# Pattern Coupled Sparse Bayesian Learning with Fixed Point Iterations for DOA and Amplitude Estimation

Dogan, Didem; Leus, Geert

**DOI**
10.1109/IEEECONF59524.2023.10477000

**Publication date**
2023

**Document Version**
Final published version

**Published in**
Proceedings of the 2023 57th Asilomar Conference on Signals, Systems, and Computers

# Pattern Coupled Sparse Bayesian Learning with Fixed Point Iterations for DOA and Amplitude Estimation

Didem Dogan[1] and Geert Leus[1]
[1]Signal Processing Systems (SPS), Delft University of Technology, The Netherlands



***Abstract*—We consider the problem of recovering block-sparse signals with unknown boundaries. Such signals arise naturally in various applications. Recent literature introduced a pattern-coupled or clustered Gaussian prior, in which each coefficient involves its own hyperparameter as well as its immediate neighbors' hyperparameters. Some methods use a hierarchical distribution making the solution vulnerable to the parameter choice. Besides, these methods mainly rely on the expectation-maximization (EM) algorithm and either require a suboptimal solution or an approximation of the hyperparameters. To address these difficulties, we propose to solve the pattern coupling problem via fixed point iterations instead of the EM algorithm. The proposed algorithm does not require any further assumptions on the hyperparameters and provides a simple update rule for the hyperparameters. Although the fixed point iterations method is an empirical strategy, it provides a fast convergence rate. The proposed algorithm is tested on a simple direction of arrival (DOA) and amplitude estimation problem. From our simulations, we see that the proposed method achieves similar reconstruction results with the state-of-the-art; however, the proposed method is faster than the existing counterparts.**

***Index Terms*—sparse Bayesian learning, block sparse signals, fixed point iterations, pattern coupling**

## I. Introduction

Block sparsity has been observed for signals in numerous applications, such as the cluster structure of scatterers on radar images [1], fetal ECG [2], DOA estimation, and so on [3]. The block sparse model can be naturally exploited by further including the relation between sparse coefficients, such as the dependence of the sparsity patterns. Under noisy environments, correlated settings, or with very compressive measurements, algorithms properly leveraging such an underlying structure could achieve a robust recovery compared to their counterparts which merely exploit the sparsity.

The pattern-coupled sparse Bayesian learning (PCSBL) algorithm incorporates a pattern-coupled hierarchical Gaussian prior where each coefficient depends on its own hyperparameter and its immediate neighbors' hyperparameters to exploit interactions between neighboring coefficients [4], [5]. For this problem, a suboptimal solution is attained for the hyperparameters; however, the performance of the PCSBL depends on a proper selection of the hyperparameters. Clustered sparse Bayesian learning (CSBL) takes on a similar idea as the pattern-coupled prior used in PCSBL yet without relying on the hierarchical distribution over the hyperparameters [6]. Both algorithms use an EM-based update rule.

This publication is part of the project TOUCAN (with project number 17208) of the research programme TTW-OTP which is financed by the Dutch Research Council (NWO).

We propose to solve the pattern coupling problem via fixed point iterations instead of the EM algorithm [7]. The fixed point iterations method is an empirical strategy, but it provides a fast convergence rate in most applications [8], [9]. In [4], PCSBL uses a lower bound to approximate the optimal hyperparameter, whose performance always depends on a careful setting of the auxiliary parameter. Instead, our algorithm does not need such a bound since it does not require any selection of auxiliary parameters. Our algorithm can be seen as a fixed point update rule-based version of the EM update rule-based CSBL algorithm [6]. However, CSBL assumes neighboring sparse coefficients to share the same variance or precision. We do not make such an assumption. Finally, although a theoretical analysis of the convergence behavior is unavailable, the proposed algorithm demonstrates a fast convergence rate.

The remainder of this paper is organized as follows. In the next section, we define the signal model for the DOA and amplitude estimation problem. In Section III, we present the proposed Bayesian formulation with a short discussion on the existing literature. Then, we compare the performance of our proposed method with the state-of-the-art. In the final section, we discuss the results and conclude this work.

## II. DOA and Amplitude Estimation

In DOA and amplitude estimation, we employ a multiple measurement vector (MMV) model, which exploits different pulse periods in synergy [10]. Such an MMV model can be expressed as

$$\mathbf{Z} = \mathbf{A}\mathbf{S} + \mathbf{N}, \tag{1}$$

where the system matrix $\mathbf{A} = [\mathbf{a}_1 ... \mathbf{a}_L] \in \mathbb{C}^{N \times L}$ contains the array steering vectors for all hypothetical DOAs as columns, with the $(n,l)$-th element given by $\exp(-j(n-1)\frac{\omega d}{c} sin\theta_l)(d$ is the element spacing, $c$ the sound speed, $\omega$ the frequency, and $\theta_l$ the $l$th DOA, e.g., $\theta_l = -90^\circ + \frac{l-1}{L} 180^\circ$); $\mathbf{S} \in \mathbb{C}^{L \times M}$ represents the complex source amplitudes $s_{l,m}$; $\mathbf{Z} = [\mathbf{z}_1 ... \mathbf{z}_M] \in \mathbb{C}^{N \times M}$ represents the concatenation of the measurements $\mathbf{z}_m$ at the $m$th snapshot; and the additive white noise $\mathbf{N}$ is assumed independent across sensors and snapshots, where each element has a complex Gaussian distribution with zero mean and variance $\sigma^2$.

