Direction-of-Arrival Estimation through Exact Continuous l20-Norm Relaxation

Emmanuel Soubies, Adilson Chinatto, Pascal Larzabal, João M T Romano, Laure Blanc-Féraud

Direction-of-Arrival Estimation through Exact Continuous $\ell_{2,0}$-Norm Relaxation

Emmanuel Soubies, Adilson Chinatto, Pascal Larzabal, João M. T. Romano, and Laure Blanc-Féraud

***Abstract*—On-grid based direction-of-arrival (DOA) estimation methods rely on the resolution of a difficult group-sparse optimization problem that involves the $\ell_{2,0}$ pseudo-norm. In this work, we show that an exact relaxation of this problem can be obtained by replacing the $\ell_{2,0}$ term with a group minimax concave penalty with suitable parameters. This relaxation is more amenable to non-convex optimization algorithms as it is continuous and admits less local (not global) minimizers than the initial $\ell_{2,0}$-regularized criteria. We then show on numerical simulations that the minimization of the proposed relaxation with an iteratively reweighted $\ell_{2,1}$ algorithm leads to an improved performance over traditional approaches.**

***Index Terms*—DOA, MMV-sparse optimization, Exact relaxations, $\ell_{2,0}$-norm minimization.**

I. Introduction

DIRECTION-OF-ARRIVAL (DOA) estimation is of fundamental importance in array signal processing. It refers to the process of retrieving the incident angles of signals reaching an antenna array. Conventional estimation techniques [1] include beamforming methods such as Bartlett or Capon's [2] beamformers, subspace methods like the MUSIC [3] or ESPRIT [4] algorithms, as well as maximum likelihood approaches [5]. Because subspace methods exploit the statistical properties of the observations, accurate DOA estimation is only made possible at the price of a large number of snapshots and sufficiently uncorrelated sources. Maximum likelihood approaches are, for their part, very sensitive to initialization.

During the last decade, these limitations have been overcome with the advent of sparse optimization. Many innovative DOA estimation approaches have been proposed in this context. They come in many flavors: *on-grid*, *off-grid*, or *gridless*, according to the strategy adopted to deal with the non-linearity of the DOA model [6]. On-grid methods make the assumption that the incident angles belong to a prescribed grid. DOA estimation is then transformed into a challenging linear group-sparse optimization problem involving the $\ell_{2,0}$ pseudo norm that can be tackled through $\ell_{2,1}$ (or group-LASSO) relaxation [7], [8], [9], $\ell_{2,q}$ relaxation ($0 \leqslant q < 1$) [10], [11], smoothed $\ell_{2,0}$-norm approximation [12], [13], or greedy methods [14], [15]. Although still relying on a grid, off-grid methods do not constraint estimated DOAs to be on that grid [16], [17]. This mitigates the grid mismatch problem [18] at the price of the introduction of an auxiliary variable to the sparse optimization problem. Finally, gridless approaches work directly in the continuous domain [19], [20], [21], [22], thus avoiding the grid mismatch problem. However, they may be computationally intensive as they rely on the resolution of a semi-definite program. For more details on sparse-based methods for DOA estimation, we refer the reader to the comprehensive reviews [6][23].

Contributions: We show that the challenging group-sparse optimization problem that defines *on-grid* DOA estimation methods can be exactly relaxed by replacing the $\ell_{2,0}$ term by a group minimax concave penalty (group-MCP) [24]. More precisely, we prove that for a suitable choice of the group-MCP parameters the relaxation preserves the global minimizers of the $\ell_{2,0}$ penalized least-squares criteria while removing some of its local minimizers (Theorem 2). Moreover, we propose a new dimensionality reduction technique to decrease the computational burden of the estimation when the number of snapshots is larger than the number of antennas (Proposition 1). Finally, we deploy an iteratively reweighted $\ell_{2,1}$ algorithm to minimize the proposed relaxation and compare its performance against previously proposed *on-grid* methods.

Notations: We use the notation $\mathbb{I}_N = \{1, \dots, N\}$. For a matrix $\mathbf{S} \in \mathbb{C}^{M\times N}$ and a set of indices $\omega \subset \mathbb{I}_N$, $\mathbf{S}_{\omega\cdot} \in \mathbb{C}^N$ denotes the restriction of $\mathbf{S}$ to its rows indexed by ω while $\mathbf{S}_{\cdot\omega} \in \mathbb{C}^M$ stands for its column counterpart. The Frobenius norm is denoted $\|\cdot\|_{\mathrm{F}}$. The indicator function of the subset Ω is defined by $\mathbb{1}_{\{x\in\Omega\}} := \{1 \text{ if } x \in \Omega,\ 0 \text{ otherwise}\}$. $\mathbf{u}\otimes\mathbf{v} \in \mathbb{C}^{M\times N}$ stands for the tensor product between $\mathbf{u} \in \mathbb{C}^M$ and $\mathbf{v} \in \mathbb{C}^N$. Finally, $\bar{x}$ denotes the conjugate of $x \in \mathbb{C}$ and $\mathbf{A}^{\mathrm{H}}$ the conjugate transpose of $\mathbf{A} \in \mathbb{C}^{M\times N}$.

II. Group-Sparse Formulation of DOA Estimation

The general equation that describes an antenna array is

$$\mathbf{Y} = \mathbf{A}(\bar{\boldsymbol{\theta}})\mathbf{S} + \mathbf{N}, \tag{1}$$

where $\mathbf{S} \in \mathbb{C}^{K\times L}$ is a matrix formed out of the L samples of the K incident signals, $\mathbf{Y} \in \mathbb{C}^{M\times L}$ is the observation matrix containing the L snapshots of the M antennas outputs, and $\mathbf{N} \in \mathbb{C}^{M\times L}$ is an additive zero mean Gaussian noise with variance σ^2_{noise}. The non-linear operator $\mathbf{A} : [0, 2\pi)^K \to \mathbb{C}^{M\times K}$ is defined by

$$\mathbf{A}(\bar{\boldsymbol{\theta}}) = (\mathbf{a}(\bar{\theta}_1)\ \mathbf{a}(\bar{\theta}_2)\ \cdots\ \mathbf{a}(\bar{\theta}_K)), \tag{2}$$

The authors would like to thank warmly Cynthia Junqueira and Jean-Pierre Barbot for fruitful discussions. This work has been partly supported by Capes/Cofecub-Ph 960/20.

