A CG-TYPE METHOD FOR COMPUTING THE DOMINANT SUBSPACE OF A SYMMETRIC MATRIX [1]

Martin Kleinsteuber * **A.K. Seghouane** **

* *Institute for Data Processing, Technische Universität München*
Arcisstraße 21, 80 333 München
e-mail: kleinsteuber@tum.de
** *National ICT Australia, Canberra Research Laboratory, Canberra, Australia*
and
The Australian National University, College of Engineering and Computer Science

Article submitted to the *Special Session on Geometric Control Theory – a tribute to Professor Fátima Silva Leite*

Abstract: Computing the subspace spanned by the eigenvectors corresponding to the r largest eigenvalues of a symmetric matrix is an important subtask in many signal processing applications and statistics. Here, from the application side, we focus on subspace based algorithms for sensor node localization in wireless sensor networks. A conjugate gradient method on the Grassmann manifold is proposed to compute the r-dimensional dominant subspace of an $(n \times n)$-symmetric matrix. This leads to new subspace algorithms which avoid the time consuming eigendecomposition of the data matrix, but rather compute the signal and the noise space in $O(n^2r^3)$ flops. Some convergence aspects are discussed and numerical simulations are presented to illustrate the performance of the proposed algorithm. When applied to the problem of sensor positions estimation with M sensors, the full-set subspace algorithm is reduced from order $O(M^6)$ to $O(M^4)$ without loosing accuracy.

Keywords: dominant subspace, signal space, principle component analysis, subspace methods, eigendecomposition, Grassmann manifold.

Subspace based algorithms are widely used in signal processing applications. They are for example used in two or three dimensional sensor node localization in wireless sensor networks (WSN) using node-to-node squared distance measurements obtained from pair-wise time-of-arrival (TOA) or received signal strength (RSS) (So & Chan, 2007; Chan, *et al.*, 2009).
A typical subproblem in subspace algorithms is the task of estimating an orthogonal basis of the signal and the noise space of the so called multidimensional similarity matrix (Chan, *et al.*, 2009). Considering M sensors, this matrix is a symmetric $(M^2 \times M^2)$ matrix of rank 2 (if noise is neglected). Since a complete eigendecomposition of a symmetric $(n \times n)$ matrix is of order $O(n^3)$, it is of order $O(M^6)$ for the multidimensional similarity matrix in terms of the number of sensors. In this paper we propose a method to compute an orthogonal basis of the signal space that avoids the computation of a complete eigendecomposition by assuming that its dimension is known a priori. The algorithm is of order $O(n^2r^3)$ and hence efficient in the case when $r << n$. Implemented in the full

[1] This work has been supported by CoTeSys - Cognition for Technical Systems, Cluster of Excellence, funded by Deutsche Forschungsgemeinschaft.

set subspace algorithm, it will reduce the order from $O(M^6)$ to $O(M^4)$.
Note, that also other methods are thinkable, e.g. methods that extract in each step an approximation of the dominant eigenvector, e.g. with Rayleigh quotient iteration methods (Parlett, 1974) and then reduce the problem. In contrast to these methods, our approach focusses on the direct computation of the signal space.

Following the ideas developed in (Gabay, 1982), a conjugate gradient (CG) method on the Grassmann manifold is presented that computes the subspace corresponding to the r largest eigenvalues of the symmetric matrix A. This geometric approach is in contrast to conjugate gradient methods in the Euclidean space (Yang, *et al.*, 1989), that do not iterate on the restricted manifold and hence work with a larger number of parameters. Such an intrinsic approach has also been exploited in (Absil, *et al.*, 2002), where a Newton-type method on the Grassmann manifold is proposed. The algorithm proposed in (Absil, *et al.*, 2002) has low numerical complexity only if the initial symmetric matrix is preconditioned to be tridiagonal. However, if A changes over time, such a tridiagonalization is required at each time instance. CG-methods, in contrast, are powerful optimization techniques and suitable for tracking (Fu & Dowling, 1995). They have also been used in (Feng & Owen, 1996) for computing the smallest eigenpair of a symmetric matrix. Another disadvantage of Newton-methods is that they usually get stuck in critical points, whereas our method shows global convergence behavior.

