Low Cost Quasi Gauss-Newton Algorithm Implementation for 2D ML-DoA Estimation using a Sparse Representation of Array Covariance

Thomas Aussaguès*†, Anne Ferréol*†, Alice Delmer* and Pascal Larzabal†
*Thales, 4 avenue des Louvresses, 92230 Gennevilliers, France
†SATIE, Université Paris-Saclay, UMR CNRS 8029, 4 avenue des Sciences, 91190 Gif-sur-Yvette, France

***Abstract*—Maximum Likelihood (ML) Direction-of-Arrival (DoA) estimation under the Vectorized Covariance Matrix Model (VCMM) provides improved performance. However, the associated optimization problem remains computationally intractable due to its highly non-convex and multi-dimensional nature.**

To alleviate this issue, a sparse estimation strategy has been proposed, shown to be equivalent to the ML after a pre-whitening noise transformation when the regularization parameter is properly chosen. Yet, the resulting cost function remains challenging, with optimization limited to first-order methods such as the Proximal Gradient Algorithm (PGA), which suffers from slow convergence due to strong correlations in the dictionary matrix.

This work exploits the decorrelation induced by the pre-whitening transform to enable acceleration via a variable stepsize strategy. The transform increases the allowable stepsize by reducing the correlation of the dictionary near source directions, thus significantly improving convergence speed. Furthermore, the final iterations are shown to be equivalent with a more computationally demanding second-order algorithm, yielding an efficient approximation with reduced complexity.

Numerical simulations confirm the predicted speed improvements in the case of 2D DoA estimation.

***Index Terms*—Noise whitening, Orthogonal dictionary, Sparse optimization, Second-order**

I. Introduction

DoA estimation arises in numerous applications, including radar, sonar, and wireless communications. Classical methods such as MUSIC [1] degrade under limited snapshots or low Signal-to-Noise Ratio (SNR). Although the ML estimator [2] is statistically efficient as it achieves the Cramér-Rao Lower Bound at high SNR, its practical use is hindered by the need to solve a highly non-convex, multi-dimensional optimization problem. Several of the aforementioned limitations can be mitigated using the VCMM [3]–[6], which exploits a Virtual Array (VA) [7] of N^2 virtual antennas, thereby improving performance in adverse conditions.

DoA estimation from the VCMM can be formulated as a sparse recovery problem [8], whose effectiveness has been demonstrated in simulation studies [4]–[6], [9]. A pre-whitening transform has been introduced [4]–[6] to handle the correlated estimation error that corrupts the VCMM. Under this transformation, equivalence between ℓ_0-based sparse estimators and the ML has been established in the presence of white Gaussian estimation error [10], [11], enabling efficient ML implementation. The estimator's ℓ_0 objective function is typically optimized using the PGA [12]. Yet, being a first-order method, the PGA algorithm experiences slow convergence due to its stepsize, constrained by the high correlation between the vectors of the dictionary matrix.

Numerous research lines have been explored to accelerate the PGA algorithm. Nesterov introduced an inertial technique [13] using an extrapolation step. Gradient preconditioning [14] methods achieve acceleration at the cost of numerical evaluation of the proximity operator. Deep unfolding techniques [15] have also yielded acceleration. Yet, their interpretability is limited and their applicability is restricted by the requirement of large datasets and supervised learning [9]. Finally, pseudo-second-order [16] and backtracking techniques have also been proposed.

This paper leverages the pre-whitening transform effects [11], [17] to substantially accelerate the convergence of the sparse estimator. The spatial decorrelation between the dictionary matrix vectors after the transform is exploited to select larger stepsizes at each iteration using Ablin's accelerated PGA [18]. The acceleration is analytically justified as a direct consequence of the improved conditioning after pre-whitening transform. Furthermore, the last iterations are proven to be equivalent with a second-order descent algorithm.

This work has been partly supported by Capes/Cofecub-Ph 960/20.

II. Signal modelling

A. The Vectorized Covariance Matrix Model (VCMM)

Let us assume M impinging narrowband planes waves on an N-element array with directions $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M\}$ where $\boldsymbol{\theta}_m = \{\theta_m, \Delta_m\}$ with θ_m and Δ_m, respectively the azimuth and elevation angles of the m-th sources. Then, the array output is:

$$\mathbf{x}(t) = \sum_{m=1}^{M} \mathbf{a}(\boldsymbol{\theta}_m)\mathbf{s}_m(t) + \mathbf{n}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}_{\boldsymbol{\theta}}(t) + \mathbf{n}(t) \quad (1)$$

where $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_1), \dots, \mathbf{a}(\boldsymbol{\theta}_M)]$ is the array steering matrix formed by steering vectors $\mathbf{a}(\boldsymbol{\theta}_m) \in \mathbb{C}^N, 1 \leq m \leq M$. The source signals $s_m(t)$ are stacked into the vector $\mathbf{s}_{\boldsymbol{\theta}}(t) = [s_1(t), \dots, s_M(t)] \in \mathbb{C}^M$. Finally, $\mathbf{n}(t) \in \mathbb{C}^N$ refers to a complex circular white Gaussian noise, independent of the sources signals *ie.* $\mathbb{E}\left[\mathbf{s}_{\boldsymbol{\theta}}^H(t)\mathbf{n}(t)\right] = 0$, with covariance matrix $\mathbb{E}\left[\mathbf{n}(t)\mathbf{n}^H(t)\right] = \sigma^2\mathbf{I}_N$ where $\mathbb{E}[\cdot]$ denotes the temporal

