

ROBUST VERSIONS OF THE PAST ALGORITHM

Markus Rupp and Carolina Reyes

Vienna University of Technology, Institute of Telecommunications, Gusshausstr. 25/e389,
1040 Wien, Austria

ABSTRACT

In this contribution we analyze the Projection Approximation Subspace Tracking (PAST) algorithm by a novel analysis method based on Singular Value Decomposition (SVD). Based on the gained new insights we propose several robust algorithmic modifications that allow to guarantee stability of the algorithm run for normalized step-sizes. Furthermore we investigate the tracking behavior for such step-size choices.

Index Terms— PAST algorithm, subspace tracking, blind source separation

1. INTRODUCTION

Since its introduction by Bin Yang in 1993, the Projection Approximation Subspace Tracking (PAST) algorithm [1, 2] and its many derivatives have become quite popular as relatively simple algorithms to detect subspaces, separate them and even track them. The original analysis of the algorithm's behavior [3, 4] was based on an Ordinary Differential Equation (ODE) approach with all its pros and cons. In the ODE framework iterative approaches are interpreted as differential equations of continuous functions and based on some Ljapunov arguments, it can be deduced whether step-sizes exist for which the algorithm converges. With such approach, a relatively large amount of assumptions has to be made in order to make it work. This is one of its larger drawbacks as it often remains unclear whether all such assumptions can be satisfied in practice. The second drawback is that a practical (upper) bound on the step-size can hardly be deduced, leaving it open for experiments which step-sizes are guaranteed to work or not.

The PAST algorithm minimizes the cost function

$$J_o = \min_{\mathbf{W}} E \left[\|\mathbf{x}_k - \mathbf{W}\mathbf{W}^H \mathbf{x}_k\|^2 \right]. \quad (1)$$

Here, $\mathbf{x}_k \in \mathcal{C}^{N \times 1}$ and $\mathbf{W}_k \in \mathcal{C}^{N \times r}$. Starting with initial values $\mathbf{R}_{yy,0} = \epsilon \mathbf{I}$ and $\mathbf{W}_0 \in \mathcal{C}^{N \times r} = [\epsilon \mathbf{I}_r, \mathbf{O}]^T$ a recursive algorithm is applied on a sequence of vectors \mathbf{x}_k in order to minimize the corresponding Least-Squares (LS) cost function

of (1) on continuously incoming observations \mathbf{x}_k :

$$\mathbf{y}_k = \mathbf{W}_{k-1}^H \mathbf{x}_k \quad (2)$$

$$\hat{\mathbf{R}}_{yy,k} = \hat{\mathbf{R}}_{yy,k-1} + \alpha_k \left[\mathbf{y}_k \mathbf{y}_k^H - \hat{\mathbf{R}}_{yy,k-1} \right] \quad (3)$$

$$\mathbf{e}_k = \mathbf{x}_k - \mathbf{W}_{k-1} \mathbf{y}_k \quad (4)$$

$$\mathbf{W}_k = \mathbf{W}_{k-1} + \gamma_k \mathbf{e}_k \mathbf{y}_k^H \hat{\mathbf{R}}_{yy,k}^{-1} \quad (5)$$

In the last equation, the so-called update equation, the inverse of the estimated autocorrelation matrix $\mathbf{R}_{yy,k} = E \{ \mathbf{y}_k \mathbf{y}_k^H \}$ is applied. This is typically achieved by the matrix inversion lemma, saving complexity. Other fast variants are possible to derive and have been proposed [5] but are out of the scope of this paper. We have selected a relatively general form of the algorithm with two free parameters α_k and γ_k . The range of α_k is between zero and one and determines the time horizon over which the averaging takes place to compute the autocorrelation matrix estimate $\hat{\mathbf{R}}_{yy,k}$. A value of α_k close to one gives more emphasis on the recent value of \mathbf{y}_k while a value closer to zero averages over a longer time. While the optimal choice of α_k depends on the tracking problem, the choice of γ_k is much more crucial for the working of the algorithm and thus will be the focus of this article.

