OBLIQUE PRINCIPAL SUBSPACE TRACKING ON MANIFOLD

Florian Yger, Maxime Berar, Gilles Gasso, Alain Rakotomamonjy

LITIS EA 4108,
Université & INSA de Rouen,
76800 Saint-Etienne du Rouvray, France

ABSTRACT

This paper addresses the problem of principal subspace tracking in presence of a colored noise. We propose to extend the YAST algorithm to handle such a case. We also propose a Riemannian framework that could benefit to other classical trackers. Finally, as a proof of concept, our method is compared to the only oblique tracker of the literature on a toy dataset.

***Index Terms*—** Subspace tracking, matrix manifold, oblique subspace

1. INTRODUCTION

In signal processing, adaptive estimation of subspaces of covariance matrices has been successfully applied (among several other applications) in signal denoising and feature extraction. For a complete review of the area of subspace tracking, the reader should refer to [1] and references therein. Due to the numerical complexity of the task, eigenvalue decomposition (EVD) cannot be directly performed at every time step.

This observation motivates research to find a way to recursively compute the subspace basis. Over the past decade, several efficient algorithms were tailored for subspace tracking. Among the proposed subspace trackers, YAST [2, 3] has demonstrated an increased stability, a low complexity and good performances.

For most of the subspace trackers, the signal is assumed to live in a low dimensional subspace and to be corrupted by a white noise. They aim at outputing an orthonormal basis of the signal subspace. However in some cases, subspace tracking has to be performed in presence of a colored noise [4, 5]. In this general framework, the problem is formulated as a Generalized EVD (GEVD).

A non-orthonormal basis of the signal subspace is then the solution of the problem and is defined with respect to a metric implied by the noise. The obPAST [5] algorithm is the sole representative of oblique trackers in the literature.

This work was supported in part by the IST Program of the European Community, under the PASCAL2 Network of Excellence, IST-216886 and by grants from ASAP ANR-09-EMER-001 and from the INRIA ARC MaBi.

Both orthonormal and oblique subspace tracking problems are usually formulated as constrained optimization problems. However, in the light of Riemannian geometry [6], both problems can be formulated as a search on a matrix manifold.

We hereafter propose to extend the YAST algorithm to handle oblique subspace tracking as well, using classical tools from the Riemannian geometry. In the appendix section of [3], the authors minimize the role of Riemannian geometry as they consider it only in their proofs and miss the important role that Riemannian geometry can play from an algorithmic point of view. As we show in this paper, the YAST update scheme (as well as other classical trackers schemes) already fits in a Riemannian framework.

This paper is organized as follows : in Section 2, we introduce the oblique subspace tracking problem and then formulate it in a Riemannian framework. Section 3 presents the oblique YAST algorithm step-by-step and is followed by the numerical experiments of Section 4. Finally, the main conclusions of this paper are summarized in Section 5.

2. SUBSPACE TRACKING & MATRIX MANIFOLD

2.1. General formulation of the problem

The Karhunen-Loève Transform is a powerful tool originated in statistics that was successfully applied to signal processing. It requires the EVD of the covariance matrix C_s of a given signal $x(t)$ of size $N \times 1$. If the signal is low dimensional p (with $p < N$) and corrupted by a colored noise of covariance C_n[1], taking the projection of the signal on the first p eigenvectors will enhance the signal. The eigenvectors W are the solution of the maximization of the Generalized Rayleigh Quotient :

$$\max_{W \in \mathbb{R}^{N \times p}} tr\left((W^\top C_s W)(W^\top C_n W)^{-1}\right). \qquad (1)$$

As the complexity of EVD methods [7, Chap. 7] is around $O(N^3)$, this approach is intractable in practice for real time application. Subspace tracking thus aims at recursively computing an estimation W_t of the eigenvectors with the lowest possible complexity. By adding constraints on W_t and

[1]This problem can be adapted to the white noise case by taking $C_n = \mathbb{I}_{\mathbb{N}}$.

considering the estimation $C_s(t)$ of the covariance matrix at time t, the optimization problem becomes

$$\left\{ \begin{array}{l} \max_{W_t \in \mathbb{R}^{N\times p}} tr\left(W_t^\top C_s(t) W_t\right) \\ \text{s.t.} \quad W_t^\top C_n W_t = \mathbb{I}_p \end{array} \right. \tag{2}$$