## III. Bayesian Formulation

### A. Priors on the Sources

Here, Bayesian interference is used. This involves determining the posterior distribution of the complex source amplitudes from the likelihood and a prior model. In (1), the probability density function is given by

$$p(\mathbf{Z}|\mathbf{S};\sigma^2) = \frac{\exp\left(-\frac{1}{\sigma^2}||\mathbf{Z}-\mathbf{AS}||_2^2\right)}{(\pi\sigma^2)^{ML}}. \tag{2}$$

as the noise is assumed to be Gaussian. In classical SBL, the complex source amplitudes $s_{l,m}$ are assumed independent across different snapshots and each other as

$$p_m(s_{l,m},\gamma_l) = \mathcal{CN}(0,\gamma_l). \tag{3}$$

On the other hand, the pattern-coupled model is proposed in this work to cope with block-sparse signals with unknown block-sparse structures. This model utilizes the fact that the sparsity patterns of neighboring coefficients are statistically dependent. Specifically, in this model, the Gaussian prior for each coefficient involves its own hyperparameter and its immediate neighbors' hyperparameters [4]. For a one-dimensional set-up, such as the presented DOA and amplitude estimation problem, the pattern-coupled prior is given as follows:

$$p_m(s_{l,m},\gamma_l,\gamma_{l-1},\gamma_{l+1}) = \mathcal{CN}(0,\gamma_l+\beta\gamma_{l-1}+\beta\gamma_{l+1}). \tag{4}$$

Then we can express their joint distribution as

$$p(\mathbf{S};\mathbf{\Gamma}) = \prod_{m=1}^{M}\prod_{l=1}^{L} p_m(s_{l,m}),\gamma_l+\beta\gamma_{l-1}+\beta\gamma_{l+1}) \tag{5}$$

$$= \prod_{m=1}^{M}\mathcal{CN}(\mathbf{0},\mathbf{\Gamma}) \tag{6}$$

where $\mathbf{\Gamma} = \text{diag}(\gamma_1+\beta\gamma_0+\beta\gamma_2,...,\gamma_L+\beta\gamma_{L-1}+\beta\gamma_{L+1})$. Note that $\gamma_0 = \gamma_{l+1} = 0$. In this approach, the sparsity is controlled by the hyperparameters. If the $\gamma_l$ is non-zero, then $s_{l,m}$ is also non-zero. Therefore, if any of the neighboring elements (i.e., $s_{l-1}$ and $s_{l+1}$) is non-zero, then the center element $s_l$ is likely to be non-zero. It might not give an exact sparse reconstruction; however, it provides the continuity of the sparsity patterns and hence the block sparsity.

Note that in PCSBL [4], the pattern-coupling prior has the following form:

$$p_m(s_{l,m},\gamma_l,\gamma_{l-1},\gamma_{l+1}) = \mathcal{CN}(\mathbf{0},(\gamma_l+\beta\gamma_{l-1}+\beta\gamma_{l+1})^{-1}) \tag{7}$$

Here, the motivation is the 'zero-coupling' effect. In this approach, if any of the neighboring elements (i.e. $s_{l-1}$ and $s_{l+1}$) are zero, then the center element $s_l$ is likely to be zero. In other words, if the $\gamma_l \to \infty$ for an element is zero, the neighboring elements are also zero. On the other hand, in CSBL [6], the prior has the following form:

$$p_m(s_{l,m},\gamma_l,\gamma_{l-1},\gamma_{l+1}) = \mathcal{CN}(\mathbf{0},\gamma_l+\gamma_{l-1}+\gamma_{l+1}). \tag{8}$$

which has a similar formulation as the proposed one in (4), without the $\beta$ term. They also provided an extension of that model, which takes the same formulation as given in (4). It is also possible to provide an update rule for $\beta$; however, it is out of the scope of this work. We select $\beta = 0.5$, resulting in a better reconstruction quality in most scenarios.

Note that apart from the DOA estimation problem, this method can be extended to reconstruct two-dimensional data in other problems, such as ultrasound imaging. However, our algorithm is not applied to two-dimensional problems in this work and is limited to the one-dimensional DOA and amplitude estimation problem.