In this contribution we present a different analysis approach based on Singular Value Decomposition (SVD) that removes most of the assumptions mentioned in [4] and provides practical step-size bounds for γ_k . It furthermore yields a new insight in the algorithmic behavior, allowing for alterations which further improve its performance. The paper is organized as follows: After this introduction, we analyze the first order moments algorithm in Section 2 and draw first conclusions. We then proceed to investigate step-size bounds in Section 3, which in turn leads to new insight of the algorithm and allows to propose some modifications, guaranteeing robust behavior and practical step-size bounds. In Section 4 we present simulation results to corroborate our findings. Finally, some concluding remarks in Section 5 round up the paper.

2. FIRST ORDER ANALYSIS

Let $\mathbf{x}_k \in \mathcal{C}^{N \times 1}$ and $\mathbf{W}_k \in \mathcal{C}^{N \times r}$, where N denotes the number of observations and $r < N$ the dimension of the subspace. The goal of the algorithm in (2)-(5) is thus to end up with a unitary matrix \mathbf{W}_k , i.e., $\lim_{k \rightarrow \infty} \mathbf{W}_k^H \mathbf{W}_k = \mathbf{I}_r$.

In a first order analysis we study the algorithmic behavior in the mean. For this we assume the sensor data \mathbf{x}_k to be of random nature with an autocorrelation matrix $\mathbf{R}_{xx,k} = E\{\mathbf{x}_k\mathbf{x}_k^H\}$. The task of the matrix \mathbf{W}_k is to extract r individual subspace signals out of \mathbf{x}_k .

Assumptions: We start our analysis with the estimated autocorrelation matrix $\hat{\mathbf{R}}_{yy,k}$. As this matrix is being averaged over time as shown in (3), we assume it to be at least short time ergodic. More specifically we assume that the ensemble average $\mathbf{R}_{yy,k}$ of the given process is determined from a time average $\hat{\mathbf{R}}_{yy,k}$ over past samples. As a consequence we have that $\mathbf{R}_{yy,k} = E\{\hat{\mathbf{R}}_{yy,k}\} = \mathbf{V}_1\mathbf{\Lambda}_{yy,k}\mathbf{V}_1^H$, where $\mathbf{\Lambda}_{yy,k}$ contains the time-variant eigenvalues of $\mathbf{R}_{yy,k}$ while \mathbf{V}_1 remains constant. After some substitutions $\mathbf{R}_{yy,k} = E\{\mathbf{y}_k\mathbf{y}_k^H\} = E\{\mathbf{W}_{k-1}^H\mathbf{x}_k\mathbf{x}_k^H\mathbf{W}_{k-1}\} = E\{\mathbf{W}_{k-1}^H\mathbf{R}_{xx,k}\mathbf{W}_{k-1}\}$ and correspondingly $\mathbf{R}_{yy,k}^{-1} = (E\{\mathbf{W}_{k-1}^H\mathbf{R}_{xx,k}\mathbf{W}_{k-1}\})^{-1}$. We have applied here the Independence Assumption (IA) to treat \mathbf{W}_{k-1} independently of \mathbf{x}_k . The IA is a common tool in adaptive filter theory [6]. Due to the averaging effect we also assume the expectation values to be independent of other terms in \mathbf{x}_k . As $\hat{\mathbf{R}}_{yy,k}$ is a time-averaged version, this will also have an impact on its decomposition: $E\{\hat{\mathbf{R}}_{yy,k}\} = \mathbf{V}_1\mathbf{\Lambda}_{yy,k}\mathbf{V}_1^H = \mathbf{V}_1\bar{\mathbf{\Sigma}}_k\mathbf{\Lambda}_{xx,k}\bar{\mathbf{\Sigma}}_k\mathbf{V}_1^H$. Here we emphasize a subtle but very important difference: the autocorrelation matrix $\hat{\mathbf{R}}_{yy,k}$ presented in (3) is a time averaged matrix whose decomposition leads to likewise temporally-averaged singular values $\bar{\mathbf{\Sigma}}_k$. These singular values should not be confused with the instantaneous singular values $\mathbf{\Sigma}_k$ originating from decomposing $E\{\mathbf{W}_{k-1}\}$.

