Low Cost Subspace Tracking Algorithms for Sparse Systems

Nacerredine Lassami[1], Karim Abed-Meraim[2] and Abdeldjalil Aïssa-El-Bey[1]

[1] IMT Atlantique, UMR CNRS 6285 Lab-STICC, Université Bretagne Loire
29238 Brest, France
[2] PRISME Laboratory, University of Orleans
12 Rue de Blois, 45067 Orleans, France

ABSTRACT

In this paper, we focus on tracking the signal subspace under a sparsity constraint. More specifically, we propose a two-step approach to solve the considered problem whether the sparsity constraint is on the system weight matrix or on the source signals. The first step uses the OPAST algorithm for an adaptive extraction of an orthonormal basis of the principal subspace, then an estimation of the desired weight matrix is done in the second step, taking into account the sparsity constraint. The resulting algorithms: SS-OPAST and DS-OPAST have low computational complexity (suitable in the adaptive context) and they achieve both good convergence and estimation performance as illustrated by our simulation experiments for different application scenarios.

Index Terms— Principal subspace tracking, sparse subspace, adaptive estimation, sparse source separation.

1. INTRODUCTION

Subspace techniques play a fundamental role in statistical estimation and array signal processing. Indeed, these techniques are widely used in applications such as data compression, system identification, filtering and parameter estimation.

Many algorithms have been proposed for tracking the principal subspace in the literature. In particular, the Oja method [1] has been given a special consideration and used in some real life applications [2]. It has been established in [3] that the Oja method can be seen as an approximate gradient technique for the minimization of a squared error (MSE) function. Following this approximate gradient method, other faster methods have been developed. Among the most robust and efficient methods, we find the PAST (Projection Approximation Subspace Tracking) method proposed in [3] and its orthogonal variant OPAST [4] well known for their linear complexity.

Most recently, some applications needed to solve the principal subspace problem but under a sparsity constraint. This sparsity can be over the data as for example in the blind separation of sparse sources [5] or over the subspace weight matrix as in the case of sparse principal component anal-

ysis [6]. Surprisingly, even though several solutions have been introduced for batch sparse subspace estimation, little has been done so far for the adaptive scheme. Among the existing adaptive solutions one can find the application proposed in [7] is the STAP (space-time adaptive processing) for Airborne Phased Array Radar, where they developed an algorithm referred to as ℓ_1-PAST which solves the principal subspace tracking problem under a sparsity constraint on the weight matrix.

In this paper, we are interested in solving the same problem, however we drop the orthogonality constraint. In fact, we figured out that looking for the best orthogonal subspace and the sparsest one at the same time is not that optimal and can lead to sub-optimal solutions. We propose at first a two-stage algorithm which gives the same subspace performance as OPAST while providing the sparsest weight matrix generating the principal subspace. For sparse source signals, we propose a second algorithm that searches in an adaptive scheme for the principal subspace weight matrix which leads to the reconstruction of source signals with maximum sparsity. The latter algorithm is illustrated in the context of adaptive blind separation of sparse sources.

2. PROBLEM FORMULATION

Let $\underline{x}(t) \in \mathbb{R}^n$ be a random data vector observed at the tth snapshot over n sensors $\underline{x}(t) = [x_1(t), ..., x_n(t)]^T$ and let's denote by $C_{xx} = E[\underline{x}\,\underline{x}^T]$ its covariance matrix. We assume that the sorted eigenvalues of C_{xx} satisfy the conditions:

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p > \lambda_{p+1} \geq ... \geq \lambda_n \tag{1}$$

p being the rank of the desired principal (signal) subspace of C_{xx}. In the adaptive scheme, the standard updating form of the covariance matrix using an exponential window is:

$$C_{xx}(t) = \sum_{i=1}^{t} \beta^{t-i} \underline{x}(i)\underline{x}(i)^T = \beta C_{xx}(t-1) + \underline{x}(t)\underline{x}(t)^T \tag{2}$$

where $0 < \beta < 1$ is the forgetting factor used to afford the tracking capability when the system operates in a nonstationary environment. We seek to resolve the problem of tracking

the signal subspace of dimension $p < n$ with a sparsity constraint over the weight matrix $W(t) \in R^{n \times p}$ or eventually a sparsity constraint on the source signals (output of the filtering matrix $W(t)$), i.e. $\underline{y}(t) = W^T(t)\underline{x}(t)$.

