Multipliers Correction Methods for Optimization Problems Over Stiefel Manifold

Lei Wang* Bin Gao† Xin Liu‡

Abstract

We propose a class of multipliers correction methods to minimize a differentiable function over the Stiefel manifold. The proposed methods combine a function value reduction step with a proximal correction step. The former one searches along an arbitrary descent direction in the Euclidean space instead of a vector in the tangent space of the Stiefel manifold. Meanwhile, the latter one minimizes a first-order proximal approximation of the objective function in the range space of the current iterate to make Lagrangian multipliers associated with orthogonality constraints symmetric at any accumulation point. The global convergence has been established for the proposed methods. Preliminary numerical experiments demonstrate that the new methods significantly outperform other state-of-the-art first-order approaches in solving various kinds of testing problems.

AMS subject classifications: 15A18, 65F15, 65K05, 90C06 , 90C30

Key words: Stiefel manifold, orthogonality constraints, multipliers correction, proximal approximation.

1 Introduction

We focus on the matrix-variable optimization problems with orthogonality constraints:

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times p}} \quad & f(X) \\ \text{s.t.} \quad & X^\top X = I_p, \end{aligned} \tag{1.1}$$

*State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, China (wlkings@lsec.cc.ac.cn). Research is supported by the National Natural Science Foundation of China (No. 11971466).

†Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Université catholique de Louvain, Belgium ICTEAM institute at UCLouvain (bin.gao@uclouvain.be). Research is supported in part by the Fonds de la Recherche Scientifique – FNRS and the Fonds Wetenschappelijk Onderzoek – Vlaanderen under EOS Project (No. 30468160).

‡State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, China (liuxin@lsec.cc.ac.cn). Research is supported in part by the National Natural Science Foundation of China (No. 11991021, 11991020 and 11971466), Key Research Program of Frontier Sciences, Chinese Academy of Sciences (No. ZDBS-LY-7022), the National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences and the Youth Innovation Promotion Association, Chinese Academy of Sciences.

where $p \leq n$, I_p is the $p \times p$ identity matrix, and $f : \mathbb{R}^{n\times p} \longrightarrow \mathbb{R}$ is a continuously differentiable function. The feasible region, denoted by $\mathcal{S}_{n,p} := \left\{X \in \mathbb{R}^{n\times p} \mid X^\top X = I_p\right\}$, is called the Stiefel manifold.

Optimization problems over the Stiefel manifold have wide applications in scientific computing and data science. For example, in linear eigenvalue problems [9, 25, 26], energy minimization in electronic structure calculations [23, 24, 36], matrix completion [8], independent component analysis [31], Bose–Einstein condensates [34], discriminant analysis [22], dictionary learning [18], and nearest low-rank correlation matrix problems [16]. Beyond that, one can find other applications in [4, 12] and the references therein.

1.1 Existing works

Optimization problems over the Stiefel manifold have been adequately studied in recent decades. There emerge quite a few algorithms and solvers, such as, geodesic–based approaches [12, 27, 28], retraction–based approaches [1–3, 5, 19, 20, 32, 33, 36], and splitting and alternating approaches [21, 30]. We refer the interested readers to the monograph [4] and survey [17] on these methods. Recently, the authors in [15] developed two orthonormalization-free approaches, called PLAM and PCAL, which are based on the augmented Lagrangian penalty function [29] but adopt an explicit expression to update Lagrangian multipliers instead of the dual ascent step. Such approaches are particularly suitable for parallel computing due to their high scalability. PCAL was further applied to solve the energy minimization problem in electronic structure calculations [13]. More recently, an exact penalty model, which shares the same global minimizers as the original problem (1.1), was proposed in [35]. In order to solve this model, they also proposed first-order and second-order approaches which subsume PCAL as a specific implementation.

In [14], the authors proposed a new algorithmic framework which consists of two steps: the function value reduction step, which preserves the feasibility, is conducted in the Euclidean space; the correction step is nothing but a rotation on the previously obtained step. As the Lagrangian multipliers associated with orthogonality constraints are symmetric and enjoy an explicit expression $X^\top \nabla f(X)$ at any first-order stationary point of (1.1) (see [15, (2.2)]), the purpose of this correction step is to guarantee the symmetry of $X^\top \nabla f(X)$ at each iteration. In summary, three algorithms were introduced in [14] to fulfill the framework; extensive numerical results illustrated their great potential. However, this framework strictly depends on the following assumption.

Assumption 1.1. *$f(X) = h(X) + \operatorname{tr}\left(G^\top X\right)$, where $G \in \mathbb{R}^{n\times p}$ is a constant matrix and $h(X)$ is orthogonal invariant, i.e., $h(XQ) = h(X)$ holds for any $Q \in \mathcal{S}_{p,p}$. Moreover, $\nabla h(X) = H(X)X$, where $H : \mathbb{R}^{n\times p} \longrightarrow \mathbb{S}^n$ and $\mathbb{S}^n$ refers to the set of $n \times n$ symmetric matrices.*

Assumption 1.1 restricts the objective to a class of composite functions. In this case, the explicit expression $X^\top \nabla f(X)$ can be divided into two parts, including a symmetric term $X^\top H(X)X$ and a linear term $X^\top G$. Hence, it is sufficient to guarantee the symmetry of $X^\top \nabla f(X)$ in the correction step by making $X^\top G$ symmetric. To this end, one can minimize $\operatorname{tr}\left(G^\top X\right)$ in the range space of X whose finding its global minimizer is equivalent to computing a singular value decomposition.

Although quite a few practical problems—such as linear eigenvalue problem and energy minimization in electronic structure calculations—satisfy this assumption, there exist important scenarios in which Assumption 1.1 does not hold; e.g., minimizing the Brockett function

(weighted sum of eigenvalues) [4,6], joint diagonalization problems [31], and dictionary learning [18] over the Stiefel manifold.

1.2 Motivation and contribution

In this paper, we intend to address the restriction of Assumption 1.1. Specifically, we solve optimization problems over the Stiefel manifold with a general objective function. To this end, we propose multipliers correction algorithmic framework, and it contains two steps. The first step is to minimize the objective function in the Euclidean space. Gradient reflection, gradient projection and column-wise block coordinate descent algorithms proposed in [14] are similarly introduced in this step. Then we propose a novel multipliers correction step whose essential idea is to minimize a first-order proximal approximation of the objective function in the range space of the current iterate. The main computational cost of such correction step is calculating the singular value decomposition of a $p \times p$ matrix, which shares the same cost with the correction step introduced in [14]. This correction step can further reduce the function value and guarantee the symmetry of Lagrangian multipliers at any accumulation point. Remarkably, the new methods work for a much wider range of problems than those proposed in [14].

In addition, we prove the global convergence and worst case complexity of the proposed methods. Numerical experiments illustrate their effectiveness. Note that the new methods outperform some state-of-the-art first-order algorithms for optimization over the Stiefel manifold, and also work well in those instances which are out of the scope of the algorithms proposed in [14].

1.3 Notation

The Euclidean inner product of two matrices $Y_1 \in \mathbb{R}^{n\times m}$ and $Y_2 \in \mathbb{R}^{n\times m}$ is defined as $\langle Y_1,\ Y_2\rangle = \mathrm{tr}\left(Y_1^\top Y_2\right)$, where $\mathrm{tr}(B)$ is the trace of a square matrix $B \in \mathbb{R}^{m\times m}$. The Frobenius norm and 2-norm of a matrix $C \in \mathbb{R}^{n\times m}$ are denoted by $\|C\|_{\mathrm{F}}$ and $\|C\|_2$, respectively. We use $C^\dagger$ to represent the pseudo-inverse of C. C_i and C_{ij} stand for the i-th column and (i,j)-th element of C, respectively. $C_{\bar{i}} \in \mathbb{R}^{n\times(m-1)}$ refers to the matrix C removing its i-th column, namely, $C_{\bar{i}} = [C_1,\ \ldots,\ C_{i-1},\ C_{i+1},\ \ldots,\ C_m]$. $C_{i,v} \in \mathbb{R}^{n\times m}$ stands for the matrix whose i-th column of C is replaced with a vector $v \in \mathbb{R}^n$, i.e., $C_{i,v} = [C_1,\ \ldots,\ C_{i-1},\ v,\ C_{i+1},\ \ldots,\ C_m]$. The ball centered at $C \in \mathbb{R}^{n\times m}$ with radius $r > 0$ is denoted by $\mathcal{B}(C,r) = \{P \in \mathbb{R}^{n\times m} \mid \|P - C\|_{\mathrm{F}} \le r\}$. $\mathbf{qr}\,(C)$ refers to the Q-matrix of reduced QR decomposition of C. The projection of a matrix $W \in \mathbb{R}^{n\times p}$ to the Stiefel manifold $\mathcal{S}_{n,p}$ is denoted by $\mathcal{P}_{\mathcal{S}_{n,p}}(W)$. $\mathrm{Diag}(\xi) \in \mathbb{R}^{n\times n}$ denotes the diagonal matrix with entries of $\xi \in \mathbb{R}^n$ in its diagonal.

1.4 Organization

The rest of this paper is organized as follows. In Section 2, we introduce our multipliers correction methods. Then we establish the theoretical analysis in Section 3. Furthermore, numerical experiments are presented in Section 4. In the end, we summarize this paper in Section 5.

2 Multipliers correction method

In this section, we present the framework of our new approaches. We start with the first-order optimality condition of the optimization problem over the Stiefel manifold (1.1). According to [14, Lemma 2.2], a point $X \in \mathbb{R}^{n\times p}$ is a first-order stationary point of (1.1), if and only if it satisfies the following equalities:

$$\begin{cases} (I_n - XX^\top)\nabla f(X) = 0, \\ X^\top \nabla f(X) = \nabla f(X)^\top X, \\ X^\top X = I_p. \end{cases} \tag{2.1}$$

The first equality in (2.1) stands for the stationarity of the gradient in the null space of $X^\top$. The second equality determines the symmetry of Lagrangian multipliers associated with orthogonality constraints. For convenience, we call these three equalities "sub-stationarity", "symmetry" and "feasibility", respectively.