### B. Stochastic Likelihood

The posterior distribution of the sources can be attained using the Bayes rule conditioned on $\gamma$ and $\sigma^2$:

$$p(\mathbf{S}|\mathbf{Z};\gamma,\sigma^2) = \frac{p(\mathbf{Z}|\mathbf{S};\sigma^2)p(\mathbf{S};\mathbf{\Gamma})}{p(\mathbf{Z};\mathbf{\Gamma},\sigma^2)} \propto p(\mathbf{Z}|\mathbf{S};\sigma^2)p(\mathbf{S};\mathbf{\Gamma})$$
$$\propto \frac{e^{-tr((\mathbf{S}-\mu_{\mathbf{S}})^H\mathbf{\Sigma}_{\mathbf{s}}^{-1}(\mathbf{S}-\mu_{\mathbf{S}}))}}{(\pi^N\det\Sigma_{\mathbf{s}})^M} = \mathcal{CN}(\mu_{\mathbf{S}},\mathbf{\Sigma}_{\mathbf{s}}). \tag{9}$$

Since both $p(\mathbf{Z}|\mathbf{S};\sigma^2)$ and $p(\mathbf{S};\mathbf{\Gamma})$ are Gaussian $p(\mathbf{S}|\mathbf{Z};\mathbf{\Gamma},\sigma^2)$ is also Gaussian with mean $\mu_{\mathbf{S}}$ and covariance $\mathbf{\Sigma}_{\mathbf{s}}$ :

$$\mu_{\mathbf{S}} = E\{\mathbf{S}|\mathbf{Z};\mathbf{\Gamma},\sigma^2\} = \mathbf{\Gamma}\mathbf{A}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{Z} \tag{10}$$

$$\mathbf{\Sigma}_{\mathbf{s}} = \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{A}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{A}\mathbf{\Gamma}. \tag{11}$$

Here, the data covariance matrix is given by

$$\mathbf{\Sigma}_{\mathbf{z}} = \sigma^2\mathbf{I}_N + \mathbf{A}\mathbf{\Gamma}\mathbf{A}^H, \tag{12}$$

and its inverse can be computed using the matrix inversion lemma

$$\mathbf{\Sigma}_{\mathbf{z}}^{-1} = \sigma^{-2}\mathbf{I}_N - \sigma^{-2}\mathbf{A}\mathbf{\Sigma}_{\mathbf{s}}\mathbf{A}^H\sigma^{-2}. \tag{13}$$

The denominator $p(\mathbf{Z};\mathbf{\Gamma},\sigma^2)$ is neglected here as it is only a normalization term and does not depend on $\mathbf{S}$. So in conclusion, we have $\mathbf{S} \sim \mathcal{CN}(\mu_{\mathbf{S}},\mathbf{\Sigma}_{\mathbf{s}})$. Using MAP estimation, we obtain

$$\hat{\mathbf{S}}^{MAP} = \mu_{\mathbf{S}} = E[\mathbf{S}|\mathbf{Z};\mathbf{\Gamma},\sigma^2] = \mathbf{\Gamma}\mathbf{A}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{Z}. \tag{14}$$

Here the diagonal elements of $\mathbf{\Gamma}$ control the row sparsity of $\hat{\mathbf{S}}^{MAP}$. The hyperparameters $\mathbf{\Gamma}$, $\sigma^2$ are estimated by a type-II maximum likelihood, i.e., by maximizing the evidence that was treated as constant in (9). The evidence is the product of the likelihood $p(\mathbf{Z}|\mathbf{S};\sigma^2)$ and the prior $p(\mathbf{S};\mathbf{\Gamma})$ integrated over the complex source amplitudes $\mathbf{S}$. The resulting $p(\mathbf{Z};\mathbf{\Gamma},\sigma^2)$ is given by

$$p(\mathbf{Z};\mathbf{\Gamma},\sigma^2) = \int_{\mathbb{R}^{2ML}} p(\mathbf{Z}|\mathbf{S};\sigma^2)p(\mathbf{S};\mathbf{\Gamma})d\mathbf{S} = \frac{e^{-tr(\mathbf{Z}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{Z})}}{(\pi^N\det\Sigma_{\mathbf{z}})^M} \tag{15}$$

where $d\mathbf{Z} = \prod_{m=1}^{M}\prod_{l=1}^{N}\text{Re}(d\mathbf{Z}_{mn})\text{Im}(d\mathbf{Z}_{mn})$ and $\mathbf{\Sigma}_{\mathbf{z}}$ is the data covariance matrix. We can derive

$$\log p(\mathbf{Z};\mathbf{\Gamma},\sigma^2) \propto -\text{tr}(\mathbf{Z}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{Z}) - M\log\det\mathbf{\Sigma}_{\mathbf{z}}. \tag{16}$$

Finally, the hyperparameters $\mathbf{\Gamma}$ and $\sigma^2$ are estimated from the log-likelihood function as

$$(\hat{\mathbf{\Gamma}},\hat{\sigma}^2) = \text{argmax}_{\mathbf{\Gamma}\geq 0,\sigma^2\geq 0}\ \log\ p(\mathbf{Z};\mathbf{\Gamma},\sigma^2). \tag{17}$$

The parameters $\hat{\mathbf{\Gamma}}$, $\hat{\sigma}^2$ and $\Sigma_{\mathbf{z}}^{-1}$ are iteratively estimated and then finally $\mathbf{S}$ is attained by (14).

