A Projection Method for Optimization Problems on the Stiefel Manifold

Oscar Dalmau-Cedeño and Harry Oviedo(✉)

Mathematics Research Center, CIMAT A.C., Guanajuato, Mexico
{dalmau,harry.oviedo}@cimat.mx

Abstract. In this paper we propose a feasible method based on projections using a curvilinear search for solving optimization problems with orthogonality constraints. Our algorithm computes the SVD decomposition in each iteration in order to preserve feasibility. Additionally, we present some convergence results. Finally, we perform numerical experiments with simulated problems; and analyze the performance of the proposed methods compared with state-of-the-art algorithms.

Keywords: Constrained optimization · Orthogonality constraints · Non-monotone algorithm · Stiefel manifold · Optimization on manifolds

1 Introduction

In this paper we consider the following optimization problem with orthogonality constraints:

$$\min_{X \in \mathbb{R}^{n \times p}} \mathcal{F}(X) \quad \text{s.t.} \quad X^\top X = I_p, \tag{1}$$

where $\mathcal{F} : \mathbb{R}^{n \times p} \to \mathbb{R}$ is a differentiable function and $I_p \in \mathbb{R}^{p \times p}$ represents the identity matrix. The feasible set $Stf(n,p) := \{X \in \mathbb{R}^{n \times p} | X^\top X = I\}$ is known as the "Stiefel Manifold". This manifold is simplified to the unit sphere when $p = 1$ and in the case $p = n$ is called "Orthogonal group". The Stiefel manifold can be seen as an embedded sub-manifold of $\mathbb{R}^{n \times p}$ with dimension equals to $np - \frac{1}{2}p(p+1)$, see [1].

Problem (1) admits many applications such as, linear eigenvalue problem [14], sparse principal component analysis [4], Kohn-Sham total energy minimization [16], orthogonal procrustes problem [5], weighted orthogonal procrustes problem [6], nearest low-rank correlation matrix problem [7,12], joint diagonalization (blind source separation) [8], among others. In addition, some problems such as PCA, LDA, multidimensional scaling, orthogonal neighborhood preserving projection can be formulated as problem (1) [9].

On the other hand, the Stiefel manifold is a compact set, which ensures that (1) has a global optimum at least. However, this manifold is not a convex set, which transforms (1) in a hard optimization problem. For example, the *quadratic assignment problem* (QAP) and the *leakage interference minimization* are NP-hard [10].

© Springer International Publishing AG 2017
J.A. Carrasco-Ochoa et al. (Eds.): MCPR 2017, LNCS 10267, pp. 84–93, 2017.
DOI: 10.1007/978-3-319-59226-8_9

In this paper we propose a new method based on projections onto the Stiefel manifold. In particular, we study two algorithms to solve problem (1). At each iteration of the algorithms, we project the corresponding update onto the Stiefel manifold using the singular value decomposition (SVD) which guarantees to obtain a feasible sequence. Although, the SVD decomposition is computationally expensive, this is less expensive than building a geodesic. In the literature, we can find other feasible methods that solve problem (1), for example, the ones based on retractions methods use projections that involve QR factorization, polar decomposition, Gram-Schmidt process or SVD decomposition [1].

This paper is organized as follows. In Subsect. 2.1 we present some standard notation and in Subsect. 2.2 we give the optimality conditions of the problem (1), Subsect. 2.3 describes the proposed update scheme, where we present a linear search monotone algorithm and a globally convergent non-monotone algorithm for solving problem (1), Subsect. 2.4 shows different strategies to choose the step size according to Armijo-Wolfe like condition, and a non-monotone search using the Barzilai Borwein step size. Some theoretical results are presented in Sect. 3. Section 4 is dedicated to numerical experiments in order to demonstrate the efficiency and robustness of the proposed algorithms.

2 Algorithms

In the first two subsections, we introduce some standard notation and the optimality conditions of problem (1) respectively. Next subsections are devoted to introduce our proposed method.