In order to solve the problem (1.1), we adopt the similar algorithmic framework proposed in [14], which consists of two steps: reduce the function value in proportion to the "sub-stationarity" violation and preserve the "symmetry". During the calculations of these two steps, we maintain the "feasibility" all the time.

In Subsection 2.1, we first review the function value reduction step in [14] based on Assumption 2.1 on the differentiability of the objective function. Then, in Subsection 2.2, we introduce a new proximal correction strategy, which can further reduce the function value in proportion to the "symmetry" violation. In the end, we present the complete algorithmic framework in Subsection 2.3.

Assumption 2.1. *$f(X)$ is twice differentiable. Then we can define $\rho \geq 0$ as*

$$\rho := \sup_{X\in\tilde{\mathcal{S}}} \left\|\nabla^2 f(X)\right\|_2,$$

where $\tilde{\mathcal{S}} = \{Y \in \mathbb{R}^{n\times p} \mid \|Y\|_{\mathrm{F}}^2 < p+1\}$. In fact, $\tilde{\mathcal{S}}$ can be replaced by any given bounded open set which contains $\mathcal{S}_{n,p}$.

2.1 Function value reduction step

Let $X^{(k)} \in \mathcal{S}_{n,p}$ be the current iterate. The function value reduction step is trying to find a feasible intermediate point $\bar{X}^{(k)} \in \mathcal{S}_{n,p}$ satisfying the following sufficient function value reduction condition:

$$f(X^{(k)}) - f(\bar{X}^{(k)}) \geq c_1 \left\| \left(I_n - X^{(k)}(X^{(k)})^\top\right) \nabla f(X^{(k)}) \right\|_{\mathrm{F}}^2, \tag{2.2}$$

where $c_1 > 0$ is a constant. The right hand side of (2.2) is in proportion to the squared Frobenius norm of "sub-stationary" violation at $X^{(k)}$. Note that it can also be viewed as the projected gradient at $X^{(k)}$ in the Euclidean space. In [14], the authors introduce three algorithms to achieve the sufficient function value reduction (2.2). We list them below.

Gradient reflection (GR) method. It takes the reflection point of the current iterate $X^{(k)}$ on the null space of $X^{(k)} - \tau\nabla f(X^{(k)})$, which can be calculated by the Householder transformation.

$$\begin{cases} V = X^{(k)} - \tau\nabla f(X^{(k)}) \ \text{ for a fixed chosen } \ \tau \in (0, \rho^{-1}), \\ \bar{X}_{\mathrm{GR}}^{(k)} = \left(-I_n + 2V(V^\top V)^\dagger V^\top\right) X^{(k)}. \end{cases}$$

Gradient projection (GP) method. It directly projects $X^{(k)} - \tau \nabla f(X^{(k)})$ onto the Stiefel manifold, which can be calculated by the following projection.

$$\begin{cases} V = X^{(k)} - \tau \nabla f(X^{(k)}) \ \text{ for a fixed chosen } \ \tau \in (0, \rho^{-1}), \\ \bar{X}^{(k)}_{\mathrm{GP}} = \mathcal{P}_{\mathcal{S}_{n,p}}(V). \end{cases}$$

Indeed, the projection $\mathcal{P}_{\mathcal{S}_{n,p}}$ is equivalent to the singular value decomposition, namely, $\mathcal{P}_{\mathcal{S}_{n,p}}(W) = RT^\top$, where $W = RST^\top$ is the reduced singular value decomposition of W.

Column-wise block coordinate descent (CBCD) method. We minimize the objective function with respect to the i-th column of the variable X, and keep the remaining $p-1$ columns fixed as those X. Specifically, we sequentially solve the following subproblem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & f_{i,X}(x) := f(X_{i,x}) \\ \text{s. t.} \ & \|x\|_2 = 1, \\ & X_{\bar{i}}^\top x = 0. \end{aligned} \tag{2.3}$$

The detailed procedure is described in Algorithm 1.

Algorithm 1 Column–wise block coordinate descent method.

1: Set $W^{(0)} = X^{(k)}$ and $i = 1$.
2: **while** $i \le p$ **do**
3: Solve the subproblem (2.3) with X replaced by $W^{(i-1)}$, and obtain a feasible point x^+ satisfying the following sufficient function value descent and asymptotic small step size safeguard:

$$f_{i,W^{(i-1)}}(X_i^{(k)}) - f_{i,W^{(i-1)}}(x^+) \ge k_1 \left\| X_i^{(k)} - x^+ \right\|_2^2,$$

$$\left\| X_i^{(k)} - x^+ \right\|_2 \ge k_2 \left\| \left(I_n - W^{(i-1)} (W^{(i-1)})^\top \right) \nabla f_{i,W^{(i-1)}}(X_i^{(k)}) \right\|_2,$$

where $k_1 > 0$ and $k_2 > 0$ are constants.
4: Set $W^{(i)} = W^{(i-1)}_{i,x^+}$ and $i \leftarrow i + 1$.
5: **end while**
6: Return $\bar{X}^{(k)}_{\mathrm{CBCD}} = W^{(p)}$.

According to [14, Lemmas 3.2, 3.3 and 3.8], these three methods—GR, GP and CBCD—provide an intermediate point satisfying the sufficient function value reduction condition (2.2).

2.2 Proximal correction step

The intermediate point $\bar{X}^{(k)} \in \mathcal{S}_{n,p}$ obtained in the previous subsection does not necessarily satisfy the "symmetry" equality $(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) = \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)}$ in (2.1). In [14], the authors introduce a correction step to obtain $X^{(k+1)}$ through a rotation on $\bar{X}^{(k)}$. The validity of this correction step highly depends on Assumption 1.1, and can not be extended to the general case.

In order to address this issue, we introduce a new proximal strategy. We still calculate the next iterate by a rotation $X^{(k+1)} = \bar{X}^{(k)} Q$ with $Q \in \mathcal{S}_{p,p}$. Ideally, we expect a minimization on

$f(\bar{X}^{(k)}Q)$ and desire to satisfy the "symmetry" equality for $X^{(k+1)}$. However, it is intractable to "cheaply" minimize a general objective function in the range space of $\bar{X}^{(k)}$:

$$\min_{Q\in\mathcal{S}_{p,p}} \quad f(\bar{X}^{(k)}Q). \tag{2.4}$$

On the other side, even if a global solution Q^* of (2.4) is obtained, the corresponding $X^{(k+1)} = \bar{X}^{(k)}Q^*$ does not necessarily satisfy the "symmetry" equality in general.

To this end, we replace the objective function $f(X)$ with its proximal linear approximation $\tilde{f}(X)$ at $\bar{X}^{(k)}$ in the problem (2.4), where

$$\tilde{f}(X) := f(\bar{X}^{(k)}) + \left\langle \nabla f(\bar{X}^{(k)}),\ X - \bar{X}^{(k)} \right\rangle + \frac{\gamma}{2}\left\|X - \bar{X}^{(k)}\right\|_{\mathrm{F}}^2,$$

and $\gamma > 0$ is a proximal parameter. Accordingly, we can construct the approximation problem:

$$\min_{Q\in\mathcal{S}_{p,p}} \quad \tilde{f}(\bar{X}^{(k)}Q). \tag{2.5}$$

In view of the orthogonality of $\bar{X}^{(k)}$ and Q, it is straightforward to obtain the following equivalent problem for (2.5):

$$\min_{Q\in\mathcal{S}_{p,p}} \quad g(Q) := \mathrm{tr}\left(Q^\top Z^{(k)}\right), \tag{2.6}$$

where $Z^{(k)} := (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \gamma I_p$. If $Z^{(k)} = 0$, the problem (2.6) is trivial and we choose $X^{(k+1)} = \bar{X}^{(k)}$. Otherwise, it is known that the global solution of (2.6) is

$$Q^{(k)} := -UV^\top,$$

where $U \in \mathbb{R}^{p\times p}$ and $V \in \mathbb{R}^{p\times p}$ come from the singular value decomposition $Z^{(k)} = U\Sigma V^\top$. In summary, we can construct a new iterate as follows,

$$X^{(k+1)} = \begin{cases} \bar{X}^{(k)}, & \text{if } (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) = \gamma I_p; \\ \bar{X}^{(k)}Q^{(k)}, & \text{otherwise.} \end{cases} \tag{2.7}$$

We call (2.7) the proximal correction step. This step can further reduce the objective function value in proportion to $\left\|(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)}\right\|_{\mathrm{F}}^2$, which will be proved in Section 3.

2.3 Complete algorithmic framework

We denote

$$c(X) := \nabla f(X) - X\nabla f(X)^\top X.$$

Note that it measures the stationarity violation of (2.1) which represents the combination of "sub-stationarity" violation and "symmetry" violation since

$$\|c(X)\|_{\mathrm{F}}^2 = \left\|\left(I_n - XX^\top\right)\nabla f(X)\right\|_{\mathrm{F}}^2 + \left\|X^\top \nabla f(X) - \nabla f(X)^\top X\right\|_{\mathrm{F}}^2 \tag{2.8}$$

holds for any $X \in \mathcal{S}_{n,p}$. The complete algorithmic framework is described in Algorithm 2.

As $X^\top \nabla f(X)$ is nothing but the explicit expression of Lagrangian multipliers associated with orthogonality constraints at any first-order stationary point of (1.1), we call our framework applying the proximal correction step as the multipliers correction methods (MCM). For the algorithms taking GR, GP and CBCD in the Step 3 of Algorithm 2, we call them GRP, GPP and CBCDP, respectively.

Algorithm 2 Multipliers correction methods.