### C. Update of $\mathbf{\Gamma}$

Since $\mathbf{\Gamma}$ represents the source powers and controls the sparsity of $\mathbf{S}$, the most significant step is the estimation of $\mathbf{\Gamma}$. We need to find $\mathbf{\Gamma}$ for the estimation of $\mathbf{S}$ in (14). To iteratively compute $\mathbf{\Gamma}$, we form the derivatives of (16) with respect to the elements $\gamma_l$ as follows:

$$\begin{aligned}&\frac{\partial \text{tr}(\mathbf{Z}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{Z})}{\partial\gamma_l} = \text{tr}\{(\frac{\partial \text{tr}(\mathbf{Z}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{Z})}{\partial \mathbf{\Sigma}_{\mathbf{z}}})^T\frac{\partial\mathbf{\Sigma}_{\mathbf{z}}}{\partial\gamma_l}\} = \\ &\mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{Z}\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l + \beta\mathbf{a}_{l-1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{Z}\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1} + \\ &\beta\mathbf{a}_{l+1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{Z}\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1}\end{aligned} \tag{18}$$

$$\begin{aligned}&\frac{\partial \log\det(\mathbf{\Sigma}_{\mathbf{z}})}{\partial\gamma_l} = \text{tr}({\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\frac{\partial\mathbf{\Sigma}_{\mathbf{z}}}{\partial\gamma_l}) = \\ &\mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l + \beta\mathbf{a}_{l-1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1} + \beta\mathbf{a}_{l+1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1}.\end{aligned} \tag{19}$$

After inserting (18) and (19) into (16), the derivative is formed as follows:

$$\begin{aligned}\frac{\partial \log p(\mathbf{Z};\mathbf{\Gamma},\sigma^2)}{\partial\gamma_l} &= ||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l||_2^2 + \beta||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1}||_2^2 \\ &+ \beta||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1}||_2^2 - M\beta\mathbf{a}_{l-1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1} \\ &- M\mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l - M\beta\mathbf{a}_{l+1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1}.\end{aligned} \tag{20}$$

Here (20) is forced to be zero and we obtain the following equality:

$$\begin{aligned}&||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l||_2^2 = M\mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l + \\ &M\beta\mathbf{a}_{l-1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1} - \beta||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1}||_2^2 + \\ &M\beta\mathbf{a}_{l+1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1} - \beta||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1}||_2^2.\end{aligned} \tag{21}$$

Thereafter, we introduce $(\frac{\gamma_l^{new}}{\gamma_l^{old}})^b$ and multiply the right side of (21) to obtain an iterative equation for $\gamma_l$ [8], [9]. In this paper, $b = 2$. This leads to the following fixed-point update rule for $\gamma_l$:

$$\gamma_l^{new} = \frac{\gamma_l^{old}}{\sqrt{M}}\frac{||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l||_2}{\sqrt{\mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l + \beta(v_{l-1}+v_{l+1})}} \tag{22}$$

where

$$\begin{aligned}v_{l-1} &= \mathbf{a}_{l-1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1} - (1/M)||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1}||_2^2, \\ v_{l+1} &= \mathbf{a}_{l+1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1} - (1/M)||\mathbf{Z}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1}||_2^2.\end{aligned}$$

Note that $\gamma_l^{old}$ and $\mathbf{\Sigma}_{\mathbf{z}}$ are given from previous iterations. For the convergence of the fixed point iterations, we need to attain $\gamma_l^{new} = \gamma_l^{old}$.

### D. Noise Variance Estimation

For fast convergence of the FP PCSBL method, it is important to develop a good noise variance estimate as it controls the sharpness of the peaks. In (14) and (12), we need to obtain $\sigma^2$ for the estimation of $\mathbf{S}$. This section estimates the noise variance $\sigma^2$, iteratively. We assume that the number of non-zero elements in the sparse vector is approximately known. Here stochastic maximum likelihood provides an asymptotically efficient estimate of $\sigma^2$ if the set