Here, we modify and specify the algorithm by using an approximation of the Riemannian exponential via QR-decomposition and a step size which is computationally efficient to compute. Ultimately, this leads to an algorithm that maintains the geometric structure of the problem and is computationally cheap. Exploiting the intrinsic structure of the problem lets us expect that the computation of the signal and noise spaces is very accurate and robust to rounding errors. Our approach has the advantage of leading to efficient subspace estimation algorithms, which are more adapted to real time applications. Besides the aforementioned improvement of the most accurate localization method, the so called full-set subspace algorithm (Chan, *et al.*, 2009), the multidimensional scaling localization method (MDS) (Shang, *et al.*, 1984) is reduced from order $O(M^3)$ to order $O(M^2)$ while maintaining accuracy.

The rest of the paper is organized as follows. The mathematical framework is provided in Section 2 and the algorithm is proposed in Section 3, including a discussion of convergence properties and aspects of implementation. Simulation results to illustrate the performance of the proposed method are presented in Section 4. Concluding remarks are given in Section 5.

1. MATHEMATICAL FRAMEWORK ON THE GRASSMANN MANIFOLD

Let $\mathcal{P}$ be the r-dimensional dominant subspace of a symmetric $(n \times n)$-matrix A. In many applications, A is a symmetric positive semidefinite matrix and $\mathcal{P}$ is called the signal space. In this context, its orthogonal complement $\mathcal{P}^\perp$ is referred to as the noise space. We follow the ideas proposed in (Kleinsteuber & Hüper, 2007) and consider a conjugate gradient algorithm on the set of r-dimensional subspaces of $\mathbb{R}^n$, the so-called Grassmann manifold, which is diffeomorphic to the set of symmetric rank-r projectors $Gr_{r,n} := \{uu^\top \mid u \in \mathbb{R}^{n\times r}, u^\top u = I\}$, where I is the identity matrix. We endow $Gr_{r,n}$ with the Riemannian structure inherited from the surrounding space of symmetric matrices with inner product $\mathrm{tr}(\mathrm{AB})$, where $\mathrm{tr}(X)$ is the trace of the matrix X. In contrast to (Kleinsteuber & Hüper, 2007), a more sophisticated approximation of the Riemannian exponential map and the parallel transport is proposed. Although here we restrict to the real case, an extension to Hermitian matrices is straightforward.

At first we have to introduce some notations. Let $\mathfrak{so}_n := \{\Omega \in \mathbb{R}^{n\times n} \mid \Omega = -\Omega^\top\}$ denote the set of all skew-symmetric $(n \times n)$-matrices and $SO_n := \{Q \in \mathbb{R}^{n\times n} \mid Q^\top Q = I\}$ the set of orthogonal matrices. Let $\mathcal{R}_n$ be the set of invertible upper triangular $(n \times n)$-matrices with positive entries on the diagonal and let $Gl_n := \{X \in \mathbb{R}^{n\times n} \mid X \text{ is invertible}\}$. It follows from the Gram-Schmidt orthogonalization procedure that the map

$$SO_n \times \mathcal{R}_n \to Gl_n, \quad (Q, R) \mapsto QR \tag{1}$$

is a diffeomorphism. According to Eq. (1) every $X \in Gl_n$ decomposes uniquely into

$$X =: X_Q X_R \tag{2}$$

with $X_Q \in SO_n$ and $X_R \in \mathcal{R}_n$. Every $(n \times n)$-matrix decomposes uniquely into the sum of a skew-symmetric and an upper triangular matrix. We write

$$X =: X_{\mathrm{skew}} + X_{\mathrm{upp}} \tag{3}$$

with $X_{\mathrm{skew}} \in \mathfrak{so}_n$ and X_{upp} upper triangular. The map

$$q_\Omega \colon \mathbb{R} \to \mathcal{U}_n, \quad t \mapsto (I + t\Omega)_Q \tag{4}$$

is smooth for all $\Omega \in \mathfrak{so}_n$. Moreover,

$$\dot{q}_\Omega(0) = \Omega, \quad \ddot{q}_\Omega(0) = (\Omega^2)_{\mathrm{upp}} \tag{5}$$

cf. (Kleinsteuber & Hüper, 2007). Now, for fixed $\theta \in SO_n$, the identities

$$\dot{q}_{\theta^\top\Omega\theta}(0) = \theta^\top\Omega\theta, \quad \ddot{q}_{\theta^\top\Omega\theta}(0) = (\theta^\top\Omega^2\theta)_{\mathrm{upp}} \tag{6}$$