2.1 Notation

We say that a matrix $W \in \mathbb{R}^{n\times n}$ is skew-symmetric if $W = -W^\top$. The trace of X is defined as the sum its diagonal elements, and we will denote by $Tr[X]$. The Euclidean inner product of two matrices $A, B \in \mathbb{R}^{m\times n}$ is defined as $\langle A, B\rangle := \sum_{i,j} A_{i,j}B_{i,j} = Tr[A^\top B]$. The Frobenius norm is defined using the previous inner product, i.e., $||A||_F = \sqrt{\langle A, A\rangle}$. Let $\mathcal{F} : \mathbb{R}^{n\times p} \to \mathbb{R}$ be a differentiable function, then the derivative of $\mathcal{F}$ with respect to X is denoted as $G := \mathcal{DF}(X) := (\frac{\partial \mathcal{F}(X)}{\partial X_{ij}})$ and the derivative of the function $\mathcal{F}$ in X in the direction Z is defined as:

$$\mathcal{DF}(X)[Z] := \frac{\partial \mathcal{F}(X + tZ)}{\partial \tau}\Big|_{t=0} = \lim_{t\to 0} \frac{\mathcal{F}(X + tZ) - \mathcal{F}(X)}{t} = \langle \mathcal{DF}(X), Z\rangle. \quad (2)$$

2.2 Optimality Conditions

The Lagrangian function associated to the optimization problem (1) is given by:

$$\mathcal{L}(X, \Lambda) = \mathcal{F}(X) - \frac{1}{2} Tr[\,\Lambda(X^\top X - I_p)\,], \quad (3)$$

where I_p is the identity matrix and Λ is the Lagrange multipliers matrix, which is symmetric due to the matrix $X^\top X$ is also symmetric. The Lagrangian function leads to the first order optimality conditions for problem (1):

$$G - X\Lambda = 0 \tag{4a}$$
$$X^\top X - I_p = 0. \tag{4b}$$

Lemma 1 *(cf. Wen and Yin [15]). Suppose that X is a local minimizer of problem (1). Then X satisfies the first order optimality conditions (4a) and (4b) with the associated Lagrangian multiplier $\Lambda = G^\top X$. Defining $\nabla\mathcal{F}(X) := G - XG^\top X$ and $A := GX^\top - XG^\top$. Then $\nabla\mathcal{F}(X) = AX$. Moreover, $\nabla\mathcal{F} = 0$ if and only if $A = 0$.*

Proof. See [15].

The Lemma 1 establishes an equivalence to the (4a) and (4b) conditions, i.e., if $X \in Stf(n,p)$ satisfies that $\nabla\mathcal{F}(X) = 0$ then X also satisfies (4a) and (4b), so we can use this result as a stopping criterion for our algorithms.

2.3 Update Schemes

In this subsection we present a linear combination based algorithm. As the new iterated of our proposals does not necessarily belong to the Stiefel Manifold, we use a projection operator, in order to force the feasibility of the new iterated. Specifically, we use the classical projection operator which is defined as $\pi(X) := \arg\min_{Q \in Stf(n,p)} ||X - Q||_F^2$, it is known that the solution of this problem is given by $\pi(X) = UI_{n,p}V^\top$ where $X = U\Sigma V^\top$ is the SVD decomposition of X, for details of the demonstration of this result see [11].

In our updating formula, we use the previous result for obtaining a new point that satisfies the constraints of the problem (1). For example, if $Y_k(\tau)$ is obtained from our proposal, i.e., the linear combination scheme, then the new test point is:

$$X_{k+1} := Z_k(\tau) := \pi(Y_k(\tau)). \tag{5}$$

In the next subsections we explain in more detail our updating formula $Y_k(\tau)$.