1: Set tolerance $\epsilon > 0$, proximal parameter $\gamma > 0$, and initial point $X^{(0)} \in \mathcal{S}_{n,p}$; Set $k \leftarrow 0$.
2: **while** $\|c(X^{(k)})\|_{\mathrm{F}} > \epsilon$ **do**
3: Based on $X^{(k)}$, find a feasible point $\bar{X}^{(k)}$ satisfying (2.2);
4: Based on $\bar{X}^{(k)}$, compute $X^{(k+1)}$ by (2.7);
5: Set $k \leftarrow k + 1$;
6: **end while**
7: Return $X^{(k)}$.

3 Convergence analysis

In this section, we establish the global convergence and worst case complexity of Algorithm 2. First of all, using the compactness of $\mathcal{S}_{n,p}$, we can define the following two constants.

$$\underline{f} := \min_{X \in \mathcal{S}_{n,p}} f(X), \qquad M := \max_{X \in \mathcal{S}_{n,p}} \|\nabla f(X)\|_2 .$$

Now we evaluate the sufficient function value reduction in the multipliers correction step.

Lemma 3.1. *Suppose Assumption 2.1 holds and $\gamma > \rho$. Let $\bar{X}^{(k)} \in \mathcal{S}_{n,p}$ and $X^{(k+1)}$ be computed by (2.7). Then we have $X^{(k+1)} \in \mathcal{S}_{n,p}$. In addition, it holds that*

$$f(\bar{X}^{(k)}) - f(X^{(k+1)}) \geq \frac{1}{8c_\gamma} \left\| (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)} \right\|_{\mathrm{F}}^2 , \tag{3.1}$$

where $c_\gamma = M + \gamma > 0$ is a constant.

Proof. The feasibility $X^{(k+1)} \in \mathcal{S}_{n,p}$ is obvious. Next, we only focus on the inequality (3.1). If $Z^{(k)} = 0$, we have $X^{(k+1)} = \bar{X}^{(k)}$ and $(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) = \gamma I_p$ is symmetric, which implies (3.1) immediately. Otherwise, since $\gamma > \rho$, we can use Taylor's Theorem and obtain

$$f(X^{(k+1)}) \leq f(\bar{X}^{(k)}) + \left\langle \nabla f(\bar{X}^{(k)}),\ X^{(k+1)} - \bar{X}^{(k)} \right\rangle + \frac{\gamma}{2} \left\| X^{(k+1)} - \bar{X}^{(k)} \right\|_{\mathrm{F}}^2 .$$

Due to the updating rule (2.7) and decomposition $Z^{(k)} = (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \gamma I_p = U \Sigma V^\top$, we have

$$\begin{aligned} f(\bar{X}^{(k)}) - f(X^{(k+1)}) \geq & -\left\langle (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}),\ Q^{(k)} - I_p \right\rangle - \frac{\gamma}{2} \left\| Q^{(k)} - I_p \right\|_{\mathrm{F}}^2 \\ = & \operatorname{tr}(\Sigma) - \gamma \operatorname{tr}\left(Q^{(k)}\right) + \operatorname{tr}\left((\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) \right) - \gamma p + \gamma \operatorname{tr}\left(Q^{(k)}\right) \\ = & \operatorname{tr}\left(\Sigma + U \Sigma V^\top\right) . \end{aligned} \tag{3.2}$$

Let $\hat{\Sigma} = \Sigma V^\top U$ and $\Gamma = (\hat{\Sigma} + \hat{\Sigma}^\top)/2$. It is easy to show that $\operatorname{tr}\left(U \Sigma V^\top\right) = \operatorname{tr}(\Gamma)$. On the other side, after simple calculations, we can obtain that

$$\begin{aligned} \left\| (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)} \right\|_{\mathrm{F}}^2 = & \left\| U \Sigma V^\top - V \Sigma U^\top \right\|_{\mathrm{F}}^2 \\ = & 2\operatorname{tr}\left(\Sigma^2\right) - 2\operatorname{tr}\left(\Sigma V^\top U \Sigma V^\top U\right) \\ = & 2\operatorname{tr}\left(\Sigma^2\right) - 2\operatorname{tr}\left(\hat{\Sigma}^2\right) . \end{aligned} \tag{3.3}$$

It follows from the equality $\Gamma = \left(\hat{\Sigma} + \hat{\Sigma}^\top\right)/2$ that $2\mathrm{tr}\left(\Gamma^2\right) = \mathrm{tr}\left(\Sigma^2\right) + \mathrm{tr}\left(\hat{\Sigma}^2\right)$. Together with (3.3), we arrive at

$$\left\|(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)}\right\|_{\mathrm{F}}^2 = 4\mathrm{tr}\left(\Sigma^2\right) - 4\mathrm{tr}\left(\Gamma^2\right). \tag{3.4}$$

Moreover, we have $\mathrm{tr}\left(\Gamma^2\right) = \mathrm{tr}\left(\Gamma^\top\Gamma\right) = \sum_{i=1}^{p} \Gamma_i^\top \Gamma_i \geq \sum_{i=1}^{p} \Gamma_{ii}^2$. Hence, it holds that

$$\mathrm{tr}\left(\Sigma^2\right) - \mathrm{tr}\left(\Gamma^2\right) \leq \sum_{i=1}^{p} \left(\Sigma_{ii}^2 - \Gamma_{ii}^2\right) = \sum_{i=1}^{p} \left(\Sigma_{ii} - \Gamma_{ii}\right)\left(\Sigma_{ii} + \Gamma_{ii}\right).$$

According to the definition of Γ, we can obtain $|\Gamma_{ii}| = \Sigma_{ii}\left(V_i^\top U_i\right) \leq \Sigma_{ii} \left\|V_i\right\|_2 \left\|U_i\right\|_2 = \Sigma_{ii}$, which implies

$$\mathrm{tr}\left(\Sigma^2\right) - \mathrm{tr}\left(\Gamma^2\right) \leq \sum_{i=1}^{p} 2\Sigma_{ii}\left(\Sigma_{ii} + \Gamma_{ii}\right) \leq 2\left\|\Sigma\right\|_2 \mathrm{tr}\left(\Sigma + \Gamma\right) \leq 2c_\gamma \mathrm{tr}\left(\Sigma + \Gamma\right), \tag{3.5}$$

where the last inequality follows from $\left\|\Sigma\right\|_2 = \left\|Z^{(k)}\right\|_2 \leq \left\|(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)})\right\|_2 + \gamma \leq M + \gamma = c_\gamma$. Combing (3.4) and (3.5), we can deduce that

$$8c_\gamma \mathrm{tr}\left(\Sigma + U\Sigma V^\top\right) = 8c_\gamma \mathrm{tr}\left(\Sigma + \Gamma\right) \geq \left\|(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)}\right\|_{\mathrm{F}}^2, \tag{3.6}$$

which together with (3.2) infers that

$$8c_\gamma\left(f(\bar{X}^{(k)}) - f(X^{(k+1)})\right) \geq \left\|(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)}\right\|_{\mathrm{F}}^2.$$

This completes the proof. □

The convergence of the function value can be a direct corollary.

Corollary 3.2. *Suppose Assumption 2.1 holds, $\gamma > \rho$, and $\{X^{(k)}\}$ is the iterate sequence generated by Algorithm 2. Then $\{f(X^{(k)})\}$ is convergent.*

Proof. According to Lemma 3.1, we have

$$\begin{aligned} & f(X^{(k)}) - f(X^{(k+1)}) = f(X^{(k)}) - f(\bar{X}^{(k)}) + f(\bar{X}^{(k)}) - f(X^{(k+1)}) \\ \geq\ & c_1 \left\|\left(I_n - X^{(k)}(X^{(k)})^\top\right)\nabla f(X^{(k)})\right\|_{\mathrm{F}}^2 + \frac{1}{8c_\gamma}\left\|(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)}\right\|_{\mathrm{F}}^2 \\ \geq\ & c_1 \left\|\left(I_n - X^{(k)}(X^{(k)})^\top\right)\nabla f(X^{(k)})\right\|_{\mathrm{F}}^2 \geq 0. \end{aligned} \tag{3.7}$$

Consequently, $\{f(X^{(k)})\}$ is a monotonically non-increasing sequence. On the other hand, it follows from the compactness of the Stiefel manifold $\mathcal{S}_{n,p}$ that $\{f(X^{(k)})\}$ has a lower bound $\underline{f}$. Therefore, we conclude that $\{f(X^{(k)})\}$ is convergent, which completes the proof. □

Then we show that the "symmetry" violation can be controlled by the distance between $X^{(k+1)}$ and $\bar{X}^{(k)}$.

Lemma 3.3. *Suppose Assumption 2.1 holds and $\{X^{(k)}\}$ is the iterate sequence generated by Algorithm 2. Then it can be verified that*

$$\left\|(X^{(k+1)})^\top \nabla f(X^{(k+1)}) - \nabla f(X^{(k+1)})^\top X^{(k+1)}\right\|_{\mathrm{F}} \leq 2\left(\rho+\gamma\right)\left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}}. \tag{3.8}$$

Proof. If $Z^{(k)} = 0$, we have $X^{(k+1)} = \bar{X}^{(k)}$. Hence, the matrix $(X^{(k+1)})^\top \nabla f(X^{(k+1)}) = (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) = \gamma I_p$ is symmetric, which infers (3.8) immediately. Next, we investigate the case that $Z^{(k)} \neq 0$. It follows from the definition of $Q^{(k)} = -UV^\top$ and decomposition $Z^{(k)} = U\Sigma V^\top$ that $(Q^{(k)})^\top Z^{(k)} = (Z^{(k)})^\top Q^{(k)}$. In view of $Z^{(k)} = (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \gamma I_p$ and $X^{(k+1)} = \bar{X}^{(k)} Q^{(k)}$, it further holds that

$$(X^{(k+1)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top X^{(k+1)} = \gamma (X^{(k+1)})^\top \bar{X}^{(k)} - \gamma (\bar{X}^{(k)})^\top X^{(k+1)}.$$