mean and $\mathbf{I}_N$ the identity matrix of size N. Throughout this paper, the m-th source direction $\boldsymbol{\theta}_m = \{\theta_m, \Delta_m\}$ is rewritten as $\boldsymbol{\theta}_m = \{u_m, v_m\}$ where $u_m = \cos(\theta_m)\cos(\Delta_m)$ and $v_m = \sin(\theta_m)\cos(\Delta_m)$.

The array covariance matrix is then given by:

$$\mathbf{R}_x = \mathbb{E}\left[\mathbf{x}(t)\mathbf{x}^H(t)\right] = \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_N \quad (2)$$

where $\mathbf{R}_s = \mathbb{E}\left[\mathbf{s}_{\boldsymbol{\theta}}(t)\mathbf{s}_{\boldsymbol{\theta}}^H(t)\right]$ stands for the sources covariance matrix. Assuming temporally decorrelated sources ($\forall i \neq j, \mathbb{E}\left[s_i^*(t)s_j(t)\right] = 0$), the VCMM observation $\mathbf{r}$ [3]–[6] is obtained from (2) by applying the column-wise vectorization operator $\mathrm{vec}(\cdot)$:

$$\mathbf{r} = \mathrm{vec}\left(\mathbf{R}_x - \sigma^2\mathbf{I}_N\right) = \sum_{m=1}^{M}\mathbf{b}(\boldsymbol{\theta}_m)\gamma_m = \mathbf{B}(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} \quad (3)$$

where $\mathbf{B}(\boldsymbol{\theta})$ is the VA [7] steering matrix formed by the vectors $\mathbf{b}(\boldsymbol{\theta}_m) = \mathbf{a}^*(\boldsymbol{\theta}_m) \otimes \mathbf{a}(\boldsymbol{\theta}_m)$ where $\otimes$ is the Kronecker product and $\boldsymbol{\gamma}_{\boldsymbol{\theta}} = \mathrm{diag}(\mathbf{R}_s) = [\gamma_1, \ldots, \gamma_M]^T$ the sources powers vector.

As the true covariance matrix $\mathbf{R}_x$ is not accessible, $\mathbf{R}_x$ is replaced by its sample estimate $\widehat{\mathbf{R}}_x$ obtained using K *identically* and *independently distributed* array snapshots $\mathbf{x}(t_k), 1 \leq k \leq K$. Under temporally uncorrelated noise (*ie.* $\forall i \neq j, \mathbb{E}\left[\mathbf{n}^H(t_i)\mathbf{n}(t_j)\right] = 0$), $\widehat{\mathbf{R}}_x$ can be written as follows:

$$\widehat{\mathbf{R}}_x = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}(t_k)\mathbf{x}^H(t_k) = \mathbf{R}_x + \boldsymbol{\Delta}\mathbf{R}_x \quad (4)$$

with $\boldsymbol{\Delta}\mathbf{R}_x$ a complex Wishart estimation error matrix induced by the finite number of samples. As a consequence of (4), the VCMM (3) is corrupted by an estimation error $\boldsymbol{\delta} = \mathrm{vec}\left(\Delta\mathbf{R}_x\right)$:

$$\mathbf{r} = \mathbf{B}(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \boldsymbol{\delta} \quad (5)$$

According to the Central Limit Theorem, the complex Wishart distribution of $\boldsymbol{\delta}$ (5) converges towards a complex Gaussian law $\mathbb{CN}\left(\mathbf{0}_{N^2\times 1}, \boldsymbol{\Gamma}, \mathbf{C}\right)$ for K sufficiently large with the following moments [19]:

$$\boldsymbol{\Gamma} = \mathbb{E}\left[\boldsymbol{\delta}\boldsymbol{\delta}^H\right] = \tfrac{1}{K}\left(\mathbf{R}_x^T \otimes \mathbf{R}_x\right) \quad \mathbf{C} = \mathbb{E}\left[\boldsymbol{\delta}\boldsymbol{\delta}^T\right] = \mathbf{K}\boldsymbol{\Gamma} \quad (6)$$

where $\mathbf{K}$ denotes the permutation matrix.