are easily verified. For two squared matrices A, B, we denote the commutator by $[A, B] := AB - BA$. The above results allow us to give an accurate approximation of the geodesics

$$\gamma_\xi(t) = \mathrm{e}^{t[\xi,P]} P \mathrm{e}^{-t[\xi,P]}, \tag{7}$$

where $\xi \in T_P Gr_{r,n} = \{[P,\Omega] \mid \Omega \in \mathfrak{so}_n\}$, the tangent space at $P \in Gr_{r,n}$. Indeed, for all $\theta \in SO_n$ the curve

$$\alpha_\xi \colon \mathbb{R} \to Gr_{r,n},$$
$$t \to \theta q_{\theta^\top [\xi,P]\theta}(t)\theta^\top P \theta \Big(q_{\theta^\top [\xi,P]\theta}(t)\Big)^\top \theta^\top, \tag{8}$$

where $q_{[\xi,P]}(t)$ is defined as in Eq. (4), is a first order approximation of the geodesic (7) around P, i.e. the identities

$$\alpha_\xi(0) = \gamma_\xi(0), \qquad \dot{\alpha}_\xi(0) = \dot{\gamma}_\xi(0) \tag{9}$$

hold. The parallel transport of $\eta \in T_P Gr_{r,n}$ with respect to the Levi-Civita connection along the geodesic $\gamma_\xi(t)$ is given by

$$\eta(t) = \mathrm{e}^{t[\xi,P]} \eta \mathrm{e}^{-t[\xi,P]} \tag{10}$$

and, according to the above results, will be approximated by

$$\tau(\eta,t) := \theta q_{\theta^\top [\xi,P]\theta}(t)\theta^\top \eta \theta \Big(q_{\theta^\top [\xi,P]\theta}(t)\Big)^\top \theta^\top. \tag{11}$$

In general, a conjugate gradient method for optimizing a smooth function on $\mathbb{R}^N$ requires a restart after N iterations in order to achieve N-step local quadratic convergence, cf. (Nocedal & Wright, 1999). This also holds true for the generalization of CG-methods to geodesically complete Riemannian manifolds, cf. (Gabay, 1982), where in this case N is the dimension of the manifold. For our purposes, however, it turns out that the algorithm converges with much fewer iterates, hence we will not stress the issue of restarting.

2. COST FUNCTION AND IMPLEMENTATION

In the following, the cost function that is to be optimized is presented. We refer to (Helmke, *et al.*, 2007) for a derivation of the stated results. Given some $u \in \{u \in \mathbb{R}^{n\times r} \mid u^\top u = I\}$ we denote by $u^\perp \in \mathbb{R}^{n\times(n-r)}, u^\top u^\perp = 0, (u^\perp)^\top u^\perp = I$ an orthonormal basis of the orthogonal complement of the vectorspace spanned by the columns of u. Note, that $u^\perp$ is unique up to multiplication with some orthogonal $(n-r)\times(n-r)$ matrix and that the matrix $[u,u^\perp]$ (not to be confounded with a commutator) is orthogonal. Let

$$f \colon Gr_{r,n} \to \mathbb{R}, \quad P \mapsto \mathrm{tr}(PA). \tag{12}$$

The Riemannian gradient of f is given by

$$\nabla f(P) = [P,[P,A]]. \tag{13}$$

Moreover, cf. (Helmke & Moore, 1994),

(a) $P = uu^\top$ is a critical point of f if and only if $\tilde{A} = [u,u^\perp]^\top A [u,u^\perp]$ is blockdiagonal, i.e.

$$\tilde{A} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix},$$

where A_{11} is $(r\times r)$ and A_{22} is $(n-r)\times(n-r)$.

(b) P is a maximum of f if and only if $\mathrm{trA}_{11} = \sum_{\mathrm{i}=1}^{\mathrm{r}} \lambda_{\mathrm{i}}$, where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of A. This maximum is unique on $Gr_{r,n}$ if $\lambda_r > \lambda_{r+1}$.

In its abstract form, the CG-method for computing the r-dimensional dominant subspace of the symmetric $(n\times n)$-matrix A now reads as follows. We will provide an implementable version subsequently. Abbreviatory, denote the Riemannian gradient at P_i by $G_i := \nabla f(P_i) = [P_i,[P_i,A]]$.

