Traian Abrudan, Jan Eriksson, and Visa Koivunen. 2007. Efficient line search methods for Riemannian optimization under unitary matrix constraint. In: Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers (ACSSC 2007). Pacific Grove, CA, USA. 4-7 November 2007, pages 671-675.

Efficient Line Search Methods for Riemannian Optimization Under Unitary Matrix Constraint

Traian Abrudan, Jan Eriksson, Visa Koivunen
SMARAD CoE, Signal Processing Laboratory, Helsinki University of Technology, P.O.Box 3000, Espoo, FINLAND 02015
e-mail: {tabrudan, jamaer, visa}@wooster.hut.fi

***Abstract*— This paper proposes a Riemannian geometry approach for optimization under unitary matrix constraint. We introduce two novel line search methods which are used together with a steepest descent algorithm on the Lie group of unitary matrices $U(n)$. The proposed approach fully exploits special properties that only appear on $U(n)$, and do not appear on the Euclidean space or arbitrary Riemannian manifolds. These properties induce an almost periodic behavior of the cost function along geodesics. Consequently, the resulting step size selection rule outperforms efficient methods such as Armijo rule [1] in terms of complexity.**

We test the proposed optimization algorithm in a blind source separation application for MIMO systems by using the joint diagonalization approach [2]. The proposed algorithm converges faster than the classical JADE algorithm [2].

I. Introduction

Unitary matrices play a crucial role in many array and signal processing applications. A common problem is optimizing a certain cost function w.r.t. a matrix whose columns are mutually orthogonal. Some typical applications include blind beamforming, high-resolution direction finding, and generally all subspace-based methods. Other straight-forward applications are Independent Component Analysis (ICA) and space-time processing algorithms for Multiple-Input Multiple-Output (MIMO) communication systems. For a recent summary of such applications we refer to [3]–[5].

Most of the existing algorithms formulate the problem of optimization under orthogonal/unitary matrix constraint on the Euclidean space of $n \times n$ matrices. Therefore, classical Steepest Descent (SD) algorithm is used together with separate orthogonalization procedure applied after each iteration. The method of Lagrange multipliers has also been used to solve such problems. These methods are characterized in general by slow convergence, and/or deviation from the unitary constraint, as shown in [4], [5]. Major improvements over the classical methods are obtained by taking into account the geometrical aspects of the optimization problem. Pioneering work by Luenberger [6] and Gabay [7] convert the constrained optimization problem into an unconstrained one, on an appropriate differentiable manifold. An extensive treatment of the optimization with orthogonality constraints is given by Edelman et al. [8] in a Riemannian context. A general framework for optimization under unitary matrix constraint is proposed by Manton [9].

[1]This work was supported in part by the Academy of Finland and GETA (Graduate School in Electronics, Telecommunications and Automation).

In this paper we propose a Riemannian geometry approach for minimizing a real-valued function $\mathcal{J}(\mathbf{W})$ under unitary matrix constraint, i.e., $\mathbf{W}^H\mathbf{W} = \mathbf{W}\mathbf{W}^H = \mathbf{I}$, where is the $n \times n$ identity matrix. The Steepest Descent (SD) algorithm considered in this paper moves towards the optimum along so-called geodesics, i.e., the locally length minimizing paths on the Riemannian manifold. We introduce two novel line search methods on $U(n)$ for selecting the step size parameter. In general, step size selection is crucial for the performance of the gradient-type of algorithms. Most of the Riemannian algorithms in the literature dealing with optimization under orthogonal constraints do not provide a practical rule for adaptively selecting the step size [8], [10], [11]. They either consider a empirical small fixed step, or they are too expensive for practical applications [12]. Step size adaptation may require expensive computations even for gradient algorithms operating on Euclidean spaces. This is because most of the existing line search methods [1] require multiple cost function evaluations. On Riemannian manifolds the problem becomes even harder, because every cost function evaluation requires expensive computations of the local parametrization, which in our case is the exponential map.

The main contribution of this paper consists of two inexpensive and accurate methods for searching along geodesics on $U(n)$. They exploit the Lie group properties such as the skew-Hermitian property of the tangent vectors at group identity. The exponential map induces an *almost periodic* behavior of the cost function along geodesics. This property leads to considerable complexity reduction compared to the Armijo rule [1], which is one of the most efficient line search methods on $U(n)$ [4]. The proposed methods require *only one* evaluation of the local parametrization. This benefit is achieved only on $U(n)$, $O(n)$ (the orthogonal group), or Lie groups whose Lie algebra is comprised of matrices having purely imaginary eigenvalues. The proposed line search methods on $U(n)$ are valid for all common optimization algorithms (steepest descent, conjugate gradient or other gradient-based methods), with the condition that they move along geodesics.