A Scheme Based on a Linear Combination. Our proposal uses the following update formula:

$$Y_k^{CL}(\tau) := X_k - \tau\,(\,\lambda B_k L + \mu C_k R\,), \tag{6}$$

where $G_k = \mathcal{D}\mathcal{F}(X_k)$, $B_k = G_k L^\top - LG_k^\top$, $C_k = G_k R^\top - RG_k^\top$, $L, R \in \mathbb{R}^{n\times p}$, τ is the step size and $(\lambda,\, \mu)$ are any two scalars satisfying:

$$\lambda||B_k||_F^2 + \mu||C_k||_F^2 \;>\; 0.$$

The following lemma shows that the curve $Y_k^{CL}(\cdot)$ defined by Eq. (6) is a descent curve at $\tau = 0$.

Lemma 2. *Let $Y_k^{CL}(\tau)$ be defined by Eq. (6), then $Y_k^{CL}(\tau)$ is a descent curve at $\tau = 0$, i.e.,*

$$\mathcal{DF}(X_k)[\dot{Y}_k^{CL}(0)] = -\frac{\lambda}{2}||B_k||_F^2 - \frac{\mu}{2}||C_k||_F^2 < 0. \tag{7}$$

Proof. The proof is straightforward, and it can be obtained by using trace properties and using Eq. (2).

Remark 1. Note that in the updating formula (6), we can select any matrix L or R, in particular one can use matrices L, R with random entries. The parameters (λ, μ) can appropriately selected, for example, we can choose both positive. This ensures that the method will descent and may eventually converge to a local minimum. In our implementation, we select $L = X_k$, $R = X_{k-1}$ and $(\lambda, \mu) = (2/3, 1/3)$.

2.4 Strategies to Select the Step Size

From now on, $Y_k(\tau)$ represents our proposal, i.e., the based on the linear combination method.

A Descent Condition. In our method, we will choose the biggest step size τ that satisfies the following condition:

$$\mathcal{F}(Z_k(\tau)) \leq \mathcal{F}(X_k) + \sigma\tau Tr[G_k^\top \dot{Y}_k(0)], \tag{8}$$

with $0 < \sigma < 1$.

Note that Eq. (8) is not exactly the classic "*Armijo condition*", since we use $\dot{Y}_k(0)$ instead of $\dot{Z}_k(0)$. However, if we only use the condition (8) for computing the step size, it ensures the descent of the objective function as long as the directional derivative $Tr[\mathcal{DF}(X_k)^\top \dot{Y}_k(0)]$ is negative. In this work, we also study the behavior of our algorithms calculating the step size as satisfying (8).

Nonmonotone Search with Barzilai Borwein Step Size. It is known that the *Barzilai-Borwein* (BB) step size, see [2], can sometimes improve the performance of linear search algorithms such as the steepest descent method without adding too much computational cost. This technique considers the classic *steepest descent method* and proposes to use any of the following step sizes:

$$\alpha_k^{BB1} = \frac{||S_k||_F^2}{Tr[S_k^\top R_k]} \quad \text{and} \quad \alpha_k^{BB2} = \frac{Tr[S_k^\top R_k]}{||R_k||_F^2}. \tag{9}$$

where $S_k = X_{k+1} - X_k$, $R_k = \mathcal{DF}(X_{k+1}) - \mathcal{DF}(X_k)$ and the matrix $B(\alpha) = (\alpha I)^{-1}$, is considered an approximation of the Hessian of the objective function. For more details see [2,13].

Since the quantities α_k^{BB1}, α_k^{BB2} could be negatives, the absolute value of these step sizes is usually considered. On the other hand, the BB-steps do not

necessarily guarantee the descent of the objective function at each iteration, this may imply that the method does not converge. In order to solve this problematic, we use a technique that guarantees global convergence, see Refs. [3,13] for details. In particular, we use a non-monotone line search algorithm, see [17], combined with the BB-step in order to select the step size, see Algorithm 1.