According to the triangular inequality, we have

$$\begin{aligned}
&\left\|(X^{(k+1)})^\top \bar{X}^{(k)} - (\bar{X}^{(k)})^\top X^{(k+1)}\right\|_{\mathrm{F}} \\
\leq\ &\left\|(X^{(k+1)})^\top \bar{X}^{(k)} - (\bar{X}^{(k)})^\top \bar{X}^{(k)}\right\|_{\mathrm{F}} + \left\|(\bar{X}^{(k)})^\top \bar{X}^{(k)} - (\bar{X}^{k})^\top X^{(k+1)}\right\|_{\mathrm{F}} \\
\leq\ &\left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}} \left\|\bar{X}^{(k)}\right\|_2 + \left\|\bar{X}^{(k)}\right\|_2 \left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}} = 2\left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}},
\end{aligned}$$

which immediately implies that

$$\left\|(X^{(k+1)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top X^{(k+1)}\right\|_{\mathrm{F}} \leq 2\gamma \left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}}. \tag{3.9}$$

On the other hand, according to Assumption 2.1, it follows that

$$\left\|\nabla f(Y_1) - \nabla f(Y_2)\right\|_{\mathrm{F}} \leq \rho \left\|Y_1 - Y_2\right\|_{\mathrm{F}}, \quad \text{for all} \ \ Y_1, Y_2 \in \mathcal{S}_{n,p}.$$

Thus, we can obtain that

$$\begin{aligned}
\left\|(X^{(k+1)})^\top \nabla f(X^{(k+1)}) - (X^{(k+1)})^\top \nabla f(\bar{X}^{(k)})\right\|_{\mathrm{F}} &\leq \left\|X^{(k+1)}\right\|_2 \left\|\nabla f(X^{(k+1)}) - \nabla f(\bar{X}^{(k)})\right\|_{\mathrm{F}} \\
&\leq \rho \left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}},
\end{aligned}$$

and similarly,

$$\left\|\nabla f(\bar{X}^{(k)})^\top X^{(k+1)} - \nabla f(X^{(k+1)})^\top X^{(k+1)}\right\|_{\mathrm{F}} \leq \rho \left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}}.$$

Together with (3.9), we can conclude that

$$\begin{aligned}
&\left\|(X^{(k+1)})^\top \nabla f(X^{(k+1)}) - \nabla f(X^{(k+1)})^\top X^{(k+1)}\right\|_{\mathrm{F}} \\
\leq\ &\left\|(X^{(k+1)})^\top \nabla f(X^{(k+1)}) - (X^{(k+1)})^\top \nabla f(\bar{X}^{(k)})\right\|_{\mathrm{F}} + \left\|(X^{(k+1)})^\top \nabla f(\bar{X}^{(k)})\right. \\
&\left. - \nabla f(\bar{X}^{(k)})^\top X^{(k+1)}\right\|_{\mathrm{F}} + \left\|\nabla f(\bar{X}^{(k)})^\top X^{(k+1)} - \nabla f(X^{(k+1)})^\top X^{(k+1)}\right\|_{\mathrm{F}} \\
\leq\ &2\left(\rho+\gamma\right)\left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}},
\end{aligned}$$

and complete the proof. □

Next we show the distance between $X^{(k+1)}$ and $\bar{X}^{(k)}$ converges to 0.

Lemma 3.4. *Suppose Assumption 2.1 holds, $\gamma > \rho$, and $\{X^{(k)}\}$ is the iterate sequence generated by Algorithm 2. Then it holds that*

$$\lim_{k\to\infty} \left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}} = 0.$$

Proof. Firstly, it follows from the inequality (3.6) that

$$8c_\gamma \mathrm{tr}\left(\Sigma + U\Sigma V^\top\right) \geq \left\|(\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)}) - \nabla f(\bar{X}^{(k)})^\top \bar{X}^{(k)}\right\|_{\mathrm{F}}^2 \geq 0.$$

Then by simple calculations, we can obtain that

$$\begin{aligned} & \left\langle X^{(k+1)} - \bar{X}^{(k)},\ X^{(k+1)} - \bar{X}^{(k)} + 2\gamma^{-1}\nabla f(\bar{X}^{(k)})\right\rangle \\ = & \left\langle X^{(k+1)} - \bar{X}^{(k)},\ X^{(k+1)} - \bar{X}^{(k)}\right\rangle + 2\gamma^{-1}\left\langle Q^{(k)} - I_p,\ (\bar{X}^{(k)})^\top \nabla f(\bar{X}^{(k)})\right\rangle \\ = & -2\gamma^{-1}\mathrm{tr}\left(\Sigma + U\Sigma V^\top\right) \leq 0. \end{aligned}$$

This relationship can guarantee that

$$\left\|X^{(k+1)} - \bar{X}^{(k)} + \gamma^{-1}\nabla f(\bar{X}^{(k)})\right\|_{\mathrm{F}} \leq \gamma^{-1}\left\|\nabla f(\bar{X}^{(k)})\right\|_{\mathrm{F}},$$

which implies that

$$X^{(k+1)} \in \mathcal{B}\left(\bar{X}^{(k)} - \gamma^{-1}\nabla f(\bar{X}^{(k)}),\ \gamma^{-1}\left\|\nabla f(\bar{X}^{(k)})\right\|_{\mathrm{F}}\right).$$

We recall [14, Lemma 3.1] and obtain that

$$\left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}}^2 \leq \frac{2}{\gamma - \rho}\left(f(\bar{X}^{(k)}) - f(X^{(k+1)})\right) \leq \frac{2}{\gamma - \rho}\left(f(X^{(k)}) - f(X^{(k+1)})\right). \quad (3.10)$$

Since $\{f(X^{(k)})\}$ is convergent, we conclude that

$$\lim_{k\to\infty}\left\|X^{(k+1)} - \bar{X}^{(k)}\right\|_{\mathrm{F}} = 0.$$

This completes the proof. □

Finally, we are ready to present our main convergence result.

Theorem 3.5. *Suppose Assumption 2.1 holds, $\gamma > \rho$, and $\{X^{(k)}\}$ is the iterate sequence generated by Algorithm 2. Then there exists at least one convergent subsequence of $\{X^{(k)}\}$. Furthermore, each accumulation point X^* of $\{X^{(k)}\}$ satisfies the first-order stationarity condition* (2.1)*. More precisely, the following inequality*

$$\min_{1\leq k\leq K}\left\|\nabla f(X^{(k)}) - X^{(k)}\nabla f(X^{(k)})^\top X^{(k)}\right\|_{\mathrm{F}} \leq \sqrt{\frac{c_2\left(f(X^{(0)}) - \underline{f}\right)}{K}},$$

holds for any $K \geq 1$, where $c_2 > 0$ is a constant defined by

$$c_2 = \frac{1}{c_1} + \frac{8\left(\gamma + \rho\right)^2}{\gamma - \rho}. \quad (3.11)$$

Proof. It follows from the compactness of the Stiefel manifold $\mathcal{S}_{n,p}$ that $\{X^{(k)}\}$ is bounded, which implies $\{X^{(k)}\}$ has at least one convergent subsequence. Suppose X^* is an accumulation point of $\{X^{(k)}\}$. It is clear that $X^* \in \mathcal{S}_{n,p}$ due to the feasibility of $\{X^{(k)}\}$.

Recalling the convergence of $\{f(X^{(k)})\}$ and (3.7), we have

$$\lim_{k\to\infty}\left\|\left(I_n - X^{(k)}(X^{(k)})^\top\right)\nabla f(X^{(k)})\right\|_{\mathrm{F}} = 0,$$

which directly implies

$$\left(I_n - X^*(X^*)^\top\right)\nabla f(X^*) = 0. \tag{3.12}$$

On the other hand, it follows from Lemma 3.3 and Lemma 3.4 that

$$\lim_{k\to\infty} \left\|(X^{(k)})^\top \nabla f(X^{(k)}) - \nabla f(X^{(k)})^\top X^{(k)}\right\|_{\mathrm{F}} \le 2\left(\rho+\gamma\right) \lim_{k\to\infty} \left\|X^{(k)} - \bar{X}^{(k-1)}\right\|_{\mathrm{F}} = 0,$$

which yields that

$$(X^*)^\top \nabla f(X^*) = \nabla f(X^*)^\top X^*. \tag{3.13}$$

Combining "feasibility", "sub-stationarity" (3.12) and "symmetry" (3.13), we conclude that X^* satisfies the first-order stationarity condition (2.1).

Furthermore, it follows from Lemma 3.3 and (3.10) that

$$\left\|(X^{(k+1)})^\top \nabla f(X^{(k+1)}) - \nabla f(X^{(k+1)})^\top X^{(k+1)}\right\|_{\mathrm{F}}^2 \le \frac{8(\gamma+\rho)^2}{\gamma-\rho}\left(f(X^{(k)}) - f(X^{(k+1)})\right).$$

Together with the relationships (2.2) and (2.8), we can arrive at

$$\begin{aligned}
&\left\|\nabla f(X^{(k)}) - X^{(k)}\nabla f(X^{(k)})^\top X^{(k)}\right\|_{\mathrm{F}}^2 \\
\le\ & \frac{1}{c_1}\left(f(X^{(k)}) - f(X^{(k+1)})\right) + \frac{8\left(\gamma+\rho\right)^2}{\gamma-\rho}\left(f(X^{(k-1)}) - f(X^{(k)})\right).
\end{aligned}$$

