, *Fellow, IEEE*

Abstract—In some applications, blind source separation can be performed by computing an approximate block-term tensor decomposition (BTD), under much milder constraints than matrix-based techniques. However, choosing the BTD model structure (*i.e.*, the number of blocks and their ranks) is a difficult problem, and the standard least-squares formulation can be ill-posed. This paper proposes an alternating group lasso algorithm to compute approximate low-rank BTDs. It solves, in a provably convergent manner, a well-posed mixed-norm regularized tensor approximation problem which allows jointly estimating the model parameters and its structure. A variant is also put forward for dealing with linearly constrained blocks, motivated by the problem of blind separation of sums of complex exponentials, which can be cast as a low-rank Hankel-structured block-term tensor approximation problem. An experimental comparison with a standard nonlinear least-squares algorithm on synthetic tensor data indicates that the proposed algorithm is much more robust with respect to initialization. We also apply the constrained variant to the extraction of atrial activity from semi-synthetic and real-world electrocardiogram recordings during atrial fibrillation episodes. Our results show its ability to consistently select an adequate structure and to extract multiple signals which can be physiologically interpreted as atrial fibrillation patterns.

Index Terms—Tensors, block term decomposition, group lasso, structured low-rank approximation, atrial fibrillation.

I. INTRODUCTION

RECORDED signals in biomedical applications, such as electroencephalography [1] and electrocardiography [2], [3], [4], [5] can be modeled as instantaneous unknown linear mixtures of R sources. To separate them, the unknown sources are assumed to be statistically independent and orthogonal, in independent component analysis (ICA) and principal component analysis (PCA), respectively. Although this renders the underlying blind source separation (BSS) well-posed, such stringent assumptions may lack physiological grounds, hindering results interpretation. A less constraining approach is to assume that sources can be approximated by sums of complex exponentials (SCE), which can model narrowband and transient signals that can be linked to specific clinical conditions. This leads to an identifiable low-rank matrix factorization involving a Vandermonde matrix [6], but the total

number of exponentials that can be identified by this matrix approach is bounded by the number of sensors.

Alternatively, by forming a matrix with shifted versions of each observed signal (*e.g.*, each channel output) and by stacking these matrices in a third-order tensor, a process called “Hankelization,” BSS of SCE signals can be cast as an approximate block-term tensor decomposition (BTD) problem with Hankel-structured blocks [7]. Each SCE source contributes in the decomposition with a structured term given by a tensor product of a rank- L_r Hankel matrix containing its samples with a vector containing its spatial signature (*i.e.*, its weights for each channel), where L_r is the number of complex exponentials (poles) in the SCE source. The main benefit of this approach is that identifiability of the decomposition is guaranteed under mild conditions which do not involve stringent constraints such as orthogonality or independence and can hold even when the sum of the block ranks exceeds the number of sensors and the dimensions of the Hankel matrices as well [8], [7].

When implementing this approach, however, two problems arise. The first one refers to the computation of the approximate BTD. Some methods have been made available for this task, such as the nonlinear least-squares methods implemented in Tensorlab [9] or the alternating least squares method with enhanced line search (ALS-ELS) [10]. However, the performance of such techniques depends considerably on the choice of structural parameters (R and L_r), which is difficult to make in practice. Furthermore, they usually exhibit strong sensitivity with respect to the initialization and address an optimization problem which may lack a global minimizer [11]; this can lead to the estimation of almost collinear blocks with no physical interpretation, a phenomenon often termed model degeneracy. The second problem refers to the low-rank Hankel structure constraint, which is hard to enforce for real-valued data. To our knowledge, no existing BTD algorithm can reliably impose Hankel constraints in this case.

In [12], a functional promoting group sparsity of the decomposition factor columns was minimized to estimate appropriate structural parameters of an (unconstrained) BTD model, but not the model itself. Moreover, the authors only considered the case where all L_r are equal. In this work, we show that (i) the same principle can be used to *jointly* estimate the model structure and its parameters, and that (ii) the general case with different block ranks can also be addressed.

To achieve this goal, we formulate the approximate BTD

P. Comon is with Univ. Grenoble Alpes, CNRS, Gipsa-Lab, F-38000 Grenoble, France (pierre.comon@gipsa-lab.fr).

J. H. de M. Goulart, P. M. R. de Oliveira, R. C. Farias and V. Zarzoso are with the Univ. Côte d’Azur, CNRS, I3S Laboratory, 06900 Sophia-Antipolis, France ({goulart, marinho, cabral, zarzoso}@i3s.unice.fr).

computation as the minimization of a (least-squares) fitting term plus a regularization term given by the sum of the $\ell_{2,1}$ -norms of the matrices containing the model parameters. The latter enforces (column-wise) group sparsity of the factor matrices, thus penalizing models with high R and L_r . In this way, we are able to find an approximate BTD without assumptions on R and L_r , effectively achieving a trade-off between data fitting and model complexity. Furthermore, the regularization renders the BTD approximation problem well-posed.

To solve this problem, we propose an algorithm called *alternating group lasso* (AGL), which is simpler than the algorithm of [12] and provably convergent. We also devise a variant termed constrained AGL (CAGL) to deal with linear (subspace) constraints over block matrices. For this purpose, a structured low-rank approximation method is applied on the block matrices at each iteration to ensure that they have low rank and belong to the specified subspace.

As an application of CAGL, we consider the problem of extracting the atrial activity (AA) in electrocardiograms (ECGs) of atrial fibrillation (AF). AF is the most common sustained cardiac arrhythmia encountered in clinical practice, a major public health and economical concern. The importance given to this challenging cardiac condition has increased in the past few years, since its mechanisms are not completely understood. Accurate analysis of the fibrillatory waves (f-waves) is then necessary for better understanding the arrhythmia. Noninvasive AA extraction from ECG recordings is, therefore, a key problem that motivates the development of signal processing techniques such as the one we propose in this paper.

Other BSS techniques such as PCA and ICA have been applied to noninvasive AA extraction [13]–[14] and provided satisfactory results. However, as previously stated, results are difficult to interpret due to the imposed constraints. Motivated by identifiability and interpretability requirements, BTD modeling using “Hankelization” for AA extraction has been recently proposed and studied in [2], [15], [3], [4], [5]. Experimental results in synthetic and real AF ECG data showed that BTD can outperform the matrix-based techniques for AA extraction in short and long segments of AF ECG recordings. The application in this work follows the same line of [2], [15], [3], [4], [5], but differs from [2], [15], [4], [5], since we do not impose the structural parameters of the model. It also differs from [3], where structural parameters are chosen fitting different BTD models and selecting the best one with respect to an information-theoretical criterion.

The outline of the paper is as follows: in Section II we present the problem of SCE source separation using the BTD with Hankel structure. Section III formulates a regularized BTD approximation problem and the AGL and CAGL algorithms. A numerical evaluation of AGL is shown in Section IV. The application of CAGL to AA extraction is then presented in Section V, where both semi-synthetic and real-world ECG datasets are considered. Finally, concluding remarks are made in Section VI.

Notation: Tensors are denoted in uppercase bold script letters \mathcal{X} , matrices in uppercase bold letters \mathbf{X} , vectors in

lowercase bold letters \mathbf{x} and scalars in lowercase letters x . The notations $\|\cdot\|_F$ and $\|\cdot\|_{2,1}$ stand for the Frobenius norm and the matrix mixed $\ell_{2,1}$ -norm, respectively. The latter is defined as the sum of the norms of its argument’s columns, as in $\|\mathbf{X}\|_{2,1} = \sum_{r=1}^R \|\mathbf{x}_r\|_2$. The symbols \otimes , \boxtimes and \odot are used for outer (tensor), Kronecker, and Khatri-Rao (column-wise Kronecker) product, respectively. The block Khatri-Rao product \odot_L is such that, given $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_R] \in \mathbb{C}^{K \times R}$ and $\mathbf{Y} = [\mathbf{y}_1 \cdots \mathbf{y}_{LR}] \in \mathbb{C}^{M \times LR}$, it yields $\mathbf{Z} = \mathbf{X} \odot_L \mathbf{Y} = [\mathbf{Z}_1 \cdots \mathbf{Z}_R] \in \mathbb{C}^{KM \times LR}$ where $\mathbf{Z}_r = \mathbf{x}_r \boxtimes [\mathbf{y}_{(r-1)L+1} \cdots \mathbf{y}_{(r-1)L+L}] \in \mathbb{C}^{KM \times L}$. $\text{Diag}(\mathbf{x})$ denotes a diagonal with the components of \mathbf{x} on its diagonal, and the $\text{vec}(\cdot)$ operator maps a matrix $[\mathbf{x}_1 \cdots \mathbf{x}_R] \in \mathbb{C}^{K \times R}$ into the vector $[\mathbf{x}_1^T \cdots \mathbf{x}_R^T]^T \in \mathbb{C}^{RK}$. The superscript $(\cdot)^T$ denotes matrix transpose, and a hat $(\hat{\cdot})$ denotes an estimate.

II. TENSOR-BASED SEPARATION OF SUMS OF COMPLEX EXPONENTIALS

This section reviews the method proposed in [7] for separation of SCE sources by means of a block-term tensor decomposition, under the assumption that each source has a small number of poles. We also discuss existence and uniqueness of solutions to the approximate decomposition problem which is addressed in practice.

A. Low-rank Hankel source model

A celebrated result in signal processing states that if a discrete-time signal is a linear combination of L damped complex exponentials, say

$$s(n) = \sum_{\ell=1}^L \alpha_\ell \exp(\zeta_\ell n), \quad n = 0, \dots, N-1, \quad (1)$$

where $\alpha_\ell, \zeta_\ell \in \mathbb{C}$, then the $M \times M$ Hankel matrix

$$\mathbf{H}_s \triangleq [\mathbf{s}(0) \quad \mathbf{s}(1) \quad \dots \quad \mathbf{s}(M-1)],$$

with $\mathbf{s}(n) \triangleq [s(n) \quad s(n+1) \quad \dots \quad s(n+M-1)]^T$ and $N = 2M-1$, has rank at most $\min\{L, M\}$. In fact, this property follows immediately from the decomposition [16]

$$\mathbf{H}_s = \mathbf{V}_s \text{Diag}(\alpha_1, \dots, \alpha_L) \mathbf{V}_s^T, \quad (2)$$

where \mathbf{V}_s is the Vandermonde matrix

$$\mathbf{V}_s \triangleq \begin{bmatrix} 1 & \dots & 1 \\ \exp(\zeta_1) & \dots & \exp(\zeta_L) \\ \vdots & & \vdots \\ \exp(\zeta_1(M-1)) & \dots & \exp(\zeta_L(M-1)) \end{bmatrix} \in \mathbb{C}^{M \times L},$$

and is at the heart of classical modal analysis methods [17]. It implies that a “simple” signal of the form (1) can be mapped into a low-rank Hankel matrix, where simple here means being constituted by a small number L of exponentials. We will see next how signal separation can be performed by relying on this relation.