The mean of update equation (5) can then be written as:

$$\begin{aligned} E\{\mathbf{W}_k\} &= E\{\mathbf{W}_{k-1}\} \\ &+ \gamma_k E\{(\mathbf{I} - \mathbf{W}_{k-1}\mathbf{W}_{k-1}^H)\mathbf{x}_k\mathbf{x}_k^H\mathbf{W}_{k-1}\} \\ &\quad \times E\{\hat{\mathbf{R}}_{yy,k}^{-1}\}. \end{aligned} \quad (6)$$

Applying the expectation only with respect to the random process \mathbf{x}_k and applying the IA we obtain

$$\begin{aligned} E\{\mathbf{W}_k\} &= E\{\mathbf{W}_{k-1}\} + \gamma_k E\{\mathbf{I} - \mathbf{W}_{k-1}\mathbf{W}_{k-1}^H\} \\ &\quad \times \mathbf{R}_{xx,k} E\{\mathbf{W}_{k-1}\} E\{\hat{\mathbf{R}}_{yy,k}^{-1}\}. \end{aligned} \quad (7)$$

Since the autocorrelation matrix $\mathbf{R}_{xx,k} = \mathbf{Q}\mathbf{\Lambda}_{xx,k}\mathbf{Q}^H$, we recognize that a steady-state solution only exists if $E\{\mathbf{W}_k\} = [\mathbf{U}_1, \mathbf{U}_2][\mathbf{\Sigma}_k, \mathbf{O}][\mathbf{V}_1\mathbf{V}_2]^H$ with $\mathbf{U}_1 = \mathbf{Q}$ where we applied a SVD on \mathbf{W} and partitioned it into the significant part $\mathbf{U}_1\mathbf{\Sigma}_k\mathbf{V}_1^H$ and a zero block. With such finding we can now rewrite (7) into its SVD components and recognize that all terms comprise of the same left matrix \mathbf{U}_1 and the right matrix \mathbf{V}_1^H , while all terms in the center become diagonal:

$$\begin{aligned} \mathbf{\Sigma}_k &= \mathbf{\Sigma}_{k-1} + \gamma_k(\mathbf{I} - \mathbf{\Sigma}_{k-1}\mathbf{\Sigma}_{k-1})\mathbf{\Lambda}_{xx,k}\mathbf{\Sigma}_{k-1} \\ &\quad \times (\bar{\mathbf{\Sigma}}_k\mathbf{\Lambda}_{xx,k}\bar{\mathbf{\Sigma}}_k)^{-1}. \end{aligned} \quad (8)$$

As the entire equation has become diagonal, we can now much simpler describe it by its diagonal terms:

$$\sigma_{i,k} = \sigma_{i,k-1} + \gamma_k \frac{(1 - \sigma_{i,k-1}^2)\sigma_{i,k-1}}{\bar{\sigma}_{i,k}^2}; i = 1, \dots, r. \quad (9)$$

Although being a complicated term in $\sigma_{i,k-1}$, we already can recognize here that the actual values $\lambda_{i,k}^{xx}$ in the autocorrelation matrix $\mathbf{\Lambda}_{xx,k}$, namely the power contributions of the observed sensor signals, are irrelevant to the algorithm. Knowing that the steady state solution is given if all signals are perfectly decorrelated, that is when all singular values are one, we aim to have $\sigma_{i,k-1} \rightarrow 1$. We can rewrite this for $i = 1, 2, \dots, r$ as

$$1 - \sigma_{i,k} = (1 - \sigma_{i,k-1}) \left[1 - \gamma_k \frac{(1 + \sigma_{i,k-1})\sigma_{i,k-1}}{\bar{\sigma}_{i,k}^2} \right] \quad (10)$$

As all singular values are positive, we have thus just proven the following theorem.