Algorithm 1. Non-monotone linear search algorithm for solve optimization problems on Stiefel manifold

Require: $X_0 \in Stf(n,p)$, $\tau > 0$, $0 < \tau_m \ll \tau_M$, $\sigma, \epsilon, \eta, \delta \in (0,1)$, $X_{-1} = X_0$, $C_0 = \mathcal{F}(X_0)$, $Q_0 = 1$, $k = 0$.
Ensure: X^* a local minimizer.
1: **while** $||\nabla \mathcal{F}(X_k)||_F > \epsilon$ **do**
2: **while** $\mathcal{F}(Z_k(\tau)) \geq C_k + \sigma\tau\mathcal{D}\mathcal{F}(X_k)[\dot{Y}_k(0)]$ **do**
3: $\tau = \delta\tau$,
4: **end while**
5: $X_{k+1} = Z_k(\tau) := \pi(Y_k(\tau))$, with $Y_k(\tau)$ using (6).
6: Calculate $Q_{k+1} = \eta Q_k + 1$ and $C_{k+1} = (\eta Q_k C_k + \mathcal{F}(X_{k+1}))/Q_{k+1}$.
7: Choose $\tau = |\alpha_k^{BB1}|$ or well $\tau = |\alpha_k^{BB2}|$, where α_k^{BB1} and α_k^{BB2} are defined as in (9).
8: Set, $\tau = \max(\min(\tau, \tau_M), \tau_m)$.
9: $k = k + 1$.
10: **end while**
11: $X^* = X_k$.

Note that when $\eta = 0$, Algorithm 1 is reduced to a monotonous algorithm which generates points satisfying the descent condition (8).

3 Theoretical Results

In this section we prove some convergence results of our Algorithm 1 when it's use with $\eta = 0$.

Lemma 3. *Let $\{X_k\}$ be an infinite sequence generated by Algorithm 1. Then $\{\mathcal{F}(X_k)\}$ is a convergent sequence. Moreover any accumulation point X_* of $\{X_k\}$ is feasible, i.e., $X_*^\top X_* = I$.*

Proof. By construction of the Algorithm 1 we have,

$$\mathcal{F}(X_{k+1}) \leq \mathcal{F}(X_k) + \sigma\tau_k Tr[G_k^\top \dot{Y}_k(0)], \qquad \forall k \tag{10}$$

or equivalently,

$$\begin{aligned} \mathcal{F}(X_k) - \mathcal{F}(X_{k+1}) &\geq -\sigma\tau_k Tr[G_k^\top \dot{Y}_k(0)], \qquad \forall k \\ &> 0 \qquad \text{(due } Y_k(\tau) \text{ is a descent curve at } \tau = 0\text{)}, \end{aligned}$$

so, $\{\mathcal{F}(X_k)\}$ is a monotonically decreasing sequence. Now, since Stiefel manifold is a compact set and $\mathcal{F}$ is a continuous function, we obtain that $\mathcal{F}$ has maximum and minimum on *Stf(n,p)*. Therefore, $\{\mathcal{F}(X_k)\}$ is bounded, and then $\{\mathcal{F}(X_k)\}$ is a convergent sequence.

On the other hand, let $\{X_k\}_{k\in\mathcal{K}}$ be a convergent subsequence of $\{X_k\}$ and suppose that this subsequence converges to X_*, that is $\lim_{k\in\mathcal{K}} X_k = X_*$, since X_k is a feasible point for all $k \in \mathcal{K}$ and *Stf(n,p)* is a compact set, then we have $X_* \in$ *Stf(n,p)*, i.e.,

$$X_*^\top X_* = I,$$

therefore every accumulation point is feasible.

Theorem 1. *Let $\{X_k\}$ be an infinite sequence generated by Algorithm 1. Then any accumulation point X_* of $\{X_k\}$ satisfies the the first order optimality conditions.*

The proof of Theorem 1 is obtained by following the ideas of the demonstration of Theorem 4.3.1 in [1] except for slight adaptations.