Remark 1. For real-valued signals, usual conjugacy conditions must hold in (1). Correspondingly, complex-valued columns of \mathbf{V}_s as well as coefficients α_ℓ must arise in complex conjugate pairs.

B. Separation of linear mixture via block term decomposition

Consider now a linear instantaneous mixture $y(n) = \sum_{r=1}^R x_r s_r(n)$, with

$$s_r(n) = \sum_{\ell=1}^{L_r} \alpha_\ell^{(r)} \exp\left(\zeta_\ell^{(r)} n\right), \quad L_r < M, \quad (3)$$

and assume one wants to recover the signals $s_r(n)$ from knowledge of y (and of the above model) only. By linearity of the map discussed above, we have $y \mapsto \mathbf{H}_y = \sum_{r=1}^R x_r \mathbf{H}_{s_r}$, so that $\text{rank } \mathbf{H}_y \leq \sum_{r=1}^R L_r$. Without further information, though, this linear combination of matrices is not of much help for separation.

The situation changes upon introduction of spatial diversity, meaning we now observe $y(k, n) = \sum_{r=1}^R x_{k,r} s_r(n)$ for $k = 1, \dots, K$. In matrix notation, we have

$$\mathbf{Y} = \mathbf{X} \mathbf{S}^T, \quad (4)$$

where $\mathbf{S} = (s_{n,r}) = (s_r(n-1))$ is an $N \times R$ matrix containing the source signals and $\mathbf{X} = (x_{k,r})$ is a $K \times R$ mixture matrix specifying how the sources are combined to yield the channels' outputs. Each such output $y_k(n) = y(k, n)$ for a fixed k (i.e., each row of \mathbf{Y}) can be mapped into an $M \times M$ Hankel matrix as before, say $y_k \mapsto \mathbf{Y}_k$. Hence, $\mathbf{Y}_k = \sum_{r=1}^R x_{k,r} \mathbf{H}_r$, where \mathbf{H}_r is the rank- L_r Hankel matrix associated with s_r . The matrices \mathbf{Y}_k can be viewed as slices of an $M \times M \times K$ tensor \mathcal{Y} satisfying

$$\begin{aligned} \mathcal{Y} &= \sum_{k=1}^K \mathbf{Y}_k \otimes \mathbf{e}_k = \sum_{k=1}^K \left(\sum_{r=1}^R x_{k,r} \mathbf{H}_r \right) \otimes \mathbf{e}_k \\ &= \sum_{r=1}^R \mathbf{H}_r \otimes \left(\sum_{k=1}^K x_{k,r} \mathbf{e}_k \right) \\ &= \sum_{r=1}^R \mathbf{H}_r \otimes \mathbf{x}_r, \end{aligned} \quad (5)$$

where \mathbf{e}_k is the k th canonical basis vector of \mathbb{C}^K , \mathbf{x}_r is the r th column of \mathbf{X} and \otimes is the tensor product. The data tensor thus consists of a sum of blocks, each one given by the tensor product of a low-rank matrix and a vector. We refer to the parameters $(R, \{L_r\}_{r=1}^R)$ as structural parameters or simply the structure of the model in (5).

It turns out that the tensor decomposition (5), known as block-term decomposition (BTD) and introduced by [8], is essentially¹ unique under relatively mild assumptions. Its uniqueness properties have been first studied in [8], and further results were given in [7]. In particular, Theorem 2.4 of [7] states that if \mathbf{X} has full column rank and $\text{rank} \sum_{r=1}^R w_r \mathbf{H}_r > \max_r \text{rank } w_r \mathbf{H}_r$ for all $\mathbf{w} = [w_1 \dots w_R]^T$ having at least two nonzero components, then the BTD in (5) is essentially unique. (It is thus necessary that $L_r < M$.) Other uniqueness results have been derived in [18], [19], including some that apply to coupled BTDs.

C. Approximate block term decomposition

In practice, the data matrix \mathbf{Y} is only approximately given by (4), due to noise and imperfect modeling. Hence, one can only approximate tensor \mathcal{Y} by a low-rank BTD model of the form in (5).

¹Note that (5) can only be unique modulo a permutation of the summands and a joint rescaling of the components of each summand as in $(\mathbf{H}_r, \mathbf{x}_r) \mapsto (\alpha \mathbf{H}_r, (1/\alpha) \mathbf{x}_r)$ for some $\alpha \neq 0$.

Since the approximate BTD problem is important in its own right, we hereby discuss it from a general perspective, momentarily leaving aside the Hankel constraints in (5) and considering a third-order tensor $\mathcal{Y} \in \mathbb{C}^{I \times J \times K}$ (in model (5) we had $I = J = M$). Typically, an approximate BTD is computed by minimizing a measure of distance between the data tensor and a model of fixed structure with respect to the model components. Mostly often, a least-squares criterion is adopted (as in, e.g., [10]), leading to

$$\min_{(\mathbf{A}, \mathbf{B}, \mathbf{X}) \in \mathcal{S}} f(\mathbf{A}, \mathbf{B}, \mathbf{X}) \triangleq \left\| \mathcal{Y} - \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \otimes \mathbf{x}_r \right\|_F^2 \quad (6)$$

with $\mathcal{S} \triangleq \mathbb{C}^{I \times \sum_{r=1}^R L_r} \times \mathbb{C}^{J \times \sum_{r=1}^R L_r} \times \mathbb{C}^{K \times R}$, where \mathbf{A}_r contains the columns of indices $1 + \sum_{m=1}^{r-1} L_m$ to $\sum_{m=1}^r L_m$ of \mathbf{A} , and likewise for \mathbf{B}_r . Observe that the factorization $\mathbf{A}_r \mathbf{B}_r^T$ is employed to bound each block rank as $\text{rank } \mathbf{H}_r \leq L_r$.

We discuss next the existence and uniqueness of solutions to (6).

1) *Existence*: Problem (6) may lack a global minimizer, because the set of tensors having a given BTD structure is not necessarily closed. An example of this phenomenon has been known since the introduction of the BTD [10]. As recently shown in [11], spaces of real-valued tensors can contain sets with nonempty interior whose elements do not admit a best approximate BTD having a given structure. This fact has practical consequences, since it implies that a random tensor drawn from an absolutely continuous distribution has nonzero probability of falling into such a set. For complex-valued tensors, [20] shows that this issue only affects tensors from sets of zero volume, and thus is of lesser practical concern.

2) *Uniqueness*: In fact, the results of [20] not only imply that a closest tensor having a specified BTD structure exists for almost all complex-valued tensors, but also that it is unique (see [20, Corollary 7.4]). However, there is a subtlety: this simply means that for a random complex-valued tensor \mathcal{Y} the problem

$$\min_{\hat{\mathcal{Y}} \in \mathcal{B}_{L_1, \dots, L_R}} \|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2$$

has a unique solution almost surely, where $\mathcal{B}_{L_1, \dots, L_R}$ is the set of all complex-valued tensors which can be written in the form $\sum_{r=1}^R \mathbf{H}_r \otimes \mathbf{x}_r$ with $\text{rank } \mathbf{H}_r \leq L_r$. This does *not* imply that the *BTD components* themselves are unique, which requires additional conditions over these components, such as those stated in Section II-B.

For real-valued tensors, though, no analogue of the above mentioned result is known.

D. Structured low-rank approximation

In the special case of interest (5), \mathbf{H}_r must belong to the subspace of $M \times M$ Hankel matrices, \mathcal{H} . Although the slices \mathbf{Y}_k are Hankel by construction, there is no reason why a solution $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{X}})$ of (12) or (6) should satisfy $\hat{\mathbf{A}}_r \hat{\mathbf{B}}_r^T \in \mathcal{H}$. In fact, even if the sum $\sum_{r=1}^R \hat{x}_{k,r} \hat{\mathbf{A}}_r \hat{\mathbf{B}}_r^T$ is Hankel, the matrices $\hat{\mathbf{A}}_r \hat{\mathbf{B}}_r^T$ do not need to be (though the opposite is certainly true). In other words, the solution may lack temporal structure, not being interpretable as a mixture of sources of the form (1).

There are basically two ways of dealing with this issue:

- 1) *Projected unconstrained solution*: Ignore the Hankel structure to compute an unconstrained approximate low-rank BTD and then project each estimated block $\hat{\mathbf{H}}_r$ onto the set \mathcal{H}_{L_r} of Hankel matrices having rank up to L_r .
- 2) *Constrained solution*: Use a constrained optimization algorithm for imposing the Hankel structure while computing the solution of (5).

More generally, a similar choice is faced in related problems such as harmonic retrieval [21] or the so-called shape-from-moments problem [22], which can be cast as the computation of a low-rank approximation of a Hankel data matrix \mathbf{Y} . In practice, one either ignores the Hankel structure or solves a structured low-rank approximation problem of the form

$$\min_{\mathbf{H} \in \mathcal{H}_L} \|\mathbf{Y} - \mathbf{H}\|_F, \quad (7)$$

which can be addressed by structured low-rank approximation methods such as those thoroughly discussed in [23].

In principle, one would expect that more accurate results are attained when the Hankel structure is imposed, for this introduces an additional prior in the optimization process. However, this is not necessarily true: for instance, the experiments in [22] concerning the shape-from-moments problem show that a constrained algorithm can outperform an unconstrained one in a low-noise regime, even if the inverse holds at a higher level of noise. This behavior is probably explained by the fact that the local optimization algorithms used to solve the nonconvex problem (7) may attain a suboptimal local minimum. Furthermore, these constrained optimization algorithms are typically more computationally demanding than unconstrained ones.

Due to these difficulties, in practice unconstrained algorithms are often employed, as seen in [22], [21]. In the particular case of our BTD problem, to date only the projected unconstrained approach has been followed, as in [1], [4]. However, in spite of the above mentioned difficulties, the use of a constrained approach can lead to more accurate solutions, as our computer experiments will show in Section V.

Finally, note that the above discussion also applies to arbitrary constraints of the form $\mathbf{H}_r \in \mathcal{U}$ where \mathcal{U} is any² subspace of \mathbb{C} .