Theorem 2.1 *The PAST algorithm converges in the mean for a sufficiently small step-size $\gamma_k > 0$, if $\frac{(1 + \sigma_{i,k-1})\sigma_{i,k-1}}{\bar{\sigma}_{i,k}^2}$ is bounded.*

As the term $\frac{(1 + \sigma_{i,k-1})\sigma_{i,k-1}}{\bar{\sigma}_{i,k}^2}$ in (10) can take on arbitrary values in the range $[0, \infty]$, it is difficult to bound the step-size at this point. We recognize that in particular small singular values are decisive. We will return to this problem in Section 3 ahead and treat it in detail.

Note that similar forms of (7) have been analyzed in [7] in the context of blind source separation and [8] as a means to compute robustly matrix inverses. Note further that in his article [2] Bin Yang also proposed a simpler gradient term algorithm, that is simply omitting the matrix inverse of $\hat{\mathbf{R}}_{yy,k}$. The analysis method presented here can be applied to such algorithm, revealing now that the step-size γ_k depends strongly on the eigenvalues $\lambda_{i,k}^{xx}$ of the observation process \mathbf{x}_k , details will be provided in Section 3.

Often the PAST algorithm is being run with a step-size $\gamma_k = 1/k$ or $\gamma_k = 1/[k+1]$. In this case after a few iterations, the step-size satisfies the stability condition and the algorithm converges. The price, however, for this convergence is a lack of tracking capability as now the step-size becomes so small that the algorithm cannot adjust to a new situation any more.

In particular for adaptive filters we are also interested in analyzing the second order moment describing the quantitative behavior of the algorithm. We have investigated the evolution of terms of the form $\mathbf{W}_k^H\mathbf{W}_k$ and found out that the result is very similar to that in (10) for the first order moment. The main difference relies on a term that appears equally for all diagonal entries but does not influence the stability of the system. On the other hand, it does influence the values for the step size γ_k . Due to space constraints, the present work focuses only on the first order moment analysis.

3. ANALYSIS OF STEP-SIZE BOUNDS

Let us return to Equation (10). We learn here that

$$0 < \gamma_k < \frac{2\bar{\sigma}_{i,k}^2}{(1 + \sigma_{i,k-1})\sigma_{i,k-1}}; i = 1, \dots, r, \quad (11)$$

and we assume that it has an upper bound

$$0 < \gamma_k < \frac{2\bar{\sigma}_{\min,k}^2}{(1 + \sigma_{\max,k-1})\sigma_{\max,k-1}}. \quad (12)$$

If we use the relation $0.25 + 2x^2 \geq x + x^2$ for $x \geq 0$ we obtain an even lower bound

$$0 < \gamma_k < \frac{2\bar{\sigma}_{\min,k}^2}{0.25 + 2\sigma_{\max,k-1}^2}. \quad (13)$$

The term $\sigma_{\max,k-1}^2$ is related to $\mathbf{W}_{k-1}^H \mathbf{W}_{k-1}$. If λ_k^{xx} and \mathbf{v}_k are an eigenvalue and eigenvector pair of the matrix $\mathbf{W}_k^H \mathbf{W}_k$, the following condition holds true:

$$|\lambda_k^{xx}|^m \|\mathbf{v}_k\| = \|\lambda_k^{xx,m} \mathbf{v}_k\| \quad (14)$$

$$= \|(\mathbf{W}_k^H \mathbf{W}_k)^m \mathbf{v}_k\| \quad (15)$$

$$\implies |\lambda_k^{xx}| \leq \|(\mathbf{W}_k^H \mathbf{W}_k)^m\|^{1/m}. \quad (16)$$