**Output:** $\mathbf{S}$: unknown data
Initialize $\sigma^2 = 0.1$, diag($\mathbf{\Gamma}$) $= \mathbf{1}$, $\epsilon_{min} = 0.001$, $J_{max} = 500$
**while** $j < J_{max}$ *and* $\epsilon_{min} < \epsilon$ **do**
    $j = j+1$, $\mathbf{\Gamma}^{old} = \mathbf{\Gamma}^{new}$
    calculate $\mathbf{\Sigma}_{\mathbf{z}} = \sigma^2\mathbf{I}_N + \mathbf{A}\mathbf{\Gamma}\mathbf{A}^H$
    update $\gamma_l^{new}$ with (22)
    $\mathbf{\Gamma} = \text{diag}(\gamma_1 + \beta\gamma_0 + \beta\gamma_2, ..., \gamma_L + \beta\gamma_{L-1} + \beta\gamma_{L+1})$
    $\mathcal{K} = \{l \in \mathbb{N} | K$ largest peaks in $\mathbf{\Gamma}\} = \{l_1, \ldots l_K\}$
    $\mathbf{A}_{\mathcal{K}} = (\mathbf{a}_{l_1}, \ldots, \mathbf{a}_{l_K})$
    update $(\sigma^2)^{new} = \frac{tr((\mathbf{I}_N - \mathbf{A}_{\mathcal{K}}\mathbf{A}_{\mathcal{K}}^{\dagger})\mathbf{S}_{\mathbf{z}})}{N-K}$
    $\epsilon = ||\text{diag}(\mathbf{\Gamma}^{new} - \mathbf{\Gamma}^{old})||_1 / ||\text{diag}(\mathbf{\Gamma}^{old})||_1$
**end**
$\tilde{\mathbf{S}} = \mathbf{\Gamma}\mathbf{A}^H\mathbf{\Sigma}_{\mathbf{z}}^{-1}\mathbf{Z}$

**Algorithm 1:** FP PCSBL

of active sources is known. Note that we do not need to know the exact number of sources. A rough guess for the number of sources is also sufficient to obtain a good performance.

Let $\mathbf{\Gamma}_{\mathcal{K}}$ be the covariance matrix of the $K$ estimated sources with corresponding steering matrix $\mathbf{A}_{\mathcal{K}}$. The corresponding data covariance matrix is

$$\mathbf{\Sigma}_{\mathbf{z}} = \sigma^2\mathbf{I}_N + \mathbf{A}_{\mathcal{K}}\mathbf{\Gamma}_{\mathcal{K}}\mathbf{A}_{\mathcal{K}}^H. \tag{23}$$

Note that the data covariance matrix (12) and (23) are identical. We first force (20) to zero as follows:

$$\begin{aligned}&\mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}(\mathbf{S}_{\mathbf{z}} - \mathbf{\Sigma}_{\mathbf{z}}){\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l + \\ &\beta\mathbf{a}_{l-1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}(\mathbf{S}_{\mathbf{z}} - \mathbf{\Sigma}_{\mathbf{z}}){\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l-1} + \\ &\beta\mathbf{a}_{l+1}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}(\mathbf{S}_{\mathbf{z}} - \mathbf{\Sigma}_{\mathbf{z}}){\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_{l+1} = 0\end{aligned} \tag{24}$$

for all sources $l$. Here, the data sample covariance matrix is $\mathbf{S}_{\mathbf{z}} = \mathbf{Z}\mathbf{Z}^H/M$. Note that (24) holds for the values of $l = 1, \ldots, L$ and results in the following:

$$\begin{bmatrix}1 & \beta & \ldots & 0 \\ \beta & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \beta & 1\end{bmatrix}\begin{bmatrix}u_1 \\ u_2 \\ \vdots \\ u_L\end{bmatrix} = \mathbf{0} \tag{25}$$

where $u_l = \mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}(\mathbf{S}_{\mathbf{z}} - \mathbf{\Sigma}_{\mathbf{z}}){\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l$. By solving (25), we obtain the following equality

$$\mathbf{a}_l^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}(\mathbf{S}_{\mathbf{z}} - \mathbf{\Sigma}_{\mathbf{z}}){\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{a}_l = 0. \tag{26}$$

Jaffer [11] shows that for a full $\mathbf{\Gamma}$ matrix, the derivative in (20) is given by

$$\frac{\partial \log p(\mathbf{Z};\mathbf{\Gamma},\sigma^2)}{\partial\mathbf{\Gamma}} = \mathbf{A}_{\mathcal{K}}^H{\mathbf{\Sigma}_{\mathbf{z}}}^{-1}(\mathbf{S}_{\mathbf{z}} - \mathbf{\Sigma}_{\mathbf{z}}){\mathbf{\Sigma}_{\mathbf{z}}}^{-1}\mathbf{A}_{\mathcal{K}} = 0. \tag{27}$$

and using the matrix inversion lemma, it is equivalent to the following condition:

$$\mathbf{A}_{\mathcal{K}}^H(\mathbf{S}_{\mathbf{z}} - \mathbf{\Sigma}_{\mathbf{z}})\mathbf{A}_{\mathcal{K}} = \mathbf{0}. \tag{28}$$

In [8], Jaffer's condition is assumed to be correct, even though $\mathbf{\Gamma}$ is a diagonal matrix. On the other hand, when $\mathbf{\Gamma}$ is diagonal,