B. The pre-whitening transform of the VCMM

As outlined by eq. (6), the VCMM (5) is corrupted by a non-white Gaussian estimation error $\boldsymbol{\delta}$. Following the methodology of [10], [11], [17], a pre-whitening transform is performed on the VCMM (5) converting the initially non-white estimation error $\boldsymbol{\delta}$ into a white Gaussian estimation error. Further details on this transform can be found in [11]. The transformed observation is:

$$\mathbf{y} = \widehat{\mathbf{W}}\mathbf{r} = \widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\theta})\boldsymbol{\gamma}_{\boldsymbol{\theta}} + \widehat{\mathbf{W}}\boldsymbol{\delta} = \mathbf{B}_w(\boldsymbol{\theta})\boldsymbol{\gamma} + \boldsymbol{\delta}_w \quad (7)$$

where $\widehat{\mathbf{W}} = \widehat{\boldsymbol{\Gamma}}^{-1/2} = \sqrt{K}\left(\widehat{\mathbf{R}}_x^{-T/2} \otimes \widehat{\mathbf{R}}_x^{-1/2}\right)$ is the whitening matrix, $\mathbf{B}_w(\boldsymbol{\theta}) = \widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\theta})$ denotes the transformed dictionary and $\boldsymbol{\delta}_w = \widehat{\mathbf{W}}\boldsymbol{\delta}$ the whitened estimation error such that $\mathbb{E}\left[\boldsymbol{\delta}_w\boldsymbol{\delta}_w^H\right] = \mathbf{I}_{N^2}$.

III. Sparse estimation of the DoA

A. Sparse modelling & sparse estimation

Assuming that all sources directions $\boldsymbol{\theta}_m = \{u_m, v_m\}$ are comprised within a grid $\boldsymbol{\varphi} = \{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_G\}$ of G directions with $\boldsymbol{\varphi}_g = \{\upsilon_g, \nu_g\}$ the g-th grid direction, a sparse equivalent of the transformed VCMM (7) can be found:

$$\mathbf{y} = \mathbf{B}_w(\boldsymbol{\varphi})\boldsymbol{\gamma}_0 + \boldsymbol{\delta}_w \quad (8)$$

where $\mathbf{B}_w(\boldsymbol{\varphi}) = \widehat{\mathbf{W}}\mathbf{B}(\boldsymbol{\varphi})$ with $\mathbf{B}(\boldsymbol{\varphi}) = [\mathbf{b}(\boldsymbol{\varphi}_1), \ldots, \mathbf{b}(\boldsymbol{\varphi}_G)]$ is an overcomplete dictionary of size $N^2 \times G, G \gg N^2$. $\boldsymbol{\gamma}_0$ is an M-sparse vector which has only M non-zero components corresponding to sources directions. Therefore, an estimate of $\boldsymbol{\gamma}_0$ is required to estimate the DoAs. Using the sparsity prior on $\boldsymbol{\gamma}_0$, the estimation problem is formulated as the minimization of an ℓ_0-penalized objective:

$$\min_{\boldsymbol{\gamma}\in\mathbb{C}^G}\left\{\mathcal{J}_{\ell_0}(\lambda, \boldsymbol{\gamma}) = \mathcal{J}_{LS}(\boldsymbol{\gamma}) + \lambda\|\boldsymbol{\gamma}\|_0\right\} \quad (9)$$

where $\mathcal{J}_{LS}(\boldsymbol{\gamma}) = \frac{1}{2}\|\mathbf{y} - \mathbf{B}_w(\boldsymbol{\varphi})\boldsymbol{\gamma}\|_2^2$ and the hyperparameter $\lambda > 0$ is introduced to balance the solution sparsity towards data fidelity. Throughout this paper, the approach of [10], [11] is employed to select λ thereby ensuring equivalence between sparse and ML estimators *ie.* both criteria share the same global minimizer.

B. The Proximal Gradient Algorithm

Problem (9) is solved using the well-known PGA algorithm [12]. One PGA iteration reads:

$$\boldsymbol{\gamma}^{(i+1)} = \mathrm{prox}_{\lambda\beta\ell_0}\left(\boldsymbol{\gamma}^{(i)} - \beta\mathbf{B}_w^H(\boldsymbol{\varphi})(\mathbf{B}_w(\boldsymbol{\varphi})\boldsymbol{\gamma}^{(i)} - \mathbf{y})\right) \quad (10)$$

with $\mathrm{prox}_{\lambda\beta\ell_0}(\cdot)$ the ℓ_0-norm proximity operator and β the stepsize. The convergence towards a local minimizer of $\mathcal{J}_{LS}$ is ensured for stepsize values satisfying $\beta \leq 1/\mathcal{L}$ where $\mathcal{L} = \lambda_{max}(\boldsymbol{\mathcal{H}}_w)$ is the Lispchitz constant of the gradient of $\mathcal{J}_{LS}$ *ie.* the largest eigenvalue of Hessian matrix:

$$\boldsymbol{\mathcal{H}}_w = \mathbf{B}_w^H(\boldsymbol{\varphi})\mathbf{B}_w(\boldsymbol{\varphi}) \quad (11)$$

IV. Accelerated PGA through pre-whitening

A. Ablin's Accelerated PGA

According to [20], PGA iterations (10) can cast as a Majorization-Minimization (MM) steps. Formally, a second-order Taylor expansion of the data fidelity term $\mathcal{J}_{LS}$ in the neighbourhood of $\boldsymbol{\gamma}^{(i)}$ yields:

$$\begin{aligned}\mathcal{J}_{LS}(\boldsymbol{\gamma}) = \mathcal{J}_{LS}(\boldsymbol{\gamma}^{(i)}) &+ \left(\mathbf{B}_w^H(\boldsymbol{\varphi})(\mathbf{B}_w(\boldsymbol{\varphi})\boldsymbol{\gamma}^{(i)} - \mathbf{y})\right)^H \\ &\times (\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)}) + \tfrac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)})^H\boldsymbol{\mathcal{H}}_w(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)}) \\ &+ o(\|(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)})\|_2^2)\end{aligned} \quad (12)$$