Thus, for $m = 1$ we obtain the simplified but practically feasible lower bound

$$0 < \gamma_k < \frac{2\bar{\sigma}_{\min,k}^2}{0.25 + 2\|\mathbf{W}_{k-1}^H \mathbf{W}_{k-1}\|} = \gamma_{\max,k} \bar{\sigma}_{\min,k}^2. \quad (17)$$

The smallest singular value $\bar{\sigma}_{\min,k}$ is thus decisive for convergence and a step-size γ_k needs to be selected based on its knowledge. In the classic PAST algorithm such knowledge is not present and thus only very small step-sizes can be selected in the hope to have sufficiently large values of $\bar{\sigma}_{\min,k}$, which in turn results in slow convergence and poor tracking.

As $\bar{\sigma}_{\min,k}$ is typically not available, we may estimate it. Nowadays, low-complexity methods are available to estimate the smallest singular value [9, 10]. Such estimation technique however, can easily lead to too small step-sizes, resulting in a very slow convergence as well as poor tracking. It is thus of further interest to modify the algorithm in such a way that the dependency on the smallest singular value disappears. In order to prevent such undesired behavior of the classic PAST algorithm, not offering a feasible step-size bound, we are proposing to alter update equation (5) into a generic update

$$\mathbf{W}_k = \mathbf{W}_{k-1} + \gamma_k \mathbf{e}_k \mathbf{y}_k^H \mathbf{B}, \quad (18)$$

with the following options:

$$\text{PAST-I: } \mathbf{B} = \hat{\mathbf{R}}_{yy,k}^{-1} \bar{\sigma}_{\min,k}^2 \quad (19)$$

$$\text{PAST-II: } \mathbf{B} = \hat{\mathbf{R}}_{yy,k}^{-1} \mathbf{R}_{\mathbf{W},k} \quad (20)$$

$$\text{PAST-III: } \mathbf{B} = [\hat{\mathbf{R}}_{yy,k} + \beta \mathbf{I}_r]^{-1} \quad (21)$$

$$\text{PAST-IV: } \mathbf{B} = \mathbf{I} \quad (22)$$

for some small but positive $\beta > 0$. Such regularization in PAST-III is not required for computing the inverse of $\hat{\mathbf{R}}_{yy,k}$ but prevents the smallest singular value to have a decisive impact on the stability. PAST-IV is the gradient-type version of the PAST algorithm [2]. In PAST-II an average of $\mathbf{W}_{k-1}^H \mathbf{W}_{k-1}$ is computed by $\mathbf{R}_{\mathbf{W},k} = \mathbf{R}_{\mathbf{W},k-1} + \alpha_k [\mathbf{W}_{k-1}^H \mathbf{W}_{k-1} - \mathbf{R}_{\mathbf{W},k-1}]$ and applied to compensate for the inverse singular values of $\hat{\mathbf{R}}_{yy,k}$, as we would find $\mathbf{R}_{\mathbf{W},k} = \mathbf{U}_1 \bar{\Sigma}_k^2 \mathbf{U}_1^H$. A simpler version of this is PAST-I where we assume knowledge of the smallest singular value only, for example by simple tracking algorithms [10].

Applying the same analysis technique as before, we find

$$0 < \gamma_{\text{I},k} < \gamma_{\max,k} \leq \min_i \frac{2\bar{\sigma}_{i,k}^2}{(1 + \sigma_{i,k-1})\sigma_{i,k-1}\bar{\sigma}_{\min,k}^2} \quad (23)$$

$$0 < \gamma_{\text{II},k} < \gamma_{\max,k} \leq \min_i \frac{2}{(1 + \sigma_{i,k-1})\sigma_{i,k-1}} \quad (24)$$

$$0 < \gamma_{\text{III},k} < \gamma_{\max,k} \delta_{\min} \leq \min_i \frac{2(\bar{\sigma}_{i,k}^2 + \delta)}{(1 + \sigma_{i,k-1})\sigma_{i,k-1}} \quad (25)$$