CG-Sweep. Let $P_0 \in Gr_{r,n}$ be given. Set $H_0 := G_0$ and $\Omega_0 := [P_0, H_0]$. Then for $i = 0,1,\ldots$

Line-Search
The line search is done via a modified one dimensional Newton step along the curve $\alpha_{H_i}(t)$, cf. Eq. (8), where θ is chosen such that $\theta^\top \Omega \theta$ is blockdiagonal with nonzero entries only in the upper $(2r\times 2r)$-block (See later for more details). The proposed stepsize is

$$\lambda_i := \frac{\frac{d}{dt}|_{t=0} f\circ\alpha_{H_i}(t)}{|\frac{d^2}{dt^2}|_{t=0} f\circ\alpha_{H_i}(t)|}, \tag{14}$$

where, using $\Omega_i = [H_i, P_i]$,

$$\begin{aligned} &\tfrac{d}{dt}|_{t=0} f\circ\alpha_{H_i}(t) = 2\mathrm{tr}(A[\Omega_i,P_i]), \\ &\tfrac{d^2}{dt^2}|_{t=0} f\circ\alpha_{H_i}(t) = \\ &= 2\mathrm{tr}(A(\theta(\theta^\top \Omega_i^2 \theta)_{\mathrm{upp}}\theta^\top P_i - \Omega_i P_i \Omega_i)). \end{aligned} \tag{15}$$

Set $P_{i+1} = \alpha_{H_0}(\lambda_i)$.

Direction-Update
We compute the new direction $H_{i+1} \in T_{P_{i+1}} Gr_{r,n}$ according to a Riemannian adaption of the Hestenes-Stiefel-Formula, cf. (Nocedal & Wright, 1999), namely

$$H_{i+1} = -G_{i+1} + \gamma_i \tau H_i, \tag{16}$$

$$\text{where} \quad \gamma_i = \frac{\mathrm{tr}\Big(G_{i+1}(G_{i+1} - \tau G_i)\Big)}{\mathrm{tr}\Big(\tau H_i (G_{i+1} - \tau G_i)\Big)} \tag{17}$$

$$\text{and} \quad \tau H_i := \tau(H_i,\lambda_i), \tau G_i := \tau(G_i,\lambda_i).$$

2.1 *On the convergence of the dominant subspace algorithm*

It is easily seen that every critical point of the cost function is a stationary point of the algorithm. However, the step size is chosen in a way, that all critical points except the maximum are not attractive. To see this, we first investigate a regular Newton-step along the curve $g(t) := f\circ\alpha_H(t)$, where $\alpha_H(0) =: P$ and $H \in T_P Gr_{r,n}$. Let $Z := g(t_*)$ be a critical point of g. The following lemma translates the well known local convergence results for the one dimensional Newton method to the problem considered here.

Lemma 1. Let $|t_*|$ be sufficiently small. A regular Newton-step along the curve α_H, given by

$$N(\alpha_H(t)) = \alpha_H\left(t - \frac{g'(t)}{g''(t)}\right), \tag{18}$$

reduces the distance from P to Z. Under the assumption that $g''(t_*) \neq 0$, this reduction is even quadratic, i.e. $\|N(P) - Z\| \leq C\|P - Z\|^2$ for some constant $C \geq 0$.

Proof. The derivative of N is given by

$$\dot{N}(\alpha_H(t)) = \dot{\alpha}_H\left(t - \frac{g'(t)}{g''(t)}\right)\left(1 - \frac{(g''(t))^2 + g'(t)g'''(t)}{(g''(t))^2}\right). \tag{19}$$

Thus, by a Taylor argument, using $N(Z) = Z$,

$$\begin{aligned}\|N(P) - Z\| &= \|N(\alpha_H(0)) - N(\alpha_H(t_*))\| \\ &= \|\dot{N}(\alpha_H(t_*)) + \tfrac{1}{2}\ddot{N}(\alpha_H(\xi))t_*^2\| \\ &\leq C_1 t_*^2,\end{aligned} \tag{20}$$

where $C_1 := \max_{\xi\in[0,t_*]} \|\frac{1}{2}\ddot{N}(\alpha_H(\xi))\|$. The approximation property of α_H, cf. Eq. (9), now yields

$$\begin{aligned}|t_*| &= \|\gamma_H(0) - \gamma_H(t_*)\| \\ &= \|P - \alpha_H(t_*) + O(t_*^2)\| \leq \|P - Z\| + O(t_*^2).\end{aligned} \tag{21}$$