4 Numerical Experiments

In this section we analyze the performance of our method by solving several simulated experiments with the format of the problem (1), for different objective functions and different sizes of problems. We also make comparisons between some state of the art methods and our proposal, in order to measure the performance and efficiency of our algorithms.

4.1 Implementation Details

All our experiments were performed using Matlab R2013a on an Intel processor i3-380M, 2.53 GHz CPU with 500 Gb HD and 8 Gb of Ram. For the different parameters of our two algorithms, we use the following values: initial step size $\tau = 1\text{e}{-2}$, $\sigma = 1\text{e}{-4}$, $\eta = 0.85$, $\delta = 0.1$. Moreover, as the convergence of the first-order methods (methods using the first derivative of the objective function) can be very slow we will use several stop criteria:

$$||\nabla\mathcal{F}(X_k)||_F < \epsilon, \quad \text{and} \quad (tol_k^x < xtol \wedge tol_k^f < ftol), \tag{11}$$

and a maximum of K iterations, where

$$tol_k^x := \frac{||X_{k+1} - X_k||_F}{\sqrt{n}}, \quad \text{and} \quad tol_k^f := \frac{\mathcal{F}(X_k) - \mathcal{F}(X_{k+1})}{|\mathcal{F}(X_k)| + 1}.$$

In the experiments, we used the following default values: $xtol = 1\text{e}{-6}$, $ftol = 1\text{e}{-12}$, $T = 5$ and $\epsilon = 1\text{e}{-4}$.

In all experiments presented in the following subsections we use the following notation:

- *Nfe*: The number of evaluations of the objective function.
- *Nitr*: The number of iterations performed by the algorithm to convergence.
- *Time*: The time (in seconds) used by the algorithm to converge.
- *NrmG*: The gradient norm of the Lagrangian function with respect to primal variables evaluated at the estimated "optimal".
- *Fval*: Evaluation of the objective function at the estimated "optimal".
- *Feasi*: Corresponds to the following error $||\hat{X}^\top \hat{X} - I_p||_F$, where $\hat{X}$ denotes the "optimal" estimated by the algorithm.

In addition, we denote by the Steepest Descent *Steep-Dest*, the Trust-Region method *Trust-Reg* and the Conjugate Gradient method *Conj-Grad* from "*manopt*" toolbox[1], and *PGST* the algorithm presented in [6]. On the other hand, *Linear-Co* denote our Algorithm 1.

4.2 Weighted Orthogonal Procrustes Problem (WOPP)

Let $X \in \mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times q}$ and $C \in \mathbb{R}^{n\times q}$. The *Weighted Orthogonal Procrustes Problem* (WOPP) consists in solving the following constrained optimization problem:

$$\begin{aligned} \min_{X\in\mathbb{R}^{m\times n}} \ & \tfrac{1}{2}||AXC - B||_F^2 \\ \text{s. t.} \quad & X^\top X = I_n. \end{aligned} \tag{12}$$

When C is the identity matrix with appropriate dimensions, this problem is known as *Unbalanced Orthogonal Procrustes Problem* (UOPP), for more details see [1].

Experiments with WOPP Problems. The problems in this subsection were taken from [18]. In particular, we considered $n = q$, $p = m$, $A = PSR^\top$ and $C = Q\Lambda Q^\top$, where P, Q and R are orthogonal matrices generated randomly with $Q \in \mathbb{R}^{n\times n}$, $R, P \in \mathbb{R}^{m\times m}$, $\Lambda \in \mathbb{R}^{n\times n}$ is a diagonal matrix with entries generated from a uniform distribution in the range $[\frac{1}{2}, 2]$ and S is a diagonal matrix defined for each type of problem, see below for details. As a starting point $X_0 \in \mathbb{R}^{m\times n}$, we generated random matrices on the Stiefel manifold. When not specified, the entries of the matrix were generated using a *standard Gaussian* distribution.