This paper is organized as follows. In Section II, we present the proposed approach for optimization under unitary matrix constraint. Two novel line search methods are introduced together with a Riemannian steepest descent on $U(n)$ [4]. Simulation results are presented in Section III. The proposed algorithm is used to solve the unitary matrix optimization problem of the JADE algorithm which is applied to blind

978-1-4244-2110-7/08/$25.00 ©2007 IEEE

source separation in a MIMO system. Finally, Section IV concludes the paper.

II. Algorithm

In this section we propose an efficient Riemannian approach for unitary optimization. First, a steepest descent (SD) algorithm on the Lie group of unitary matrices $U(n)$ is considered in Subsection II-A. In Subsection II-B we present a key feature of $U(n)$ necessary in our derivations. The novel line search methods are introduced in Subsection II-C

A. SD algorithm on $U(n)$

The cost function $\mathcal{J}(\mathbf{W})$ may be minimized iteratively by using the steepest descent (SD) algorithm on $U(n)$ derived in [4]. The Riemannian gradient of $\mathcal{J}(\mathbf{W})$ at $\mathbf{W}_k$, translated to the group identity element is given by

$$\mathbf{G}_k = \mathbf{\Gamma}_{\mathbf{W}} \mathbf{W}_k^H - \mathbf{W}_k \mathbf{\Gamma}_{\mathbf{W}}^H, \tag{1}$$

where by $\mathbf{\Gamma}_{\mathbf{W}} = \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}(\mathbf{W})$ we denote the Euclidean gradient of $\mathcal{J}(\mathbf{W})$ at $\mathbf{W}_k$. The kth iteration corresponding to a SD algorithm along geodesics on $U(n)$ is given by

$$\mathbf{W}_{k+1} = \exp(-\mu_k \mathbf{G}_k)\mathbf{W}_k, \quad k = 0, 1, \ldots \tag{2}$$

A typical initial value is $\mathbf{W}_0 = \mathbf{I}$. The rotational update (2) maintains $\mathbf{W}_{k+1}$ unitary at each iteration. The step size $\mu_k > 0$ controls the convergence speed and needs to be computed at each iteration. Two novel inexpensive step size selection methods are proposed in Subsection II-C.

B. Almost periodic cost function along geodesics on $U(n)$

An unexploited feature of $U(n)$ is that smooth cost functions are almost periodic along geodesics. This is a key property which we exploit in this paper. The rotational update given in (2) is written in terms of an exponential of a skew-Hermitian matrix. The corresponding geodesic curve emanating form $\mathbf{W}_k \in U(n)$, is

$$\mathcal{W}(\mu) = \exp(-\mu \mathbf{G}_k)\mathbf{W}_k, \quad \mathbf{G}_k \in \mathfrak{u}(n), \quad \mu \in \mathbb{R}. \tag{3}$$

It is important to note that the gradient $\mathbf{G}_k$ (1) is a skew-Hermitian, i.e., $\mathbf{G}_k = -\mathbf{G}_k^H$. This is due to the fact that the tangent space at the identity element of $U(n)$ is the Lie algebra of skew-Hermitian matrices $\mathfrak{u}(n)$ [4]. Skew-Hermitian matrices have purely imaginary eigenvalues of form $\jmath\omega_i$, $i = 1, \ldots, n$. Therefore, the eigenvalues of the matrix exponential $\exp(\mu \mathbf{G}_k)$ are complex exponentials of form $e^{\jmath\omega_i\mu}$. Consequently, the cost function along geodesics given by

$$\hat{\mathcal{J}}(\mu) = \mathcal{J}(\mathcal{W}(\mu)) \tag{4}$$

is an *almost periodic function* [13], and therefore it may be expressed as a sum of periodic functions of μ. Also its derivatives are almost periodic functions of μ. The almost periodic property of the cost function and its derivatives appears only in the case of exponential map, unlike other parametrizations. This property motivates the fact that the optimization on $U(n)$ should be carried along geodesics. Moreover, this special property appears only on certain manifolds, such as the unitary group $U(n)$ and the orthogonal group $O(n)$ and it does not appear on Euclidean spaces or on general Riemannian manifolds. The almost-periodic behavior of the cost function along geodesics may be used to perform geodesic search on $U(n)$. This it will be shown in Subsection II-C where two novel step size selection methods are introduced.