It follows, that for small $|t_*|$ there is a constant $C_2 > 0$ (independent of t_*) such that

$$|t_*| \leq C_2\|P - Z\| \tag{22}$$

and hence

$$\begin{aligned}\|N(P) - Z\| &\leq C_1 t_*^2 \\ &\leq C_1 C_2^2 \|P - Z\|^2 \\ &\leq C\|P - Z\|^2,\end{aligned} \tag{23}$$

where $C := C_1C_2^2$. In the case where t_* is a higher order zero of g, i.e. $g^{(l)}(t_*) = 0$ for $l = 2, \ldots, n$ and $g^{(n+1)}(t_*) \neq 0$, a standard argument from the one-dimensional Newton iteration applies in a similar way, leading to linear reduction of the distance between P and Z.

The non-attractiveness of the critical points Z that are not maxima is easily verified: Since Z is not a maximum, there exists a direction $H \in T_Z Gr_{r,n}$ and some $\delta > 0$ such that $g(t) := \alpha_H(t)$, $\alpha_H(0) = Z$ is strictly convex for $t \in (-\delta, \delta)$. Now let $P := \alpha_H(\delta/2)$. Since $g''(\delta/2) > 0$, the stepsize (14) is exactly the negative of the regular Newton step. Since the latter one reduces the distance between P and Z if δ is chosen sufficiently small, stepsize (14) enlarges the distance. Note that the above statements do not guarantee global convergence of the algorithm, nor do they allow to deduce results of the local speed of convergence. However, the non-attractiveness of the non-maximal critical points seems to be essential for practical purposes. Indeed, if a regular Newton step is implemented instead of the one that is proposed here, the algorithm get stuck in non-maximal critical points, whereas in all experiments with stepsize (14), global convergence has been observed.

2.2 *Implementation*

Let $u \in \mathbb{R}^{n\times r}$ with $u^\top u = I$. Then $u^\perp$ is computed via a Householder QR-decomposition, i.e.

$$Q[I \quad 0]^\top = u. \tag{24}$$

This computation requires $2/3r(r^2-3rn+6n^2)$ flops, cf. (Golub & van Loan, 1996), and $Q = [u, u^\perp]$.

Lemma 2
Let $P = uu^\top$, let Q as in Eq. (24) and $H \in T_P Gr_{r,n}$. Then there exist r Householder transformations $\tilde{\theta} := \theta_r\theta_{r-1}\ldots\theta_1$ such that for $\Omega := [H, P]$ we obtain the block structure

$$\Omega = Q\tilde{\theta}\begin{bmatrix}0 & -R^\top & 0\\ R & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\tilde{\theta}^\top Q^\top, \tag{25}$$

with $R \in \mathbb{R}^{r\times r}$ upper triangular. The computation of the sequence $\theta_r\theta_{r-1}\ldots\theta_1$ requires $2rn^2 + r^2n - 8/3r^3$ flops.

Proof
Let $s := Hu \in \mathbb{R}^{n\times r}$ (requires n^2r flops), $\tilde{s} := (u^\perp)^\top s$ (requires $(n-r)nr$ flops) and let

$$\tilde{\theta}[R \quad 0]^\top = \tilde{s} \tag{26}$$

be the QR-decomposition of $\tilde{s}$. Storing $\tilde{\theta}$ as a sequence of r Householder transformations, the computation of Eq. (26) requires $2r^2((n-r) - r/3)$ flops. Now

$$\begin{aligned}Q^\top\Omega Q &= Q^\top(Huu^\top - uu^\top H)Q \\ &= Q^\top(su^\top - us^\top)Q \\ &= [u, u^\perp]^\top su^\top[u, u^\perp] - [u, u^\perp]^\top us^\top[u, u^\perp] \\ &= \begin{bmatrix}u^\top s & 0\\ (u^\perp)^\top s & 0\end{bmatrix} - \begin{bmatrix}s^\top u & s^\top u^\perp\\ 0s & 0\end{bmatrix} \\ &= \begin{bmatrix}0 & -s^\top u^\perp\\ (u^\perp)^\top s & 0\end{bmatrix}\begin{bmatrix}0 & -\tilde{s}^\top\\ \tilde{s} & 0\end{bmatrix},\end{aligned} \tag{27}$$

since $s^\top u = u^\top s$. Hence altogether $2rn^2 + r^2n - 8/3r^3$ flops are needed to compute

$$\Omega = \theta\begin{bmatrix}0 & -R^\top & 0\\ R & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\theta^\top, \quad \text{with} \quad \theta := Q\begin{bmatrix}I & 0\\ 0 & \tilde{\theta}\end{bmatrix}. \quad \square \tag{28}$$