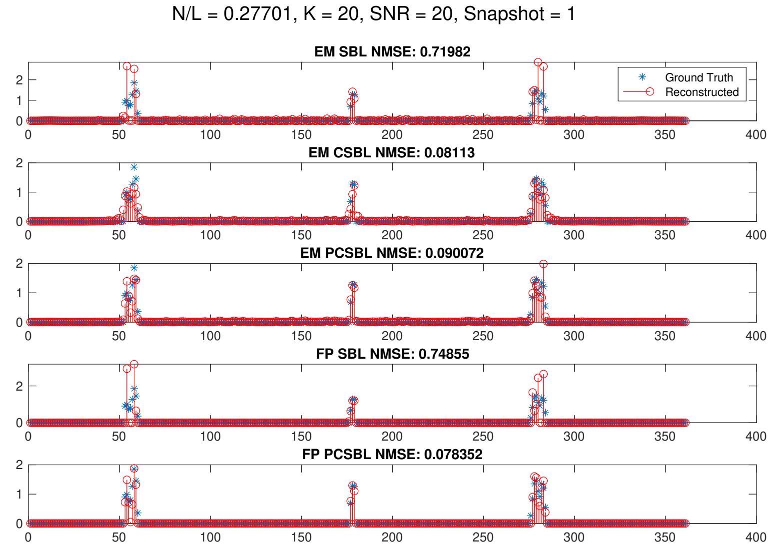



Fig. 1: The reconstructions and NMSE performance of the sparse Bayesian algorithms for a single snapshot with correlated data under 20 dB SNR

we cannot guarantee that the following is always true

$$\mathbf{a}_l^H \mathbf{\Sigma_z}^{-1}(\mathbf{S_z} - \mathbf{\Sigma_z})\mathbf{\Sigma_z}^{-1}\mathbf{a}_{l'} = 0 \text{ for } l \neq l', \tag{29}$$

as this comes from the derivative $\frac{\partial \log p(\mathbf{Z};\mathbf{\Gamma},\sigma^2)}{\partial \mathbf{\Gamma}_{ll'}}$ for the off-diagonal elements of $\mathbf{\Gamma}$. However, we do not have any off-diagonal terms in $\mathbf{\Gamma}$. Therefore, Jaffer's condition might not hold for diagonal matrices. Unlike [8], instead of using Jaffer's condition, we estimate $\sigma^2$ by using the approximation $\text{tr}(\mathbf{S_z}) \approx \text{tr}(\mathbf{\Sigma_z})$. Then, by using (23) we attain

$$\begin{aligned} &\epsilon = \text{tr}(\mathbf{S_z} - \mathbf{\Sigma_z}) = \text{tr}(\mathbf{S_z} - \sigma^2\mathbf{I}_N - \mathbf{A}_\mathcal{K}\mathbf{\Gamma}_\mathcal{K}\mathbf{A}_\mathcal{K}^H) \Rightarrow \\ &\text{tr}(\mathbf{S_z}) - \text{tr}(\sigma^2\mathbf{I}_N) - \epsilon = \text{tr}(\mathbf{A}_\mathcal{K}\mathbf{\Gamma}_\mathcal{K}\mathbf{A}_\mathcal{K}^H) = \\ &\text{tr}(\mathbf{P}\mathbf{A}_\mathcal{K}\mathbf{\Gamma}_\mathcal{K}\mathbf{A}_\mathcal{K}^H\mathbf{P}) = \text{tr}(\mathbf{P}(\mathbf{\Sigma_z} - \sigma^2\mathbf{I}_N)\mathbf{P}) \end{aligned} \tag{30}$$

where $\mathbf{P}$ is the projection matrix onto the subspace spanned by the active components and is written by

$$\mathbf{P} = \mathbf{A}_\mathcal{K}\mathbf{A}_\mathcal{K}^\dagger = \mathbf{A}_\mathcal{K}(\mathbf{A}_\mathcal{K}^H\mathbf{A}_\mathcal{K})^{-1}\mathbf{A}_\mathcal{K}^H = \mathbf{P}^H = \mathbf{P}\mathbf{P}. \tag{31}$$

Thereafter we obtain

$$\text{tr}(\mathbf{S_z}) - \sigma^2\text{tr}(\mathbf{I}_N) - \epsilon = \text{tr}(\mathbf{P}\mathbf{\Sigma_z}\mathbf{P}) - \sigma^2\text{tr}(\mathbf{P}\mathbf{P}). \tag{32}$$

Then, the trace in (32) is evaluated and it leads to $\text{tr}(\mathbf{P}\mathbf{P}) = \text{tr}(\mathbf{P}^H\mathbf{P}) = K$, $\text{tr}(\mathbf{I}_N) = N$ and $\text{tr}(\mathbf{P}\mathbf{\Sigma_z}\mathbf{P}) = \text{tr}(\mathbf{P}^2\mathbf{\Sigma_z}) = \text{tr}(\mathbf{P}\mathbf{\Sigma_z})$. That gives

$$\text{tr}(\mathbf{S_z}) - N\sigma^2 - \epsilon = \text{tr}(\mathbf{P}\mathbf{\Sigma_z}) - K\sigma^2. \tag{33}$$