In the literature, several ways are used to estimate the covariance matrix. We focus in this paper on exponential windows estimation :

$$C_s(t) = \beta C_s(t-1) + x(t)x^\top(t) \qquad 0 < \beta < 1 \tag{3}$$

Note that, due to space limitation, C_n is assumed to be fixed through time. The non trivial case of an evolving noise will be studied in an extended version of this work. A classical scheme of subspace tracking [1] consists of an update step and eventually of an orthonormalization step to insure that the constraints are satisfied. Hereafter, we show that both steps may fit in a Riemannian framework under certain conditions and we present the tools of this framework.

2.2. Generalized Stiefel manifold framework

Following the trends summed up in the inspiring book [6], we propose to make use of the manifold structure of the problem to solve it efficiently. In this section, we recall the basic notions on matrix manifolds needed for solving a subspace tracking problem and give insights on Riemannian geometry. The author of [8] foresaw the interest of Riemannian geometry approaches in signal processing. He also highlighted the conceptual differences between constrained optimization and Riemannian approaches.

We now define the Generalized Stiefel manifold as (for a given matrix $B \succ 0$) : $St_B(N \times p) = \{X \in \mathbb{R}^{N\times p} : X^\top B X = \mathbb{I}_p\}$. It can be seen as a p-dimensional B-orthonormal subspace of $\mathbb{R}^{N\times N}$. Providing that $C_n \succ 0$, our optimization problem is now formulated as :

$$\left\{ \begin{array}{l} \max_{W_t} tr\left(W_t^\top C_s(t) W_t\right) \\ \text{s.t.} \quad W_t \in St_{C_n}(N \times p) \end{array} \right. \tag{4}$$

In a Riemannian view, we should express our optimization problem as a search along a geodesic curve in the manifold. Such a search being in practice intractable, we chosed to approximate it by a search along another smooth curve on the manifold. This smooth curve is defined by a function that transforms any displacement in the tangent space to a point on the manifold. Such a function (that also obeys to some technical conditions, see [6, Sec 4.1]) is called a retraction.

Figure 1 depicts a Riemannian manifold $\mathcal{M}$ and a tangent space at a point X on this manifold. The tangent space is a vector space that locally approximates the manifold. Then the retraction is a mapping that locally transforms a search in the manifold into a search in the tangent space.

As explained in the next section, our algorithm finds a direction[2] in the tangent space, calculates a step length and applies a retraction back on the manifold to obtain the next iterate.

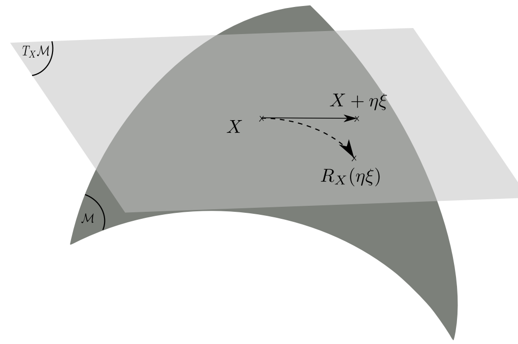

Fig. 1: Insights on Riemannian geometry : the link between a manifold $\mathcal{M}$, its tangent space $T_X\mathcal{M}$ at a point X and the retraction R_X that locally maps a displacement $\eta\xi$ in $T_X\mathcal{M}$ to $\mathcal{M}$.

The definitions of the tangent space and of the retraction can be found in [6, p. 59] for the orthonormal metric case of the Stiefel manifold ($B = \mathbb{I}_N$). Deriving the same calculus for the generalized case, we obtain the following definitions :

- Tangent space at a point X on a Generalized Stiefel manifold :

$$T_X St_B(N \times p) = \{Z \in \mathbb{R}^{N\times p} : \; X^\top B Z + Z^\top B X = 0\}$$

- Polar retraction :

$$\forall \xi \in T_X St_B(N \times p) \text{ and } X \in St_B(N \times p) \; :$$
$$R_X(\eta\xi) = (X + \eta\xi)(\mathbb{I}_p + \eta^2 \xi^\top B \xi)^{-\frac{1}{2}}$$

There exists several other retractions that can be applied for (Generalized) Stiefel manifold. Among them, the retraction based on QR-decomposition could also be applied. In regard to the classical scheme of subspace tracking defined earlier, any tracker that uses an additive update living in the tangent space can be followed by a retraction step. Conversely, any orthonormalization scheme applied to an update living in the tangent space should be a valid retraction.