We use the above results and implement an intrinsic CG-method of order $O(n^2r^3)$ for computing the r-dimensional dominant subspace of a symmetric $(n \times n)$-matrix A. The algorithm stops if the norm of the gradient is small. As an easy norm to compute for a matrix G, we choose

$$\|G\|_\infty := \max_{i,j} |G_{ij}|. \tag{29}$$

Let I_r denote the $(r \times r)$-identity matrix and let the blocks of A be denoted as

$$A = \begin{bmatrix}A_{11} & A_{21}^\top\\ A_{21} & A_{22}\end{bmatrix},$$

where $A_{11} \in \mathbb{R}^{r\times r}$, $A_{22} \in \mathbb{R}^{(n-r)\times(n-r)}$. For some squared matrix X, $\mathrm{diag}(X)$ is a diagonal matrix

having the same diagonal entries than X. Moreover, X_{ij} denotes the (i,j)-entry of X.

Algorithm
function $[u, u^\perp] = \mathbf{domsub}(A, r)$

$u \leftarrow \begin{bmatrix} I_r \\ 0 \end{bmatrix}$; $u^\perp \leftarrow \begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}$; $H \leftarrow \begin{bmatrix} 0 & A_{21}^\top \\ A_{21} & 0 \end{bmatrix}$;
$b \leftarrow \begin{bmatrix} 0 \\ A_{21} \end{bmatrix}$;
while 1
$s \leftarrow Hu$; $\tilde{s} \leftarrow (u^\perp)^\top s$; $\left(\tilde{\theta}, \begin{bmatrix} R \\ 0 \end{bmatrix}\right) \leftarrow QR$-dec. of $\tilde{s}$; $\theta \leftarrow [u, u^\perp] \begin{bmatrix} I_r & 0 \\ 0 & \tilde{\theta} \end{bmatrix}$;
% Next, we compute $\Delta := (\theta^\top \Omega^2 \theta)_{\mathrm{upp}}$ *as in Eq. (15).*

$$S \leftarrow \begin{bmatrix} -R^\top R & 0 & 0 \\ 0 & -RR^\top & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad \Delta \leftarrow \mathrm{diag}(S)$$

for $i = 1 : 2r$
for $j = i + 1 : 2r$
$\Delta_{ij} \leftarrow 2S_{ij}$;
end
end

$$\lambda \leftarrow \frac{\mathrm{tr}\Big((b^\top (s - us^\top u)\Big)}{\Big|\mathrm{tr}\Big(b^\top \theta \Delta \theta^\top u\Big) - \mathrm{tr}\Big((s - us^\top u)^\top (bs^\top u - As)\Big)\Big|};$$

$(\theta_2, R_2) \leftarrow QR$-decomposition of $I_{2r} + \lambda \begin{bmatrix} 0 & -R^\top \\ R & 0 \end{bmatrix}$;
$\Theta \leftarrow \theta \begin{bmatrix} \theta_2 & 0 \\ 0 & I_{n-2r} \end{bmatrix} \theta^\top$;
$\tau G \leftarrow \Theta(ub^\top + bu^\top - 2uu^\top bu^\top)\Theta^\top$; $u \leftarrow \Theta u$;
$\left([u, u^\perp], \begin{bmatrix} I_r \\ 0 \end{bmatrix}\right) \leftarrow QR$-dec. of u;
$b \leftarrow Au$; $\tau\Omega \leftarrow \Theta s u^\top - u(\Theta s)^\top$;
$\tau H \leftarrow \tau\Omega uu^\top + u(\tau\Omega u)^\top$;
$G \leftarrow ub^\top + bu^\top - 2uu^\top bu^\top$;
if $\|G\|_\infty < thresh$; **break**;

$$\gamma \leftarrow \frac{\mathrm{tr}\Big(G(G - \tau G)\Big)}{\mathrm{tr}\Big(\tau H(G - \tau G)\Big)}; \quad H \leftarrow -G + \gamma \tau H;$$

end

Note, that it is not required to explicitly compute θ nor θ_2, but rather storing the product of $[u, u^\perp]$ and a sequence of Housholder matrices, cf. Eq. (25).