$$0 < \gamma_{\text{IV},k} < \frac{\gamma_{\max,k}}{\text{tr}[\mathbf{R}_{xx,k}]} \leq \min_i \frac{2}{(1 + \sigma_{i,k-1})\sigma_{i,k-1}\lambda_{i,k}^{xx}} \quad (26)$$

with $\delta = \beta/\lambda_{i,k}^{xx}$ and $\delta_{\min} = \beta/\lambda_{\max,k}^{xx}$. Thus, the knowledge of δ_{\min} is sufficient to provide a conservative step-size bound. Indirectly the choice of β determines now also the convergence speed; larger values typically offering higher speed. Bounds for (24) and (26) can also be derived, following the approach explained at the beginning of this section. If in (26) $\lambda_{i,k}^{xx}$ is not known, it may be feasible to replace it by $\text{tr}(\mathbf{R}_{xx,k})$. Now it only depends on the matrix norm to compute a safe upper bound of γ_k which is a feasible operation. In practice it turns out that a matrix one norm provides tight results. The upper bounds define some time-variant maximal value $\gamma_{\max,k}$ and we select fractions $\alpha \gamma_{\max,k}$, where $\alpha \in [0, 1]$.

4. SIMULATION RESULTS

The following experiments apply the PAST algorithm to a typical beam steering experiment, that is the observation vector \mathbf{x}_k is given by

$$\mathbf{x}_k = \sum_{i=1}^r \mathbf{a}(\omega_i) s_{i,k} + \mathbf{v}_k = \mathbf{A} \mathbf{s}_k + \mathbf{v}_k. \quad (27)$$

Here $s_{i,k}$ is a random process forming the random vector \mathbf{s}_k and \mathbf{v}_k is the observation noise. The steering vectors are of the form $\mathbf{a}(\omega_i) = [1, e^{j\omega_i}, e^{2j\omega_i}, \dots, e^{(n-1)j\omega_i}]^T$. We selected $r = 3$ frequencies $\omega_i = [0.01, 0.03, 0.2]$, the variance of the signal $s_{i,k}$ is a unit norm and the additive noise variance $\sigma_v^2 = 10^{-6}$. Thus, the obtained three eigenvalues of the autocorrelation matrix \mathbf{R}_{xx} are $\{118.8, 57.8, 3.3\}$. We further

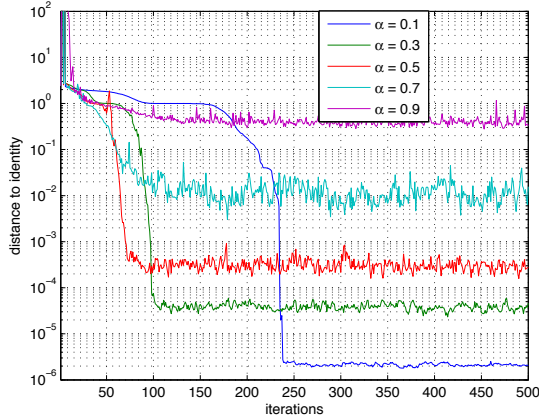


Fig. 1. Classic PAST algorithm under various constant step-sizes.

chose $N = 60$ observations per time instant of such a process, and set $\alpha_k = \alpha$ in (3), $\gamma_k = \alpha$ for the classic PAST algorithm and $\gamma_k = \alpha\gamma_{\max,k}$ for the robust versions. Finally, we averaged over 100 Monte Carlo runs.

4.1. First experiment: classic PAST algorithm

In the first experiment we ran the classic PAST algorithm with a range of constant step-sizes $\gamma_k = \alpha = [0.1, 0.2, 0.5, 0.7, 0.9]$. Figure 1 depicts the distance measure $\sum_{i=1}^r (1 - \sigma_{i,k})^2$. The fastest convergence is obtained for values around $\alpha = 0.5$ and for $\alpha = 0.9$ the algorithm shows first signs of instability. When investigating such smallest singular values we find, however, a very large variation of them from $[10^{-5}, 1]$ as depicted in Figure 2. Using a normalized step-size according to (17) shows that the range of the smallest singular value can be reduced considerably but at the expense of a very slow learning rate.