To sum up both sides of the above inequality from $k = 1$ to K, we can obtain

$$\begin{aligned}
&\sum_{k=1}^{K} \left\|\nabla f(X^{(k)}) - X^{(k)}\nabla f(X^{(k)})^\top X^{(k)}\right\|_{\mathrm{F}}^2 \\
\le\ & \frac{1}{c_1}\sum_{k=1}^{K}\left(f(X^{(k)}) - f(X^{(k+1)})\right) + \frac{8\left(\gamma+\rho\right)^2}{\gamma-\rho}\sum_{k=1}^{K}\left(f(X^{(k-1)}) - f(X^{(k)})\right) \\
=\ & \frac{1}{c_1}\left(f(X^{(1)}) - f(X^{(K+1)})\right) + \frac{8\left(\gamma+\rho\right)^2}{\gamma-\rho}\left(f(X^{(0)}) - f(X^{(K)})\right) \le c_2\left(f(X^{(0)}) - \underline{f}\right),
\end{aligned}$$

where c_2 is defined by (3.11). Together with the fact that

$$\sum_{k=1}^{K} \left\|\nabla f(X^{(k)}) - X^{(k)}\nabla f(X^{(k)})^\top X^{(k)}\right\|_{\mathrm{F}}^2 \ge K \min_{1\le k\le K} \left\|\nabla f(X^{(k)}) - X^{(k)}\nabla f(X^{(k)})^\top X^{(k)}\right\|_{\mathrm{F}}^2,$$

we complete the proof. □

Remark 3.6. *According to the stopping criterion, Theorem 3.5 guarantees the termination of Algorithm 2 in at most $O(1/\epsilon^2)$ iterations.*

4 Numerical experiments

In this section, we report the numerical performance of the algorithms based on Algorithm 2. Two types of testing problems are introduced in Subsection 4.1. The implementation details including the selection of algorithm parameters and stopping criterion are presented in Subsection 4.2. The numerical comparison among our algorithms and those introduced in [14] is presented in Subsection 4.3. Finally, we compare our algorithms with other two state-of-the-art approaches, and numerical results are shown in Subsection 4.4. All experiments are performed on a workstation with one Intel(R) Xeon(R) Silver 4110 CPU (at 2.10GHz×32) and 384GB of RAM running in MATLAB R2018a under Ubuntu 18.10.

4.1 Testing problems

Problem 1. The first class of testing problems is a quadratic objective minimization over the Stiefel manifold:

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times p}} \ & f_1(X) = \frac{1}{2}\mathrm{tr}\left(X^\top M X\right) + \mathrm{tr}\left(N^\top X\right) \\ \text{s. t.} \quad & X^\top X = I_p. \end{aligned}$$

In the experiments, $M \in \mathbb{R}^{n\times n}$ and $N \in \mathbb{R}^{n\times p}$ are randomly generated by

$$M = E\Psi E^\top, \quad N = \alpha Q D,$$

where $E = \mathbf{qr}\left(\mathbf{randn}(n,n)\right) \in \mathbb{R}^{n\times n}$, $\tilde{Q} = \mathbf{randn}(n,p) \in \mathbb{R}^{n\times p}$, and $Q \in \mathbb{R}^{n\times p}$ with $Q_i = \tilde{Q}_i / \left\|\tilde{Q}_i\right\|_2$ $(i = 1,\ldots,p)$. The notation $\mathbf{randn}(n,m)$ represents an $n \times m$ matrix randomly generated by i.i.d. standard Gaussian distribution. Moreover, $\Psi \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{p\times p}$ are diagonal matrices with, respectively,

$$\Psi_{ii} = \begin{cases} \eta^{1-i}, & \text{if } \omega_i < 0.5, \\ -\eta^{1-i}, & \text{otherwise,} \end{cases} \quad \text{for all } \ i = 1,2,\ldots,n,$$

$$D_{ii} = \zeta^{1-i}, \ \text{ for all } \ i = 1,2,\ldots,p,$$

where $\omega_i \in [0,1]$ for $i = 1,2,\ldots,n$ are randomly generated numbers. Here, $\eta \geq 1$ is a parameter determining the decay of eigenvalues of M, and $\zeta \geq 1$ is a parameter referring to the growth rate of the columns norm of N. The parameter $\alpha > 0$ represents the scale difference between the quadratic term and the linear term. Unless otherwise stated, the default values of these parameters are $\eta = 1.01$, $\zeta = 1.01$, $\alpha = 1$. This class of testing problems is also used in [14], which satisfies Assumption 1.1.

Problem 2. The second class of testing problems is Brockett function minimization over the Stiefel manifold:

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times p}} \ & f_2(X) = \frac{1}{2}\mathrm{tr}\left(DX^\top A X\right) \\ \text{s. t.} \quad & X^\top X = I_p. \end{aligned}$$

The data matrix $A \in \mathbb{R}^{n\times n}$ is randomly generated by

$$A = E\Psi E^\top.$$

Here, $E = \mathbf{qr}\left(\mathbf{randn}(n,n)\right) \in \mathbb{R}^{n\times n}$, $\Psi \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{p\times p}$ are diagonal matrices with, respectively,

$$\Psi_{ii} = \begin{cases} \eta^{1-i} + \beta, & \text{if } \omega_i < 0.5, \\ -\eta^{1-i} - \beta, & \text{otherwise,} \end{cases} \quad \text{for all } \ i = 1,2,\ldots,n,$$

$$D_{ii} = \begin{cases} \alpha\zeta^{1-i}, & \text{if } \theta_i < 0.5, \\ -\alpha\zeta^{1-i}, & \text{otherwise,} \end{cases} \quad \text{for all } \ i = 1,2,\ldots,p,$$

where $\omega_i \in [0,1]$ for $i = 1,2,\ldots,n$ and $\theta_i \in [0,1]$ for $i = 1,2,\ldots,p$ are randomly generated numbers. Two parameters $\eta \geq 1$ and $\beta \geq 1$ determine the difference of eigenvalues of A. Moreover, $\zeta \geq 1$ is a parameter referring to the decrease rate of diagonal entries of D. The parameter $\alpha > 0$ represents the scale difference between A and D. Unless otherwise stated, the default values of these parameters are $\eta = 1.05$, $\zeta = 1.05$, $\beta = 2$, $\alpha = 0.1$. This class of testing problems does not satisfy Assumption 1.1.

4.2 Implementation details

All of the three algorithms GRP, GPP and CBCDP have a common parameter γ. Although in the theoretical analysis, γ should be larger than the constant ρ, we set $\gamma = 10^{-3}s$ in practice, where s is an estimation of $\|\nabla^2 f(0)\|_2$. More specifically, we choose $s = \|M\|_2$ and $s = \|A\|_2 \|D\|_2$ for Problems 1 and 2, respectively.

In practice, we recommend to use the following alternating BB stepsize introduced in [10]:

$$\tau_{\mathrm{ABB}}^{(k)} = \begin{cases} \tau_{\mathrm{BB1}}^{(k)} & \text{if } k \text{ is odd,} \\ \tau_{\mathrm{BB2}}^{(k)} & \text{if } k \text{ is even.} \end{cases}$$

Here, two Barzilai-Borwein (BB) stepsizes were first introduced in [7]:

$$\tau_{\mathrm{BB1}}^{(k)} = \frac{|\langle J_k,\ K_k\rangle|}{\langle K_k,\ K_k\rangle}, \quad \text{or} \quad \tau_{\mathrm{BB2}}^{(k)} = \frac{\langle J_k,\ J_k\rangle}{|\langle J_k,\ K_k\rangle|},$$

where $J_k = X^{(k)} - X^{(k-1)}$, $K_k = c(X^{(k)}) - c(X^{(k-1)})$.

As for the CBCDP method, the subproblem (2.3) can be solved globally if our testing problems are quadratic, which has been elaborately introduced in [14] and hence omitted here. For the updating order of the block coordinate descent scheme, we simply choose the Gauss–Seidel manner.

The stopping criterion can be described as follows,

$$\left\|\nabla f(X^{(k)}) - X^{(k)}\nabla f(X^{(k)})^\top X^{(k)}\right\|_{\mathrm{F}} \le \epsilon_g \left\|\nabla f(X^{(0)}) - X^{(0)}\nabla f(X^{(0)})^\top X^{(0)}\right\|_{\mathrm{F}}, \tag{4.1}$$

where $\epsilon_g > 0$ is a tolerance constant. In addition, we also adopt the following stopping rules based on the relative error:

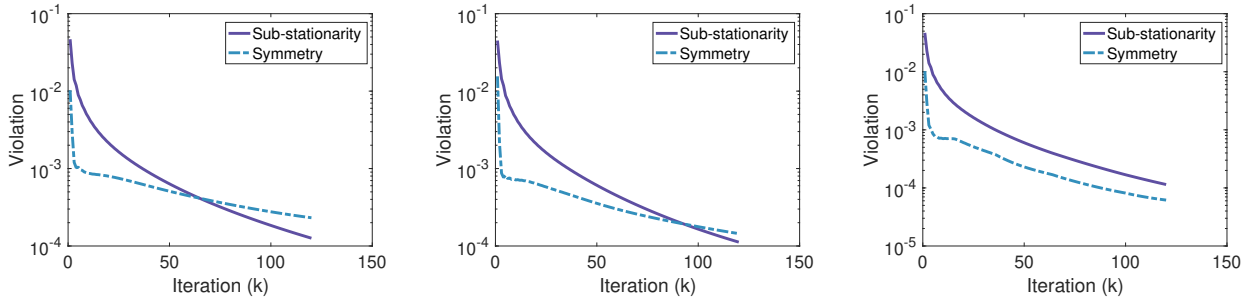
$$\mathrm{tol}_x^{(k)} = \frac{\left\|X^{(k)} - X^{(k-1)}\right\|_{\mathrm{F}}}{\sqrt{n}} \le \epsilon_x, \quad \mathrm{tol}_f^{(k)} = \frac{\left|f(X^{(k)}) - f(X^{(k-1)})\right|}{|f(X^{(k-1)})| + 1} \le \epsilon_f, \tag{4.2}$$

and

$$\mathrm{mean}(\mathrm{tol}_x^{(k-\min\{k,T\}+1)}, \ldots, \mathrm{tol}_x^{(k)}) \le 10\epsilon_x, \quad \mathrm{mean}(\mathrm{tol}_f^{(k-\min\{k,T\}+1)}, \ldots, \mathrm{tol}_f^{(k)}) \le 10\epsilon_f, \tag{4.3}$$

where $\epsilon_x > 0$ and $\epsilon_f > 0$ are also tolerance constants, and $\mathrm{mean}(a_1, \ldots, a_m)$ denotes the mean value of numbers $a_1, \ldots, a_m$. We terminate the algorithm when it satisfies one of the above three stopping criteria (4.1)-(4.3), or reaches a preset maximum iteration number MaxIter. Unless otherwise stated, we set the tolerance parameters $\epsilon_x = 10^{-6}$, $T = 5$ and MaxIter = 3000. For Problems 1 and 2, we set $\epsilon_g = 10^{-5}$, $\epsilon_f = 10^{-10}$ and $\epsilon_g = 10^{-3}$, $\epsilon_f = 10^{-8}$, respectively.