Inserting $\theta = \text{tr}(\mathbf{P}(\mathbf{S_z} - \mathbf{\Sigma_z}))$ and solving (33) for $\sigma^2$ results in

$$\sigma^2 = \frac{\text{tr}(\mathbf{S_z} - \mathbf{P}\mathbf{S_z}) + \theta - \epsilon}{N - K} \approx \frac{\text{tr}(\mathbf{S_z} - \mathbf{P}\mathbf{S_z})}{N - K} = \hat{\sigma}^2, \tag{34}$$

which is the same variance estimator as in [8], although it is derived differently. This approximation makes the noise power estimation error-free if $\text{tr}(\mathbf{\Sigma_z}) = \text{tr}(\mathbf{S_z})$ and $\text{tr}(\mathbf{P}\mathbf{\Sigma_z}) = \text{tr}(\mathbf{P}\mathbf{S_z})$ or $\text{tr}(\mathbf{P}(\mathbf{S_z} - \mathbf{\Sigma_z})) = \text{tr}(\mathbf{S_z} - \mathbf{\Sigma_z})$, unbiased as $E[\epsilon] = 0$ and $E[\theta] = 0$ , consistent since its variance tends to zero for $M \to \infty$ and asymptotically efficient as it approaches the Cramér-Rao lower bound (CRLB) as $M \to \infty$.

## IV. Numerical Results

The proposed algorithm is tested on a DOA and amplitude estimation problem with block sparse sources. It should be noted that most analysis in the literature has been done with randomly designed sensing matrices [4], [12]. However, such a random design is not realistic to evaluate the performance of our algorithm. Hence, we tested and compared different algorithms for a simple DOA and amplitude estimation problem. The performance of the proposed algorithm is close to both PCSBL and CBSL; however, it is faster than both algorithms. Here, we consider an array with various numbers of elements and various numbers of snapshots and SNR values. The DOAs are on an angular grid $[-90 : 0.5 : 90]^\circ$, and $L = 361$. The noise is modeled as i.i.d. complex Gaussian. The single snapshot array signal-to-noise ratio (SNR) is $\text{SNR} = 10\log_{10}[E[||\mathbf{A}\mathbf{s}_m||_2^2]/E[||\mathbf{n}_m||_2^2]]$. For $M$ snapshots the noise power becomes

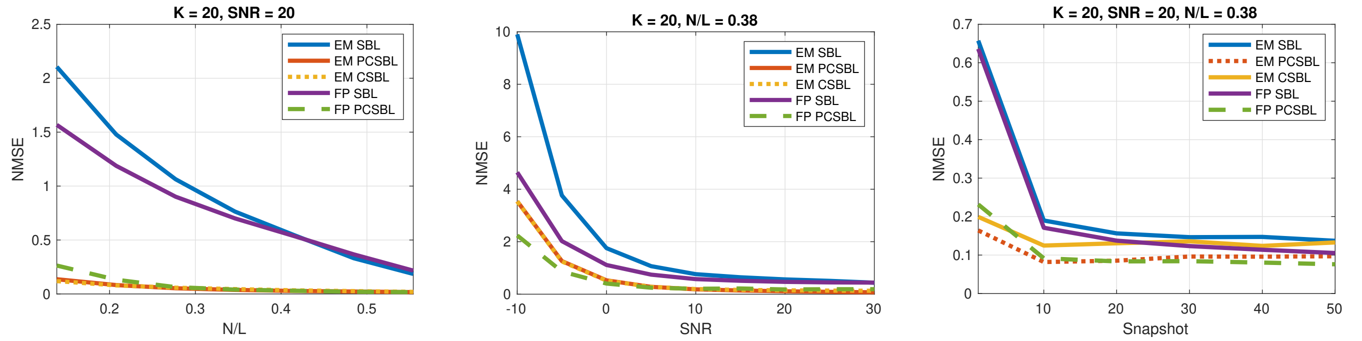
$$E[||\mathbf{N}||_\mathcal{F}^2]/(MN) = 10^{-SNR/10}[E[||\mathbf{A}\mathbf{S}||_2^2]/(MN). \tag{35}$$

Here, we examine a scenario with $K = 20$ random sources at random three DOA groups with $s_{l,m}$ having random complex amplitudes. The sources are chosen to be correlated to see the robustness of the algorithms with such a setting. The correlated sources are created as $\mathbf{S} = \mathbf{R}^{1/2}\mathbf{W}$ where $\mathbf{W}$ is complex random noise with unit variance. We choose

$$\mathbf{R} = \begin{bmatrix} 1 & a & \dots & 0 \\ a & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{bmatrix} \tag{36}$$

where $a = 0.5$. $\mathbf{R}$ is chosen to be a tridiagonal correlation matrix; hence, only the neighboring correlations are considered.