For this end, we propose a two-step approach where the first is for the adaptive extraction of an orthonormal basis of the principal subspace of $C_{xx}(t)$. The second step is dedicated to the estimation of the weight matrix with the desired sparsity of the system or the data matrix according to the considered application assumption.

In the first stage, the objective function[1] to minimize is:

$$J_{PS}(W(t)) = \sum_{i=1}^{t} \beta^{t-i} \|\underline{x}(i) - W(t)W^T(t)\underline{x}(i)\|_2^2 \quad (3)$$

Under the unitary constraint of the weight matrix, i.e. $W^T(t)W(t) = I_p$. The previous term represents the mean square error of the projected data vector $(I_n - W(t)W^T(t))\underline{x}(t)$. The optimization of (3) would be achieved by the OPAST algorithm as shown next, and hence we refer to the extracted weight matrix by $W_{opast}(t)$.

For the second stage, we seek for the desired weight matrix in the form $W(t) = W_{opast}(t)Q(t)$ where the $p \times p$ nonsingular matrix Q is introduced in order to optimize the chosen sparsity criterion. In the case the sparsity objective is for the system matrix, Q is computed in such a way we minimize the cost function:

$$\arg \min_{Q(t)} \|W(t)\|_1 = \arg \min_{Q(t)} \|W_{opast}(t)Q(t)\|_1$$

where $\|W(t)\|_1 \triangleq \sum_{i=1}^{n} \sum_{j=1}^{p} |W_{ij}(t)|$. Note that the ℓ_0 norm is more appropriate to represent the sparsity of $W(t)$ but that will make the objective function non convex and hard to optimize. Therefore, we use the relaxation to ℓ_1 norm which is one of the best convex approximation of ℓ_0 norm. We refer to the previous cost function to as $J_{SS}(Q)$ where SS stands for System matrix Sparsity.

• **Remark:** Note that the orthogonality is not necessarily aligned with the sparsity constraint of the weight matrix for which reason we have considered a two-step approach with two different weight matrices W_{opast} and W.

Now, for sparse source signals, the objective function minimization becomes:

$$\arg \min_{Q(t)} \sum_{i=1}^{t} \beta^{t-i} \|Q(t)^T W_{opast}(t)^T \underline{x}(i)\|_1$$

where matrix $W(t)$ is a separation (beamforming) matrix which extracts the sparse source signals. The latter cost function is referred to as $J_{DS}(Q)$ where DS stands for Data Sparsity.

---

[1]The index 'PS' stands for Principal Subspace.

Next we introduce two algorithms dedicated to the two previously mentioned problems where the minimization of J_{SS} and J_{DS} is obtained via a natural gradient approach.

3. SUBSPACE TRACKING ALGORITHMS

In this section, we introduce the different algorithms used for tracking the principal subspace with sparsity constraint. At first, the OPAST algorithm is briefly revisited together with the projection approximation technique.

3.1. OPAST algorithm

Let's consider the function :

$$J_{PS}(W(t)) = \sum_{i=1}^{t} \beta^{t-i} \|\underline{x}(i) - W(t)W^T(t)\underline{x}(i)\|_2^2 \quad (4)$$

It is shown in [3] that:

- W is a stationary point of $J_{PS}(W)$ if and only if $W = V_p Q$, where V_p is an $n \times p$ matrix containing p distinct eigenvectors of C_{xx}, and Q is any $p \times p$ unitary matrix.
- Every stationary point of $J_{PS}(W)$ is a saddle point, except when V_p contains the p dominant eigenvectors of C_{xx}. In this case $J_{PS}(W)$ attains the global minimum.