Then, an upper bound of the second-order term in (12) is introduced yielding $\mathcal{Q}_{\mathcal{L}}$ an upper bound on $\mathcal{J}_{LS}$ satisfying $\mathcal{J}_{LS}(\boldsymbol{\gamma}) \leq \mathcal{Q}_{\mathcal{L}}(\boldsymbol{\gamma})$:

$$\begin{aligned}\mathcal{Q}_{\mathcal{L}}(\boldsymbol{\gamma}) = \mathcal{J}_{LS}(\boldsymbol{\gamma}^{(i)}) &+ \left(\mathbf{B}_w^H(\boldsymbol{\varphi})(\mathbf{B}_w(\boldsymbol{\varphi})\boldsymbol{\gamma}^{(i)} - \mathbf{y})\right)^H \\ &\times (\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)}) + \tfrac{\mathcal{L}}{2}\|\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)}\|_2^2\end{aligned} \quad (13)$$

Finally, assuming that the stepsize is $\beta = 1/\mathcal{L}$, $\boldsymbol{\gamma}^{(i+1)}$ in (10), minimizing $\mathcal{Q}_{\mathcal{L}}$ directly gives (10) prior the application of the proximity operator. Note that with the particular setting $\beta = 1/\mathcal{L}$, (10) corresponds to a Gauss-Newton [2] step on $\mathcal{Q}_{\mathcal{L}}$ whose Hessian matrix is $\mathcal{L}\mathbf{I}_G$.

Ablin [18] proposed an accelerated algorithm FBS exploiting tighter upper bounds of (12) by leveraging the sparsity of $\boldsymbol{\gamma}^{(i)}$. Let us introduce:

$$\mathcal{S}^{(i)} = \text{supp}(\boldsymbol{\gamma}^{(i)}) = \begin{cases} \left\{ i \mid \gamma_g^{(i)} \neq 0 \right\} & i > 0 \\ \{1, \dots, G\}, i = 0 \end{cases} \tag{14}$$

the support of $\boldsymbol{\gamma}^{(i)}$ at iteration i. Then, a variable stepsize $\beta_{\mathcal{S}}^{(i)}$ is considered at each iteration in (10). The stepsize $\beta_{\mathcal{S}}^{(i)} = 1/\mathcal{L}_{\mathcal{S}}^{(i)}$ is computed using $\mathcal{L}_{\mathcal{S}}^{(i)} = \lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})$ the largest eigenvalue of the Hessian matrix $\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)} = \mathbf{B}_{w,\mathcal{S}}^{(i)H}(\boldsymbol{\varphi})\mathbf{B}_{w,\mathcal{S}}^{(i)}(\boldsymbol{\varphi})$ (11) computed using $\mathbf{B}_{w,\mathcal{S}}^{(i)}(\boldsymbol{\varphi})$ the restriction of $\mathbf{B}_w(\boldsymbol{\varphi})$ to column indices associated to non-null components in $\boldsymbol{\gamma}^{(i)}$. Then, $\beta_{\mathcal{S}}^{(i)}$ is a valid stepsize if $\text{supp}(\boldsymbol{\gamma}^{(i+1)}) \subset \mathcal{S}^{(i)}$ [18] where $\boldsymbol{\gamma}^{(i+1)}$ is obtained by minimizing $\mathcal{Q}_{\mathcal{L}_{\mathcal{S}}}^{(i)}$ the corresponding upper bound obtained using $\mathcal{L}_{\mathcal{S}}^{(i)}$ which is tighter than $\mathcal{Q}_{\mathcal{L}}$. Using $\mathcal{Q}_{\mathcal{L}_{\mathcal{S}}}^{(i)}$, larger stepsizes $\beta_{\mathcal{S}}^{(i)} = 1/\mathcal{L}_{\mathcal{S}}^{(i)} \geq 1/\mathcal{L}$ can be taken thus enabling faster convergence. Indeed, in the first iterations $\mathcal{S}^{(i)} = \{1, \dots, G\}$ leading to the standard stepsize value whereas in the last iterations $\mathcal{S}^{(i)}$ contains only the indices of the sources directions yielding munch larger stepsize than the intial stepsize $\beta_{\mathcal{S}}^{(i)} \geq \beta$. Hence, tighter upper bounds of (12) can be exploited and thus larger stepsizes $\beta_{\mathcal{S}}^{(i)}$ can be taken which enables faster convergence. Algorithm 1 summarizes the method[1].