C. Efficient geodesic search methods on $U(n)$ – the step size selection

In general, the step size adaptation may be computationally expensive even in the case of gradient algorithms operating on Euclidean spaces. This is due to the fact that most of the methods [1] require multiple cost function evaluations. On Riemannian manifolds the step size selection problem becomes even harder because every cost function evaluation requires expensive computations of the local parametrization. In [4], Armijo method is efficiently used since only few computations of the matrix exponential are needed in average per iteration. The complexity issues of the Armijo method are treated in detail in [4].

In this section, we propose two novel methods for performing high-accuracy one-dimensional search along geodesics on $U(n)$. They rely on the fact that smooth functions are *almost periodic* along geodesics on $U(n)$. The first method is based on a polynomial approximation of the first-order derivative of the cost function along geodesics. The second one is a Discrete Fourier Transform (DFT) based approach. We emphasize the fact that the proposed methods require *only one* evaluation of the matrix exponential. Therefore, they outperform significantly the Armijo method in terms of complexity.

The main goal is find a step size $\mu_k > 0$ along the geodesic curve $\mathcal{W}(\mu)$ (3) which minimizes the composed cost function $\hat{\mathcal{J}}(\mu) = \mathcal{J}(\mathcal{W}(\mu))$. The direction $-\mathbf{G}_k \in \mathfrak{u}(n)$ in (3) corresponds to a steepest descent algorithm. The proposed line search methods are also valid for conjugate gradient, or any other gradient-type of algorithm along geodesics on $U(n)$. Therefore, instead of limiting our line search method to the steepest descent, we consider a general search direction $-\mathbf{H}_k \in \mathfrak{u}(n)$.

Consider two successive points on $U(n)$ such that $\mathbf{W}_k = \mathcal{W}(0)$ and $\mathbf{W}_{k+1} = \mathcal{W}(\mu_k)$. Finding the step value $\mu = \mu_k$ that minimizes $\hat{\mathcal{J}}(\mu)$ may be done by computing the first-order derivative $d\hat{\mathcal{J}}/d\mu$ and setting it to zero. By using the chain rule for the composed function $\mathcal{J}(\mathcal{W}(\mu))$, we get

$$\frac{d\hat{\mathcal{J}}}{d\mu}(\mu) = -2\Re\{\mathrm{trace}\Big\{\frac{\partial \mathcal{J}}{\partial \mathcal{W}^*}(\mathcal{R}(\mu)\mathbf{W}_k)\ \mathbf{W}_k^H \mathcal{R}^H(\mu)\mathbf{H}_k^H\Big\}\}. \tag{5}$$

The first-order derivative (5) is also an almost periodic function. Periodicity may be exploited in many ways in order to find the zeros of the derivative corresponding to local minima of the cost function along geodesics. We present two different approaches. The first one finds only the first positive zero-crossing value of the first-order derivative by using a polynomial approximation of the derivative. The second one finds several zero-crossing values of the derivative and is based