An iterative optimisation of (4) leads to the solution

$$W(t) = C_{xx}(t)W(t-1)(W^T(t-1)C_{xx}(t)W(t-1))^{-1} \quad (5)$$

PAST algorithm uses (5) and the projection approximation

$$C_{xx}(t)W(t) \approx C_{xx}(t)W(t-1) \quad (6)$$

to solve iteratively (4) with a linear complexity $3np + O(p^2)$.

Algorithm 1 OPAST

1: $\underline{y} = W^T(t-1)\underline{x}(t)$
2: $\underline{q} = \frac{1}{\beta}Z(t-1)\underline{y}$
3: $\gamma = \frac{1}{1+\underline{y}^T\underline{q}}$ and $\tau = \frac{1}{\|\underline{q}\|^2}\left(\frac{1}{\sqrt{1+\|\underline{q}\|^2\gamma^2(\|\underline{x}(t)\|^2-\|\underline{y}\|^2)}} - 1\right)$
4: $\underline{e} = W(t-1)(\tau\underline{q} - \gamma(1+\tau\|\underline{q}\|^2)\underline{y}) + (1+\tau\|\underline{q}\|^2)\gamma\underline{x}(t))$
5: $Z(t) = \frac{1}{\beta}Z(t-1) - \gamma\underline{q}\underline{q}^T$
6: $W(t) = W(t-1) + \underline{e}\underline{q}^T$

The OPAST algorithm is based on the same principle as PAST followed by a fast orthonormalization of $W(t)$ (see [4] and the above algorithm's summary for details).

3.2. Sparsity constraint on the system matrix

We look now for the full rank matrix $Q(t) \in \mathbb{R}^{p \times p}$ that minimizes $\|W_{opast}(t)Q(t)\|_1$ where $Q(t)$ is the matrix that transforms the orthonormal principal subspace basis given by OPAST into a sparse one.

The optimization of the ℓ_1-norm criterion is achieved here by using a natural gradient approach. Hence, we seek the updated matrix $Q(t)$ in the form $Q(t) = Q(t-1)(I+\epsilon)$ where matrix $\epsilon \in \mathbb{R}^{p \times p}$ has small valued entries that can be computed using a first order approximation according to:

$$\hat{\epsilon} = \arg\min_{\epsilon} \|W_{opast}(t)Q(t-1) + W_{opast}(t)Q(t-1)\epsilon\|_1 \quad (7)$$

Let's define $M = W_{opast}(t)Q(t-1)$. Now we can rewrite (7) as

$$\hat{\epsilon} = \arg\min_{\epsilon} \sum_{i=1}^{n}\sum_{j=1}^{p}\left|M_{ij} + \sum_{k=1}^{p} M_{ik}\epsilon_{kj}\right| \quad (8)$$

Under the assumption $z << |x|$, one can approximate $|x+z|$ by the value $|x| + \text{sign}(x)z$. Using this approximation in (8) leads to:

$$\hat{\epsilon} \approx \arg\min_{\epsilon} \sum_{i=1}^{n}\sum_{j=1}^{p}|M_{ij}| + \text{sign}(M_{ij})\sum_{k=1}^{p} M_{ik}\epsilon_{kj}$$

$$\approx \arg\min_{\epsilon} \|M\|_1 + Tr(\epsilon R^T) \quad (9)$$

with $R = M^T\text{sign}(M)$ where $\text{sign}(M)$ refers to the matrix with $ij-th$ entry equal to $\text{sign}(M_{ij})$.

Algorithm 2 SS-OPAST

1: Compute $W_{opast}(t)$ using OPAST in [4]
2: $M = W_{opast}(t)Q(t-1)$
3: $R = M^T\text{sign}(M)$ then $R = \frac{R}{\|R\|^2}$
4: $Q(t) = Q(t-1)(I_p - \mu R)$ then normalize its columns
5: $W(t) = W_{opast}(t)Q(t)$.

So by choosing $\hat{\epsilon} = -\mu\frac{R}{\|R\|^2}$ with $\mu > 0$, we ensure the local decrease of the cost function according to $\|M\|_1 - \mu < \|M\|_1$. The full algorithm is summarized in **Algorithm 2**. After every iteration, we need to normalize the columns of $Q(t)$ to better control the conditioning of matrix $W(t)$.