In Algorithm 2, the proximal correction step is performed once in each iteration. A special test on GPP employed in solving Problem 2 with $n = 5000$ and $p = 50$ demonstrates that the decay rate of the "symmetry" violation is worse than that of the "sub-stationarity" violation. Such unbalance affects the overall performance of our algorithms. Hence, we consider multiple proximal correction steps in each iteration. From Figure 1, we can learn that three times proximal correction can accelerate the decay of "symmetry" violation. Heuristically, we recommend $\delta_k = 2\lceil\sqrt{k}/2\rceil - 1$ times proximal correction steps in the k-th iteration, which substantially makes the two decay rates close to each other. Therefore, in the following comparison, we use δ_k as the default number of proximal correction steps in each iteration.



(a) Single proximal correction step (b) 3 proximal correction steps (c) δ_k proximal correction steps

Figure 1: Comparison of multiple proximal correction steps.

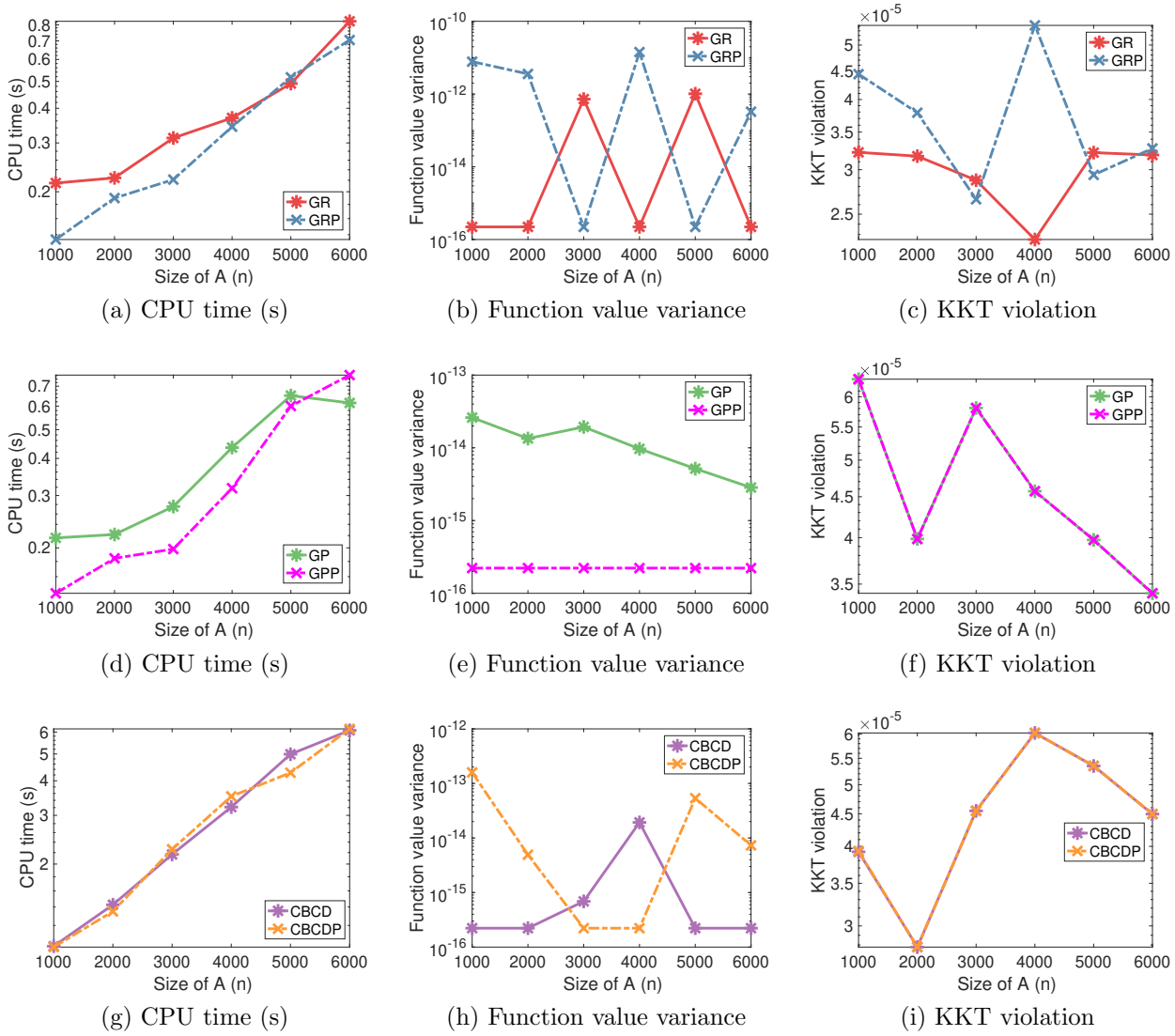
We use three measurements in the numerical comparison, including CPU time in seconds, KKT violation ($\left\|\nabla f(X) - X\nabla f(X)^\top X\right\|_{\mathrm{F}}$) and function value variance, which is defined as $|f_s - f_{\min}| / (1 + |f_{\min}|) + \mathrm{eps}$. Here, f_s and $f_{\min}$ refer to the final objective function value returned by solver s and the smallest one of those obtained by all solvers in the comparison, respectively. We add $\mathrm{eps} = 2.2204 \times 10^{-16}$, the machine precision in MATLAB, to the relative variance of function value for the sake of logarithmic scale demonstration. Finally, all the tested algorithms are initiated from the same point $X^{(0)}$, which is randomly generated by $X^{(0)} = \mathbf{qr}\left(\mathbf{randn}(n, p)\right) \in \mathcal{S}_{n,p}$.

4.3 Comparison with GR, GP and CBCD

In this subsection, we mainly compare our GRP, GPP and CBCDP with GR, GP, and CBCD, respectively. In the test, all of GR, GP, and CBCD are taken their default settings introduced in [14], which are almost the same as our algorithms, except for completely different multipliers correction step.

For this purpose, we perform on a set of problems based on Problem 1 with n ranging from 1000 to 6000 increment 1000 and $p = 60$. Other parameters take their default values. We demonstrate the numerical results in Figure 2. We observe that these six algorithms all reach comparable KKT violations and final function values. In most cases, GRP, GPP, and CBCDP require less CPU time than GR, GP, and CBCD, respectively. In this sense, our multipliers correction methods are comparable with those proposed in [14] for the problems satisfying Assumption 1.1.

We also make a comparison among GRP, GPP and CBCDP, when they are employed to solve Problem 1 and Problem 2. In this test, we set $n = 3000$ and p ranging from 20 to 120 increment 20. Other parameters take their default values. The numerical results are illustrated in Figure 3. We observe that GPP outperforms GRP and CBCDP in most cases. Therefore, we choose GPP to represent our new multipliers correction methods in the following numerical experiments.



(a) CPU time (s) (b) Function value variance (c) KKT violation

(d) CPU time (s) (e) Function value variance (f) KKT violation

(g) CPU time (s) (h) Function value variance (i) KKT violation

Figure 2: Comparison between multipliers correction methods with their original versions.

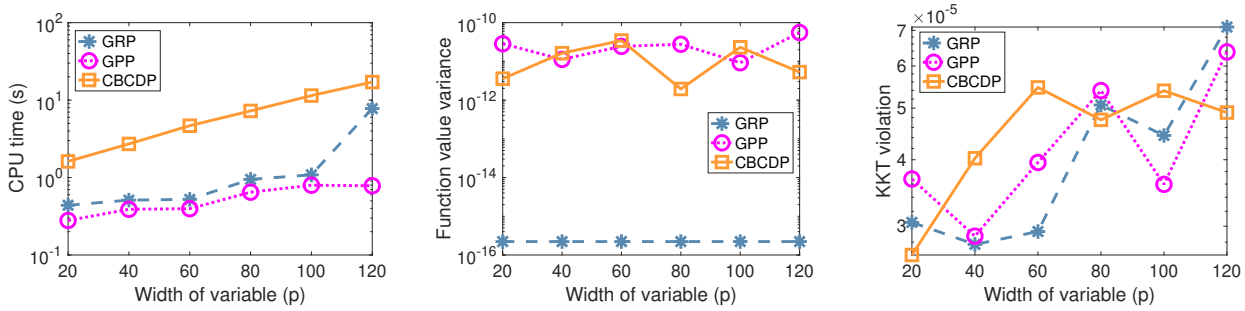
4.4 Performance comparison with other algorithms

In this subsection, we compare the performance of GPP with other two state-of-the-art algorithms for optimization problems over the Stiefel manifold. One is OptM[1] proposed in [33]. The other one is MOptQR from the package MANOPT[2] which is proposed in [4]. The original version is MOptQR-LS (manifold QR method with line search). For fair comparison, we implement the same alternating BB step size strategy to MOptQR-LS, which can significantly accelerate the algorithm as our GPP.