Algorithm 1 Ablin's [18] accelerated PGA

Input: $\mathbf{y}, \boldsymbol{\gamma}^{(0)}, \mathbf{B}_w(\boldsymbol{\varphi}), \lambda > 0$
Output: $\boldsymbol{\gamma}^{(i)}$

```
while stopping criterion is not satisfied do
    Compute S^(i) = supp(γ^(i)) and β_S = 1/L_S
    Compute γ̃^(i+1) using PGA step (10) with β_S
    if supp(γ̃^(i+1)) ⊂ S^(i) then
        γ^(i+1) ← γ̃^(i+1)
    else
        Redo the PGA step 3 using β = 1/L to obtain γ^(i+1)
    end if
end while
return γ^(i)
```

Fig.1 presents the different upper bounds $\mathcal{Q}_{\mathcal{L}_{\mathcal{S}}}^{(i)}$ prior (observation $\mathbf{r}$ (5)) and after the pre-whitening transform (observation $\mathbf{y}$ (7)). It follows that all upper bounds are tighter than the initial ones. Upper bounds obtained after the pre-whitening transform are substantially flatter. Thereby, Fig.1 indicates that significantly larger stepsizes can be taken after the transform leading to consequent speed improvements.

[1]Note that algorithm 1 can be interpreted as a Variable Forward Backward Metric (VMFB) [21] scheme as it each iteration consists in a preconditionned PGA step with preconditionner $\mathbf{U} = (\beta_{\mathcal{S}^{(i)}})^{-1}\mathbf{I}_G$

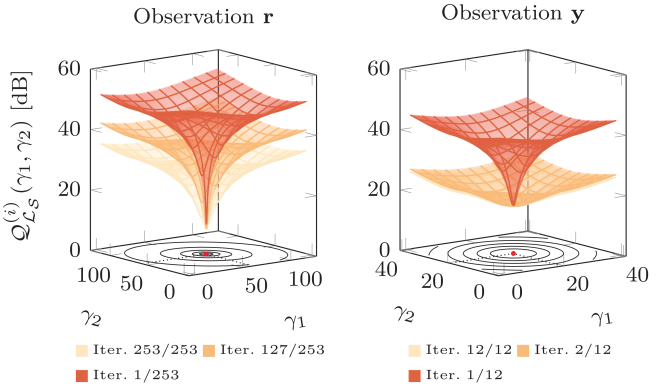


Fig. 1. Upper bounds $\mathcal{Q}_{\mathcal{L}_{\mathcal{S}}}^{(i)}$ projections onto the sources directions prior and after the pre-whitening transform for several PGA iterations. Dashed lines and the red circle represent the global minimum $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \hat{\gamma}_2]^T$ coordinates for each criterion. $M = 2$ sources of directions $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\} = \{(0.5, -0.30), (-0.75, 0.29)\}$ (leading to spatial correlation $r_{\mathbf{A}}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 0.3$) are considered with $K = 400$ and $\text{SNR} = 10\,\text{dB}$ along the array defined in section V.

B. Distance between upper bounds and the true function

The subsection objective is to show how the pre-whitening transform (7) leads to tighter upper bounds and so provides substantial acceleration. To this end, the distance:

$$\varepsilon_1^{(i)} = |\tfrac{\mathcal{L}_{\mathcal{S}}^{(i)}}{2}\|\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)}\|_2^2 - \tfrac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)})^H \boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)})| \tag{15}$$

between the true quadratic term in (12) and its upper bound is computed and shown to be significantly reduced after the transform. Let us assume that $\boldsymbol{\gamma}^{(i)}$ has support $\mathcal{S}^{(i)}$ and that $\beta_{\mathcal{S}}^{(i)}$ is a valid stepsize. According to IV-A and [18], $\text{supp}(\boldsymbol{\gamma}) \subset \mathcal{S}^{(i)}$. Then, an Eigenvalue Decomposition (EVD) on $\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}$ yields:

$$\begin{aligned} &\tfrac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)})^H \boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)}) = \tfrac{1}{2}\sum_{g \in \mathcal{S}} \lambda_g |\gamma_g - \gamma_g^{(i)}|^2 \\ &= \tfrac{1}{2}\sum_{g \in \mathcal{S}\setminus\{g_{max}\}} \lambda_g |\gamma_g - \gamma_g^{(i)}|^2 + \tfrac{\mathcal{L}_{\mathcal{S}}^{(i)}}{2}|\gamma_{g_{max}} - \gamma_{g_{max}}^{(i)}|^2 \end{aligned} \tag{16}$$

where λ_g is the g-th eigenvalue of $\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}$ and g_{max} the index corresponding to $\mathcal{L}_{\mathcal{S}}^{(i)} = \lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})$. Since:

$$\tfrac{\mathcal{L}_{\mathcal{S}}^{(i)}}{2}\|\boldsymbol{\gamma} - \boldsymbol{\gamma}^{(i)}\|_2^2 = \tfrac{\mathcal{L}_{\mathcal{S}}^{(i)}}{2}|\gamma_{g_{max}} - \gamma_{g_{max}}^{(i)}|^2 + \tfrac{\mathcal{L}_{\mathcal{S}}^{(i)}}{2}\sum_{g \in \mathcal{S}\setminus\{g_{max}\}} |\gamma_g - \gamma_g^{(i)}|^2 \tag{17}$$

inserting (16) into (15) gives:

$$\varepsilon_1^{(i)} = \tfrac{\mathcal{L}_{\mathcal{S}}^{(i)}}{2}\sum_{g \in \mathcal{S}\setminus\{g_{max}\}} \left(1 - \tfrac{\lambda_g}{\mathcal{L}_{\mathcal{S}}^{(i)}}\right) |\gamma_g - \gamma_g^{(i)}|^2 \tag{18}$$