In what follows, we use the pragmatic tools of Riemannian geometry presented in this section and apply them to extend the YAST algorithm to oblique subspace tracking.

3. OBLIQUE IMPLEMENTATION OF YAST

The generalized eigenvectors W_t are associated to a rank p oblique projector $O_t = C_n W_t W_t^\top$, which has the following properties (besides idempotence) :

$$O_t C_n W_t = C_n W_t \text{ and } W_t^\top O_t = W_t^\top.$$

[2]Note that our algorithm is not gradient based and this direction will be deduced from geometrical properties of our problem.

At each time step, the new sample $x(t)$ can be projected onto the previous signal oblique subspace and it defines both the p-dimensional compressed data vector $y(t)$ and the residual vector $e(t)$ C_n-orthogonal to W_{t-1} :

$$y(t) = W_{t-1}^\top x(t),$$
$$e(t) = C_n^{-1} x(t) - W_{t-1} y(t).$$

Let $\sigma(t)$ be its norm (with respect to the C_n metric)

$$\sigma(t) = \sqrt{e^\top(t) C_n e(t)}.$$

If $\sigma(t) = 0$, then the new sample $x(t)$ is actually included in the old subspace spanned by W_{t-1}. Otherwise, we can defined the normalized vector $u(t) = \frac{e(t)}{\sigma(t)}$.

Following the YAST approach [2], we assume that the solution of problem 2 lives in the augmented subspace

$$\underline{W}_t = [W_{t-1}, u(t)]$$

and that the oblique projector solution of our problem can be obtained by removing a direction from the extended oblique projector $\underline{O}_t = C_n \underline{W}_t \underline{W}_t^\top$:

$$O_t = \underline{O_t} - C_n \underline{v}(t) \underline{v}^\top(t) \quad (5)$$

where $\underline{v}(t)$ is a C_n-unitary vector in the span of $\underline{W}_t$

$$\underline{v}(t) = \underline{W}_t \underline{\phi}(t) \quad (6)$$

Using algebraic manipulations, the criterion can be rewritten with respect to the oblique projector

$$tr\left(W_t^\top C_s(t) W_t\right) = tr(C_s(t) C_n^{-1} O_t). \quad (7)$$

Using Eq. 5, 6 and 7, the optimization problem is then :

$$\begin{cases} \min_{\underline{\phi}(t)} tr\left(\underline{\phi}^\top(t) \underline{W}_t^\top C_s(t) \underline{W}_t \underline{\phi}(t)\right) \\ \text{s.t.} \quad \underline{\phi}^\top(t) \underline{\phi}(t) = 1 \end{cases} \quad (8)$$

The additive update proposed in the YAST algorithm is

$$W_t = W_{t-1} - \varepsilon u(t) \phi^\top(t) = W_{t-1} + \eta \xi_t \quad (9)$$

where $\underline{\phi}(t) = \left[\begin{array}{c} \varepsilon\phi(t) \\ \hline \varphi \end{array}\right] \begin{array}{l} \} \, p \times 1 \\ \} \, 1 \times 1 \end{array}$ and $\varepsilon^2 + \varphi^2 = 1.$

As ξ_t is in the Tangent space of $St_{C_n}(N, p)$ at the point W_{t-1}, we can use a retraction operation to assure the C_n-orthonormality. The exact orthonormalization procedure stated in [1, p. 227] corresponds in our case to a polar retraction. It is known to be robust for principal subspace tracking [9]. Note that, other types of orthonormalization methods (or retraction methods in our Riemannian context) could also be applied as in the original YAST algorithm.