3.3. Sparsity constraint on the sources

Similarly to the previous subsection, we seek for the full rank matrix $Q(t) \in \mathbb{R}^{p \times p}$ that minimizes $\|Q^T(t)W_{opast}^T(t)X\|_1$ where $X = [\beta^{t-1}\underline{x}(1), \beta^{t-2}\underline{x}(2)\ldots\underline{x}(t)]$ is the exponentially windowed data matrix. We use the same assumption about $Q(t)$ and $\epsilon(t)$ to reformulate the problem as:

$$\hat{\epsilon} = \arg\min_{\epsilon} \|Q^T(t)W_{opast}^T(t)X\|_1$$

$$= \arg\min_{\epsilon} \|(\epsilon^T + I)Q^T(t-1)W_{opast}^T(t)X\|_1$$

$$= \arg\min_{\epsilon} \|\epsilon^T Y + Y\|_1 \quad (10)$$

With $Y = Q^T(t-1)W_{opast}^T(t)X$. Using the same development as in equations (9) and (10), we get:

$$\hat{\epsilon} = \arg\min_{\epsilon} \sum_{i=1}^{p}\sum_{j=1}^{t}|Y_{ij}| + \sum_{i=1}^{p}\sum_{k=1}^{p}\epsilon_{ki}(Y\text{sign}(Y^T))_{ki}$$

$$= \arg\min_{\epsilon} \|Y\|_1 + Tr(\epsilon R^T) \quad (11)$$

With $R = Y\text{sign}(Y^T)$. By choosing $\hat{\epsilon} = -\mu Y\text{sign}(Y^T)$ and adjusting μ we ensure the decreasing of ℓ_1 norm. However, since the dimension of Y grows linearly with time, a direct computation of matrix R would be prohibitive. To reduce the cost, we use the projection approximation as shown below:

$$R(t) = Q(t-1)^T W(t)^T[\beta X_{t-1}, \underline{x}_t]$$
$$\times \text{sign}([\beta X_{t-1}, \underline{x}_t]^T W(t)Q(t-1))$$
$$= \beta Q(t-1)^T W(t)^T X_{t-1}\text{sign}(\beta X_{t-1}^T W(t)Q(t-1))$$
$$+ Q(t-1)^T W(t)^T \underline{x}_t\text{sign}(\underline{x}_t^T W(t)Q(t-1)) \quad (12)$$

Under the assumption $W(t) \approx W(t-1)$, we can write that $R(t) \approx \beta R(t-1) + \underline{y}_t\text{sign}(\underline{y}_t^T)$ with $\underline{y}_t = Q(t-1)^T W(t)^T \underline{x}_t$. The full algorithm is summarized in **Algorithm 3**.

Algorithm 3 DS-OPAST

1: Compute $W_{opast}(t)$ using OPAST in [4]
2: $\underline{y}_t = Q(t-1)^T W_{opast}^T(t)\underline{x}_t$
3: $R(t) = \beta R(t-1) + \underline{y}_t\text{sign}(\underline{y}_t^T)$ then $R(t) = \frac{R(t)}{\|R(t)\|^2}$
4: $Q(t) = Q(t-1)(I_p - \mu R(t))$
5: Normalize the columns of $Q(t)$

• **Remark:** Note that the computational cost of DS-OPAST is of order $O(np)$ per iteration. Now, if the update of the weight matrix $W(t) = W_{opast}(t)Q(t)$ is needed, the latter cost would be of order $O(np^2)$ due to the previous matrix product. However, by using the updating equations in steps 3 and 4 of **Algorithm 3** together with the projection approximation $Q(t-1) \approx Q(t-2)$ one can avoid the matrix - matrix product and preserve the linear complexity $O(np)$.