2.3 *Discussion of the dominant subspace algorithm*

The above algorithm for computing the signal and the noise space of a symmetric matrix is, as well as any algorithm for computing the complete eigendecomposition, *not* a finite algorithm, i.e. the dominant subspace can not be computed in finitely many steps. We therefore have to check that the number of iterations until a certain stopping criterion is achieved is independent of the size of the initial matrix. To that end, several simulations are presented for different matrices that support this independence. In Figure 1, the dominant

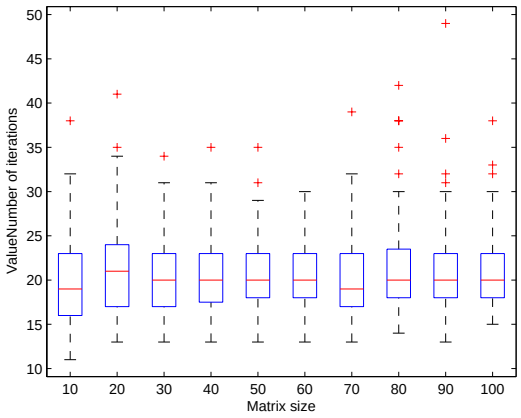

Fig. 1. Boxplot illustration of the independency of the dominant subspace algorithm of the size n of the matrix for $n = 1, 2, 3..., 10$.

subspace algorithm for computing the 2-dimensional signal space is applied to ten matrices of size $(n \times n)$, $n = 1, 2, 3, ..., 10$. In each case the matrix A has been randomly generated 100 times via

$$A = XX^\top + \varepsilon, \tag{30}$$

where X is randomly chosen in $\mathbb{R}^{n\times 2}$ and ε is a symmetric matrix with entries uniformly distributed on the open interval $(0, 1)$. As a measure of quality for how good the computed subspaces u and $u^\perp$ approximate the signal and the noise space, we make use of the fact that for the true noise space u_s

$$[u_s, u_s^\perp] A [u_s, u_s^\perp]^\top = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix}. \tag{31}$$

Let

$$[u, u^\perp] A [u, u^\perp]^\top =: \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ \tilde{B}_{21} & \tilde{B}_{22} \end{bmatrix} \tag{32}$$

As a test function we choose

$$t(u) = \log(\|\tilde{B}_{12}\|), \tag{33}$$

where $\|\tilde{B}_{12}\|$ is the Frobenius norm of $\tilde{B}_{12}$. The simulations support that, independently of the size of A, the change in the number of iteration to compute the signal and noise subspace is very small. Indeed, it can be observed from the boxplot in Figure 1 that the variations in the integer median, lower quartile (25^{th} percentiles) and upper quartile (75^{th} percentiles) of the number of iterations is very small. In Figure 2, the dominant subspace algorithm for computing the k-dimensional signal space for $k = 2, 5$ is applied to three matrices of size $(10 \times 10), (100 \times 100)$ and (1000×1000) to illustrate the speed of convergence. This corresponds to $t(u)$ versus the number of iteration averaged over 10 cases. Again, it can be observed that the grow in required iterations is very slow.

3. CONCLUSION

A number of subspace algorithms have been proposed in signal processing applications. Some of these algorithms are based on diagonalizing an $(M \times M)$

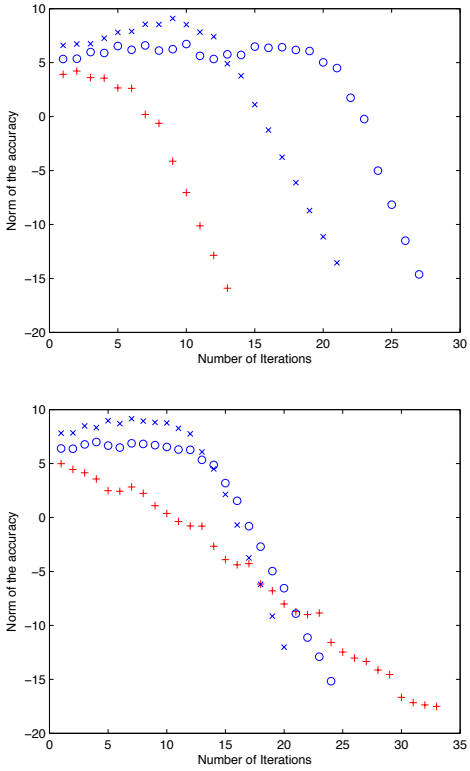

Fig. 2. Illustration of the convergence speed of the proposed dominant subspace algorithm for computing the k-dimensional signal space for $k = 2$ top, 3, 4, $k = 5$ bottom for matrices of size: +: $n = 10$, o: $n = 100$, x: $n = 1000$ averaged over 10 cases.