For comparison purposes, we created problems with a known solution $Q_* \in \mathbb{R}^{m\times n}$ randomly selected on the Stiefel manifold. Then, we built the matrix B as $B = AQ_*C$. Finally, for the different tested problems the diagonal matrix S is described below.

Problem 1: The diagonal elements of S were generated by a normal distribution in the interval [10,12].

[1] The tool-box manopt is available in http://www.manopt.org/.

Problem 2: The diagonal of S is given by $S_{ii} = i + 2r_i$, where r_i was a random number uniformly distributed in the interval $[0, 1]$.

For each experiment, a total of 300 WOOP's problems were built with the matrix S generated according to problems **Problem 1** and **Problem 2** respectively. The maximum number of iterations, for all methods, was $K = 8000$.

The results of the previous experiments are presented in Tables 1 and 2. We denote by *Error* to the standard error with respect to the global solution Q_*, i.e., $||\hat{X} - Q_*||_F$ where $\hat{X}$ is the optimum estimated by the algorithms. Furthermore, *min*, *mean*, *max* denote the minimum, maximum and average obtained by each algorithm in the 300 runs.

According to Table 1 for well-conditioned problems, i.e., **Problem 1**, all the algorithms present similar results. Note that **PGST** obtained a lower number of iterations. In general, all the methods presented a similar performance for this type of problems. On the other hand, for ill-conditioned problems, i.e., **Problem 2**, we observe that all the method arrived to the solution Q_*, according to **NrmG**, **Fval** and **Error** measures. Moreover, our **Linear-Co** procedure obtained similar results compared with the **PGST** algorithm when $n < m$, and when $m = n$ **Linear-Co** method achieved better results that the **PGST**, see Table 2.

Table 1. Performance of the methods for well conditioned WOPP problems (**Problem 1**)

Method		**Problem 1** with m = 500 and n = 70					
		Nitr	Nfe	Time	NrmG	Fval	Error
Linear-Co	Min	48	49	2.60	1.33e−05	7.13e−13	1.44e−07
	Mean	59.7	60.7	3.72	6.10e−05	3.63e−11	1.27e−06
	Max	71	72	5.06	9.95e−05	1.34e−10	2.92e−06
PGST	Min	36	35	1.87	9.63e−06	7.96e−13	1.75e−07
	Mean	41.6	40.0	2.37	7.85e−05	3.68e−11	1.12e−06
	Max	49	42	3.22	2.23e−04	1.35e−10	2.98e−06
Method		**Problem 1** with m = 200 and n = 200					
		Nitr	Nfe	Time	NrmG	Fval	Error
Linear-Co	Min	46	47	1.77	1.35e−05	9.35e−14	4.51e−08
	Mean	53.0	54.1	2.64	6.16e−05	1.08e−11	6.40e−07
	Max	63	65	3.81	9.97e−05	3.94e−11	1.58e−06
PGST	Min	33	36	1.86	1.64e−04	2.25e−11	5.48e−07
	Mean	38.2	42.0	2.75	6.56e−04	9.62e−10	5.95e−06
	Max	43	45	3.73	9.99e−04	3.83e−09	1.55e−05

Table 2. Performance of the methods for ill-conditioned WOPP problems (**Problem 2**)

Method		**Problem 2** with m = 300 and n = 20					
		Nitr	Nfe	Time	NrmG	Fval	Error
Linear-Co	Min	2078	2133	16.36	5.20e−04	4.17e−09	2.38e−05
	Mean	4732.2	4861.3	40.25	1.01e−02	9.57e−02	8.04e−02
	Max	8000	8229	72.37	3.40e−01	9.91e−01	4.89e−01
PGST	Min	3118	2080	18.15	6.52e−05	1.59e−13	1.46e−07
	Mean	6373.1	4142.3	37.38	4.67e−01	8.66e−02	8.14e−02
	Max	8000	8478	53.75	2.62e+01	1.22	4.96e−01
Method		**Problem 2** with m = 150 and n = 150					
		Nitr	Nfe	Time	NrmG	Fval	Error
Linear-Co	Min	576	775	13.48	1.20e−04	6.74e−11	2.66e−06
	Mean	1164.1	1210.6	20.80	1.30e−03	1.06e−08	3.71e−05
	Max	1881	1945	33.56	1.29e−02	2.56e−07	2.53e−04
PGST	Min	1125	962	27.16	1.67e−04	3.66e−12	6.59e−08
	Mean	2039.6	1921.1	50.62	8.52e−04	5.50e−09	2.85e−05
	Max	3521	3558	116.36	1.00e−03	1.98e−08	8.50e−05