The analysis of [11] revealed that the pre-whitening transform reduces the conditioning of the Hessian matrix $\eta(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}) = \frac{\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}{\lambda_{min}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}$ such that $\eta(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}) \to 1$. This yields $\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}) \to \lambda_{min}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})$ and so $1 - \frac{\lambda_g}{\mathcal{L}_{\mathcal{S}}^{(i)}} \to 0$. Hence

$\varepsilon_1^{(i)} \to 0$ which makes the upper bound closer to the true function thereby enhancing the algorithm convergence.

In the special case of $M = 2$ impinging sources, assuming without loss of generality that $g_{max} = 1$, (18) becomes:

$$\varepsilon_1^{(i)} = \frac{r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{1+r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)} \times |\gamma_2 - \gamma_2^{(i)}|^2 \tag{19}$$

which connects $\varepsilon_1^{(i)}$ to the sources spatial correlation $r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \frac{\mathbf{b}_w^H(\boldsymbol{\theta}_1)\mathbf{b}_w(\boldsymbol{\theta}_2)}{\|\mathbf{b}_w(\boldsymbol{\theta}_1)\|_2\|\mathbf{b}_w(\boldsymbol{\theta}_2)\|_2}$ computed for steering vectors $\mathbf{b}_w(\boldsymbol{\theta}_1), \mathbf{b}_w(\boldsymbol{\theta}_2)$. Again, the spatial correlation $r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ after the transform (using $\mathbf{B}_w(\boldsymbol{\varphi})$) tends to 0 and so $\varepsilon_1^{(i)}$ which is not the case prior the transform (using $\mathbf{B}(\boldsymbol{\varphi})$ and so $r_{\mathbf{B}}$ in (19) instead of $r_{\mathbf{B}_w}$).

Proof: Inserting $\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}) = 1 + r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ and $\lambda_{min}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}) = 1 - r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ [11] directly yields (19). ■

C. Equivalence with a second-order algorithm

This subsection purpose is to demonstrate equivalence between last iterations of the accelerated PGA algorithm after the pre-whitening transform and Gauss-Newton iterations [2] in the special case of $M = 2$ impinging sources. Assuming that the support $\mathcal{S}^{(i)}$ contains only the two indices corresponding to sources directions $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, Gauss-Newton iterations can be written as:

$$\boldsymbol{\gamma}^{(i+1)} = \boldsymbol{\gamma}^{(i)} - (\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})^{-1}\boldsymbol{\nabla}\mathcal{J}_{LS,\mathcal{S}}^{(i)}(\boldsymbol{\gamma}^{(i)}) \tag{20}$$

with $\boldsymbol{\nabla}\mathcal{J}_{LS,\mathcal{S}}^{(i)} = \left(\mathbf{B}_{w,\mathcal{S}^{(i)}}^H(\boldsymbol{\varphi})(\mathbf{B}_{w,\mathcal{S}^{(i)}}(\boldsymbol{\varphi})\boldsymbol{\gamma}^{(i)} - \mathbf{y})\right)$ the gradient and $\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}$ the Hessian matrix (11) both obtained using only the indices associated to sources directions. The gradient step of PGA iterations (10) becomes:

$$\boldsymbol{\gamma}^{(i+1)} = \boldsymbol{\gamma}^{(i)} - \frac{1}{\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}\mathbf{I}_2\boldsymbol{\nabla}\mathcal{J}_{LS,\mathcal{S}}^{(i)}(\boldsymbol{\gamma}^{(i)}) \tag{21}$$

Eq. (21) is similar to (20) except that the inverse Hessian matrix is replaced with $\frac{1}{\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}\mathbf{I}_2$ in (21). Hence, equivalence between (21) and (20) relies on the distance between $\frac{1}{\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}\mathbf{I}_2$ and $(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})^{-1}$. Then, the following can be derived:

$$\varepsilon_2^{(i)} = \|(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})^{-1} - \frac{1}{\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}\mathbf{I}_2\| = \frac{2r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{(1+r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2))^2} \tag{22}$$

where $\|\cdot\| = \lambda_{max}(\cdot)$ is the matrix spectral norm.

Since the pre-whitening transform decorrelates the dictionary vectors in the sources directions [11], [17] *ie.* $r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \approx 0$, the last iterations of the accelerated PGA after the pre-whitening transform are equivalent to Gauss-Newton as $\varepsilon_2^{(i)} \to 0$.