symmetric data matrix of known rank to generate the signal and noise subspaces which are essential in computing the parameter estimates of interest. This leads in the best case to a computational complexity of order $O(M^3)$. A novel approach to compute the signal and noise subspaces of a data matrix of known rank r has been proposed. Based on the Knowledge of the dimension of the signal subspace, the proposed approach uses concept from geometrical optimization to compute the basis that spans the noise subspace rather than using diagonalization. This has advantage of reducing the computational complexity of these subspace algorithms. Indeed, when applied to the problem of sensor positions estimation with the full-set subspace algorithm (Chan, *et al.*, 2009) and the multidimensional scaling (MDS)(Shang, *et al.*, 1984) the proposed approach reduces the computational complexity order from $O(M^6)$ to $O(M^4)$ and from $O(M^3)$ to $O(M^2)$ respectively without loosing accuracy. This makes this approach particularly suited for real-time applications. The performance of the proposed approach applied to the full-set subspace algorithm was analyzed in a simulation example in term of CPU computation time.

4. REFERENCES

Absil, P.A., R. Mahony, R. Sepulchre, and P. van Dooren (2002). A Grassmann–Rayleigh quotient iteration for computing invariant subspaces. *SIAM Rev.*, **44(1)**, 57–73.

Chan, F.K.W., H.C. So, and W.K. Ma (2009). A novel subspace approach for cooperative localization in wireless sensor networks using range measurements. *IEEE Transactions on Signal Processing*, **57(1)**, 260–269.

Fu, Z. and E.M. Dowling (1996). Conjugate gradient methods for solving the smallest eigenpair of large symmetric eigenvalue problems. *Int. J. Numerical Methods in Engineering*, **39(13)**, 2209–2229.

Fu, Z. and E.M. Dowling (1995). Conjugate Gradient Eigenstructure Tracking for Adaptive Spectral Estimation. *IEEE Trans. on Signal Processing*, **43(5)**, 1151–1160.

Gabay, D. (1982). Minimizing a Differentiable Function over a Differential Manifold. *J. of Optimization Theory and Applications*, **37(2)**, 177–219.

Golub, G.H. and C.F. van Loan (1996). *Matrix Computations*, The John Hopkins University Press, Baltimore.

Helmke, U. and J.B. Moore (1994). *Optimization and Dynamical Systems.* Springer, London.

Helmke, U., K. Hüper, and J. Trumpf (2007). Newton's method on Graßmann manifolds. *arXiv.org*, arXiv:0709.2205v2.

Kleinsteuber, M. and K. Hüper (2007). An intrinsic CG-algorithm for computing dominant subspaces. *Proc. of the IEEE International Conference on Acoustic, Speech and Signal Processing*, **4**, 1405–1408.

Nocedal, J. and S.J. Wright (1999). *Numerical Optimization. Springer Series in Operations Research.* Springer, New York.

Parlett, B.N. (1974). The Rayleigh quotient iteration and some generalizations for nonnormal matrices. *Math. Comp.*, **28**, 679–693.

Shang, Y., W. Ruml, Y. Zhang, and M. Fromherz (1984). Localization from connectivity in sensor networks. *IEEE Transactions on Parallel Distributed Systems*, **15(11)**, 961–974.

So, H.C. and F.K.W. Chan (2007). A generalized subspace approach for mobile positioning with time-of-arrival measurements. *IEEE Transactions on Signal Processing*, **55(10)**, 5103–5107.

Wang, Q., Y.J. Luo, W.L. Yang, J. Xu, J. Tang, and Y.N. Peng (2005). Mobile localization method based on multidimansional similarity analysis. *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, **4**, 1081–1084.

Yang, X., T.K. Sarkar, and E. Arvas (1989). A Survey of Conjugate Gradient Algorithms for Solution of Extreme Eigen-Problems of a Symmetric Matrix. *IEEE Transactions on Acoustic, Speech and Signal Processing*, **37(10)**, 1550–1556.