Journal of Computational Mathematics
Vol.39, No.2, 2021, 207–226.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1911-m2018-0242

ACCELERATED OPTIMIZATION WITH ORTHOGONALITY CONSTRAINTS[*]

Jonathan W. Siegel
Department of Mathematics, Pennsylvania State University, University Park, PA, USA
Email: jus1949@psu.edu, jwsiegel2510@gmail.com

Abstract

We develop a generalization of Nesterov's accelerated gradient descent method which is designed to deal with orthogonality constraints. To demonstrate the effectiveness of our method, we perform numerical experiments which demonstrate that the number of iterations scales with the square root of the condition number, and also compare with existing state-of-the-art quasi-Newton methods on the Stiefel manifold. Our experiments show that our method outperforms existing state-of-the-art quasi-Newton methods on some large, ill-conditioned problems.

Mathematics subject classification: 65K05, 65N25, 90C30, 90C48.
Key words: Riemannian optimization, Stiefel manifold, Accelerated gradient descent, Eigenvector problems, Electronic structure calculations.

1. Introduction

Optimization problems over the set of orthonormal matrices arise naturally in many scientific and engineering problems. Most notably, eigenfunction and electronic structure calculations involve minimizing functions over the set of orthonormal matrices [1,3,9,21]. In these applications, the objective functions are smooth but often ill-conditioned. There are also more recent applications which involve non-smooth objectives, most notably the calculation of compressed modes [13,14], which involve an L^1 penalization of variational problems arising in physics.

In this paper, we consider optimization problems with orthogonality constraints, i.e. problems of the form

$$\arg\min_{X^TX=I_k} f(X), \tag{1.1}$$

where X is an $n \times k$ matrix, I_k is the identity matrix, and f is a smooth function. The manifold of orthonormal matrices over which we are optimizing is referred to as the Stiefel manifold in the literature. Many methods have been proposed for solving (1.1), including variants of gradient descent, Newton's method, quasi-Newton methods, and non-linear conjugate gradient methods [1,3,4,18,22]. However, existing methods can suffer from slow convergence when the problem is ill-conditioned, by which we mean that the Hessian of f at (or near) the minimizer is ill-conditioned [3]. Such problems are of particular interest, since they arise when doing electronic structure calculations, or when solving non-smooth problems by smoothing the objective, for instance. Moreover, preconditioning such problems can be very difficult due to the manifold constraint.

In an attempt to solve ill-conditioned problems more efficiently, we develop an extension of the well-known Nesterov's accelerated gradient descent algorithm [10] designed for optimizing

[*] Received October 29, 2018 / Revised version received April 11, 2019 / Accepted November 6, 2019 / Published online May 5, 2020 /

functions on the Stiefel manifold. For the class of smooth, strongly convex functions on $\mathbb{R}^n$, accelerated gradient descent obtains an asymptotically optimal iteration complexity of $O(\sqrt{\kappa})$, compared to $O(\kappa)$ for gradient descent with optimal step size selection (see [2], section 3.7, here κ is the condition number of the problem). Our method extends this convergence behavior from $\mathbb{R}^n$ to the Stiefel manifold, thus providing an efficient method for solving ill-conditioned optimization problems with orthogonality constraints.

Other work on accelerated gradient methods on manifolds includes [8] and [20]. In [8] an accelerated gradient method on general manifolds is presented. However, their algorithm involves solving a non-linear equation involving both the metric on the manifold and the objective function f. Unfortunately, solving this equation is only feasible for the special type of model problem which they consider and cannot be generally applied to arbitrary optimization problems on the Stiefel manifold. In [20], a theory is developed which shows that a certain type of accelerated method can achieve accelerated convergence locally. However, their method involves calculating a geodesic logarithm in every iteration and has not yet been implemented, although an iterative method for calculating the geodesic logarithm has been developed in [23]. In constrast, our method only involves very simple linear algebra calculations in each iteration and can be run efficiently on large problems.

The paper is organized as follows. In Section 2, we briefly introduce the necessary notation and ideas from differential geometry. In Section 3, we discuss accelerated gradient descent on $\mathbb{R}^n$. We recall results which are relevant to our work. In Section 4, we detail the design of our method. One of the key ingredients is an efficient procedure for performing approximate extrapolation and interpolation on the manifold, which we believe could be useful in developing other optimization methods. In Section 5, we show numerical results which provide evidence that our method achieves the desired iteration complexity. Finally, in Section 6, we present comparisons with other optimization methods on the Stiefel manifold. We show that our method outperforms existing state of the art methods on some large, ill-conditioned problems.

2. Riemannian Manifolds

In this section, we briefly introduce the notation we will use in the rest of the paper concerning the Stiefel manifold and differential geometry in general. We also collect some formulas for calculating on the Stiefel manifold which will be used later. Some references for differential geometry include [6, 16] and for the geometry of the Stiefel manifold, see [3].

Let M be a smooth manifold and $x \in M$. We denote the tangent space of M at x by T_xM and the dual tangent space by $(T_xM)^*$. We denote the tangent bundle of $M$by TM, and likewise the dual tangent bundle by $(TM)^*$.

Suppose f is a C^1 function on M. We denote the derivative of f, by $\nabla f(x) \in (T_xM)^*$. Notice that the derivative of f is naturally an element of the dual tangent space $(T_xM)^*$. If M is a Riemannian manifold, then each tangent space T_xM is equipped with a positive definite inner product $g : T_xM \times T_xM \to \mathbb{R}$. We denote the norm induced by g as $\|v\|_g^2$ and the norm induced by g on the dual space as $\|w\|_{g*}$.

Additionally, the inner product g provides an isometry between the tangent space and its dual, which we denote by $\phi_g : (T_xM)^* \to T_xM$ and $\phi_g^{-1} : T_xM \to (T_xM)^*$, and which are also known as raising and lowering indices. Given a C^1 function f on a Riemannian manifold M, an object which is often considered is the Riemannian gradient, obtained by raising the indices of the gradient $\nabla f(x) \in (T_xM)^*$ to obtain a tangent vector (instead of a dual tangent vector).