Proof: Using an EVD, $\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}$ can be rewritten as:

$$\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)} = \lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})\mathbf{e}_1\mathbf{e}_1^H + \lambda_{min}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})\mathbf{e}_2\mathbf{e}_2^H \tag{23}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the eigenvectors of $\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}$. Inserting $\lambda_{min}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}) = \lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)}) - 2r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ into (23) yields:

$$\begin{aligned}\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)} &= \lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})\left(\mathbf{e}_1\mathbf{e}_1^H + \mathbf{e}_2\mathbf{e}_2^H\right) - 2r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\mathbf{e}_2\mathbf{e}_2^H \\ &= \lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})\mathbf{I}_2 - 2r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\mathbf{e}_2\mathbf{e}_2^H\end{aligned} \tag{24}$$

since $\mathbf{e}_1\mathbf{e}_1^H + \mathbf{e}_2\mathbf{e}_2^H = \mathbf{I}_2$. A first-order Taylor expansion of $(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})^{-1}$ (24) then gives:

$$(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})^{-1} \approx \frac{1}{\lambda_{max}(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}\mathbf{I}_2 + \frac{2r_{\mathbf{B}_w}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{\lambda_{max}^2(\boldsymbol{\mathcal{H}}_{w,\mathcal{S}}^{(i)})}\mathbf{e}_2\mathbf{e}_2^H \tag{25}$$

Finally, substituting (25) in (22) completes the proof as $\|\mathbf{e}_2\mathbf{e}_2^H\| = 1$ since it is an orthogonal projector. ■

V. Numerical simulations

A. Experimental setup

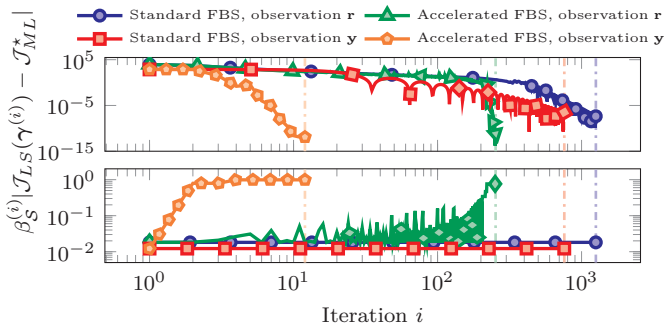
An 6-element array with 5 antennas distributed around a circle of radius $r = 0.8\lambda_0$ where λ_0 is the wavelength and one central antenna is considered. $M = 2$ sources with $\mathrm{SNR} = 10\,\mathrm{dB}$ and $K = 400$ array snapshots are considered. An oracle grid containing the ML estimates is employed to remove any grid bias. To alleviate the computational cost of computing the stepsize $\beta_{\mathcal{S}} = 1/\mathcal{L}_{\mathcal{S}}$ in (**??**), the EVD is replaced by the power methods. The ℓ_0 penalty is replaced with the Continuous Exact ℓ_0 penalty (CEL0) [22] which preserves the ℓ_0 criterion global minimizer while drastically reducing the number of local minima. λ is chosen using [23] for the non-transformed observation $\mathbf{r}$ (5) whereas [10] is employed when using the transformed observation $\mathbf{y}$ (7). Finally, the PGA is initialized with $\boldsymbol{\gamma}^{(0)} = \mathbf{0}$.

B. Convergence analysis

Fig.2 presents the convergence curves obtained prior and after the pre-whitening transform using both standard and accelerated PGA with a grid step of 0.1 on both directions. Corresponding stepsizes are displayed on the bottom subfigure. $M = 2$ sources of directions $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\} = \{(0.5, -0.30), (-0.75, 0.29)\}$ are considered leading to leading to spatial correlation $r_{\mathbf{A}}^2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = 0.3$. First, the transform slightly accelerates the standard algorithm as 759 iterations are required instead of 1257 to reach the stopping criterion on $\|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\|_2^2/\|\boldsymbol{\gamma}^{(i)}\|_2^2 \leq 10^{-6}$. Using the accelerated algorithm, the transform allows extremely fast and stable convergence in 12 iterations compared to the highly unstable 253 iterations without the transform. Fig.2 confirms the speed improvement obtained combining the transform with the accelerated FBS algorithm as 12 iterations are required to achieve convergence 118.5 compared to without the transform.

C. Number of FLOPS

To investigate the connection between the spatial correlation and the convergence revealed by subsections IV-B and IV-C, sources directions are chosen to satisfy a specified spatial correlation $r_{\mathbf{A}}^2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ defined using the array steering matrix. Then, the number of FLOPS is measured both prior and after the transform as a function of $r_{\mathbf{A}}^2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$. As shown by Fig.3, without the transform, the algorithm can not be accelerated as the FLOPS number is greater than 10^9 even for spatially decorrelated sources. Using the transform and the standard algorithm, a first improvement in the computational cost is observed. Finally, the transform combined with the accelerated algorithm yields an efficient algorithm with less



Algorithm	Observation $\mathbf{r}$ (5)	Observation $\mathbf{y}$ (7)
PGA [12]	1380.0	633.5
Accelerated PGA [18]	118.5	**11.3**

Fig. 2. Top: Convergence curves prior and after the pre-whitening transform with $\mathcal{J}_{ML}^\star$ the optimal value of J_{LS}. Middle: corresponding stepsizes at each iteration. Bottom: Mean number of iterations on 100 trials.

