CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2020-0008

Vol. **2**, No. 3, pp. 508-531
September 2021

# Multipliers Correction Methods for Optimization Problems over the Stiefel Manifold

Lei Wang[1,2], Bin Gao[3] and Xin Liu[1,2,*]

[1] *State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China.*
[2] *University of Chinese Academy of Sciences, Beijing, P.R. China.*
[3] *ICTEAM Institute, UCLouvain, 1348 Louvain-la-Neuve, Belgium.*

Received 30 November 2020; Accepted 3 March 2021

**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**: Multipliers correction, proximal approximation, orthogonality constraint, Stiefel manifold.

## 1 Introduction

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

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

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 $\mathcal{S}_{n,p}:=\left\{X\in\mathbb{R}^{n\times p}\,|\,X^\top X=I_p\right\}$, is called the Stiefel manifold.

*Corresponding author. *Email addresses:* `wlkings@lsec.cc.ac.cn` (L. Wang), `gaobin@lsec.cc.ac.cn` (B. Gao), `liuxin@lsec.cc.ac.cn` (X. Liu)

 

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.** $f(X)=h(X)+\mathrm{tr}\left(G^\top X\right)$, 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 $\mathrm{tr}\left(G^\top X\right)$ in the range space of $X$ where finding its global minimizer is equivalent to computing a singular value decomposition.