E. Soubies is with IRIT, Université de Toulouse, CNRS, Toulouse, France (emmanuel.soubies@irit.fr).

A. Chinatto is with School of Electrical and Computer Engineering, University of Campinas and Espectro Ltd., Campinas, Brazil.

J.M.T. Romano is with School of Electrical and Computer Engineering, University of Campinas, Campinas, Brazil.

P. Larzabal is with Université Paris-Saclay, Laboratoire SATIE, Gif-sur-Yvette, France.

L. Blanc-Féraud is with Université Côte dAzur, CNRS, INRIA, I3S, Sophia-Antipolis, France.

where $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1 \cdots \bar{\theta}_K)^T \in [0, 2\pi)^K$ is the vector of incident angles. The steering vectors $(\mathbf{a}(\bar{\theta}_k))_{k=1}^K$ depend on the geometry of the antenna array. Then, given $\mathbf{Y}$, DOA estimation amounts to retrieve the number of signals K and their incident angles $\bar{\boldsymbol{\theta}}$. This is a challenging non-linear inverse problem.

By considering a set of predefined possible DOA angles $\boldsymbol{\theta} = (\theta_1 \cdots \theta_N)^T \in [0, 2\pi)^N$ ($N \gg K$), we obtain a linearized version of model (1) as

$$\mathbf{Y} \approx \mathbf{A}\mathbf{Z} + \mathbf{N}, \tag{3}$$

where $\mathbf{Z} \in \mathbb{C}^{N\times L}$ is a row sparse matrix with $K \ll N$ nonzero rows. Here, the matrix $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}) = (\mathbf{a}(\theta_1)\ \mathbf{a}(\theta_2)\ \cdots\ \mathbf{a}(\theta_N)) \in \mathbb{C}^{M\times N}$ is formed out of the candidate steering vectors $(\mathbf{a}(\theta_n))_{n=1}^N$. It follows that the nonzero rows of $\mathbf{Z}$ (*i.e.,* its support) encode the incident angles $\bar{\boldsymbol{\theta}}$ up to the fineness of the grid $(\theta_n)_{n=1}^N$. Hence, with (3), DOA estimation is converted into a group-sparse estimation problem also referred to as multiple measurement vectors (MMV) sparse estimation problem.

A natural measure of the row-sparsity of a matrix $\mathbf{Z}$ is given by the mixed $\ell_{2,0}$ pseudo norm [10], [12]

$$\|\mathbf{Z}\|_{2,0} = \sum_{n\in\mathbb{I}_N} |\, \|\mathbf{Z}_{n\cdot}\|_2 \,|_0, \tag{4}$$

where $|z|_0 = \{0 \text{ if } z = 0\ ;\ 1 \text{ otherwise}\}$ and $\mathbf{Z}_{n\cdot}$ denotes the nth row of $\mathbf{Z}$. Then, DOA estimation can be addressed through the following (ℓ_2-$\ell_{2,0}$) optimization problem

$$\hat{\mathbf{Z}} \in \arg\min_{\mathbf{Z}\in\mathbb{C}^{N\times L}} J(\mathbf{Z}) := \frac{1}{2}\|\mathbf{A}\mathbf{Z} - \mathbf{Y}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{Z}\|_{2,0}, \tag{5}$$

where $\lambda > 0$ balances between data-fidelity and sparsity. This problem is nonconvex, noncontinuous, and NP hard due to its combinatorial nature. Yet, the single measurement vector (SMV) case (*i.e.,* $L = 1$) has been widely studied, driven by the compressed sensing paradigm. Naturally, many of these approaches have been extended to the MMV setting, such as those mentioned in the introduction. Such extensions are essential as the resolution of MMV problems leads to an improvement in the size of the recoverable support [25].

III. Dimensionality Reduction

The computational cost of the algorithms deployed to minimize J in (5) grows with the size of the problem (*i.e.,* $N \times L$). It is thus of practical interest to reduce this size. Inspired by the ℓ_1-SVD method [7], [26], we show in Proposition 1 that, when $M < L$, minimizing $J : \mathbb{C}^{N\times L} \to \mathbb{R}$ is equivalent to minimizing $F_0 : \mathbb{C}^{N\times M} \to \mathbb{R}$ defined by

$$F_0(\mathbf{X}) = \frac{1}{2}\|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_{2,0}. \tag{6}$$

where $\mathbf{V}$ comes from the singular value decomposition of $\mathbf{Y}$ ($\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}$) and $\mathbf{D} = [\mathbf{I}_M, \mathbf{0}_{M\times(L-M)}]$. This shows that the dimension of (5) can be reduced from $(N \times L)$ to $(N \times M)$.