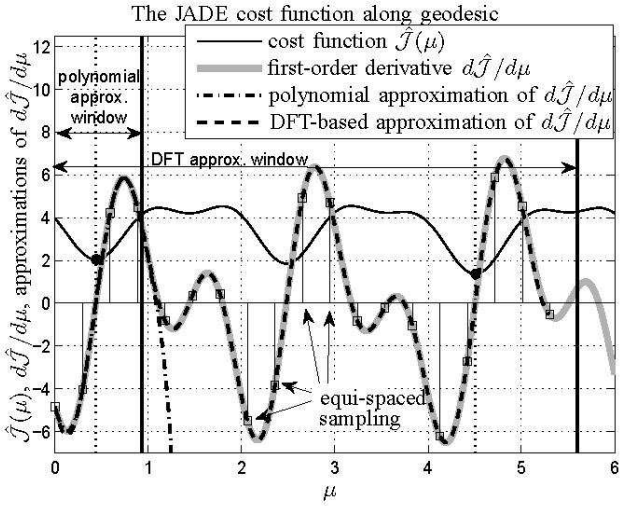

Fig. 1. Performing the geodesic search for the JADE [2] cost function. The almost periodic behavior of the function $\hat{\mathcal{J}}(\mu)$ and its first-order derivative $d\hat{\mathcal{J}}/d\mu$ (5) along geodesic $\mathcal{W}(\mu)$ (3) may be noticed. A fourth-order polynomial is used to approximate $d\hat{\mathcal{J}}/d\mu$ and find the first local minimum of $\hat{\mathcal{J}}(\mu)$. This corresponds to the smallest positive zero-crossing value of $d\hat{\mathcal{J}}/d\mu$, i.e., the desired step size μ_k. A DFT-based approximation of $d\hat{\mathcal{J}}/d\mu$ (dashed line) is used to find several local minima of $\hat{\mathcal{J}}(\mu)$ along geodesic and select the best one.

on a Fourier series approximation. Other approaches may also be possible, and they have been under our investigation.

1) Method 1 - Polynomial approximation approach : The goal of the first step size selection method proposed in this paper is to find *the first local minimum* of the cost function along a given geodesic. The main argument for the present approach is the computational benefit, i.e., *only one* evaluation of the matrix exponential is needed. Finding the first minimum of the cost function along a given geodesic corresponds to finding the smallest positive zero-crossing value of the first-order derivative of the cost function, which is also almost periodic. In this purpose we use a low-order polynomial approximation of the derivative and find its smallest positive zero-crossing value. The approximation range of the derivative is determined from its spectral content. The full method is explained in detail below.

In Figure 1 we take as an example the JADE cost function used to perform the joint diagonalization for blind separation in [2]. A practical application of the proposed algorithm to blind separation by optimizing the JADE criterion will be given later in Section III. The cost function $\hat{\mathcal{J}}(\mu)$ is represented by black continuous curve in Figure 1. Its first-order derivative $d\hat{\mathcal{J}}/d\mu$ is represented by the gray continuous curve in Figure 1. The first local minimum of the cost function is attained where the first-order derivative crosses zero for the first time. Since the direction $-\mathbf{H}_k \in \mathfrak{u}(n)$ corresponds to a descent direction at $\mathbf{W}_k$, the first-order derivative $d\hat{\mathcal{J}}/d\mu$ is always negative at the origin (at $\mu = 0$). Therefore, the cost function $\hat{\mathcal{J}}(\mu)$ is monotonically decreasing up to the first positive zero-crossing value of $d\hat{\mathcal{J}}/d\mu$. This value corresponds to a local minimum of $\hat{\mathcal{J}}(\mu)$ along geodesic (3) (or seldom to a saddle point). In order to find the first positive zero-crossing value we use a low-order polynomial approximation of the first-order derivative $d\hat{\mathcal{J}}/d\mu$. The approximation range is determined from the spectral content of the derivative, as described below.

Due to differentiation, the spectrum of $d\hat{\mathcal{J}}/d\mu$ is the high-pass filtered spectrum of $\hat{\mathcal{J}}(\mu)$. The frequency components are determined by the purely imaginary eigenvalues of $-\mathbf{H}_k$, as shown in Subsection II-B. Therefore, the cost function as well as its derivative possess discrete frequency spectra. For our task at hand, we are not interested in the complete spectrum of $d\hat{\mathcal{J}}/d\mu$, we are only interested in the smallest zero-crossing value of the derivative. This is determined by the highest frequency component in the spectrum of $d\hat{\mathcal{J}}/d\mu$ in the following way. In the interval of μ which is equal to one period corresponding to the highest frequency in the spectrum, the function $d\hat{\mathcal{J}}/d\mu$ has at most one complete cycle on that frequency, and less than one on other frequencies. The highest frequency component of $d\hat{\mathcal{J}}/d\mu$ is $q|\omega_{\max}|$, where $\omega_{\max}$ is the eigenvalue of $\mathbf{H}_k$ having the highest magnitude, and q is the order of the cost function[1] in coefficients of $\mathbf{W}$. The corresponding period is

$$T_\mu = \frac{2\pi}{q|\omega_{\max}|}. \tag{6}$$

The highest frequency component is amplified the most due to differentiation (high-pass filtering). Therefore, the first-order derivative $d\hat{\mathcal{J}}/d\mu$ crosses zero at most twice within the interval $[0, T_\mu)$. The presence of the zero is detected as sign change of the derivative within the interval T_μ. Since $d\hat{\mathcal{J}}/d\mu$ varies very slowly within the interval $[0, T_\mu)$ and due to the almost periodic property of the derivative, a low-order polynomial approximation of the derivative is sufficient to determine the corresponding root (see Figure 1).