III. ALTERNATING GROUP LASSO ALGORITHM FOR BTD

In the following, we derive a provably convergent algorithm for computing an (unconstrained) approximate BTD of a given tensor. Subsequently, we show how linear (subspace) constraints can be imposed upon the block matrices, Hankel structure being a special case of these constraints.

A. Problem formulation

Instead of determining the BTD structure beforehand, one can include penalization terms promoting low-rank blocks and controlling the number of blocks in the formulation, as in

$$\min_{(\mathbf{A}, \mathbf{B}, \mathbf{X}) \in \mathcal{S}} F(\mathbf{A}, \mathbf{B}, \mathbf{X}) \triangleq f(\mathbf{A}, \mathbf{B}, \mathbf{X}) + \gamma g(\mathbf{A}, \mathbf{B}, \mathbf{X}), \quad (8)$$

²In some special cases, such as when \mathcal{U} is the space of circulant matrices, the structured low-rank matrix problem has a closed-form solution.

where $\mathcal{S} \triangleq \mathbb{C}^{I \times LR} \times \mathbb{C}^{J \times LR} \times \mathbb{C}^{K \times R}$, f is the same as in (6), $\gamma > 0$ is a regularization parameter and g is a regularization function of the form

$$g(\mathbf{A}, \mathbf{B}, \mathbf{X}) \triangleq \|\mathbf{A}\|_{2,1} + \|\mathbf{B}\|_{2,1} + \|\mathbf{X}\|_{2,1}. \quad (9)$$

Adding a mixed $\ell_{2,1}$ -norm regularization term is a well-known strategy for inducing *group sparsity* of its argument's columns. This is essentially a generalization of the lasso (least absolute shrinkage and selector operator) estimator principle, called group lasso, and is owed to geometric properties of the $\ell_{2,1}$ -norm [24].

Hence, for sufficiently high γ , minimizers of (8) will be formed by \mathbf{A} and \mathbf{B} displaying some columns made entirely of zeros, effectively yielding a BTD of low-rank blocks. The same applies to \mathbf{X} , possibly reducing the number of blocks. This allows much more flexibility compared to (6), since now the number of degrees of freedom of the model can adapt to the data \mathcal{Y} . Moreover, at least one solution is *guaranteed* to exist: because F is nonnegative, coercive³ (due to g) and continuous, existence follows from the extreme value theorem.

B. Algorithm for unconstrained blocks

To tackle the nonconvex and nonsmooth problem (8), we employ a block coordinate descent (BCD) approach. This widespread technique consists in partitioning the set of optimization variables and then sequentially solving subproblems in each subset of variables (with the others fixed) until all subsets are updated, completing one iteration of the algorithm.

Here, we have a natural partition into three blocks: \mathbf{A} , \mathbf{B} and \mathbf{X} . Let $\hat{\mathbf{A}}^{(t-1)}$, $\hat{\mathbf{B}}^{(t-1)}$ and $\hat{\mathbf{X}}^{(t-1)}$ denote the estimates obtained at iteration $t-1$. Fixing $\mathbf{B} = \hat{\mathbf{B}}^{(t-1)}$ and $\mathbf{X} = \hat{\mathbf{X}}^{(t-1)}$ in (8), the subproblem in \mathbf{A} of iteration t becomes a standard group lasso problem

$$\begin{aligned} \min_{\mathbf{A} \in \mathbb{C}^{I \times LR}} F_{\mathbf{A}}^{(t)}(\mathbf{A}), \\ \text{with } F_{\mathbf{A}}^{(t)}(\mathbf{A}) \triangleq \frac{1}{2} \|\mathcal{Y} - \mathcal{W}_{\mathbf{A}}^{(t)}(\mathbf{A})\|_F^2 + \gamma \|\mathbf{A}\|_{2,1}, \end{aligned} \quad (10)$$

where $\mathcal{W}_{\mathbf{A}}^{(t)}$ is a linear map depending on $\hat{\mathbf{B}}^{(t-1)}$ and $\hat{\mathbf{X}}^{(t-1)}$. The groups are disjoint and correspond to the columns of \mathbf{A} . Existing group lasso algorithms such as that in [25] can be readily invoked⁴ to solve (10), which is convex. Similar subproblems can be derived for \mathbf{B} and \mathbf{X} .

Now, despite being convex, subproblem (10) may not be strictly convex, because $\mathcal{W}_{\mathbf{A}}^{(t)}$ may not be injective at some iterations. It can thus fail to have a unique solution, and convergence to a stationary point cannot be guaranteed [26]. One can remedy this shortcoming by adding a proximal term, as in $\min_{\mathbf{A} \in \mathbb{C}^{I \times LR}} F_{\mathbf{A}}^{(t)}(\mathbf{A}) + \frac{\tau}{2} \|\mathbf{A} - \hat{\mathbf{A}}^{(t-1)}\|_F^2$, with $\tau > 0$. Putting $\mathbf{a} \triangleq \text{vec}(\mathbf{A})$ and $\mathbf{y}_1 \triangleq \text{vec}(\mathbf{Y}_{\langle 1 \rangle})$, where $\mathbf{Y}_{\langle 1 \rangle}$ indicates the mode-1 matrix unfolding of \mathcal{Y} (see *e.g.* [7]), this is equivalent to

$$\min_{\mathbf{a} \in \mathbb{C}^{ILR}} \frac{1}{2} \left\| \begin{bmatrix} \mathbf{y}_1 \\ \sqrt{\tau} \hat{\mathbf{a}}^{(t-1)} \end{bmatrix} - \begin{bmatrix} \mathbf{W}_{\mathbf{A}}^{(t)} \\ \sqrt{\tau} \mathbf{I} \end{bmatrix} \mathbf{a} \right\|_2^2 + \gamma \sum_{r=1}^R \sum_{l=1}^L \|\mathbf{a}_{r,l}\|_2, \quad (11)$$

³Recall that a function $F(\mathbf{A}, \mathbf{B}, \mathbf{X})$ is said to be coercive if and only if $F(x) \rightarrow \infty$ whenever $\|\mathbf{A}\|_F \rightarrow \infty$ or $\|\mathbf{B}\|_F \rightarrow \infty$ or $\|\mathbf{X}\|_F \rightarrow \infty$.

⁴This algorithm was, however, formulated for the real-valued setting only.

TABLE I: Pseudocode for constrained AGL algorithm.

Inputs:	Data tensor \mathcal{Y} , penalty parameter γ , proximal term weight τ , initial point $(\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \mathbf{X}^{(0)})$
Outputs:	Approximate BTD factors $(\mathbf{A}, \mathbf{B}, \mathbf{X})$
1: $t \leftarrow 1$	
2: while stopping criteria not met do	
3:	Solve group lasso subproblem (11) to obtain $\mathbf{A}^{(t)}$ from $\mathbf{A}^{(t-1)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$
4:	Solve group lasso subproblem in \mathbf{B} analogous to (11) to obtain $\mathbf{B}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$
5:	for $r = 1, \dots, R$ do
6:	$L_r^{(t)} \leftarrow \text{rank}(\hat{\mathbf{A}}_r^{(t)}(\hat{\mathbf{B}}_r^{(t)})^\top)$
7:	$(\mathbf{A}_r^{(t)}, \mathbf{B}_r^{(t)}) \leftarrow \text{slra}(\mathbf{A}_r^{(t)}(\mathbf{B}_r^{(t)})^\top, L_r^{(t)})$
8:	$(\mathbf{A}_r^{(t)}, \mathbf{B}_r^{(t)}) \leftarrow ([\mathbf{A}_r^{(t)} \mathbf{0}_{I \times L-L_r^{(t)}}], [\mathbf{B}_r^{(t)} \mathbf{0}_{I \times L-L_r^{(t)}}])$
9:	Solve group lasso subproblem in \mathbf{X} analogous to (11) to obtain $\mathbf{X}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t)}$ and $\mathbf{X}^{(t-1)}$
10:	$t \leftarrow t + 1$

where $\mathbf{W}_A^{(t)} \triangleq (\hat{\mathbf{X}}^{(t-1)} \odot_L \hat{\mathbf{B}}^{(t-1)}) \boxtimes \mathbf{I}_I$ and $\mathbf{a}_{r,l}$ holds components $((r-1)L+l-1)I+1$ to $((r-1)L+l)I$ of \mathbf{a} . By construction, (11) is strictly convex, since the Hessian of the least-squares term is positive definite. Analogous strictly convex subproblems can also be derived for \mathbf{B} and \mathbf{X} , with

$$\begin{aligned} \mathbf{W}_B^{(t)} &\triangleq (\hat{\mathbf{X}}^{(t-1)} \odot_L \hat{\mathbf{A}}^{(t)}) \boxtimes \mathbf{I}_J, \\ \mathbf{W}_X^{(t)} &\triangleq \left[(\hat{\mathbf{B}}^{(t)} \odot \hat{\mathbf{A}}^{(t)}) (\mathbf{I}_R \boxtimes \mathbf{1}_L) \right] \boxtimes \mathbf{I}_K. \end{aligned}$$

The AGL algorithm solves them in alternating fashion, cycling through updates of $\hat{\mathbf{A}}^{(t)}$, $\hat{\mathbf{B}}^{(t)}$ and $\hat{\mathbf{X}}^{(t)}$, in that order, at each iteration t . It can be seen as a regularized version of the alternating least squares scheme proposed in [10]. As we explain in Appendix A, AGL's formulation satisfies the conditions of [26, Theorem 2], and hence its iterates converge to a stationary point of problem (8).

It should be noted that AGL is a general approach to compute an unconstrained approximate low-rank BTD and is not specifically tailored to the application we consider later in the paper. Hence, it could also be applied, for example, to the problems presented in [1], [27] and [28].

C. Handling linear constraints in \mathbf{H}_r

In order to address the issue discussed in Section II-D, we propose a variant of AGL, termed constrained AGL (CAGL), to address the problem

$$\min_{(\mathbf{A}, \mathbf{B}, \mathbf{X}) \in \mathcal{S}} F(\mathbf{A}, \mathbf{B}, \mathbf{X}) \quad \text{subj. to} \quad \forall r, \mathbf{A}_r \mathbf{B}_r^\top \in \mathcal{U}, \quad (12)$$

with \mathcal{U} denoting a given subspace of $I \times J$ matrices. In CAGL, the constraint in (12) is enforced by applying a structured low-rank approximation (SLRA) algorithm during the iterations, after having estimated $\hat{\mathbf{A}}^{(t)}$ and $\hat{\mathbf{B}}^{(t)}$. For clarity, the algorithm is summarized in Table I, where notation was simplified by dropping the symbol (\cdot) from the computed estimates and `slra` denotes the SLRA algorithm, which takes a matrix $\hat{\mathbf{H}}_r$ and a maximum rank L_r as inputs and produces factors $\hat{\mathbf{A}}_r$ and $\hat{\mathbf{B}}_r$ having L_r columns and satisfying $\hat{\mathbf{H}}_r \approx \hat{\mathbf{A}}_r \hat{\mathbf{B}}_r^\top \in \mathcal{U}$.

