Journal of Computational Mathematics
Vol.42, No.5, 2024, 1246–1276.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2307-m2021-0331

SOLVING OPTIMIZATION PROBLEMS OVER THE STIEFEL MANIFOLD BY SMOOTH EXACT PENALTY FUNCTIONS*

Nachuan Xiao
The Institute of Operations Research and Analytics, National University of Singapore, Singapore
Email: xnc@lsec.cc.ac.cn
Xin Liu[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
Email: liuxin@lsec.cc.ac.cn

Abstract

In this paper, we present a novel penalty model called ExPen for optimization over the Stiefel manifold. Different from existing penalty functions for orthogonality constraints, ExPen adopts a smooth penalty function without using any first-order derivative of the objective function. We show that all the first-order stationary points of ExPen with a sufficiently large penalty parameter are either feasible, namely, are the first-order stationary points of the original optimization problem, or far from the Stiefel manifold. Besides, the original problem and ExPen share the same second-order stationary points. Remarkably, the exact gradient and Hessian of ExPen are easy to compute. As a consequence, abundant algorithm resources in unconstrained optimization can be applied straightforwardly to solve ExPen.

Mathematics subject classification: 90C30, 65K05.
Key words: Orthogonality constraint, Stiefel manifold, Penalty function.

1. Introduction

In this paper, we consider the following optimization problem:

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

where I_p denotes the $p\times p$ identity matrix, and $f:\mathbb{R}^{n\times p}\mapsto\mathbb{R}$ satisfies the following assumption.

Assumption 1.1 (Blank Assumption on f). *The functions f and ∇f are locally Lipschitz continuous in $\mathbb{R}^{n\times p}$.*

Recall that a mapping $T:\mathbb{R}^{n\times p}\to\mathbb{R}^m$ is locally Lipschitz continuous over $\mathbb{R}^{n\times p}$ if for any $X_0\in\mathbb{R}^{n\times p}$, there exists a constant M and $\delta>0$ such that for any $X\in\mathbb{R}^{n\times p}$ satisfying $\|X-X_0\|_{\mathrm{F}}\le\delta$, it holds that $\|T(X)-T(X_0)\|\le M\|X-X_0\|_{\mathrm{F}}$.

The feasible region of the orthogonality constraints $X^\top X=I_p$ is the Stiefel manifold embedded in the $n\times p$ real matrix space, denoted by

$$\mathcal{S}_{n,p}:=\big\{X\in\mathbb{R}^{n\times p}\mid X^\top X=I_p\big\}.$$

* Received November 20, 2021 / Revised version received November 2, 2022 / Accepted July 26, 2023 / Published online October 27, 2023 /
1) Corresponding author

We also call it as the Stiefel manifold for brevity. Optimization problems with orthogonality constraints have wide applications in statistics [17, 49], scientific computation [37, 43], image processing [6] and many other related areas [25, 41, 68]. Interested readers can refer to some recent works [21, 35, 59, 63], a recent survey [33], and several books [4, 10] for details.

1.1. Motivation

Optimization over the Stiefel manifold, which is a smooth and compact Riemannian manifold, has been discovered to enjoy a close relationship with unconstrained optimization. However, developing optimization approaches over the Stiefel manifold is inherently complicated by the nonconvexity of the manifold. Various existing unconstrained optimization approaches, i.e. the approaches for solving nonconvex unconstrained optimization problems, can be extended to their Riemannian versions by the local diffeomorphisms between the Stiefel manifold and Euclidean space. The approaches, called Riemannian optimization approaches for brevity hereinafter, include gradient descent with line-search [2, 4, 36, 58, 60], conjugate gradient methods [3], Riemannian accelerated gradient method [13, 54, 66, 67], Riemannian adaptive gradient methods [7], etc. With the frameworks and geometrical materials described in [4], theoretical results of these Riemannian optimization approaches have been established by following almost the same proof techniques as their unconstrained prototypes. These results include the global convergence, local convergence rate, worst-case complexity, and saddle-point-escaping properties, [5, 12, 13, 24, 32, 55, 69].

The Riemannian optimization approaches usually consist of two fundamental parts. The first one is the so-called retraction which maps a point from the tangent space to the manifold. Retractions can be further categorized into two classes: the geodesic-like retractions and the projection-like ones. The former ones require to calculate the geodesics along the manifold and hence are expensive. The latter ones enjoy relatively lower computational cost, but as demonstrated in various existing works [21, 61, 63], computing the projection-like retractions is still more expensive than matrix-matrix multiplication. The second part is called parallel transport which moves a tangent vector along a given curve on a Stiefel manifold parallelly. The purpose of parallel transport is to design the manifold version of some advanced unconstrained optimization approaches, such as conjugate gradient methods or gradient methods with momentum. However, as illustrated in [4], computing the parallel transport on Stiefel manifold is equivalent to finding a solution to a differential equation, which is definitely impractical in computation. To this end, the authors of [4] have proposed the concept of vector transport, which can be regarded as an approximation to parallel transport, hence is computationally affordable. Unfortunately, due to the approximation error introduced by vector transports, analyzing the convergence properties of Riemannian optimization algorithms is challenging and usually cannot directly follow the existing results for their unconstrained counterparts, see [35] for instances. As illustrated in various existing works [5, 12, 13, 66, 67, 69], both parallel transports and geodesics play an essential role in establishing convergence properties. It is still difficult to verify whether their theoretical convergence properties is valid when these approaches are built by retractions and vector transports.

To avoid computing the retractions, parallel transports, or vector transports to the Stiefel manifold, some approaches aim to find smooth mappings from the Euclidean space to the Stiefel manifold, which directly reformulates OCP to unconstrained optimization. Among them, [38,39] construct equivalent unconstrained problems for OCP by exponential function for square matrices. To efficiently compute the matrix exponential, they apply the iterative approach