4.2. Second experiment: robust PAST algorithms

In a second experiment we repeated the first experiment with the robust versions PAST-I to PAST-IV. As expected the (normalized) step-size $\gamma_k = \alpha\gamma_{\max,k}$ now ranges up to (or close to) a maximum value $\alpha = 1$ for all four algorithms. This is shown in Figure 3 on the example of the PAST-II algorithm. Again the fastest convergence is obtained for $\alpha = 0.5$, now even faster than in the first experiment.

In Figure 4 the evolution of the smallest singular value is shown again. It now depicts a substantially smaller range than before, which indicates the more robust behavior of the algorithm. Note that the other variants require additional regularization measures in order to show robust behavior, while the proposed PAST-II algorithm immediately showed such

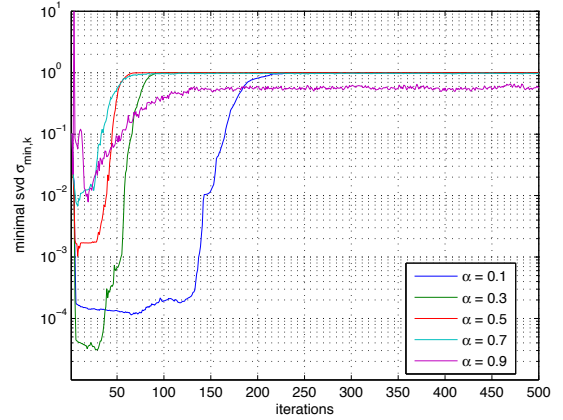


Fig. 2. Evolution of smallest singular value in classic PAST algorithm under constant step-sizes.

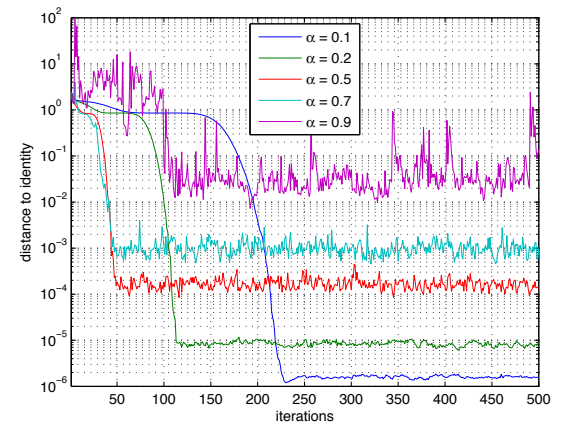


Fig. 3. PAST-II algorithm under various normalized step-sizes.

robustness. We therefore display only the results of PAST-II, as it is the most promising variant.

4.3. Third experiment: tracking behavior

In a third experiment we studied the tracking behavior. To this end, we added to the three frequencies of the previous experiment a fourth frequency that varies over time by $f_4 = f_0 + k\delta_f; k = 0, 1, \dots, k_{\max}$. We selected $\delta_f = 0.001$ such that there is a strong tracking component. The PAST-II algorithm exhibits the best tracking results in Figure 5.

5. CONCLUSIONS

We presented a novel analysis method for the well known PAST algorithm. The method allowed us to derive new properties and even alter the algorithm in such a way that we can

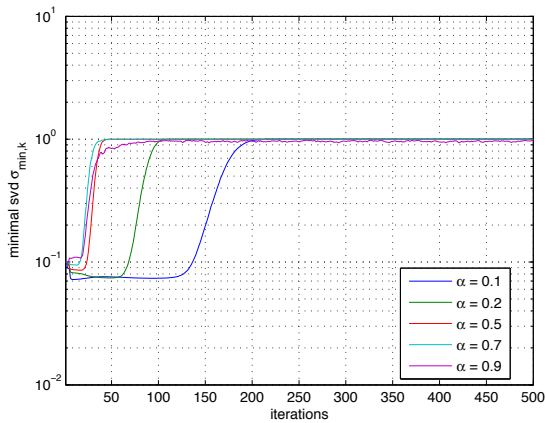


Fig. 4. Evolution of smallest singular value in the PAST-II algorithm under various normalized step-sizes.