Note that each application of `slra` re-estimates both \mathbf{A}_r and \mathbf{B}_r . Also, in practice the computation of $L_r^{(t)}$ at line 6

can be performed by counting how many of the columns of \mathbf{A}_r and \mathbf{B}_r are simultaneously nonzero. Finally, it should be noted that, since $\hat{\mathbf{A}}_r^{(t)}$ and $\hat{\mathbf{B}}_r^{(t)}$ always have L columns, zeros must be added in those blocks after applying `slra`, as done in line 8 of Table I.

As we will see in Section V, in practice CAGL is able to yield meaningful solutions to the source separation problem. However, constraining the block matrices brings a heavier computational load, due to use of the structured low-rank approximation algorithm, which is typically iterative. Furthermore, the arguments which allow showing the convergence of AGL no longer apply here.

Remark 2. In the particular case of a Hankel constraint, another way of imposing it is by estimating for each block matrix a low-rank decomposition of the form (2), where the Vandermonde factors are parameterized by their poles ζ_ℓ . In this case, a sparsity prior over the coefficients α_ℓ can be used to promote low-rank blocks. However, since ζ_ℓ and α_ℓ are generally complex-valued, conjugacy constraints would be needed to obtain real-valued block matrices, which requires fixing the (maximum) number of complex-conjugate pole pairs and of single real-valued poles of each block *a priori*. This loss of generality can be mitigated by using a high value of L for the starting rank (thus including ‘‘enough’’ conjugate pole pairs and single poles), but at the expense of increased computing cost.

IV. NUMERICAL EVALUATION ON RANDOM BLOCK TERM DECOMPOSITION MODELS

We now evaluate AGL in the approximate computation of synthetic BTD models, with and without Hankel constraint.

A. Unconstrained BTD

We generate 500 joint realizations of $(\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathcal{N})$ by drawing the (real-valued) elements of \mathbf{A} , \mathbf{B} , \mathbf{X} and \mathcal{N} in an independent and identically distributed (i.i.d.) fashion from the standard normal distribution. \mathbf{X} is then normalized column-wise and the condition $\max_{i,j} |\mathbf{x}_i^\top \mathbf{x}_j| < 0.9$ is imposed (by drawing \mathbf{X} multiple times until it is met). This prevents nearly collinear spatial signatures. Next, we construct the noisy model $\mathcal{Y} = \mathcal{Y}_0 + \sigma_{\mathcal{N}} \mathcal{N}$, where $\mathcal{Y}_0 = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{x}_r$ and $\sigma_{\mathcal{N}}$ is the standard deviation of the noise, which is adjusted to achieve a desired level of signal-to-noise ratio (SNR), defined as $\text{SNR} \triangleq \|\mathcal{Y}_0\|_F^2 \sigma_{\mathcal{N}}^{-2} \|\mathcal{N}\|_F^{-2}$.

We set $I = J = 16$ and $K = R = 3$. The block ranks are $L_1 = 8$, $L_2 = 6$ and $L_3 = 4$. These values were chosen to yield $\sum_{r=1}^3 L_r > \min\{I, J\} = I$, in which case one cannot employ an algebraic solution as described by [8, Theorem 4.1] to initialize the algorithm. Hence, in the absence of a better initialization strategy, we are forced to employ an arbitrarily chosen (in practice, a random) initial solution. Under these circumstances, the approximate BTD problem becomes particularly challenging because it is nonconvex. Moreover, as we shall see in Section V-C, the case where $\sum_{r=1}^R L_r > \min\{I, J\}$ is relevant in the application of BTD to source separation. Note that the procedure based on

simultaneous diagonalization proposed in [19] does not apply either, because we consider blocks with different ranks, which is also pertinent in the application dealt with in Section V-C.

For each realization, AGL is applied with $L = 8$ using the following procedure:

- starting from a random initial point, the algorithm is run with $\gamma = \gamma_0$, producing an initial solution;
- for $p = 1, \dots, P - 1$, it is run with $\gamma = \gamma_p = (p + 1)\gamma_0$, using the solution obtained for γ_{p-1} as the initial point.

This γ -sweeping procedure is inspired by solution-path techniques used in the statistics community [29] and produces a sequence of candidate solutions. Of all P candidate solutions, we keep the best one according to the normalized mean squared error (NMSE) over the blocks:

$$\text{NMSE}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{X}}) \triangleq \frac{1}{R} \sum_{r=1}^R \frac{\|(\mathbf{A}_r \mathbf{B}_r^T) \otimes \mathbf{x}_r - (\hat{\mathbf{A}}_r \hat{\mathbf{B}}_r^T) \otimes \hat{\mathbf{x}}_r\|_F^2}{\|(\mathbf{A}_r \mathbf{B}_r^T) \otimes \mathbf{x}_r\|_F^2}.$$

For comparison, Tensorlab's Gauss-Newton BTD algorithm⁵ (BTD-GN) [9] is used with the same rank L for all three blocks and also with the true ranks L_r .

To study sensitivity with respect to initialization, we run each algorithm starting at several random initial solutions: 12 initial points for AGL and 30 initial points for BTD-GN. (The first 12 initial points used for both algorithms are the same, and when the true ranks are used in BTD-GN, we keep only the first L_r columns of matrices $\hat{\mathbf{A}}_r^{(0)}$ and $\hat{\mathbf{B}}_r^{(0)}$ used in AGL.)

Fig. 1(a) shows the results for SNR = 10 dB. The n_i th curve (from bottom to top) of each algorithm is the empirical cumulative distribution function (ECDF) of the NMSE obtained by keeping the best solution among those given by the first n_i initializations. Clearly, AGL outperforms BTD-GN by a significant margin, even when the latter is given the true ranks (which is an unrealistic condition in practice). Indeed, a single run of AGL suffices to produce statistically better results than 30 runs of BTD-GN. The magnified portions of the plot show that the curves start to “accumulate” when n_i attains a certain value, suggesting that little or no improvement can be achieved by further increasing it. AGL is also visibly much more robust with respect to its initialization, since its curves are much less spread apart. The average computing time and standard deviation for a single run of each algorithm are: $\mu = 4.91$ s and $\sigma = 1.06$ s for AGL (taking into account the execution of AGL for the whole range of values of γ_p , as described above); $\mu = 4.44$ s and $\sigma = 3.89$ s for BTD-GN with ranks (8, 8, 8) and $\mu = 4.88$ s and $\sigma = 4.04$ s for BTD-GN with ranks (8, 6, 4).

Fig. 2(a)–(c) shows the proportion of block ranks estimated by AGL. All block ranks are well estimated most of the time, but there is a significant chance of underestimation of L_1 . This behavior probably comes from the choice of the γ parameter range, whose values may be slightly higher than necessary.

For a more comprehensive comparison, we repeat the experiment for other SNR values and then compute the median NMSE over the blocks attained by each algorithm. The results are shown in Fig. 3, where now the n_i th curve from top to

TABLE II: Pseudocode of Cadzow's algorithm [31] for a Hankel constraint.

Inputs:	Data matrix \mathbf{Y} , target rank L , tolerance ϵ_{CA} and maximum number of iterations T_{CA}
Outputs:	Low-rank Hankel approximation \mathbf{H} of \mathbf{Y}
<hr/> 1: Initialization: $t \leftarrow 1$, $\mathbf{H}^{(0)} \leftarrow \mathbf{Y}$ 2: for $t = 1, 2, \dots$ do 3: $\mathbf{H}^{(t)} \leftarrow \mathcal{P}_{\mathcal{H}}(\mathbf{H}^{(t-1)})$ 4: Compute the SVD: $\mathbf{H}^{(t)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ 5: Truncate the computed SVD at rank L : $\mathbf{H}^{(t)} \leftarrow \mathbf{U}_L \mathbf{\Sigma}_L \mathbf{V}_L^T$ 6: if $\ \mathbf{H}^{(t)} - \mathbf{H}^{(t-1)}\ _F < \epsilon_{\text{CA}} \ \mathbf{H}^{(t-1)}\ _F$ or $t = T_{\text{CA}}$ then 7: break for loop and output $\mathbf{H} = \mathbf{H}^{(t)}$ <hr/>	

bottom of each algorithm is computed by keeping the best result among the first n_i initializations. One can see that AGL is preferable over this entire SNR range, as the median NMSE of BTD-GN approaches that of AGL only for higher SNR values and when the algorithm is run several times (with the known ranks), thus incurring a high computing cost. Finally, the fact that the 12 curves of AGL are indistinguishable showcases its robustness regarding the choice of initialization.

The results obtained with BTD-GN highlights the difficulties that it faces in practice: frequent stop due to stagnation in a region of very slow convergence or failure to converge within the maximum allowed number of iterations (which is 1500, the same used for AGL). The employed multi-initialization scheme mitigates these problems, but only to some extent. These difficulties stem from the relatively large $\sum_r L_r$, but are also possibly related to the ill-posedness of problem (6): an optimal solution might not exist at all, or the algorithm may traverse regions close to tensors having no best approximation with the chosen structure, which causes numerical instability [30] and typically leads to a very slow convergence.

B. Constrained BTB

A similar evaluation was performed in the case where block matrices have Hankel structure. Apart from CAGL and BTD-GN, this evaluation also includes AGL in order to compare the performances of the constrained and unconstrained approaches. For the sake of simplicity, the structured low-rank approximation method used in CAGL is Cadzow's algorithm (CA) [31], which consists in performing alternating projections onto the Hankel subspace \mathcal{H} and onto the set of matrices with rank bounded by a prescribed value. Although the approximations provided by CA are not (locally) optimal in general [32], they are often satisfying in practice. Furthermore, CA is very simple to implement.