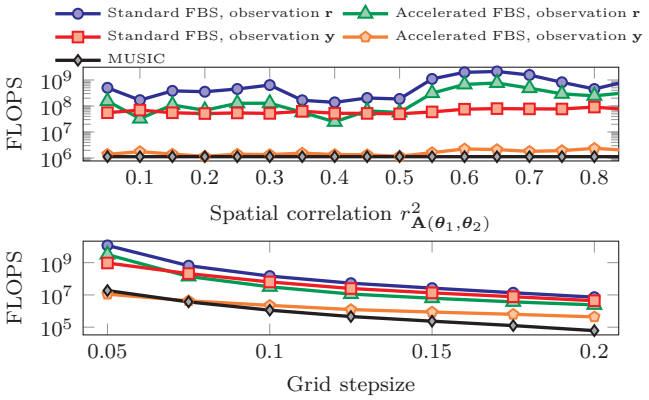


Fig. 3. Number of FLOPS as a function of the sources spatial correlation (top) and the grid stepsize (bottom). Both u and v grid stepsizes are equal.

than 10^7 FLOPS even highly for highly correlated sources. The number of FLOPS is increasing as more iterations are required to achieve convergence as predicted in subsections IV-B and IV-C. The previous experiment is repeated with a fixed correlation of $r^2_{\mathbf{A}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 0.3$ and a varying grid stepsize. Again, combining the transform and the accelerated algorithm yields an efficient 2D ML implementation even for thin grids.

VI. Conclusion

In this paper, the pre-whitening transform effects are leveraged to accelerate the sparse estimator convergence. By spatially decorrelating the dictionary vectors, a fast implementation of 2D ML-DoA estimation is proposed. Furthermore, equivalence between the algorithm's last iterations and a second-order descent scheme is proven. Numerical simulations demonstrated the proposed algorithm's efficiency.

References

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

[2] B. Ottersten, M. Viberg, P. Stoica, e. S. Nehorai, A.", J. Litva, and T. J. Shepherd, *Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, pp. 99–151.

[3] J. S. Picard and A. J. Weiss, "Direction finding of multiple emitters by spatial sparsity and linear programming," in *2009 9th International Symposium on Communications and Information Technology*, 2009, pp. 1258–1262.

[4] Z. He, Q. Liu, L. Jin, and S. Ouyang, "Low complexity method for doa estimation using array covariance matrix sparse representation," *Electronics Letters*, vol. 49, no. 3, pp. 228–230, 2013.

[5] W. Cui, T. Qian, and J. Tian, "Enhanced covariances matrix sparse representation method for doa estimation," *Electronics Letters*, vol. 51, no. 16, pp. 1288–1290.

[6] T. Qian, W. Cui, and Q. Shen, "Sparse reconstruction method for doa estimation based on dynamic dictionary and negative exponent penalty," *Chinese Journal of Electronics*, vol. 27, no. 2, pp. 386–392, 2018.

[7] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, "On the virtual array concept for higher order array processing," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1254–1271, 2005.

[8] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," 2017.

[9] H. Zhang, X. Wang, J. Lin, and Z. Du, "Sparse optimization driven deep unfolding network for doa estimation," in *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2024, pp. 1–6.

[10] T. Aussaguès, A. Ferréol, A. Delmer, and P. Larzabal, "Looking for equivalence between maximum likelihood and sparse doa estimators," in *2024 32th European Signal Processing Conference (EUSIPCO)*, 2024.

[11] T. Aussaguès, A. Ferréol, A. Delmer, and P. Larzabal, "Ml-doa estimation using a sparse representation of array covariance with a non-standard noise," *Submitted to Signal Processing*, 2025.

[12] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," 2010.

[13] Y. Nesterov, "A method for solving the convex programming problem with convergence rate $O(1/k^2)$," *Proceedings of the USSR Academy of Sciences*, vol. 269, pp. 543–547, 1983.

[14] M. Savanier, E. Chouzenoux, J.-C. Pesquet, and C. Riddell, "Unmatched preconditioning of the proximal gradient algorithm," *IEEE Signal Processing Letters*, vol. 29, pp. 1122–1126, 2022.

[15] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *International Conference on Machine Learning*, 2010.

[16] S. Becker, J. Fadili, and P. Ochs, "On quasi-newton forward–backward splitting: Proximal calculus and convergence," *SIAM Journal on Optimization*, vol. 29, 01 2018.

[17] T. Aussaguès, A. Ferréol, A. Delmer, and P. Larzabal, "Whitening effects for ml-doa estimation using a sparse representation of array covariance," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[18] P. Ablin, T. Moreau, M. Massias, and A. Gramfort, "Learning step sizes for unfolded sparse coding," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[19] R. J. Muirhead, "Aspects of multivariate statistical theory," in *Wiley Series in Probability and Statistics*, 1982.

[20] 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.

[21] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function," *Journal of Optimization Theory and Applications*, vol. 162, pp. 107–132, 07 2014.

[22] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A Continuous Exact l0 penalty (CEL0) for least squares regularized problem," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. pp. 1607–1639 (33 pages), Jul. 2015.

[23] A. Delmer, A. Ferréol, and P. Larzabal, "On regularization parameter for l0-sparse covariance fitting based doa estimation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 4552–4556.