5 Conclusions

In this paper we proposed a feasible method for solving optimization problems with orthogonality constraints. This method is very general and was based on a linear combination of descent directions and using the same manifold framework. We are currently exploring several variants of this procedure. In order to preserve feasibility, our proposal requires to project onto the Stiefel manifold. In particular, we used the SVD decomposition in each iteration. In this work, we also presented some convergence results. Finally, in numerical experiments, the proposed algorithms obtained a competitive performance compared with some state of the art algorithms.

Acknowledgments. This work was supported in part by CONACYT (Mexico), Grant 258033.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2009)
2. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**(1), 141–148 (1988)
3. Dai, Y.H., Fletcher, R.: Projected barzilai-borwein methods for large-scale box-constrained quadratic programming. Numerische Mathematik **100**(1), 21–47 (2005)

4. d'Aspremont, A., Ghaoui, L., Jordan, M.I., Lanckriet, G.R.: A direct formulation for sparse PCA using semidefinite programming. SIAM Rev. **49**(3), 434–448 (2007)
5. Eldén, L., Park, H.: A procrustes problem on the stiefel manifold. Numerische Mathematik **82**(4), 599–619 (1999)
6. Francisco, J., Martini, T.: Spectral projected gradient method for the procrustes problem. TEMA (São Carlos) **15**(1), 83–96 (2014)
7. Grubisi, I., Pietersz, R.: Efficient rank reduction of correlation matrices. Linear Algebra Appl. **422**(2), 629–653 (2007)
8. Joho, M., Mathis, H.: Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation. In: Sensor Array and Multichannel Signal Processing Workshop Proceedings, pp. 273–277. IEEE (2002)
9. Kokiopoulou, E., Chen, J., Saad, Y.: Trace optimization and eigenproblems in dimension reduction methods. Numer. Linear Algebra Appl. **18**(3), 565–602 (2011)
10. Liu, Y.F., Dai, Y.H., Luo, Z.Q.: On the complexity of leakage interference minimization for interference alignment. In: 2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), pp. 471–475. IEEE (2011)
11. Manton, J.H.: Optimization algorithms exploiting unitary constraints. IEEE Trans. Signal Process. **50**(3), 635–650 (2002)
12. Pietersz, R., Groenen, P.J.: Rank reduction of correlation matrices by majorization. Quant. Fin. **4**(6), 649–662 (2004)
13. Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optim. **7**(1), 26–33 (1997)
14. Saad, Y.: Numerical Methods for Large Eigenvalue Problems, vol. 158. SIAM, Manchester (1992)
15. Wen, Z., Yin, W.: A feasible method for optimization with orthogonality constraints. Math. Program. **142**(1–2), 397–434 (2013)
16. Yang, C., Meza, J.C., Lee, B., Wang, L.W.: KSSOLVoa MATLAB toolbox for solving the Kohn-Sham equations. ACM Trans. Math. Softw. (TOMS) **36**(2), 10 (2009)
17. Zhang, H., Hager, W.W.: A nonmonotone line search technique and its application to unconstrained optimization. SIAM J. Optim. **14**(4), 1043–1056 (2004)
18. Zhang, Z., Du, K.: Successive projection method for solving the unbalanced procrustes problem. Sci. China Ser. A **49**(7), 971–986 (2006)