For clarity, a description of the employed CA is given in Table II. The projection onto \mathcal{H} , denoted by $\mathcal{P}_{\mathcal{H}}$, is cheap to compute by averaging anti-diagonals of $\mathbf{H}^{(t)}$. By the Eckart–Young theorem, the projection of $\mathbf{H}^{(t)}$ onto \mathcal{L} can be obtained via its truncated singular value decomposition (SVD), which is easy but costly to compute. Hence, to avoid an excessive cost per iteration of CAGL, we set the stopping criteria (see Table II) to $\epsilon_{\text{CA}} = 10^{-3}$ and $T_{\text{CA}} = 10$, implying that the constraint is approximately satisfied along the iterations.

In this scenario, we also generate 500 realizations of noisy constrained BTB models for each value of SNR. To this

⁵The Tensorlab function used in our experiments is `l1l1_nls`.

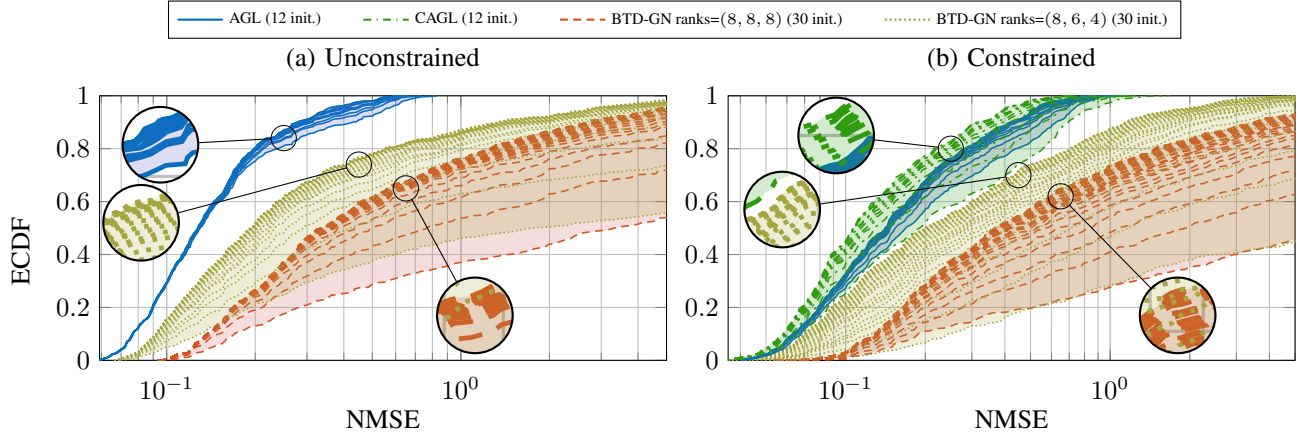


Fig. 1: Empirical CDFs of NMSE over estimated blocks attained by AGL and BTD-GN for 500 realizations of a noisy random BTD model (SNR = 10 dB), in the unconstrained and (Hankel-)constrained scenarios. The block ranks used in BTD-GN are indicated in the legend, while AGL and CAGL start with all ranks set to $L = 8$. The true ranks are $(L_1, L_2, L_3) = (8, 6, 4)$. BTD-GN and (C)AGL are run with 30 and 12 random initializations, respectively. The n_i th curve (from bottom to top) of each algorithm is the ECDF obtained by keeping the best solution among those given by the first n_i initializations. In the constrained scenario, all algorithms are followed by Cadzow's algorithm to enforce the low-rank Hankel constraint.

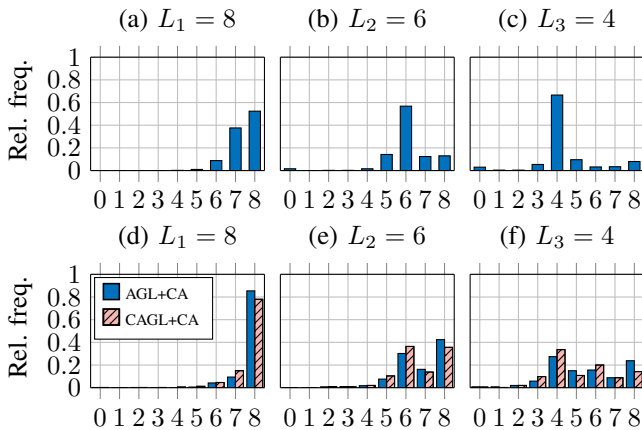


Fig. 2: Proportion of block ranks estimated by AGL and CAGL with SNR=10dB: (a)–(c) refer to the unconstrained scenario of Section IV-A and (d)–(f) refer to the constrained scenario of Section IV-B. In the constrained scenario, the rank estimates refer to the best obtained result (with respect to NMSE over blocks) among the first $n_i = 5$ runs.

end, the block matrices are random low-rank Hankel matrices generated using CA with $\epsilon_{CA} = 2.22 \times 10^{-16}$ (machine precision) and $T_{CA} = 1000$. The algorithms are run as in the previous section, but now we apply CA (with $\epsilon_{CA} = 10^{-10}$ and $T_{CA} = 1000$) to their outputs for enforcing the Hankel constraint.

The results obtained for SNR = 10 dB are depicted in Fig. 1(b) and show that, statistically, both AGL and CAGL produce more accurate results than BTD-GN. Though the performance of multi-initialization BTD-GN is now close to that of single-initialization CAGL, this only applies when the former is given the true block ranks. The curves also show that a single run of CAGL performs worse than AGL, but the opposite is true when more runs are performed. Furthermore,

CAGL is more sensitive with respect to its initialization, which is expected since the constrained problem is harder to solve. The block rank estimation is less accurate in this scenario than in the unconstrained case, as shown in Fig. 2(d)–(f): while L_1 is mostly well estimated, overestimation occurs for the other blocks, with CAGL producing slightly better estimates.

Evidently, CAGL is also considerably more costly than AGL: the average and standard deviation of computing times (for a single run) are $\mu = 8.96$ s and $\sigma = 2.51$ s for AGL; $\mu = 33.49$ s and $\sigma = 21.60$ s for CAGL; $\mu = 6.71$ s and $\sigma = 5.39$ s for BTD-GN with ranks (8, 8, 8) and $\mu = 7.10$ s and $\sigma = 5.09$ s for BTD-GN with ranks (8, 6, 4). Nonetheless, CAGL is able to attain a superior statistical performance when the multi-initialization scheme is used as the accumulation of the ECDF curves suggest that AGL's results cannot be further enhanced by more runs. Therefore, CAGL is a more costly but more accurate alternative than the other methods in this scenario.

Finally, the curves of Fig. 3(b) show that the above conclusions hold also for other values of SNR, though the advantage of CAGL and AGL with respect to BTD-GN decreases as the SNR grows.

V. EXPERIMENTAL RESULTS WITH ECG DATA

A. Tensor representation of ECG signals

ECG produces a time plot that represents the heart's electrical activity recorded from electrodes placed on the body surface. The ECG data matrix with K leads and N samples can be modeled as (4), where $\mathbf{X} \in \mathbb{R}^{K \times R}$ is the mixing matrix, that models the propagation of the cardiac electrical sources from the heart to the body surface, $\mathbf{S} \in \mathbb{R}^{N \times R}$ is the source matrix that contains the atrial, ventricular, and possibly disturbance sources, and R is the number of sources [14].

AA extraction in AF ECG recordings can be viewed as a BSS problem where the goal is to estimate the matrices \mathbf{X}

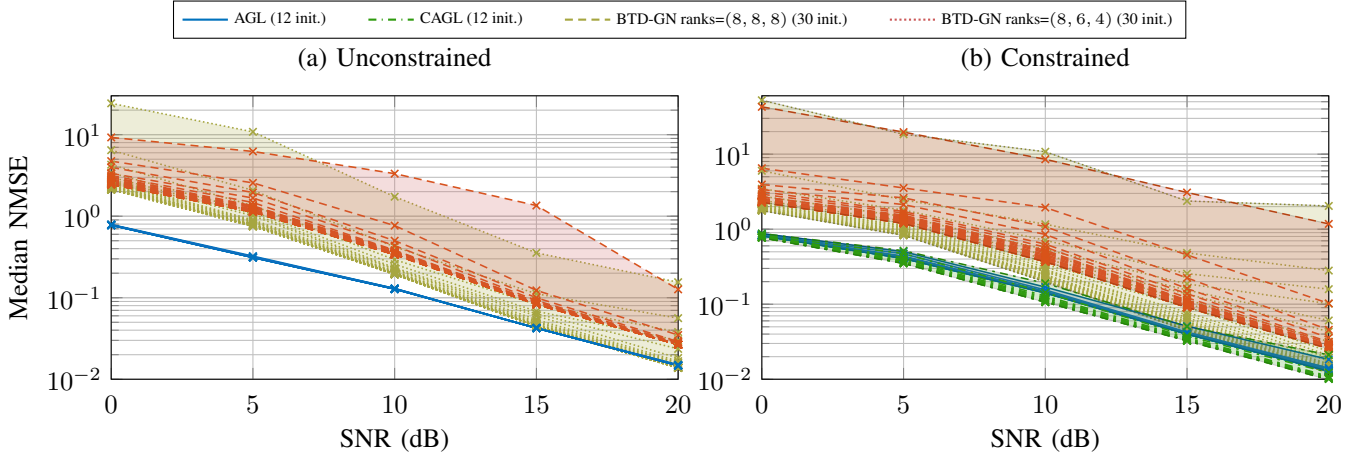


Fig. 3: Median NMSE over estimated blocks attained by AGL and BTD-GN for 500 realizations of a noisy random BTD model, in the unconstrained and (Hankel-)constrained scenarios. The true ranks are $(L_1, L_2, L_3) = (8, 6, 4)$. BTD-GN and (C)AGL are run with 30 and 12 random initializations, respectively. The n_i th curve (from top to bottom) of each algorithm is obtained by keeping the best solution among those given by the first n_i initializations. In the constrained scenario, all algorithms are followed by Cadzow’s algorithm to enforce the low-rank Hankel constraint.

and \mathbf{S} only from the observed data matrix \mathbf{Y} . The tensor built from the ECG data matrix as described in Section II-B then satisfies (5), where $\mathbf{H}_r \in \mathbb{R}^{M \times M}$ is a Hankel matrix built from the r th column of \mathbf{S} , and thus contains samples of the r th ECG source. Vector \mathbf{x}_r , which is the r th column of \mathbf{X} , quantifies the contribution of this source to each electrode’s output, and so can be thought of as its spatial signature.

Due to the quasi-periodic nature of AF, atrial sources can be represented by the SCE model (1) with a small number of exponentials. Ventricular sources, in their turn, are typically composed by a few transient components, and thus can also be well modeled by (1) with small L . Hence, these signals can be mapped into low-rank Hankel matrices, as discussed in Section II-B.