Proposition 1. *Let $M < L$ and F_0 be defined by* (6)*. Then*

1) *For each local minimizer $\hat{\mathbf{X}} \in \mathbb{C}^{N\times M}$ of F_0, $\hat{\mathbf{Z}} = \hat{\mathbf{X}}\mathbf{D}\mathbf{V}^H \in \mathbb{C}^{N\times L}$ is a local minimizer of J and $J(\hat{\mathbf{Z}}) = F_0(\hat{\mathbf{X}})$.*
2) *There is a one-to-one mapping between strict local minimizers (including global minimizers) of J and F_0.*

Proof. Let $\hat{\mathbf{X}} \in \mathbb{C}^{N\times M}$ be a local minimizer of F_0 and denote by $\omega \subseteq \mathbb{I}_N$ its support. Then, from [27, Lemma 2.4][1] we have

$$(\mathbf{A}_{\cdot\omega}^{\mathrm{H}}\mathbf{A}_{\cdot\omega})\hat{\mathbf{X}}_{\omega\cdot} = \mathbf{A}_{\cdot\omega}^{\mathrm{H}}\mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}} \tag{7}$$

$$\implies (\mathbf{A}_{\cdot\omega}^{\mathrm{H}}\mathbf{A}_{\cdot\omega})\hat{\mathbf{X}}_{\omega\cdot}\mathbf{D}\mathbf{V}^H = \mathbf{A}_{\cdot\omega}^{\mathrm{H}}\mathbf{Y}, \tag{8}$$

$$\implies (\mathbf{A}_{\cdot\omega}^{\mathrm{H}}\mathbf{A}_{\cdot\omega})(\hat{\mathbf{X}}\mathbf{D}\mathbf{V}^H)_{\omega\cdot} = \mathbf{A}_{\cdot\omega}^{\mathrm{H}}\mathbf{Y}, \tag{9}$$

showing that $\hat{\mathbf{Z}} = \hat{\mathbf{X}}\mathbf{D}\mathbf{V}^H$ is a local minimizer of J. To obtain (8), we used the fact that $\mathbf{V}$ is unitary and that, by definition of $\mathbf{V}$ and $\mathbf{D}$, $\mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{Y}\mathbf{V}$. Then, one can see from (7)–(9) that $\hat{\mathbf{X}}$ and $\hat{\mathbf{Z}}$ have the same row-support and thus that $\|\hat{\mathbf{X}}\|_{2,0} = \|\hat{\mathbf{Z}}\|_{2,0}$. Finally, we obtain the equality $J(\hat{\mathbf{Z}}) = F_0(\hat{\mathbf{X}})$ by combining the previous arguments with the equality $\|\cdot\|_{\mathrm{F}}^2 = \|\cdot\,\mathbf{D}\mathbf{V}^{\mathrm{H}}\|_{\mathrm{F}}^2$.

The second assertion of the proposition comes from the fact that $\mathbf{A}_{\cdot\omega}$ is full rank [27, Theorem 3.2] for strict local minimizers. This implies that the systems in (7)–(9) have a unique solution. Finally, the fact that global minimizers of J and F_0 are strict [27, Theorem 4.4] completes the proof. □

From Proposition 1, we get that we can easily obtain a local minimizer of J from one of F_0 (first assertion). And more importantly, that any global minimizer of J can be reached from global minimizers of F_0 (second assertion). In this respect, the two problems are equivalent.

IV. An Exact Continuous Relaxation of F_0

We consider the following relaxation[2] of F_0 in (6)

$$\tilde{F}(\mathbf{X}) = \frac{1}{2}\|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \sum_{n\in\mathbb{I}_N} \phi(\gamma_n, \lambda; \|\mathbf{X}_{n\cdot}\|_2), \tag{10}$$

where $\gamma_n > 0$ for $n \in \mathbb{I}_N$, and $\phi(\gamma, \lambda; \cdot) : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is the minimax concave penalty (MCP) [28] defined, for $x > 0$, by

$$\phi(\gamma, \lambda; x) = \lambda - \frac{1}{2\gamma}\left(x - \sqrt{2\lambda\gamma}\right)^2 \mathbb{1}_{\{x\leq\sqrt{2\lambda\gamma}\}}. \tag{11}$$

It is a piecewise quadratic function (see Figure 1) that satisfies $\phi(\gamma, \lambda; x) \leq \lambda|x|_0$ with equality for $x \in \{0\} \cup [\sqrt{2\lambda\gamma}, +\infty)$. The complete penalty term in (10) is known as group-MCP [24]. The rationale behind this choice is that, in the SMV case, it has been shown in [29], [30] that minimizing F_0 in (6) is equivalent to minimizing $\tilde{F}$ in (10) for a suitable choice of the parameters γ_n. Not only $\tilde{F}$ admits the same global minimizers as F_0, but some local (not global) minimizers of F_0 are removed by $\tilde{F}$ [31]. We extend this result to the MMV setting in Theorem 2 (proof in *Supplementary Material*).

Theorem 2. *Let $\mathcal{L}_0$ (resp., $\tilde{\mathcal{L}}$) be the set local minimizers of F_0 (resp., $\tilde{F}$). Let $\mathcal{G}_0 \subseteq \mathcal{L}_0$ (resp. $\tilde{\mathcal{G}} \subseteq \tilde{L}$) be the corresponding subset of global minimizers. Then, if $\gamma_n < 1/\|\mathbf{A}_{\cdot n}\|_2^2$ for all $n \in \mathbb{I}_N$, we have,*

$$\tilde{\mathcal{L}} \subseteq \mathcal{L}_0 \ \text{ and } \ \tilde{\mathcal{G}} = \mathcal{G}_0. \tag{12}$$

[1] One can easily extend Lemma 2.4, Theorem 3.2, and Theorem 4.4 of [27] (used in the proof of Proposition 1) to the MMV setting.

[2] It is noteworthy to mention that, as both J and F_0 are ℓ_2-$\ell_{2,0}$ functionals, all the developments that we are doing for F_0 can be transposed to J when $L < M$ (*i.e.,* when the dimensionality reduction is not relevant).