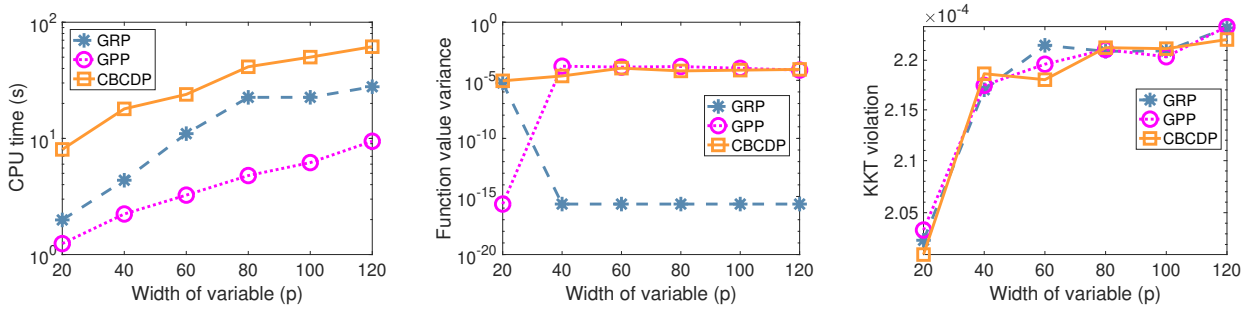
We design five groups of testing problems based on Problem 2, in each of which there is only one parameter varying with all the others fixed. More specifically, we describe the varying parameters of each group as follows.

- $n = 9000 + 1000j$ for $j = 1, 2, 3, 4, 5, 6$; $p = 100$.

[1]Downloadable from https://github.com/wenstone/OptM.
[2]Downloadable from https://www.manopt.org/.



(a) CPU time (s) (Problem 1) (b) Function value variance (Problem 1) (c) KKT violation (Problem 1)



(d) CPU time (s) (Problem 2) (e) Function value variance (Problem 2) (f) KKT violation (Problem 2)

Figure 3: Comparison of GRP, GPP, and CBCDP for different p.

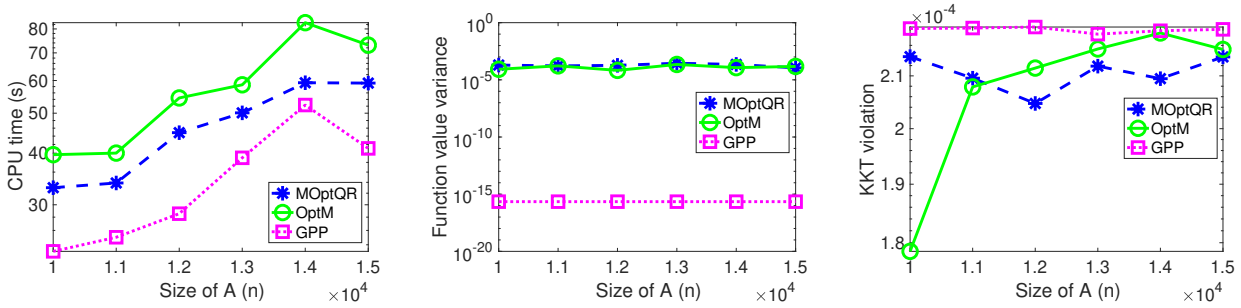
- $p = 20j$ for $j = 1, 2, 3, 4, 5, 6$; $n = 10000$.
- $\beta = 1.1 + 0.3j$ for $j = 0, 1, 2, 3, 4, 5$; $n = 10000$; $p = 60$.
- $\eta = 1.01 + 0.02j$ for $j = 0, 1, 2, 3, 4, 5$; $n = 10000$; $p = 60$.
- $\zeta = 1.01 + 0.05j$ for $j = 0, 1, 2, 3, 4, 5$; $n = 10000$; $p = 60$.

All the other parameters take their default values.

The numerical results of the above five groups of testing problems are depicted in Figures 4 to 8, respectively. We observe that these algorithms achieve comparable KKT violation, and GPP outperforms the other two algorithms in terms of CPU time and function value variance.

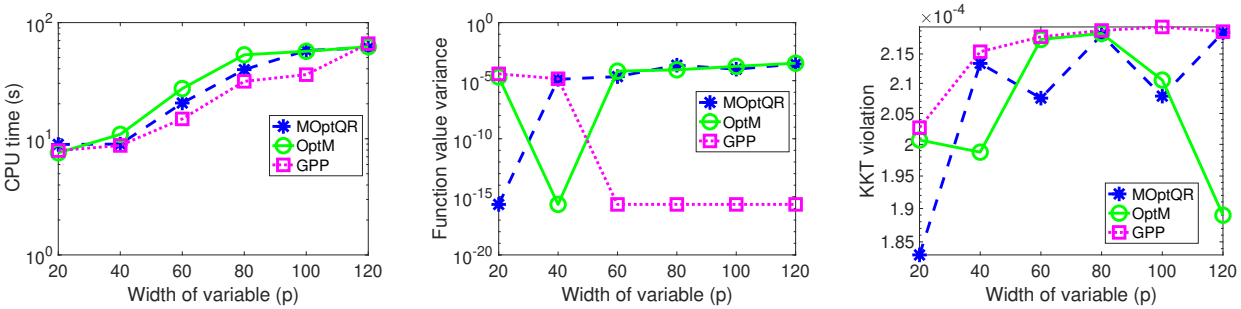
In order to make a more comprehensive comparison, we use performance profiles based on [11] to visualize the different behaviors among these solvers. For this purpose, we design a variety of random problems based on Problem 2, which can be described as follows:

- $n = 2000 + 1000j$ for $j = 1, 2, 3, 4, 5, 6$;
- $p = 20j$ for $j = 1, 2, 3, 4, 5, 6$;
- $\beta = 1 + 0.5j$ for $j = 0, 1, 2, 3$;
- $\eta = 1.01 + 0.05j$ for $j = 0, 1, 2, 3$;
- $\zeta = 1.1 + 0.05j$ for $j = 0, 1, 2, 3$.



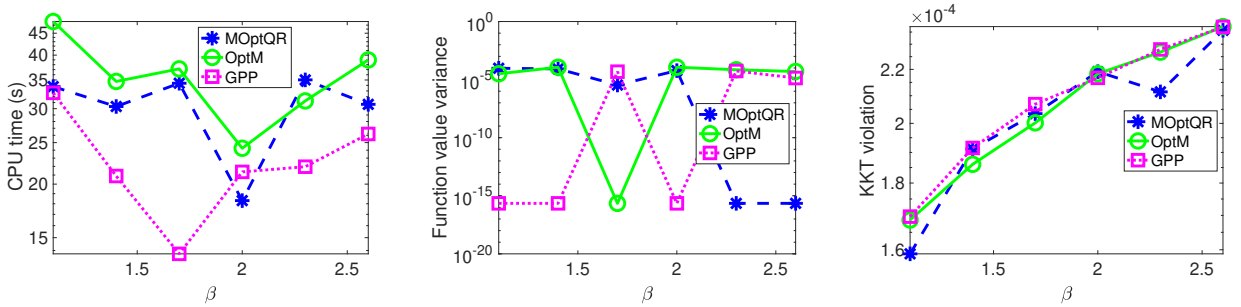
(a) CPU time (s) (b) Function value variance (c) KKT violation

Figure 4: Comparison of GPP, OptM and MOptQR for different n on Problem 2.



(a) CPU time (s) (b) Function value variance (c) KKT violation

Figure 5: Comparison of GPP, OptM and MOptQR for different p on Problem 2.



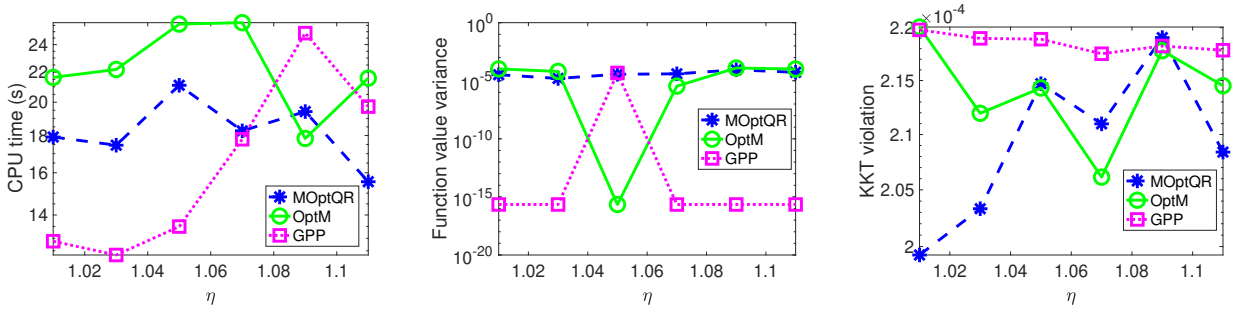
(a) CPU time (s) (b) Function value variance (c) KKT violation

Figure 6: Comparison of GPP, OptM and MOptQR for different β on Problem 2.

There are altogether $6 \times 6 \times 4 \times 4 \times 4 = 2304$ randomly generated problems. We simply explain the performance profile as the following. For problem m and solver s, we use $t_{m,s}$ to represent its CPU time. Performance ratio is defined as $r_{m,s} = t_{m,s} / \min_s \{t_{m,s}\}$. If solver s fails to solve problem m, the ratio $r_{m,s}$ is set to a preset large number. Finally, the overall performance of solver s is defined by

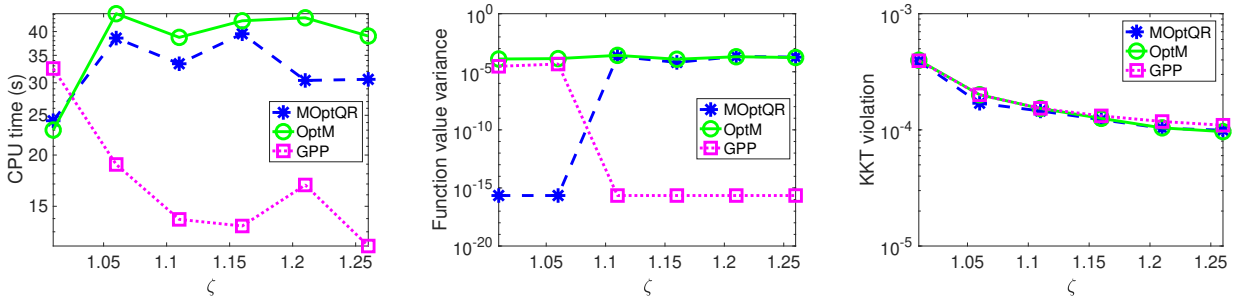
$$\pi_s(\omega) = \frac{\text{number of problems where } r_{m,s} \leq \omega}{\text{total number of problems}}.$$

It means the percentage of testing problems that can be solved in $\omega \min_s \{t_{m,s}\}$ seconds. It is clear that the closer π_s is to 1, the better performance solver s has.