4. SIMULATION RESULTS AND DISCUSSION

4.1. Sparsity constraint on the system matrix

We present here some numerical simulations to assess the performance of the proposed algorithm and we use the OPAST [4] and ℓ_1-PAST [7] algorithms for comparison. At first, we consider the data model $X = AS + N$ with $A \in \mathbb{R}^{n \times p}$ a random sparse mixing matrix (we use the function SPRANDN in MATLAB for its generation), $S \in \mathbb{R}^{p \times T}$ the source signals generated according to a Gaussian distribution, with zero mean and unit variance (T being the total number of samples) and $N \in \mathbb{R}^{n \times T}$ is a white Gaussian noise. The used performance factors are the norm ℓ_1 of the weight matrix for the

sparsity measure (with columns normalized to 1) and for the subspace performance we use a normalized quadratic error given by :

$$\rho(t) = \frac{1}{r} \sum_{i=1}^{r} \frac{trace(W_i^{\#}(t)(I_n - W_{ex}W_{ex}^T)W_i(t))}{trace(W_i^{\#}(t)W_{ex}W_{ex}^T W_i(t))} \quad (13)$$

where $r = 150$ is the number of Monte Carlo runs, $W_i(t)$ is the matrix of desired subspace at experience i and iteration t ($W_i^{\#}(t)$ represents its pseudo inverse) and W_{ex} is the exact orthogonal subspace computed from the exact matrix A. We experimented our algorithm for different system sizes and present here two cases: $(n = 16, p = 9)$ and $(n = 100, p = 30)$.
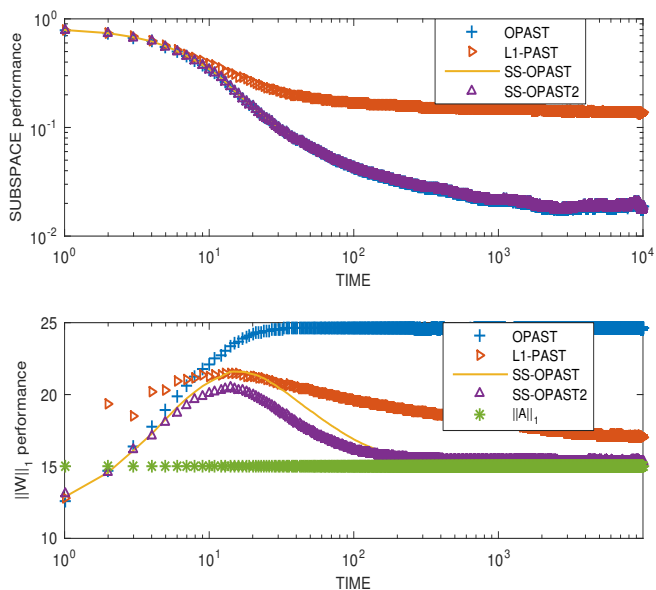

Fig. 1: Subspace performance $\rho(t)$ and norm ℓ_1 of W versus time with: $\mu = 1$ for SS-OPAST, μ is an adaptive vector for SS-OPAST2 and $(n = 16, p = 9, SNR = 15dB)$

Figure 1 illustrates the improved performance of the SS-OPAST as compared to ℓ_1-PAST and OPAST for SNR=15dB. Note that at the beginning of the adaptive process, the ℓ_1 norm of W is not decreasing due to the inappropriate first order approximation at this early stage (the decrease starts only after few tens of iterations). An adaptive step might be considered to force the decrease of the cost function.

For figure 2, we choose larger system dimensions ($n = 100, p = 30$) in which case the previous observations are confirmed. Based on these simulations, we note some general remarks