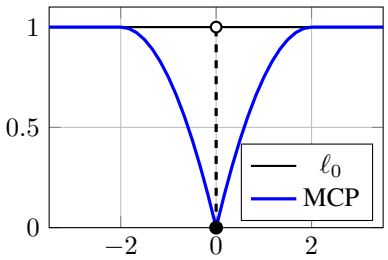

Fig. 1: Graph of ℓ_0 and MCP for $\lambda = 1$ and $\gamma = 2$

When (12) is satisfied, we say that the continuous relaxation is exact. From Theorem 2, the proposed continuous relaxation $\tilde{F}$ is exact as soon as $\gamma_n < 1/\|\mathbf{A}_{\cdot n}\|_2^2$. If the columns of the matrix $\mathbf{A}$ are normalized, this condition becomes $\gamma_n < 1$.

Remark 1. *The closer γ_n gets to the bound $1/\|\mathbf{A}_{\cdot n}\|_2^2$, the more $\tilde{F}$ is likely to eliminate local (not global) minimizers of F_0. Indeed, one gets from Lemma 5 (Supplementary Material) that $\mathbf{X} \in \tilde{\mathcal{L}}$ implies, $\forall n \in \mathbb{I}_N$, $\|\mathbf{X}_{n\cdot}\|_2 \in \{0\} \cup [\sqrt{2\lambda\gamma_n}, +\infty)$. Hence, if $\mathbf{X} \in \mathcal{L}_0$ is such that $\|\mathbf{X}_{n\cdot}\|_2 \in (0, \sqrt{2\lambda\gamma_n})$ for some $n \in \mathbb{I}_N$, then $\mathbf{X} \notin \tilde{\mathcal{L}}$. This shows that increasing γ_n can eliminate more local minimizers of F_0.*

Remark 2. *For the limit case $\gamma_n = 1/\|\mathbf{A}_{\cdot n}\|_2^2$, a similar result can be obtained, but the analysis is a bit more involved. Yet, such a result has been derived in [32] when $L = 1$, leading to the continuous exact ℓ_0 (CEL0) relaxation.*

V. Minimizing the Relaxation $\tilde{F}$

The continuity of $\tilde{F}$ allows us to deploy nonsmooth nonconvex optimization algorithms for its minimization that cannot be used directly with F_0.

A. Iteratively Reweighted $\ell_{2,1}$

We consider the iteratively reweighted $\ell_{2,1}$ algorithm (IRL1). It proceeds by minimizing a series of convex majorizations of the objective which are equal to it at the current point. To minimize $\tilde{F}$, we follow [33]. Because $\phi(\gamma, \lambda; \cdot)$ is concave on $\mathbb{R}_{\geqslant 0}$, it is majored by its tangents (or half-tangent at 0). At $\tilde{x} \in \mathbb{R}_{\geqslant 0}$, the (half) tangent of $\phi(\gamma, \lambda; \cdot)$ is

$$t(x) = w(\gamma, \lambda; \tilde{x})(x - \tilde{x}) + \phi(\gamma, \lambda; \tilde{x}), \tag{13}$$

where the expression of the slope is

$$w(\gamma, \lambda; \tilde{x}) = \begin{cases} \sqrt{2\lambda/\gamma} - \tilde{x}/\gamma & \text{if } \tilde{x} < \sqrt{2\lambda\gamma}, \\ 0 & \text{if } \tilde{x} \geqslant \sqrt{2\lambda\gamma}. \end{cases} \tag{14}$$

Given $\tilde{\mathbf{X}} \in \mathbb{C}^{N \times M}$, we can thus define a majorant of the penalty term in (10) as

$$Q(\mathbf{Z}) = \sum_{n \in \mathbb{I}_N} w(\gamma_n, \lambda; \|\tilde{\mathbf{X}}_{n\cdot}\|_2) \|\mathbf{Z}_{n\cdot}\|_2. \tag{15}$$

Note that Q in (15) is defined up to a constant (*i.e.,* ignoring the terms that are constant with respect to x in (13)). Then, the IRL1 algorithm [33] generates a sequence $(\mathbf{X}^k)_{k \in \mathbb{N}}$ as

$$\mathbf{X}^{k+1} \in \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \sum_{n \in \mathbb{I}_N} w_n^k \|\mathbf{X}_{n\cdot}\|_2, \tag{16}$$

Algorithm 1

Require: $\mathbf{X}^0 \in \mathbb{C}^{N \times M}$
1: $\mathbf{X}^1 \leftarrow \mathrm{IRL1}(\tilde{F}; \mathbf{X}^0)$
2: $k = 1$
3: **while** $\mathbf{X}^k \notin \mathcal{L}_0$ **do**
4: Select $n \in \mathbb{I}_N$ such that $\|\mathbf{X}_{n\cdot}^k\|_2 \in (0, \sqrt{2\lambda\gamma_n})$
5: Find $\alpha \in \{0, \sqrt{2\lambda\gamma_n}\}$ minimizing $\tilde{F}(\mathbf{X}_{\setminus n}^k + \alpha \mathbf{e}_n \otimes \frac{\mathbf{X}_{n\cdot}^k}{\|\mathbf{X}_{n\cdot}^k\|_2})$
6: $\mathbf{X}^{k+1} \leftarrow \mathrm{IRL1}(\tilde{F}; \mathbf{X}_{\setminus n}^k + \alpha \mathbf{e}_n \otimes \frac{\mathbf{X}_{n\cdot}^k}{\|\hat{\mathbf{X}}_{n\cdot}\|_2})$
7: $k = k + 1$
8: **end while**

where $w_n^k = w(\gamma_n, \lambda; \|\mathbf{X}_{n\cdot}^k\|_2)$. Each sub-problem (16) is a weighted $\ell_{2,1}$-norm minimization problem which can be solved using FISTA [34]. The convergence of the sequence generated by IRL1 to a critical point of the objective is proven in [33] when the objective verifies the Kurdyka-Lojasiewicz (KL) inequality. It is the case for $\tilde{F}$ as $\mathbf{X} \mapsto \frac{1}{2}\|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}}\|_{\mathrm{F}}^2$ is a polynomial function and $\phi(\gamma, \lambda; \cdot)$ has a piecewise polynomial graph, which are sufficient ingredients to conclude [35].