B. Semi-synthetic AF data

The usefulness of CAGL for ECG source separation is now assessed by resorting to a semi-synthetic AF data model. To simulate the AA signal during AF, the model proposed in [33] that mimics the sawtooth pattern (a typical characteristic of the f waves) is used. This model is given by

$$s(n) = -\sum_{p=1}^P a_p(n) \sin(p\theta(n)) \quad (13)$$

with modulated amplitude and phase respectively given by

$$a_p(n) = \frac{2}{p\pi} \left[a + \Delta a \sin\left(2\pi \frac{f_a}{F_s} n\right) \right]$$

and

$$\theta(n) = 2\pi \frac{f_0}{F_s} n + \left(\frac{\Delta f}{F_f}\right) \sin\left(2\pi \frac{F_f}{F_s} n\right),$$

where a is the sawtooth amplitude, Δa is the modulation peak amplitude, f_a is the amplitude modulation frequency, F_s is the sampling frequency, f_0 is the frequency value in which $\theta(n)$ varies sinusoidally, Δf is the maximum frequency deviation and F_f is the modulation frequency.

TABLE III: Parameters of the synthetic AA signal model (13).

Model	P	a	Δa	f_a	F_s	f_0	Δf	F_f
1	5	150	50	0.08	1000	6	0.2	0.10
2	3	60	18	0.50	1000	8	0.3	0.23

1) *One AA source:* We first consider a scenario with one AA source $s(n)$, which is generated using (13) with the parameters of Model 1 given in Table III. This signal is shown in Fig. 4(a). A random spatial signature $\mathbf{x} \in \mathbb{R}^{12}$ over all 12 ECG leads is generated for this source, having standard normal i.i.d. components. The ventricular activity (VA) source is taken from a real 12-lead ECG of a healthy person, after P wave suppression as in [2]. This ECG, that belongs to the database of [34], is acquired at a sampling rate of 1 kHz and is preprocessed by a zero-phase forward-backward type-II Chebyshev bandpass filter with cutoff frequencies of 0.5 and 30 Hz, in order to suppress high-frequency noise and baseline wandering. Additive white Gaussian noise (AWGN) with variance σ^2 was added, yielding the overall model

$$\mathbf{Y} = \mathbf{V} + \alpha \mathbf{x} \mathbf{s}^T + \mathbf{N} \in \mathbb{R}^{12 \times N}, \quad (14)$$

where \mathbf{V} holds the normalized VA signal, $\mathbf{s} \in \mathbb{R}^N$ holds the AA signal samples, \mathbf{N} contains the AWGN samples and $\alpha = 2$ is a scaling factor chosen to obtain an average atrial-to-ventricular power ratio consistent with clinical observations. A window of about 1.2 seconds is used, yielding 1221 samples. An example of the overall generated signal is shown in Fig. 4(b) (dashed curve). A direct Hankelization of this matrix yields a tensor of dimensions $611 \times 611 \times 12$, whose approximate BTD demands a large computing time. To reduce it, we downsample the signals by a factor of 10 before computing the decomposition. The resulting tensor \mathcal{Y} has dimensions $62 \times 62 \times 12$.

CAGL is run with the same γ -sweeping procedure used in Section IV, but now with γ taking 30 equispaced values in the interval $[8 \times 10^{-4}, 0.33 \times 10^{-2}]$ and keeping the last

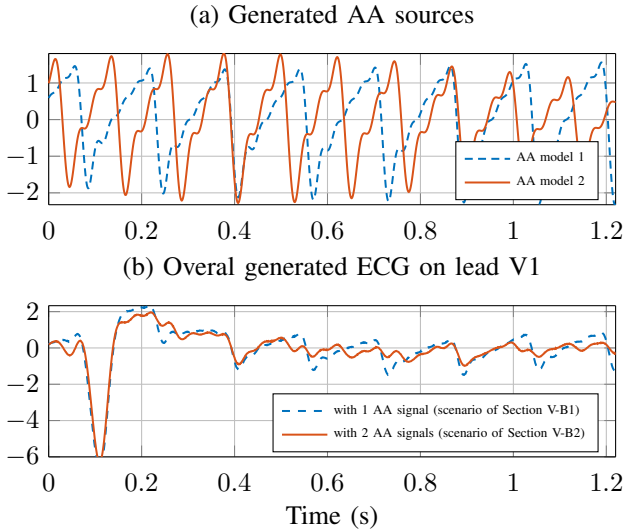


Fig. 4: Examples of generated semi-synthetic models: (a) AA sources following model (13) with the parameters shown in Table III; (b) overall synthesized ECG signals on lead V1.

solution. For γ_0 , we start the algorithm with $R = 6$ random blocks of rank $L = 40$. Among the estimated sources, the AA source is chosen as that which maximizes (in absolute value) the correlation coefficient ρ with respect to the ground truth $s(n)$.

BTD-GN is run to estimate R blocks of fixed rank L , for all combinations $(R, L) \in \{4, 5, 6\} \times \{1, \dots, 40\}$. For each rank L , the initial point is generated by filling the L columns of \mathbf{A}_r and \mathbf{B}_r with the first L columns used to initialize these variables in CAGL.

This procedure is repeated 30 times for each of 10 different realizations of (\mathbf{x}, \mathbf{N}) . We found that the results of CAGL are remarkably consistent, producing very similar estimates regardless of the chosen initial point. By contrast, the results obtained with BTD-GN are much more sensitive in this respect. Specifically, the value of L that yields the best performance for a given run is not the same across different runs, as shown by the ECDF of Fig. 5(a). For CAGL, the rank chosen for the AA signal block is almost always 10 as seen in Fig. 5(a).

Though the most adequate choice for BTD-GN seems to be $R = 4$ and $L \in \{10, \dots, 15\}$, its performance is highly variable for this range of L . This is seen in Fig. 6(a), which displays the histogram of the correlation coefficient ρ (in absolute value) between the estimated AA source and the ground truth. We have included all results produced by BTD-GN for $L \in \{10, \dots, 15\}$, and all results produced by CAGL. It can be seen that the choice of R significantly affects performance, and a large proportion of results given by BTD-GN achieves a poor ρ for every R . By contrast, ρ is very likely to be quite close to 1 for CAGL.

In conclusion, BTD-GN only produces good results with a proper combination of R , L and the initial point. By contrast, CAGL only requires choosing a reasonable range for γ and behaves much more robustly with regard to initialization.

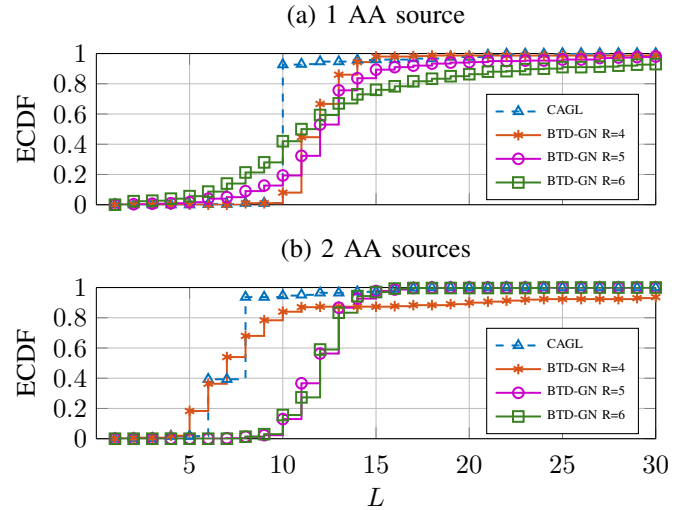


Fig. 5: Empirical distribution of rank chosen by CAGL for the AA source and of rank L yielding the best AA extraction for BTD-GN with different numbers of blocks R .

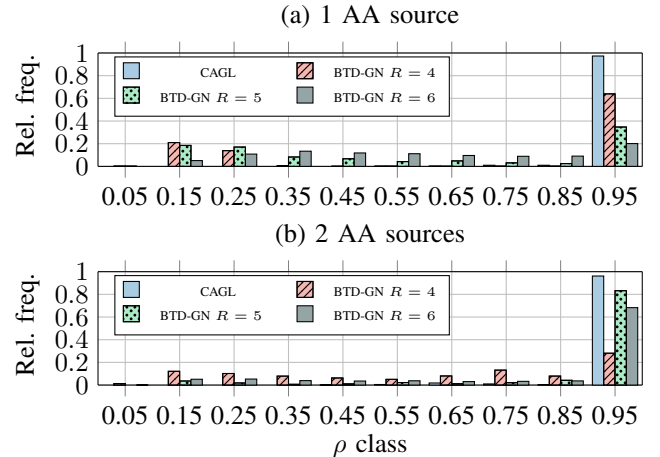


Fig. 6: Histogram of computed correlation coefficient ρ (in absolute value) between true and estimated AA sources.

2) *Two AA sources*: In this scenario, the VA source is still the same as in the previous scenario, but now two AA sources are generated, each one using the parameters of one row of Table III. These AA signals are shown in Fig. 4(a). Accordingly, the model now reads

$$\mathbf{Y} = \mathbf{V} + \mathbf{X}\mathbf{S}^T + \mathbf{N} \in \mathbb{R}^{12 \times N}, \quad (15)$$

where each column of $\mathbf{S} \in \mathbb{R}^{N \times 2}$ contains one of the AA signals and those of $\mathbf{X} \in \mathbb{R}^{12 \times 2}$ hold their respective spatial signatures. Signals are again downsampled by a factor of 10, yielding data tensors with the same dimensions as before. Fig. 4(b) displays one example of the overall generated ECG signal (solid curve).

CAGL and BTD-GN are run following the same procedure as in the case with one AA source. However, here the estimated AA sources are extracted by choosing first the block with the highest correlation coefficient (in absolute value) with one of

the reference AA sources, and secondly the block maximizing the correlation with the remaining AA source.

The corresponding results are shown in Fig. 5(b) and Fig. 6(b). The conclusions are similar to the previous scenario, with three main differences: (i) BTD-GN now performs best with $R = 5$, as expected; (ii) its performance is closer to that of CAGL for this choice of R ; (iii) two choices of rank were most often made by CAGL (rather than one), namely $L_r = 6$ and $L_r = 8$.

Note that CAGL is run with exactly the same procedure as in the previous scenario. By contrast, BTD-GN now yields best results with a different choice of R . This highlights the fact that CAGL can effectively adapt to a given dataset, typically behaving more stably than the usual approach across different circumstances.