Table 1: Pseudo-code of the obYAST algorithm

Input : $x(t), W_{t-1}, C_s(t-1), C_{yy}(t-1)$
Constant : C_n, C_n^{-1}, β

$y = W_{t-1}^\top x(t)$
$e = C_n^{-1} x(t) - W_{t-1} y$
$\sigma_2 = e^\top C_n e$
$\sigma = \sqrt{\sigma_2}$
if $\sigma > 0$
  $u = \frac{e}{\sigma}$
  $x' = C_s(t-1) C_n^{-1} x(t)$
  $y' = W_{t-1}^\top C_s(t-1) W_{t-1} y$
  $y'' = W_{t-1}^\top x'$
  $\gamma' = \frac{x^\top C_n^{-1} x' - 2y^\top y'' + y^\top y'}{\sigma_2}$
  $C'_{yy} = \beta C_{yy}(t-1) + yy^\top$
  $z = \beta \frac{y'' - y'}{\sigma} + \sigma y$
  $\gamma = \beta^2 \gamma' + \sigma_2$
  $\underline{C}_{yy} = \left[\begin{array}{c|c} C'_{yy} & z \\ \hline z^\top & \gamma \end{array}\right]$
  $(\underline{\phi}, \lambda) = \max eig(\underline{C}_{yy})$ % *with* $\underline{C}_{yy}\underline{\phi} = \lambda \underline{\phi}$
  if $(\underline{\phi}(p+1) < 0)$ then $\underline{\phi} := -\underline{\phi}$
  $\varphi' = 1 - \underline{\phi}(p+1)$
  $\varepsilon = \sqrt{1 - \underline{\phi}(p+1)^2}$
  $\phi = 1/\varepsilon \times \underline{\phi}(1:p)$
  $W_t = W_{t-1} - \varepsilon u \phi^\top - \varphi' W_{t-1} \phi \phi^\top$
endif

Instead of fixing $\eta = \varepsilon$ in the update rule, we propose to identify it after the application of the polar retraction

$$W_t = \left(W_{t-1} - \eta u(t) \phi^\top(t)\right) \left(I_p + (\eta^2) \phi(t) \phi^\top(t)\right)^{-1/2}.$$

We efficiently reformulate the retraction of the update $\eta \xi_t$ as :

$$-\left(\sqrt{\frac{\eta^2}{1+\eta^2}}\right) u(t) \phi^\top(t) - \left(1 - \sqrt{\frac{1}{1+\eta^2}}\right) W_t \phi(t) \phi^\top(t).$$

If we make sure that $O_t = C_n W_t W_t^\top$ then we identify $\left(\sqrt{\frac{\eta^2}{1+\eta^2}}\right)$ as ε and the simplified update rule follows :

$$W_t = W_{t-1} - \varepsilon u(t) \phi^\top(t) - (1-\varphi) W_{t-1} \phi(t) \phi^\top(t) \quad (10)$$

Note that the proposed update rule consists of the classical YAST update and an additive correction (ensuring the C_n-orthonormality). The algorithm can be implemented as in Table 1, where $C_{yy}(t) = W_t^\top C_{xx}(t) W_t$ is the compressed covariance matrix. For an optimized implementation in the orthogonal case, the reader should refer to [3].

4. NUMERICAL EXPERIMENTS

We compare the performances of the oblique YAST tracker to those of the oblique PAST tracker. As the update of the

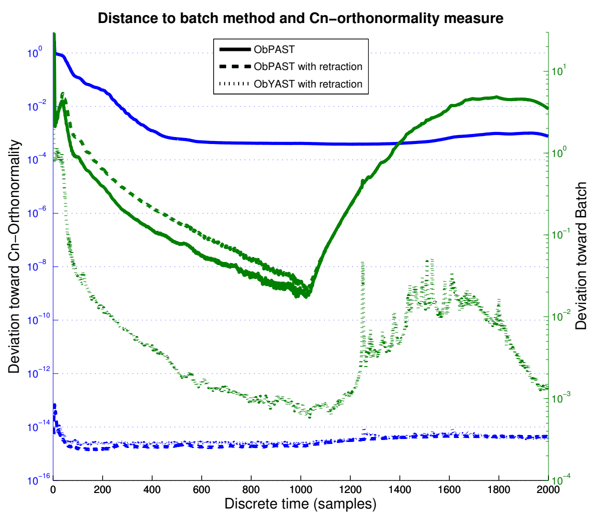

Fig. 2: Performances of the oblique trackers

obPAST algorithm can also be proved to live in the tangent space of the manifold, polar retraction can be applied to the result of the obPAST algorithm. The retracted obPAST is the third algorithm in competition.