B. Ensuring the Convergence to Local Minimizers of F_0

The IRL1 algorithm only ensures the convergence to a critical point of $\tilde{F}$ while Theorem 2 provides a relation between (local) minimizers of $\tilde{F}$ and F_0. It is thus of interest to complete the result of Theorem 2 with an analysis of the critical points of $\tilde{F}$.

Lemma 3. *Let $\gamma_n < 1/\|\mathbf{A}_{\cdot n}\|_2^2$ for all $n \in \mathbb{I}_N$ and $\hat{\mathbf{X}} \in \mathbb{C}^{N \times M}$ be a critical point of $\tilde{F}$.*

1) *If, $\forall n \in \mathbb{I}_N$, $\|\hat{\mathbf{X}}_{n\cdot}\|_2 \in \{0\} \cup \left[\sqrt{2\lambda\gamma_n}, +\infty\right)$, then $\hat{\mathbf{X}}$ is a local minimizer of F_0 (i.e., $\hat{\mathbf{X}} \in \mathcal{L}_0$).*
2) *Otherwise, $\forall n \in \mathbb{I}_N$ such that $\|\hat{\mathbf{X}}_{n\cdot}\|_2 \in (0, \sqrt{2\lambda\gamma_n})$, there exists $\alpha \in \{0, \sqrt{2\lambda\gamma_n}\}$ such that*

$$\tilde{F}\left(\hat{\mathbf{X}}_{\setminus n} + \alpha \mathbf{e}_n \otimes \hat{\mathbf{X}}_{n\cdot}/\|\hat{\mathbf{X}}_{n\cdot}\|_2\right) < \tilde{F}(\hat{\mathbf{X}}), \tag{17}$$

where $\hat{\mathbf{X}}_{\setminus n} = \hat{\mathbf{X}} - \mathbf{e}_n \otimes \hat{\mathbf{X}}_{n\cdot}$.

From the first statement of Lemma 3, one can easily check whether a critical point of the relaxation $\tilde{F}$ is a local minimizer of the initial functional F_0. Moreover, if this is not the case, one can easily obtain a new point that decreases $\tilde{F}$ (second statement of Lemma 3). This suggests to deploy the strategy described in Algorithm 1 where $\mathrm{IRL1}(\tilde{F}; \mathbf{X})$ stands for the minimization of $\tilde{F}$ using IRL1 initialized by $\mathbf{X}$. From Lemma 3, the convergence of this scheme can be obtained in the same way as [32, Theorem 5.1]. The main difference being that the 1D restriction at line 5 is linear (with nonzero slope) on $[0, \sqrt{2\lambda\gamma_n}]$ whereas its counterpart in [32] is constant (making $\alpha = 0$ always a valid choice for non-increasing $\tilde{F}$).

Remark 3. *To fully exploit the result provided by Theorem 2 an algorithm that ensures the convergence to a local minimizer of $\tilde{F}$ has to be defined. In the absence of such an algorithm, Algorithm 1 is an interesting alternative. It ensures to reach a critical point of $\tilde{F}$ which is also a local minimizer of F_0.*

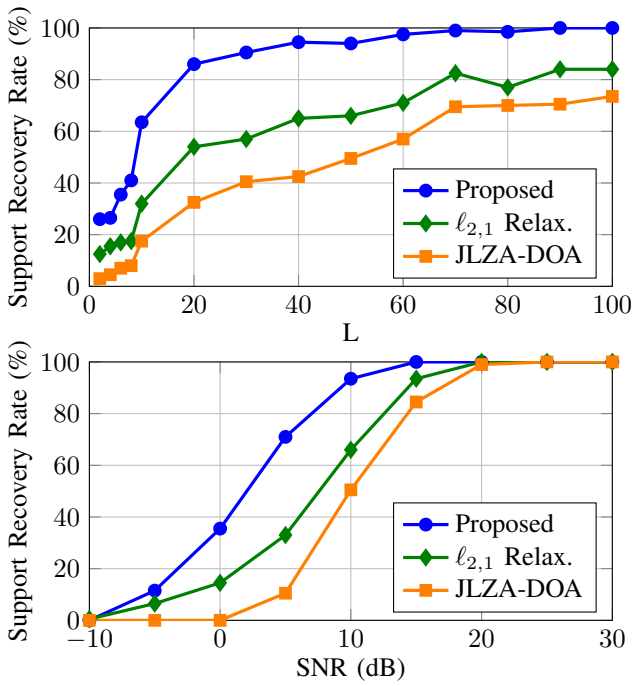

Fig. 2: Support recovery rate as a function of the number of snapshots L for SNR = 10 dB (top), and as a function of SNR for $L = 50$ (bottom).

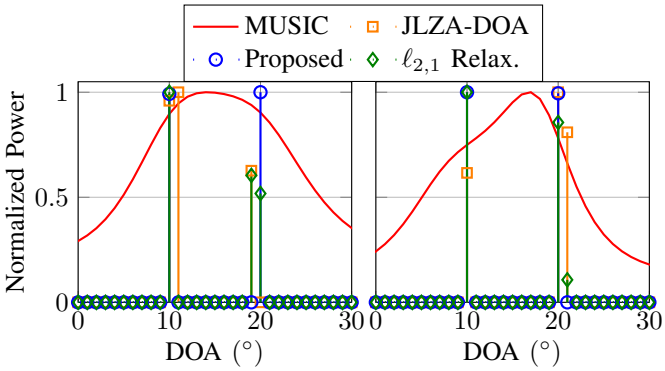

Fig. 3: Normalized power spectra for two realizations of noise with $L = 40$ and SNR=10 dB (zoom between 0° and 30°). The true DOAs are $\bar{\theta}_1 = 10°$ and $\bar{\theta}_2 = 20°$.