The approximation requires evaluating the cost function at least at P points, where P is the order of the polynomial. In order to reduce complexity, the derivative is evaluated at equi-spaced points $\{0, \frac{T_\mu}{P}, 2\frac{T_\mu}{P}, \ldots, T_\mu\}$. Consequently, only one computation of the matrix exponential $\mathcal{R}(\mu) = \exp(-\mu\mathbf{H}_k)$ is needed at $\mu = T_\mu/P$, and the next $(P-1)$ values are the powers of $\mathcal{R}(\mu)$. This is based on the desirable property of the matrix exponential that $\exp(-m\mu\mathbf{H}_k) = [\exp(-\mu\mathbf{H}_k)]^m$. The whole procedure requires one matrix exponential and $(P-1)$ matrix multiplications. Since P is very low, this procedure has reasonably low complexity. We also emphasize the fact that when evaluating the approximation interval T_μ by using (6), only the magnitude of the maximum eigenvalue $\omega_{\max}$ of $-\mathbf{H}_k$ needs to be computed and not the full eigen-decomposition. Having the first-order derivative evaluated at equi-spaced

[1]The order q corresponds to the highest degree that t appears on in the expansion $\mathcal{J}(\mathbf{W} + t\mathbf{Z})$. Due to the fact that the local parametrization is the exponential map, the order of the cost function coincides with the order of its derivative and it is assumed to be finite (most of the practical cost functions). In the case that q is not finite, a finite order of the expansion needs to be used to approximate the cost function.

points, we can approximate it by using a polynomial of order P,

$$\hat{\mathcal{J}}'(\mu) \triangleq \frac{d\hat{\mathcal{J}}}{d\mu} \approx a_0 + a_1\mu + \ldots + a_p\mu^P. \quad (7)$$

In Figure 1, a fourth order approximating polynomial is used at equi-spaced points within the interval $[0, T_\mu)$. The approximation is represented by thick dashed line. Because the function $\hat{\mathcal{J}}'(\mu)$ is very slowly varying within the interval $[0, T_\mu)$, the smallest real and positive root of the approximating polynomial is practically the step size μ_k corresponding to an exact line search, as shown in Figure 1. In terms of complexity the proposed geodesic search method is more efficient than the Armijo method [1] which requires multiple evaluations of the matrix exponential at every iteration [4].

2) Method 2 - DFT approximation approach : The goal of our second step size selection method is to find *multiple local minima* of the cost function along a given geodesic. The method requires also *only one* evaluation of the matrix exponential, but more matrix squaring operations. The basic idea is to approximate the almost periodic function $d\hat{\mathcal{J}}/d\mu$ (5), by a periodic one, using the classical Discrete Fourier Series (DFT) approach. This is shown in Figure 1, where the JADE cost function [2] is considered, analogously to the example in Figure 1.

First, the length of the DFT interval T_{DFT} needs to be set. The longer DFT interval is considered, the better approximation is obtained. In practice we have to limit the length of the DFT interval to several periods T_μ (6) corresponding to the highest frequency component. Once the DFT interval length is set, the derivative $d\hat{\mathcal{J}}/d\mu$ needs to be sampled at N_{DFT} points. According to the Nyquist sampling theorem, $K \geq 2$ samples must be taken in an interval of length T_μ. Therefore, if N_T periods T_μ are considered, the DFT length is $N_{\text{DFT}} \geq 2N_T$. Due to the fact that T_μ does not necessarily correspond to any almost period of the derivative, the values at the edges of the DFT interval may differ. In order to avoid approximation mismatches at the edges of the DFT interval, a window function may be applied [14]. The chosen window function must be strictly positive, in order to preserve the position of the zeros we are interested in. In our approach we choose a Hann window and discard the zero-values at the edges, i.e., $h(i) = 0.5 - 0.5\cos\left(2\pi\frac{i+1}{N_{\text{DFT}}+1}\right)$, $i = 0, \ldots, N_{\text{DFT}} - 1$. Therefore, instead of approximating the first-order derivative (5), it is more desirable to approximate the windowed derivative $\mathcal{D}(\mu_i) = h(i)\frac{d\hat{\mathcal{J}}}{d\mu}(\mu_i)$, $i = 0, \ldots, N_{\text{DFT}} - 1$. The approximating Fourier series of the first-order derivative after the windowing operation may be written as:

$$\mathcal{D}(\mu) \approx \sum_{k=-(N_{\text{DFT}}-1)/2}^{(N_{\text{DFT}}+1)/2} c_k \exp\left(\jmath\frac{2\pi k}{T_{\text{DFT}}}\mu\right). \quad (8)$$

The DFT-based approximation of $d\hat{\mathcal{J}}/d\mu$ is shown in Figure 1 by the thick dashed line.