The reconstructed amplitudes and their NMSE performance with correlated data for a single snapshot problem with




(a) (b) (c)

Fig. 2: Performance comparison of the sparse Bayesian algorithms for different (a) $N/L$ ratios; (b) SNR values for a single snapshot problem and (c) different number of snapshots

$N/L = 0.27$ and $SNR = 20$ dB are given in Fig. 1. FP SBL represents a sparse Bayesian learning algorithm with a Type II likelihood maximization with a fixed point update rule [8], [9]. FP PCSBL is the pattern-coupled version of that algorithm, which is proposed in this work. We compare our algorithm with EM update-based SBL algorithms which are classical SBL (EM SBL), PCSBL (EM PCSBL), and CSBL (EM CSBL). Even though we consider correlated data, FP PCSBL provides a considerable improvement compared to regular FP SBL. Note that FP PCSBL exploits the statistical dependence of the sparsity patterns of the neighboring coefficients when the data is uncorrelated. However, it still provides a huge improvement for correlated data. Similarly, EM PCSBL and EM CSBL significantly improve EM SBL. On the other hand, EM PCSBL, EM CSBL, and FP PCSBL have similar performances. However, while overall EM PCSBL and EM CSBL algorithms take around $0.5 - 1.5$ seconds with a Macbook Pro 2019 (16 GB of RAM and 6-core Intel Core i7 2.6 GHz processor), FP PCSBL converges in $0.03 - 0.05$ seconds for a single snapshot problem thanks to the fast convergence of the fixed point update rule.

The NMSE performance of the reconstructions with correlated data for a single snapshot problem with different $N/L$, SNR, and snapshot values are given in Fig. 2a,2b, and 2c, respectively. All results are averaged $100$ Monte Carlo simulations. While FP SBL performs better than EM SBL for different $N/L$ and SNR values, we observe similar performances for FP PCSBL, EM PCSBL, and EM CSBL. For small SNR values, FP PCSBL has slightly better performance than its counterparts, but we observe the opposite for small $N/L$ values. Lastly, the performance of all the algorithms is increased by increasing the number of snapshots to some extent. However, in the multi-snapshot problem, with an increasing number of snapshots, we do not see a considerable benefit of using pattern coupling.

The algorithm is also tested with different values of $a$, (i.e. $0 \leq a \leq 0.5$) and we observe similar results as in Figs. 1, 2a, 2b and 2c, including for the case of uncorrelated data (i.e. $a = 0$). Hence, the correlations in the data for given values of $a$ do not have a significant impact on the reconstruction performance.

## V. Conclusions and Discussion

In this work, we proposed a pattern coupling algorithm with fixed point iterations based on the update rule for the hyperparameters instead of using the EM algorithm. It does not require further assumptions on the hyperparameters and provides a simple update rule. The proposed algorithm is tested on a simple DOA and amplitude estimation problem, and its performance is close to both PCSBL and CBSL; however, it is faster than both counterparts thanks to the fixed point iterations.

## References

[1] L. Wang, L. Zhao, G. Bi, C. Wan, and L. Yang, “Enhanced ISAR imaging by exploiting the continuity of the target scene,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5736–5750, 2014.

[2] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, “Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 300–309, 2013.

[3] H. Huang, M. Fauß, and A. M. Zoubir, “Block sparsity-based DOA estimation with sensor gain and phase uncertainties,” in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[4] J. Fang, Y. Shen, H. Li, and P. Wang, “Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals,” *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 360–372, 2015.

[5] H. Duan, L. Zhang, J. Fang, L. Huang, and H. Li, “Pattern-coupled sparse Bayesian learning for inverse synthetic aperture radar imaging,” *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1995–1999, 2015.

[6] L. Wang, L. Zhao, S. Rahardja, and G. Bi, “Alternative to extended block sparse Bayesian learning and its relation to pattern-coupled sparse Bayesian learning,” *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2759–2771, 2018.

[7] D. P. Wipf and B. D. Rao, “An empirical Bayesian strategy for solving the simultaneous sparse approximation problem,” *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[8] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, “Multisnapshot sparse Bayesian learning for DOA,” *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1469–1473, 2016.

[9] C. F. Mecklenbriiuker, P. Gerstoft, and G. Leus, “Sparse Bayesian learning for DOA estimation of correlated sources,” in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2018, pp. 533–537.

[10] D. Malioutov, M. Cetin, and A. 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.

[11] A. Jaffer, “Maximum likelihood direction finding of stochastic sources: a separable solution,” in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*. Los Alamitos, CA, USA: IEEE Computer Society, apr 1988, pp. 2893,2894,2895,2896.

[12] Z. Zhang and B. D. Rao, “Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2009–2015, 2013.