VI. Numerical Experiment

A. Description of the Experiment

We consider an uniformly linear array (ULA) geometry composed of $M = 8$ omnidirectional elements spaced by half the electromagnetic wavelength. Given an incident angle θ, the corresponding steering vector $\mathbf{a}(\theta)$ is

$$\mathbf{a}(\theta) = (1 \; e^{j\pi \sin\theta} \; e^{j2\pi \sin\theta} \; \cdots \; e^{j(M-1)\pi \sin\theta})^T. \qquad (18)$$

We simulate $K = 2$ correlated narrowband signals with planar wave fronts and incident angles $\bar{\theta}_1 = 10°$ and $\bar{\theta}_2 = 20°$. The correlation coefficient is fixed to 0.99. The measurements are corrupted with Gaussian noise so that to reach a specified signal-to-noise ratio (SNR). Finally, we define the group-sparse estimation problem (Section II) by slicing the possible range of incident angles from $\theta_{\min} = -90°$ to $\theta_{\max} = +89°$ in steps of 1° (*i.e.,* $N = 180$).

To assess the performance of the proposed method (*i.e.,* minimization of the exact continuous relaxation $\tilde{F}$ using Algorithm 1 with $\mathbf{X}^0 = \mathbf{0}$), we compute the exact support recovery rate for the two following scenarios

- number of snapshots varying from $L = 100$ to $L = 2$ with a SNR fixed to 10 dB,
- noise levels varying from SNR = 30 dB to SNR = -10 dB with the number of snapshots fixed to $L = 50$.

For each couple (L, SNR) we perform 200 independent realizations of noise in order to determine the support recovery rate. We consider that the estimation is successful when the estimated $\hat{\mathbf{X}}$ has only two non-zero rows that correspond to the two incident angles $\bar{\theta}_1 = 10°$ and $\bar{\theta}_2 = 20°$.

Following Remark 1, we set $\gamma_n = 0.99/\|\mathbf{A}_{\cdot n}\|^2$ in (10). Then, the selection of the regularization parameter λ is made so that to maximize the recovery rate while keeping the same value for all the 200 realizations.

For comparison, we consider the minimization of the $\ell_{2,1}$ convex relaxation of F_0 using FISTA [34], as well as the JLZA-DOA[3] algorithm [12]. The latter is designed to minimize F_0 using a graduated non-convexity approach based on a smoothed $\ell_{2,0}$-norm approximation. All these methods benefit from the dimensionality reduction presented in Section III and we adopt the same strategy to select the parameter λ.

B. Discussion

From Figure 2, we see that the minimization of the proposed exact relaxation $\tilde{F}$ outperforms both JLZA-DOA and the $\ell_{2,1}$ convex relaxation in terms of support recovery. Moreover, we found that a direct minimization of F_0 using a proximal gradient algorithm [35] is unable to consistently recover the support over the different realizations of noise for the same value of λ. Hence, we do not report the corresponding curves.

On Figure 3, we depict the normalized power spectra obtained by the three methods for two realizations of noise with $L = 40$ and SNR = 10 dB. One can see that both JLZA-DOA and the minimization of the $\ell_{2,1}$ convex relaxation lead to the detection of spurious DOAs close to the true ones. Although, on the right plot, a post-processing step that extracts local maxima would allow to recover the two correct DOAs, the same computation on the left plot would result in erroneous DOAs. In contrast, the proposed approach provides a two-sparse solution that recovers the true DOAs.

For completeness, we provide the power spectra obtained by the MUSIC algorithm which is unable to resolve the two sources with only $L = 40$ snapshots. This outlines the difficulty of the considered scenario that combines highly correlated sources, few antennas, and close sources that fall within the 3 dB main beamforming lobe.

[3]We tuned the parameters of JLZA-DOA and found that, for our experiment, the best ones were $\rho = 0.78$, $\eta = 0.1$, $\sigma_0 = 0.005$, and $\gamma = 0.5$.

REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[3] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[4] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques-ESPRIT," in *Asilomar Conference on Circuits, Systems and Computers*, 1985, pp. 83–89.

[5] J. Bohme, "Estimation of source parameters by maximum likelihood and nonlinear regression," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 9, 1984, pp. 271–274.

[6] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*. Elsevier, 2018, vol. 7, pp. 509–581.

[7] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE transactions on signal processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[8] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[9] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 505–519, 2009.

[10] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[11] Y. Hu, C. Li, K. Meng, J. Qin, and X. Yang, "Group sparse optimization via $\ell_{p,q}$ regularization," *Journal of Machine Learning Research*, vol. 18, no. 30, pp. 1–52, 2017.

[12] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, 2010.

[13] J. Liu, W. Zhou, and F. H. Juwono, "Joint smoothed ℓ_0-norm DOA estimation algorithm for multiple measurement vectors in MIMO radar," *sensors*, vol. 17, no. 5, p. 1068, 2017.

[14] S. F. Cotter, "Multiple snapshot matching pursuit for direction of arrival (DOA) estimation," in *European Signal Processing Conference*, 2007, pp. 247–251.

[15] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[16] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, 2011.

[17] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658–4671, 2012.

[18] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[19] P. Pal and P. P. Vaidyanathan, "A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 737–741, 2014.

[20] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4959–4973, 2014.

[21] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2015.

[22] K. Mahata and M. M. Hyder, "Grid-less TV minimization for DOA estimation," *Signal Processing*, vol. 132, pp. 155–164, 2017.

[23] Y. Chi and M. F. Da Costa, "Harnessing sparsity over the continuum: Atomic norm minimization for superresolution," *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 39–57, 2020.