C. Real AF data

To further study the usefulness of CAGL in the target application, we perform experiments with real-world standard 12-lead ECG recordings from five patients suffering from persistent AF. These recordings belong to a database provided by the Cardiology Department of Princess Grace Hospital Center, in Monaco. They are acquired at a 977 Hz sampling rate and are preprocessed by a zero-phase forward-backward type-II Chebyshev bandpass filter with cutoff frequencies of 0.5 and 40 Hz, in order to suppress high-frequency noise and baseline wandering. For each patient, the segment with the largest TQ interval recording is chosen for the experiment. The recordings lengths range from about 1.08 to 1.40 seconds.

First, we downsample all signals by a factor of 10. This decreases computing cost, with practically negligible information loss. After normalization of each signal tensor, CAGL is applied using the same γ -sweeping strategy as in Section V-B. However, here we choose the final solution by inspection of the separated signals.

To assess AA estimation, we employ two commonly used performance parameters. The first one is spectral concentration (SC), *i.e.*, the relative amount of energy around the dominant frequency (DF), computed as in [35]:

$$SC = \left(\sum_{f_i=0.82f_p}^{1.17f_p} P_{AA}(f_i) \right) \left(\sum_{f_i=0}^{F_s/2} P_{AA}(f_i) \right)^{-1},$$

where f_p is the value of the DF, defined as $\arg \max_{f_i} P_{AA}(f_i)$, F_s is the sampling frequency, f_i is the discrete frequency and P_{AA} is the power spectrum of the AA signal computed using Welch's method as in [35]. An AA signal during AF typically should have a DF between 3 and 9 Hz with high SC. The second parameter is kurtosis of the signal in the frequency domain, acquired by a 4096-point FFT. As in [36], we use a sample-based estimate $\hat{\kappa}$ of the general expression of kurtosis valid for non-circular complex data, given by

$$\hat{\kappa} = \frac{E[|S_r(k)|^4] - 2E[|S_r(k)|^2]^2 - |E[S_r(k)]|^2}{E[|S_r(k)|^2]^2}$$

where $S_r(k)$ denotes the FFT of the r th source. As kurtosis measures peakedness and sparsity of a distribution, it naturally provides a quantitative measure of harmonicity of the signal when computed in the frequency domain. A high kurtosis is thus suggestive of a harmonic signal like AA during AF [4].

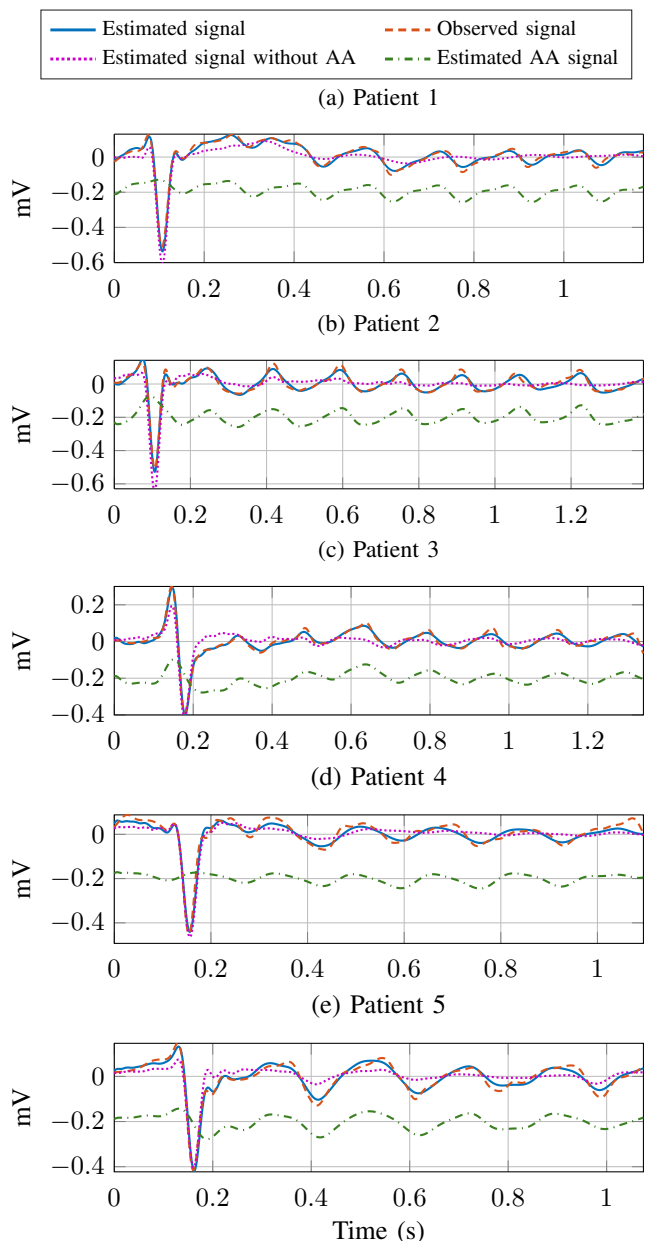


Fig. 7: Results produced by CAGL with real-world ECG data: observed and estimated signals at lead V1. The estimated AA signals are vertically shifted by -0.2 mV for ease of visualization.

Table IV displays a quantitative assessment of the potential AA sources extracted for each patient, while Fig. 7 shows the estimated overall and AA signals for each patient along with the observed signals. Note that, for Patients 2 and 5, two blocks were identified as potential AA sources; accordingly, the estimated AA signals in Fig. 7(b) and Fig. 7(e) are given by the linear combinations of these blocks multiplied by their corresponding spatial weights for lead V1. These distinct potential AA sources for Patients 2 and 5 are shown in Fig. 8, along with their respective spatial weights (*i.e.*, their associated columns in \mathbf{X}). The shown signals are normalized as $[\max_n |s_r(n)|]^{-1} s_r(n)$, and the corresponding spatial weights

TABLE IV: Block ranks of the sources extracted by CAGL and characteristics of the potential AA sources.

Patient	Non-AA source ranks	AA source	SC (%)	DF (Hz)	$\hat{\kappa}$	AA source rank
1	3, 9, 10, 16, 18	1	80.23	6.44	163.69	8
2	28, 32	1	59.36	6.20	116.77	29
		2	82.05	6.20	163.77	12
3	21, 27, 29	1	66.13	6.20	132.29	20
4	8, 16, 20	1	74.51	5.96	196.16	10
5	32, 38, 39	1	91.70	5.72	348.42	10
		2	78.63	5.01	166.95	21

are then rescaled so as to absorb the factor $\max_n |s_r(n)|$. Based on the results of Table IV and Fig. 7, it is seen that a satisfying extraction is achieved, as the potential AA sources have typical f-wave features, relatively high SC and DF between 5 and 6.5 Hz.

Regarding Patient 2, the values of SC and $\hat{\kappa}$ for its AA source 1 are not so high due to the residual of the T-wave (ventricular repolarization), which can be seen around 0.3 seconds of Fig. 8(a). Furthermore, Fig. 8(b) shows that the contribution of AA source 2 decays along the path V1–V2–V3, which suggests that this source may reflect electrical activity mainly occurring in the right atrium, and is almost null on other leads. By contrast, AA source 1 gives significant contributions to most leads, which suggests that the associated electrical propagation pattern may be harbored in a region including both atria. In fact, these sources have very different spatial signatures: the cosine of the angle between their respective columns of \mathbf{X} is around -0.13, and so they are far from being collinear. Furthermore, despite having the same DF, their observed power spectra are considerably different. The lack of temporal synchronization between the estimated atrial sources, as manifested by the time lag between the maxima of the two signals plotted in Fig. 8(a), further supports the hypothesis that the associated activities may arise from different areas of atrial tissue.

Patient 5 provides another example in which two potential AA sources with considerably different spatial signatures are extracted: their respective columns in \mathbf{X} form an angle whose cosine is about 0.09. Moreover, the dominant frequencies computed for these sources also differ significantly: 5.72 for AA source 1 and 5.01 for AA source 2. Here, it is AA source 2 that exhibits significant contributions to most leads, while AA source 1 manifests itself mostly on lead V1 and thus may correspond to more localized electrical activity taking place in the right atrium.

While the possibility of extracting more than one atrial source presents great interest for the noninvasive analysis of AF, a thorough validation of this result would be required by means of ground truth data such as a full electroanatomical mapping of the atria performed during catheter ablation interventions.

A final important observation is that for all but one patient, the sum of the estimated block ranks exceeds the dimension of the Hankel matrices: for Patient 1, $\sum_r L_r = 64 > M = 63$; for Patient 2, $\sum_r L_r = 101 > M = 74$; for Patient 3, $\sum_r L_r = 97 > M = 71$; and for Patient 5, $\sum_r L_r =$

$140 > M = 58$. This corroborates the practical importance of the case where $\sum_r L_r$ exceeds all tensor dimensions. Furthermore, for all patients the sum of block ranks far exceeds the number of leads, *i.e.*, $\sum_r L_r > K$, which showcases the benefit of using a tensor method, since a matrix decomposition approach could not possibly identify all the poles constituting each model.

VI. CONCLUSION

We have proposed a convergent alternating optimization algorithm for a well-posed penalized formulation of the approximate BTD problem that jointly estimates the model structure and its parameters. The resulting subproblems can be solved by existing group lasso methods. Moreover, linear (subspace) structure can be imposed on the block matrices by using a structured low-rank approximation method, though a study of the convergence is still needed in this case. Experimental results with random tensors show that our approach is much more robust with respect to initialization than the conventional least-squares approach (without regularization).

To illustrate its practical usefulness, our algorithm has been applied to extract atrial activity signals from ECG recordings of atrial fibrillation episodes. In this problem, Hankel constraints must be imposed on the block matrices. Our results with semi-synthetic and real-world data highlight the ability of this approach to consistently perform an effective separation without the need of choosing structural parameters *a priori*. A particularly interesting result is the occurrence in (two of the five) real-world examples of two distinct sources with f-wave characteristics and very different spatial signatures. Further study is needed to give this result a more thorough physiological interpretation.

As future developments on approximate BTD computation, we can mention studies on how to properly choose γ in practice and on the convergence of CAGL using a locally optimal structured low-rank approximation method to impose the linear constraints. Regarding the AA extraction in AF ECGs, future work would aim at analyzing the occurrence of multiple atrial sources, as well as performing the experiments in a large database of AF patients in order to provide more statistically significant results. Finally, the classification of subjects on the basis of their extracted ECG signals can also be envisaged.