• **Choice of** μ: if we choose $\mu = 0$, we get the same result as OPAST without sparsity, and if μ is too large we get better sparsity but we can lose the constraint on the solution rank (i.e. $rank(W) = p$) since the weight matrix becomes close to singular when we run the algorithm for a long period.
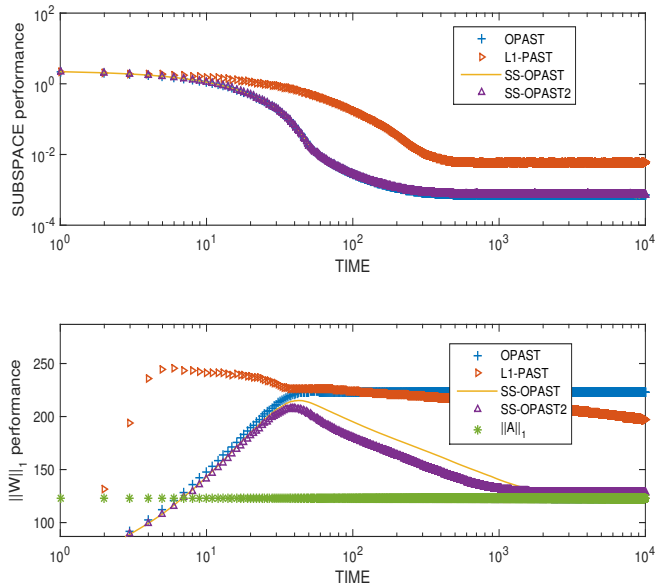

Fig. 2: Subspace performance $\rho(t)$ and norm ℓ_1 of W versus time with: $\mu = 3$ for SS-OPAST, μ is an adaptive vector for SS-OPAST2 and $(n = 100, p = 30, SNR = 15dB)$

We thought about an adaptive solution to choose μ and we figured out that by taking $\mu_j = \sum_{i=1}^{n} |M_{ij}|$ and changing the step 4 of **Algorithm 2** by $Q(t) = Q(t - 1)(I_p - R\text{diag}(\mu))$, we improve the convergence performance and the stability of SS-OPAST ($\text{diag}(\mu)$ is a diagonal matrix with the elements of $\mu = [\mu_1, ..., \mu_p]$ on the main diagonal). This is illustrated by the plots referred to as SS-OPAST2 of figures 1 and 2.

• **Changing OPAST:** The algorithm we presented here can be easily modified to fit with other adaptive subspace algorithms such as PAST [3] and FAPI [8] just by changing the first step of our algorithm.

• **Complexity:** Our algorithm is based on OPAST algorithm which costs $3np + O(p^2)$ operations plus the ℓ_1 minimization part which cost $3np^2 + O(p^3)$ so the overall complexity is $3np^2 + O(p^3)$. The ℓ_1-PAST has a complexity of $3np^2 + O(p^2)$ which is slightly less than our algorithm.

4.2. Sparsity constraint on the sources

As an illustrative example, our algorithm is applied in blind sparse source separation (BSSS). We consider the same data model but this time the matrix S of source signals is the sparse one. In the context of BSSS, the performance index is the global rejection level [9] which is defined as: $\mathcal{I}_{pq} \stackrel{def}{=} \sum_{p \neq q} \mathcal{I}_{pq}$ where $\mathcal{I}_{pq}$ measures the ratio of the power of the interference of the qth source to the power of the pth source signal and they are defined as $\mathcal{I}_{pq} = E|(\hat{A}^{\#}A)_{pq}|$ (in our case $\hat{A}^{\#} = W^T$).

We simulated 100 times the data with $p = 2$ sparse sources, $n = 5$ sensors and we choose $\mu = 0.001$. Figure

3 shows an example of the source signals used and how the separated signals fit them well (after adjusting the amplitude and put every output signal with its correspondent source signal to remove the inherent ambiguities of BSSS).
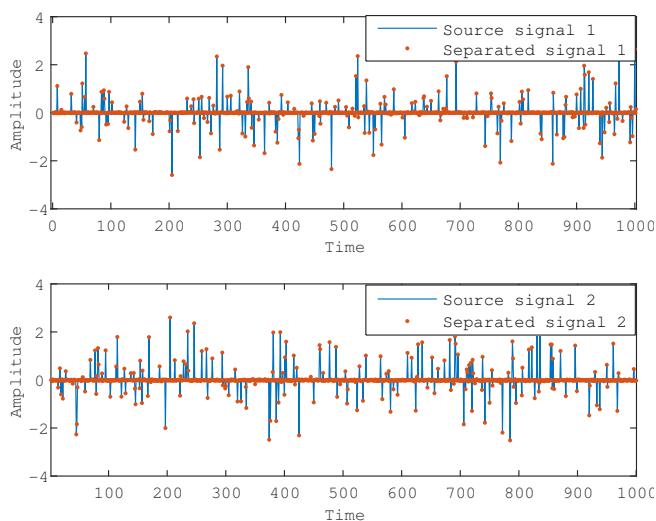