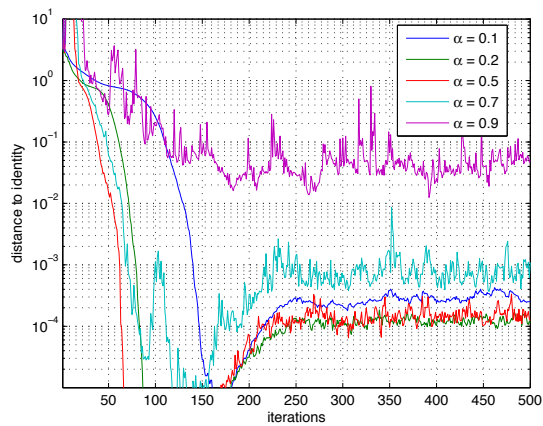


Fig. 5. Tracking behavior of the PAST-II algorithm.

guarantee convergence in the mean sense for a range of step-sizes. The authors believe that the method can be further extended to incorporate also distributed versions of the algorithm, that rely on one or more consensus algorithms for data exchange in wireless sensor networks [11, 12].

ACKNOWLEDGEMENT

Partial funding was provided by the fFORTE-Women in Technology Program in cooperation with the Vienna University of Technology, the Austrian Ministry for Science and Research. We gratefully acknowledge the cooperation with the Austrian Science Fund, through the NFN-SISE S10611-N13 project.

6. REFERENCES

[1] B. Yang, “Subspace tracking based on the projection approach and the recursive least squares method,” in *IEEE*

International Conference on Acoustics, Speech, and Signal Processing, (ICASSP’93), Apr. 1993, vol. 4, pp. 145–148.

[2] B. Yang, “Projection approximation subspace tracking,” *IEEE Transactions in signal processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[3] B. Yang, “Convergence analysis of the subspace algorithms PAST and Pastd,” *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP’96)*, pp. 3192–3193, 1996.

[4] B. Yang, “Asymptotic convergence analysis of the projection approximation subspace tracking algorithms,” *Signal Processing*, vol. 50, pp. 123–136, 1996.

[5] K. Abed-Meraim, A. Chkeif, and Y. Hua, “Fast orthonormal past algorithm,” *IEEE Signal processing letters*, vol. 7, no. 3, Mar. 2000.

[6] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, Inf. and System Sciences Series, 1986.

[7] S. C. Douglas, “Simple adaptive algorithms for Cholesky, LDL^T , QR, and eigenvalue decompositions of autocorrelation matrices for sensor array data,” in *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2001.

[8] C. Mehlführer and M. Rupp, “A robust MMSE equalizer for MIMO enhanced HSDPA,” in *Conference Record of the Fourtieth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct. 2006.

[9] P. Comon and G.H. Golub, “Tracking a few extreme singular values and vectors in signal processing,” *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1327–1343, Aug. 1990.

[10] Z. Bai et al., *Templates for the Solution of Algebraic Eigenvalue Problems: a Practical Guide*, SIAM, Philadelphia, 2000.

[11] C. Reyes, T. Hilaire, S. Paul, and C. F. Mecklenbräuker, “Evaluation of the root mean square error performance of the PAST-consensus algorithm,” in *International ITG Workshop on Smart Antennas (WSA 2010)*, Germany, Feb. 2010.

[12] C. Reyes, T. Hilaire, and C. F. Mecklenbräuker, “Distributed projection approximation subspace tracking based on consensus propagation,” in *The 3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Aruba, Dec. 2009.