(a) CPU time (s) (b) Function value variance (c) KKT violation

Figure 7: Comparison of GPP, OptM and MOptQR for different η on Problem 2.



(a) CPU time (s) (b) Function value variance (c) KKT violation

Figure 8: Comparison of GPP, OptM and MOptQR for different ζ on Problem 2.

The performance profile with respect to the CPU time is given in Figure 9. On the 2304 testing problems, GPP is of the best numerical behavior in terms of CPU time and it always solves problems in no more than twice the fastest time among these three algorithms. In addition, we also provide the average KKT violation, feasibility violation and function value variance over these 2304 random problems in Table 1, which shows that all solvers achieve a comparable average KKT violation, feasibility violation, and function value variance.

5 Conclusion

The first-order algorithmic framework proposed in [14] consists of a function value reduction step in the Euclidean space and a rotation step to guarantee the symmetry of the explicit expression of Lagrangian multipliers associate with orthogonality constraints. Three algorithms based on this framework have illustrated their efficiency in solving problems such as

	GPP	MOptQR	OptM
KKT violation	1.4917×10^{-3}	1.1534×10^{-3}	1.1516×10^{-3}
Function value variance	3.3934×10^{-4}	6.8694×10^{-4}	8.1899×10^{-4}
Feasibility violation	2.5227×10^{-15}	2.2282×10^{-15}	2.0217×10^{-15}

Table 1: Average KKT violation, feasibility violation, and function value variance.

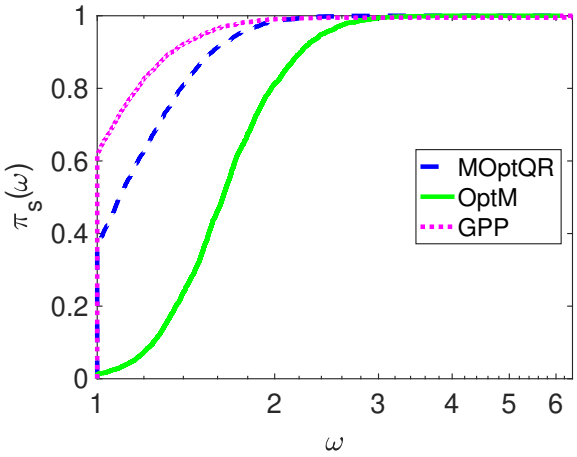


Figure 9: Performance profile on 2304 problems with respect to CPU time.

minimizing quadratic objective over the Stiefel manifold and discretized Kohn–Sham total energy minimization. However, a crucial limitation of this approach is its strict assumption on the objective. In practice, there are quite some critical instances that do not satisfy that assumption.

In this paper, we propose a novel multipliers correction strategy, which minimizes a linear approximation with a proximal term in the range space of the intermediate iterate generated by the function value reduction step. Such correction strategy can guarantee further function value reduction in proportion to the "symmetry" violation. Consequently, the convergent point satisfies the "symmetry" property. We establish the complete global convergence analysis and worst case complexity as well. Furthermore, numerical experiments illustrate that the new multipliers correction methods have better performances than those proposed in [14]. Remarkably, our multipliers correction methods can solve problems that those proposed in [14] can not solve. In solving these testing problems, our methods outperform other state-of-the-art first-order approaches.

References

[1] T. E. Abrudan, J. Eriksson, and V. Koivunen. Steepest descent algorithms for optimization under unitary matrix constraint. *IEEE T. Signal Proces.*, 56(3):1134–1147, 2008.

[2] T. E. Abrudan, J. Eriksson, and V. Koivunen. Conjugate gradient algorithm for optimization under unitary matrix constraint. *Signal Process.*, 89(9):1704 – 1714, 2009.

[3] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust–region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, 2006.

[4] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[5] P.-A. Absil and J. Malick. Projection–like retractions on matrix manifolds. *SIAM J. Optimiz.*, 22(1):135–158, 2012.

[6] K. Anstreicher and H. Wolkowicz. On Lagrangian relaxation of quadratic matrix constraints. *SIAM J. Matrix Anal. A.*, 22(1):41–55, 2000.

[7] J. Barzilai and J. M. Borwein. Two–point step size gradient methods. *IMA J. Numer. Anal.*, 8(1):141–148, 1988.

[8] N. Boumal and P.-A. Absil. Low-rank matrix completion via preconditioned optimization on the Grassmann manifold. *Linear Algebra Appl.*, 475:200–239, 2015.

[9] A. Caboussat, R. Glowinski, and V. Pons. An augmented Lagrangian approach to the numerical solution of a non-smooth eigenvalue problem. *J. Numer. Math.*, 17(1):3–26, 2009.

[10] Y.-H. Dai and R. Fletcher. Projected Barzilai–Borwein methods for large-scale box-constrained quadratic programming. *Numer. Math.*, 100(1):21–47, 2005.

[11] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2):201–213, 2002.

[12] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. A.*, 20(2):303–353, 1998.

[13] B. Gao, G. Hu, Y. Kuang, and X. Liu. An orthogonalization-free parallelizable framework for all-electron calculations in density funcitonal theory. *arXiv:2007.14228*, 2020.

[14] B. Gao, X. Liu, X. Chen, and Y.-x. Yuan. A new first-order algorithmic framework for optimization problems with orthogonality constraints. *SIAM J. Optimiz.*, 28(1):302–332, 2018.

[15] B. Gao, X. Liu, and Y.-x. Yuan. Parallelizable algorithms for optimization problems with orthogonality constraints. *SIAM J. Sci. Comput.*, 41(3):A1949–A1983, 2019.

[16] I. Grubišić and R. Pietersz. Efficient rank reduction of correlation matrices. *Linear Algebra Appl.*, 422(2-3):629–653, 2007.

[17] J. Hu, X. Liu, Z. Wen, and Y.-x. Yuan. A brief introduction to manifold optimization. *J. Oper. Res. Soc. CHN.*, 8(2):199–248, 2020.

[18] X. Hu and X. Liu. An efficient orthonormalization-free approach for sparse dictionary learning and dual principal component pursuit. *Sensors*, 20(11):3041, 2020.

[19] W. Huang, K. A. Gallivan, and P.-A. Absil. A Broyden class of quasi–Newton methods for Riemannian optimization. *SIAM J. Optimiz.*, 25(3):1660–1685, 2015.

[20] B. Jiang and Y.-H. Dai. A framework of constraint preserving update schemes for optimization on Stiefel manifold. *Math. Program.*, 153(2):535–575, 2015.

[21] R. Lai and S. Osher. A splitting method for orthogonality constrained problems. *J. Sci. Comput.*, 58(2):431–449, 2014.

[22] Z. Li, F. Nie, X. Chang, and Y. Yang. Beyond trace ratio: weighted harmonic mean of trace ratios for multiclass discriminant analysis. *IEEE T. Knowl. Data En.*, 29(10):2100–2110, 2017.

[23] X. Liu, X. Wang, Z. Wen, and Y.-x. Yuan. On the convergence of the self–consistent field iteration in Kohn–Sham density functional theory. *SIAM J. Matrix Anal. A.*, 35(2):546–558, 2014.

[24] X. Liu, Z. Wen, X. Wang, M. Ulbrich, and Y.-x. Yuan. On the analysis of the discretized Kohn–Sham density functional theory. *SIAM J. Numer. Anal.*, 53(4):1758–1785, 2015.

[25] X. Liu, Z. Wen, and Y. Zhang. Limited memory block Krylov subspace optimization for computing dominant singular value decompositions. *SIAM J. Sci. Comput.*, 35(3):A1641–A1668, 2013.

[26] X. Liu, Z. Wen, and Y. Zhang. An efficient Gauss–Newton algorithm for symmetric low-rank product matrix approximations. *SIAM J. Optimiz.*, 25(3):1571–1608, 2015.

[27] J. H. Manton. Optimization algorithms exploiting unitary constraints. *IEEE T. Signal Proces.*, 50(3):635–650, 2002.

[28] Y. Nishimori and S. Akaho. Learning algorithms utilizing quasi–geodesic flows on the Stiefel manifold. *Neurocomputing*, 67:106–135, 2005.

[29] J. Noceda and S. J. Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.

[30] G. Rosman, X. C. Tai, R. Kimmel, and A. M. Bruckstein. Augmented Lagrangian regularization of matrix-valued maps. *Methods Appl. Anal.*, 21(1):121–138, 2014.

[31] H. Sato. Riemannian Newton–type methods for joint diagonalization on the Stiefel manifold with application to independent component analysis. *Optimization*, 66(12):2211–2231, 2017.

[32] B. Savas and L.-H. Lim. Quasi–Newton methods on Grassmannians and multilinear approximations of tensors. *SIAM J. Sci. Comput.*, 32(6):3352–3393, 2010.

[33] Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Math. Program.*, 142(1-2):397–434, 2013.

[34] X. Wu, Z. Wen, and W. Bao. A regularized Newton method for computing ground states of Bose–Einstein condensates. *J. Sci. Comput.*, 73(1):303–329, 2017.

[35] N. Xiao, X. Liu, and Y.-x. Yuan. A class of smooth exact penalty function methods for optimization problems with orthogonality constraints. *Optim. Method. Softw.*, 0(0):1–37, 2020.

[36] C. Yang, J. C. Meza, and L.-W. Wang. A trust region direct constrained minimization algorithm for the Kohn–Sham equation. *SIAM J. Sci. Comput.*, 29(5):1854–1875, 2007.