First, evaluating the derivative $d\hat{\mathcal{J}}/d\mu$ at points $\mu_i \in \{0, T_{\text{DFT}}/N_{\text{DFT}}, \ldots, (N_{\text{DFT}}-1)T_{\text{DFT}}/N_{\text{DFT}}\}$ and applying the Hann window is needed. After determining the Fourier coefficients c_k, the polynomial corresponding to the Fourier series approximation (8) is set to zero. The roots of the polynomial (8) which are close to the unit circle need to be determined, i.e., $\rho_l = e^{\jmath\omega_l}$, $l \leq 2N_T$. The values of μ corresponding to those roots are:

$$\mu_l = \left(\frac{\omega_l T_{\text{DFT}}}{2\pi}\right)_{\text{modulo } T_{\text{DFT}}}, \mu_1 \leq \mu_2 \leq \ldots \quad (9)$$

Given a descent direction $-\mathbf{H}_k$, the smallest step size value μ_l corresponds to a minimum (or seldom to a saddle point). If no saddle points occur within the DFT window, all the step size values μ_l with l odd, correspond to local minima. The even ones correspond to maxima. This may be noticed in Figure 1. Within the interval T_{DFT} there are at most N_T minima, and there is the possibility to choose the best one. Therefore, the global minimum within the DFT window can be chosen in order to reduce the cost function as much as possible at every iteration. Finding the best minimum would require additional evaluations of the cost function. Therefore, it would involve computing the matrix exponential for all μ_l with odd l, which is preferable to be avoided since this is computationally rather expensive. A reasonable solution is in this case using the information on the sampled values of the cost function. Therefore, the step size is set to the root which is closest to the value that achieves a minimum of the sampled cost function.

III. Simulation Results and Applications

In this section, we test how the proposed method performs in a blind source separation application for MIMO systems. More applications may be found in [3]–[5].

Separating signals blindly in a MIMO communication system may be done by exploiting the statistical information of the transmitted signals. The JADE (Joint Approximate Diagonalization of Eigen-matrices) algorithm [2] is a reliable choice. After *pre-whitening* of the received signal a *unitary rotation* is needed. The problem is formulated as a unitary optimization which can be efficiently solved by using the proposed algorithm. The complexity per iteration is lower compared to more general optimization algorithms [8], [9] which do not exploit the additional group structure of $U(n)$.

A number of $m = 20$ independent 16-QAM signals are sent through each of the sixteen transmit antennas and they are received by $r = 20$ receive antennas. The signal-to-noise ratio is $20dB$. The frequency-flat Rayleigh MIMO channel is represented by a 20×20 mixing matrix. We perform the blind separation and co-channel signal cancellation. The goal is to determine a unitary matrix $\mathbf{W}$ such that the fourth-order cumulant matrices corresponding to the whitened signals are jointly diagonalized, i.e., the JADE criterion [2] is minimized.

The performance is studied in terms of convergence speed considering the goodness of the optimization solution (*JADE criterion*) and the goodness of the solution of the entire blind source separation problem (*Amari distance*). The two criteria versus the number of iterations are shown in Figure 2. We

compare the Riemannian SD in [4] which uses the Armijo step size selection method to the two proposed methods, i.e, the polynomial and the DFT-based approximations, respectively. The proposed DFT and polynomial-based approaches converge at the same speed as Armijo method, but their computational complexity is much lower. This is because only one matrix exponential is computed per iteration, while for the Armijo method several computations are needed. This is validated also by simulations. In Table I, we may notice that the number of matrix exponentials for the Armijo method is higher by one order of magnitude compared to the proposed methods. For more details on the computational complexity of the Armijo method, see [4].