Two criteria are used to evaluate the algorithms, the distance to the projector O_t^* computed as the result from the GEV of the estimated covariance matrix $C_s(t)$:

$$D(O_t, O_t^*) = ||O_t - O_t^*||_{\mathcal{F}} ,$$

and the C_n-orthonormality of the matrix W_t :

$$\left|\left|W_t^\top C_n W_t - \mathbb{I}_r\right|\right|_{\mathcal{F}} ,$$

where $||\cdot||_{\mathcal{F}}$ denotes the Frobenius norm. We have used a simulation protocol similar to [5]. In this toy problem, nine sensors measure three cosine signals corrupted with a colored noise. The noise covariance matrix C_n is set as $c_{ij} = 0.01 \times (0.9 + 0.1 \times \delta_{ij}) \times (m - i + 1) \times (m - j + 1)$ with $m = 9$ and do not evolve. For the three algorithms, the forgetting factor β is set to 0.995 and the SNR is equal to 6 dB. The spatial frequencies of the cosine change according to the scenario in [5]. First, the frequencies linearly change from $[-0.2\ 0.3\ 0.2]$ to $[-0.2\ 0.2\ 0.3]$ within the 1000 first samples. Then, they are constant until a sudden change to $[-0.2\ 0.2\ 0.4]$ at the 1200^{th} snapshot occurs.

We averaged the results over 100 independent runs and compared the three algorithms. The distances to the projector O_t^* are presented in green in Figure 2. The obYAST algorithms clearly outperformed the two competitors. The retracted version of obPAST shows a similar behaviour (though slightly worse) as the regular obPAST. Those results are coherent with those achieved by the orthogonal YAST algorithm [2, 3]. Blue curves in Figure 2 show that the used polar retraction is a powerful tool to ensure that the C_n-orthonormality constraint is satisfied. Indeed, unlike the obPAST method, the two retracted methods present numerically negligible deviation from the C_n-orthonormality.

5. CONCLUSION

In this paper, we extended an efficient subspace tracker to the oblique case. Moreover, we showed that mathematical tools from the Riemannian geometry are well suited for this kind of problem. As we only used the polar retraction, we plan to study the impact of other retractions on trackers and to extend this work to minor subspaces and complex signals. In future work, we plan to improve obYAST and to apply it for Brain Computer Interfaces that need efficient and adaptive algorithm for GEV. Moreover, in order to be more realistic, we intend to study the challenging problem of time-varying colored noise.

6. ACKNOWLEDGEMENT

We thank Pr. Canu for the discussions on matrix computation and Mr. Dubuisson Duplessis for his help with Figure 1.

7. REFERENCES

[1] J.-P. Delmas, "Subspace tracking for signal processing," in *Adaptive Signal Processing: Next Generation Solutions*, T. Adali and S. Haykin, Eds., chapter 4, pp. 211–270. John Wiley & Sons, Inc., 2010.

[2] R. Badeau, B. David, and G. Richard, "Yet another subspace tracker," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, 2005, vol. 4.

[3] R. Badeau, G. Richard, and B. David, "Fast and stable yast algorithm for principal and minor subspace tracking," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3437–3446, 2008.

[4] T. Gustafsson, "Instrumental variable subspace tracking using projection approximation," *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 669–681, 1998.

[5] M. Chen and Z. Wang, "Subspace tracking in colored noise based on oblique projection," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'06)*, 2006, vol. 3.

[6] P.A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2008.

[7] G.H. Golub and C.F. Van Loan, *Matrix computations*, Johns Hopkins University Press, third edition, 1996.

[8] M.D. Plumbley, "Geometry and manifolds for independent component analysis," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'07)*, 2007, vol. 4.

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