Fig. 3: Source and separated signals ($n = 5, p = 2, SNR = 30dB$)

To illustrate the adaptive separation capability of our algorithm, we present in figure 4 the mean rejection level (dB) in the first place versus time with a fixed SNR (dB) and in the second place versus SNR after $T = 3000$ iterations.
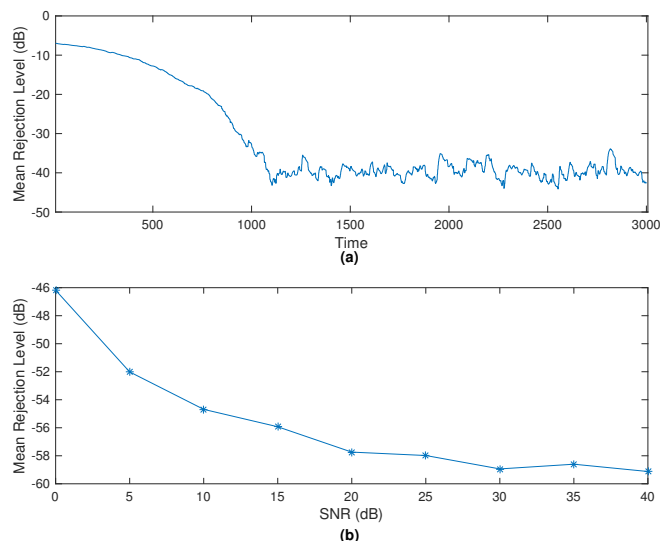

Fig. 4: Mean Rejection Level (dB) versus: (a) time for $SNR = 30dB$, (b) SNR for $T = 3000$ ($n = 5, p = 2$)

It is clear that our algorithm converges to the right solution, however, the convergence time depends on the chosen parameter μ and the size of the system. Finally, note that

one of the strengths of our algorithm is its low computational complexity.

5. CONCLUSION

In this paper, adaptive subspace algorithms for sparse systems have been studied. We have proposed a two-step approach where the first uses the OPAST algorithm to track an orthonormal basis for the principal subspace. The second step seeks to find the weight matrix corresponding to the desired sparsity constraint either on the system weight matrix or on the source signals. Many applications are possible and can be treated by our algorithms which have the advantages of low computational cost and improved performance as compared to the existing solutions.

REFERENCES

[1] E. Oja, "Simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, vol. 15, no. 3, pp. 267–273, 1982.

[2] S. Y. Kung, K. I. Diamantaras, and J. S. Taur, "Adaptive principal component extraction (APEX) and applications," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1202–1217, 1994.

[3] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.

[4] K. Abed-Meraim, A. Chkeif, and Y. Hua, "Fast orthonormal PAST algorithm," *IEEE Signal Processing Letters*, vol. 7, no. 3, pp. 60–62, 2000.

[5] Si-Mohamed Aziz Sbai, Abdeldjalil Aïssa-El-Bey, and Dominique Pastor, "Contribution of statistical tests to sparseness-based blind source separation," *EURASIP J. Adv. Sig. Proc.*, vol. 2012, pp. 169, 2012.

[6] K. Benidis, Y. Sun, P. Babu, and D. P. Palomar, "Orthogonal sparse PCA and covariance estimation via procrustes reformulation," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6211–6226, 2016.

[7] X. Yang, Y. Sun, T. Zeng, T. Long, and T. K. Sarkar, "Fast STAP method based on PAST with sparse constraint for airborne phased array radar," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4550–4561, 2016.

[8] R. Badeau, B. David, and G. Richard, "Fast approximated power iteration subspace tracking," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2931–2941, 2005.

[9] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.