In Figure 2, the three Riemannian algorithms are also compared to the classical JADE solution which performs the joint diagonalization by using Givens rotations. The Riemannian steepest descent algorithms converge faster than the classical JADE algorithm [2] at comparable cost per iteration. The Riemannian algorithms have $\mathcal{O}(m^3)$ complexity per iteration, while the classical JADE algorithm [2] has a total complexity of $\mathcal{O}(m^4)$. The number of iterations required for the Riemannian algorithms to achieve convergence remains almost constant with m, unlike the Givens rotations approach [2], whose number of iterations increases linearly with m. Therefore, using a Riemannian approach for joint diagonalization is justified when the number of signals m to be separated is relatively large.

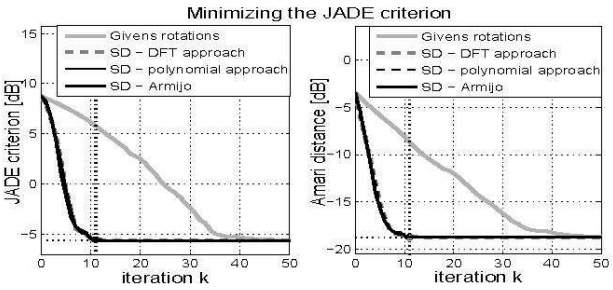

Fig. 2. The JADE criterion and the Amari distance vs. the iteration step. Riemannian steepest descent on $U(m)$ using three different methods for step size selection: the Armijo method [1], the proposed polynomial approach and the proposed DFT-based approach. The solution is achieved in few iterations all three methods, but the two proposed methods are less complex compared to the Armijo method. All three Riemannian algorithms converge faster that the classical JADE algorithm [2] based on Givens rotations, at comparable cost per iteration.

Number of matrix exponential operations/iteration		
Armijo method	Proposed methods	
	polynomial approach	DFT-based approach
9.52	1	1

TABLE I

SIMULATED AVERAGE COMPUTATIONAL COMPLEXITY (1000 RUNS).

IV. CONCLUSIONS

In this paper, a Riemannian approach for optimization under unitary matrix constraint is proposed. We introduce two inexpensive line search methods coupled with a steepest descent algorithm on the Lie group of $n \times n$ unitary matrices $U(n)$. The proposed line search methods exploit the properties of certain Lie groups such as the unitary group $U(n)$ and orthogonal group $O(n)$. This advantage may not be achieved in the Euclidean space or in arbitrary Riemannian manifolds. The *almost periodic* property of the cost function along geodesics is exploited, resulting to two low-complexity step size adaptation methods. The proposed algorithm solves efficiently the joint diagonalization problem. Other possible applications are smart antenna algorithms, wireless communications, biomedical measurements and signal separation, where unitary matrices play an important role in general.

REFERENCES

[1] E. Polak, *Optimization: Algorithms and Consistent Approximations*. New York: Springer-Verlag, 1997.

[2] J. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, 1993.

[3] J. H. Manton, "On the role of differential geometry in signal processing," in *International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 1021–1024, Philadelphia, March 2005.

[4] T. Abrudan, J. Eriksson, and V. Koivunen, "Steepest descent algorithms for optimization under unitary matrix constraint," *IEEE Transactions on Signal Processing (to appear)*.

[5] T. Abrudan, J. Eriksson, and V. Koivunen, "Optimization under unitary matrix constraint using approximate matrix exponential," in *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, 2005.*, 28 Oct. - 1 Nov. 2005.

[6] D. G. Luenberger, "The gradient projection method along geodesics," *Management Science*, vol. 18, pp. 620–631, 1972.

[7] D. Gabay, "Minimizing a differentiable function over a differential manifold," *Journal of Optimization Theory and Applications*, vol. 37, pp. 177–219, June 1982.

[8] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[9] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, pp. 635–650, Mar. 2002.

[10] M. D. Plumbley, "Geometrical methods for non-negative ICA: manifolds, Lie groups, toral subalgebras," *Neurocomputing*, vol. 67, pp. 161–197, 2005.

[11] Y. Nishimori and S. Akaho, "Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold," *Neurocomputing*, vol. 67, pp. 106–135, Jun. 2005.

[12] S. Fiori, "Quasi-geodesic neural learning algorithms over the orthogonal group: a tutorial," *Journal of Machine Learning Research*, vol. 1, pp. 1–42, Apr. 2005.

[13] A. Fischer, "Structure of Fourier exponents of almost periodic functions and periodicity of almost periodic functions," *Mathematica Bohemica*, vol. 121, no. 3, pp. 249–262, 1996.

[14] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, pp. 51–83, Jan. 1978.