[24] J. Huang, P. Breheny, and S. Ma, "A selective review of group selection in high-dimensional models," *Statistical science: a review journal of the Institute of Mathematical Statistics*, vol. 27, no. 4, 2012.

[25] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[26] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 995–1006, 2015.

[27] M. Nikolova, "Description of the minimizers of least squares regularized with ℓ_0-norm. Uniqueness of the global minimizer," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 904–937, 2013.

[28] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, vol. 38, no. 2, pp. 894–942, 2010.

[29] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A unified view of exact continuous penalties for $\ell_2 - \ell_0$ minimization," *SIAM Journal on Optimization*, vol. 27, no. 3, pp. 2034–2060, 2017.

[30] M. Carlsson, "On convex envelopes and regularization of non-convex functionals without moving global minima," *Journal of Optimization Theory and Applications*, vol. 183, no. 1, pp. 66–84, 2019.

[31] E. Soubies, L. Blanc-Féraud, and G. Aubert, "New insights on the optimality conditions of the $\ell_2 - \ell_0$ minimization problem," *Journal of Mathematical Imaging and Vision*, pp. 1–17, 2019.

[32] ——, "A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1607–1639, 2015.

[33] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 331–372, 2015.

[34] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[35] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods," *Mathematical Programming*, vol. 137, no. 1-2, pp. 91–129, 2013.

Direction-of-Arrival Estimation through Exact Continuous $\ell_{2,0}$-Norm Relaxation (Supplementary Material)

Emmanuel Soubies, Adilson Chinatto, Pascal Larzabal, João M. T. Romano, and Laure Blanc-Féraud

I. Preliminary Lemmas

Lemma 4. *Let $\gamma_n < 1/\|\mathbf{A}_{\cdot n}\|_2^2$ for all $n \in \mathbb{I}_N$. Then, for any $\mathbf{X} \in \mathbb{C}^{N\times L}$, $n \in \mathbb{I}_N$, and $\mathbf{t} \in \mathbb{C}^L$ such that $\|\mathbf{t}\|_2 = 1$, the one-dimensional restriction of $\tilde{F}$ defined by, $\forall \alpha \in \mathbb{R}_+$,*

$$f_{n,\mathbf{t}}(\alpha) = \tilde{F}(\mathbf{X}_{\setminus n} + \alpha \mathbf{e}_n \otimes \mathbf{t}), \tag{19}$$

where $\mathbf{X}_{\setminus n} = \mathbf{X} - \mathbf{e}_n \otimes \mathbf{X}_{n\cdot}$, is strictly concave on $(0, \sqrt{2\lambda\gamma_n})$.

Proof. One can see from (10) and (11) that $f_{n,\mathbf{t}}$ is twice differentiable on $(0, \sqrt{2\lambda\gamma_n})$. Then, we have, $\forall \alpha \in (0, \sqrt{2\lambda\gamma_n})$,

$$f'_{n,\mathbf{t}}(\alpha) = \alpha\|\mathbf{A}_{\cdot n}\|_2^2\|\mathbf{t}\|_2^2 - \frac{\|\mathbf{t}\|_2}{\gamma_n}\left(\alpha\|\mathbf{t}\|_2 - \sqrt{2\lambda\gamma_n}\right) + C, \tag{20}$$

where C is a constant that does not depend on α. Differentiating a second time and using the fact that $\|\mathbf{t}\|_2 = 1$, we obtain

$$f''_{n,\mathbf{t}}(\alpha) = \|\mathbf{A}_{\cdot n}\|_2^2 - \frac{1}{\gamma_n} < 0 \tag{21}$$

where the last inequality comes by assumption of the Lemma. This completes the proof. □

Lemma 5. *Let $\gamma_n < 1/\|\mathbf{A}_{\cdot n}\|_2^2$ for all $n \in \mathbb{I}_N$ and $\hat{\mathbf{X}} \in \tilde{\mathcal{L}}$ (i.e., $\hat{\mathbf{X}}$ is a local minimizer of $\tilde{F}$). Then,*

$$\forall n \in \mathbb{I}_N,\ \|\hat{\mathbf{X}}_{n\cdot}\|_2 \in \{0\} \cup \left[\sqrt{2\lambda\gamma_n}, +\infty\right) \tag{22}$$

Proof. This result is a direct consequence of Lemma 4. □

II. Proof of Theorem 2

Let us first recall that, by definition of $\tilde{F}$ in (10)–(11), we have,

$$\forall \mathbf{X} \in \mathbb{C}^{N\times M},\ \tilde{F}(\mathbf{X}) \leq F_0(\mathbf{X}) \tag{23}$$

with equality if, $\forall n \in \mathbb{I}_N$, $\|\mathbf{X}_{n\cdot}\|_2 \in \{0\} \cup \left[\sqrt{2\lambda\gamma_n}, +\infty\right)$. Hence from Lemma 5 we have that

$$\forall \mathbf{X} \in \tilde{\mathcal{L}},\ \tilde{F}(\mathbf{X}) = F_0(\mathbf{X}). \tag{24}$$

1) $\underline{\tilde{\mathcal{L}} \subseteq \mathcal{L}_0}$ Let $\hat{\mathbf{X}} \in \tilde{\mathcal{L}}$ and assume that $\hat{\mathbf{X}} \notin \mathcal{L}_0$. Then, for any neighborhood $\mathcal{V} \subset \mathbb{C}^{N\times M}$ containing $\hat{\mathbf{X}}$, there exists $\bar{\mathbf{X}} \in \mathcal{V}$ such that

$$\tilde{F}(\bar{\mathbf{X}}) \underset{(23)}{\leq} F_0(\bar{\mathbf{X}}) < F_0(\hat{\mathbf{X}}) \underset{(24)}{=} \tilde{F}(\hat{\mathbf{X}}), \tag{25}$$

which contradicts $\hat{\mathbf{X}} \in \tilde{\mathcal{L}}$ and completes the proof of this statement.