APPENDIX A

CONVERGENCE OF (UNCONSTRAINED) AGL

AGL's formulation satisfies the conditions of Theorem 2 of [26] for the following reasons:

- since the directional derivative of g (see (9)) exists at every point $(\mathbf{A}, \mathbf{B}, \mathbf{X}) \in \mathcal{S}$ and AGL's subproblems satisfy $F(\mathbf{A} | \hat{\mathbf{B}}^{(t)}, \hat{\mathbf{X}}^{(t)}) + \frac{\tau}{2} \|\mathbf{A} - \mathbf{A}^{(t)}\|_F^2 \geq F(\mathbf{A} | \hat{\mathbf{B}}^{(t)}, \hat{\mathbf{X}}^{(t)})$ with equality at $\mathbf{A} = \mathbf{A}^{(t)}$ and likewise for \mathbf{B} and \mathbf{X} , Assumption 2 of [26] holds (see [26, Proposition 2]);
- $F(\mathbf{A}, \mathbf{B}, \mathbf{X})$ is continuous and coercive, and thus its sublevel sets $\mathcal{F}_c = \{(\mathbf{A}, \mathbf{B}, \mathbf{X}) \in \mathcal{S} : F(\mathbf{A}, \mathbf{B}, \mathbf{X}) \leq c\}$, with $c \in \mathbb{R}$, are compact;

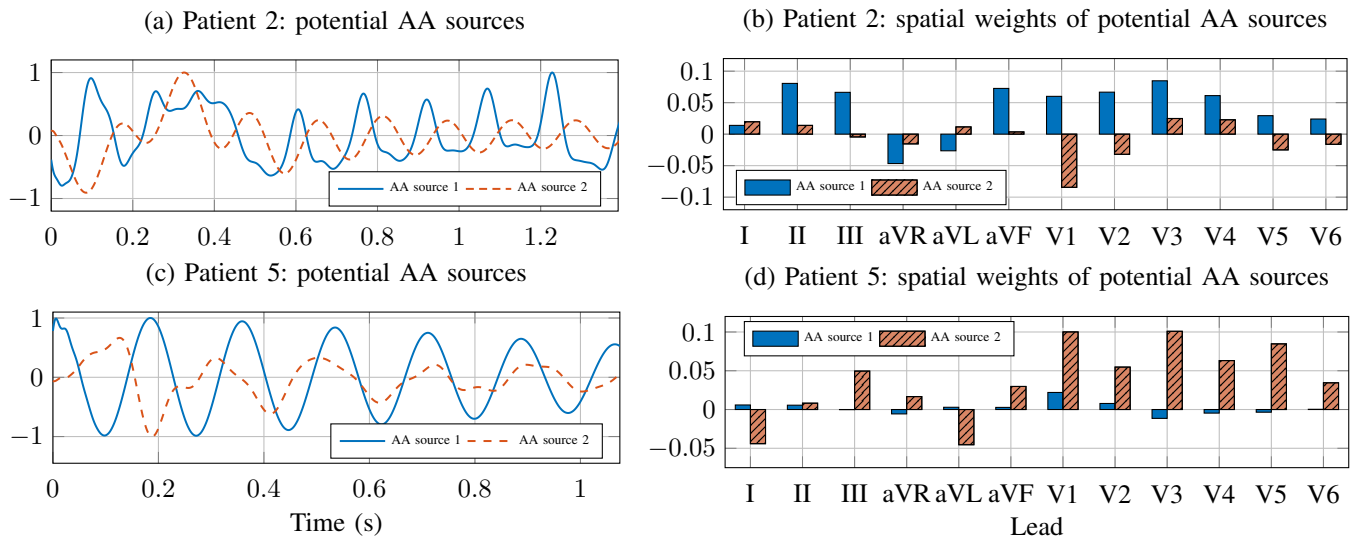


Fig. 8: Results produced by CAGL with real-world ECG data from Patients 2 and 5.

- AGL's subproblems are strictly convex, thus having unique minimizers;
- as $f(\mathbf{A}, \mathbf{B}, \mathbf{X}) < \infty$ and f is Gâteaux-differentiable on all $(\mathbf{A}, \mathbf{B}, \mathbf{X}) \in \mathcal{S}$, [37, Lemma 3.1] implies F is regular at every point $(\mathbf{A}, \mathbf{B}, \mathbf{X}) \in \mathcal{S}$.

ACKNOWLEDGMENT

The authors would like to thank O. Meste for helpful discussions on the application. J. H. de M. Goulart and P. Comon were supported by the European Research Council under the European Programme FP7/2007-2013, Grant AdG-2013-320594 "DECODA." P. M. R. de Oliveira is funded by a PhD scholarship from the IT Doctoral School (ED STIC) of the Université Côte d'Azur. V. Zarzoso is a member of the *Institut Universitaire de France*.

REFERENCES

- [1] B. Hunyadi, D. Camps, L. Sorber, W. Van Paesschen, M. De Vos, S. Van Huffel, and L. De Lathauwer, "Block term decomposition for modelling epileptic seizures," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 139, 2014.
- [2] L. N. Ribeiro, A. R. Hidalgo-Muñoz, and V. Zarzoso, "Atrial signal extraction in atrial fibrillation electrocardiograms using a tensor decomposition approach," in *Proc. IEEE Eng. Med. Biol. Soc. Conf. (EMBC)*. Milan, Italy: IEEE, aug 2015, pp. 6987–6990.
- [3] V. Zarzoso, "Parameter estimation in block term decomposition for noninvasive atrial fibrillation analysis," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*. Curaçao, Dutch Antilles: IEEE, dec 2017.
- [4] P. M. R. de Oliveira and V. Zarzoso, "Source analysis and selection using block term decomposition in atrial fibrillation," in *Proc. 14th Int. Conf. , LVA/ICA 2018*. Guildford, UK: Springer, jul 2018, pp. 46–56.
- [5] —, "Block term decomposition of ECG recordings for atrial fibrillation analysis: Temporal and inter-patient variability," *J. Commun. and Inf. Syst.*, vol. 34, no. 1, pp. 111–119, Apr. 2019.
- [6] I. Markovsky, O. Debals, and L. D. Lathauwer, "Sum-of-exponentials modeling and common dynamics estimation using tensorlab," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14 150 – 14 155, 2017, 20th IFAC World Congress.
- [7] L. De Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in rank- $(L_r, L_r, 1)$ terms," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1451–1474, 2011.
- [8] —, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008.
- [9] N. Vervliet, O. Debals, L. Sorber, M. V. Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016, available online. URL: <http://www.tensorlab.net>.
- [10] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms—Part III: Alternating least squares algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1067–1083, 2008.
- [11] J. H. de M. Goulart and P. Comon, "On the minimal ranks of matrix pencils and the existence of a best approximate block-term tensor decomposition," *Linear Algebra Appl.*, vol. 561, pp. 161–186, Jan. 2019.
- [12] X. Han, L. Albera, A. Kachenoura, H. Shu, and L. Senhadji, "Block term decomposition with rank estimation using group sparsity," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*. Curaçao, Dutch Antilles: IEEE, dec 2017.
- [13] J. J. Rieta, F. Castells, C. Sánchez, V. Zarzoso, and J. Millet, "Atrial activity extraction for atrial fibrillation analysis using blind source separation," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1176–1186, jul 2004.
- [14] V. Zarzoso, "Extraction of ECG characteristics using source separation techniques: exploiting statistical independence and beyond," in *Adv. Biosignal Process.* Springer, 2009, pp. 15–47.
- [15] L. N. Ribeiro, A. L. De Almeida, and V. Zarzoso, "Enhanced block term decomposition for atrial activity extraction in atrial fibrillation ECG," in *Proc. Sensor Array Multichannel Signal Process. Workshop (SAM), 2016 IEEE*. Rio de Janeiro, Brazil: IEEE, jul 2016, pp. 1–5.
- [16] D. L. Boley, F. T. Luk, and D. Vandevoorde, "Vandermonde factorization of a Hankel matrix," in *Proc. Workshop Scientific Comput.*, Hong Kong, 1997.
- [17] L. L. Scharf and C. Demeure, *Statistical signal processing: detection, estimation, and time series analysis*. Addison-Wesley Reading, MA, 1991, vol. 63.
- [18] M. Sørensen and L. D. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(l_r, n, l_r, n, 1)$ terms—Part I: Uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 2, pp. 496–522, 2015.
- [19] M. Sørensen, I. Domanov, and L. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(l_r, n, l_r, n, 1)$ terms—Part II: Algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 3, pp. 1015–1045, 2015.
- [20] Y. Qi, M. Michałek, and L.-H. Lim, "Complex tensors almost always have best low-rank approximations," *arXiv preprint arXiv:1711.11269v2*, sep 2018.
- [21] S.-Y. Kung, K. S. Arun, and D. V. B. Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *J. Optical Soc. of America*, vol. 73, no. 12, pp. 1799–1811, 1983.

- [22] M. Elad, P. Milanfar, and G. H. Golub, "Shape from moments—An estimation theory perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1814–1829, 2004.
- [23] I. Markovsky, *Low-rank approximation*, ser. Communications and Control Engineering. Springer, 2012.
- [24] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Stat. Soc.: Series B (Stat. Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [25] J. Liu and J. Ye, "Moreau-Yosida regularization for grouped tree structure learning," in *Proc. Adv. Neural Inf. Process. Systems*, 2010, pp. 1459–1467.
- [26] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.
- [27] L. De Lathauwer and A. de Baynast, "Blind deconvolution of DS-CDMA signals by means of decomposition in rank- $(1, l, l)$ terms," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1562–1571, 2008.
- [28] J. Spiegelberg, J. Rusz, and K. Pelckmans, "Tensor decompositions for the analysis of atomic resolution electron energy loss spectra," *Ultramicroscopy*, vol. 175, pp. 36–45, 2017.
- [29] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the lasso and generalizations*. CRC press, 2015.
- [30] P. Breiding and N. Vannieuwenhoven, "The condition number of join decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 39, no. 1, pp. 287–309, 2018.
- [31] J. A. Cadzow, "Signal enhancement—a composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, 1988.
- [32] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, "Structured low rank approximation," *Linear Algebra Appl.*, vol. 366, pp. 157–172, 2003.
- [33] M. Stridh and L. Sornmo, "Spatiotemporal QRST cancellation techniques for analysis of atrial fibrillation," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 105–111, 2001.
- [34] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [35] F. Castells, J. J. Rieta, J. Millet, and V. Zarzoso, "Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 2, pp. 258–267, Feb. 2005.
- [36] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 248–261, 2010.
- [37] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.