2) $\underline{\tilde{\mathcal{G}} = \mathcal{G}_0}$ The inclusion $\tilde{\mathcal{G}} \subseteq \mathcal{G}_0$ is obtained using the same arguments as in 1), replacing $\tilde{\mathcal{L}}$, $\mathcal{L}_0$, and $\mathcal{V}$ by $\tilde{\mathcal{G}}$, $\mathcal{G}_0$, and $\mathbb{C}^{N\times M}$, respectively. For the reciprocal inclusion $\tilde{\mathcal{G}} \supseteq \mathcal{G}_0$, let $\hat{\mathbf{X}} \in \mathcal{G}_0$ and assume that $\hat{\mathbf{X}} \notin \tilde{\mathcal{G}}$. Then, there exist $\bar{\mathbf{X}} \in \mathbb{C}^{N\times M}$, such that

$$F_0(\bar{\mathbf{X}}) = \tilde{F}(\bar{\mathbf{X}}) < \tilde{F}(\hat{\mathbf{X}}) \underset{(23)}{\leq} F_0(\hat{\mathbf{X}}). \tag{26}$$

The first equality in (26) comes from the fact that $\bar{\mathbf{X}}$ is chosen so that it does not belong to the subsets where $\tilde{F}$ is strictly concave along a given direction (*i.e.,* $\|\bar{\mathbf{X}}_{n\cdot}\|_2 \in \{0\} \cup \left[\sqrt{2\lambda\gamma_n}, +\infty\right)\ \forall n \in \mathbb{I}_N$ from Lemma 4). This is always possible because $\tilde{F}$ is continuous and these subsets where $\tilde{F}$ is strictly concave are bounded. Finally, (26) contradicts $\hat{\mathbf{X}} \in \mathcal{G}_0$ and completes the proof.

III. Proof of Lemma 3

Let $\hat{\mathbf{X}} \in \mathbb{C}^{N\times M}$ be a critical point of $\tilde{F}$ and denote by $\omega \subseteq \mathbb{I}_N$ its support (*i.e.,* indices of its non-zero rows). For $n \in \mathbb{I}_N$, denote $\mathbf{z}_n = (\mathbf{A}\hat{\mathbf{X}} - \mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}})^T\bar{\mathbf{A}}_{\cdot n}$. Then, the first order optimality conditions lead to, $\forall n \in \mathbb{I}_N$,

$$\begin{cases} \|\mathbf{z}_n\|_2 \leq \sqrt{\frac{2\lambda}{\gamma_n}} & \text{if } \|\hat{\mathbf{X}}_{n\cdot}\|_2 = 0 \\ \mathbf{z}_n - \frac{\hat{\mathbf{X}}_{n\cdot}}{\gamma_n} + \sqrt{\frac{2\lambda}{\gamma_n}}\frac{\hat{\mathbf{X}}_{n\cdot}}{\|\hat{\mathbf{X}}_{n\cdot}\|_2} = \mathbf{0} & \text{if } \|\hat{\mathbf{X}}_{n\cdot}\|_2 \in (0, \sqrt{2\lambda\gamma_n}) \\ \mathbf{z}_n = \mathbf{0} & \text{if } \|\hat{\mathbf{X}}_{n\cdot}\|_2 \geq \sqrt{2\lambda\gamma_n} \end{cases} \tag{27}$$

Under the assumption of Lemma 3 (*i.e.,* $\forall n \in \mathbb{I}_N,\ \|\hat{\mathbf{X}}_{n\cdot}\|_2 \in \{0\} \cup \left[\sqrt{2\lambda\gamma_n}, +\infty\right)$), we have that $\forall n \in \omega$ $\|\hat{\mathbf{X}}_{n\cdot}\|_2 \geq \sqrt{2\lambda\gamma_n}$. Combining that with (27) we obtain

$$(\mathbf{A}\hat{\mathbf{X}} - \mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}})^T\bar{\mathbf{A}}_{\cdot n} = \mathbf{0}, \quad \forall n \in \omega \tag{28}$$

$$\iff (\mathbf{A}_{\cdot\omega}^H\mathbf{A}_{\cdot\omega})\hat{\mathbf{X}}_{\omega\cdot} = \mathbf{A}_{\cdot\omega}^H\mathbf{Y}\mathbf{V}\mathbf{D}^{\mathrm{T}} \tag{29}$$

which shows that $\hat{\mathbf{X}}$ is a local minimizer of F_0 and completes the proof of the first statement of Lemma 3.

To prove the second statement of Lemma 3, let $n \in \mathbb{I}_N$ be such that $\|\hat{\mathbf{X}}_{n\cdot}\|_2 \in (0, \sqrt{2\lambda\gamma_n})$ and define $\mathbf{t} = \hat{\mathbf{X}}_{n\cdot}/\|\hat{\mathbf{X}}_{n\cdot}\|_2$. Then, from Lemma 4, the one-dimensional restriction $f_{n,\mathbf{t}}$ defined in (19) is strictly concave on $(0, \sqrt{2\lambda\gamma_n})$ and verifies $f_{n,\mathbf{t}}(\|\hat{\mathbf{X}}_{n\cdot}\|_2) = \tilde{F}(\hat{\mathbf{X}})$. It follows from this strict concavity that there exists $\alpha \in \{0, \sqrt{2\lambda\gamma_n}\}$ such that

$$\tilde{F}(\hat{\mathbf{X}}_{\setminus n} + \alpha\mathbf{e}_n \otimes \mathbf{t}) \underset{(19)}{=} f_{n,\mathbf{t}}(\alpha) < f_{n,\mathbf{t}}(\|\hat{\mathbf{X}}_{n\cdot}\|_2) = \tilde{F}(\hat{\mathbf{X}}), \tag{30}$$

which completes the proof.