CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.2020-0026

Vol. **1**, No. 4, pp. 766-801
December 2020

A Derivative-Free Geometric Algorithm for Optimization on a Sphere

Yannan Chen[1], Min Xi[2] and Hongchao Zhang[3,*]

[1] *School of Mathematical Sciences, South China Normal University, Guangzhou, China.*

[2] *School of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangzhou China; School of Mathematical Sciences, Jiangsu Key Laboratory for NSLSCS, Nanjing Normal University, Nanjing, China.*

[3] *Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918, USA.*

Received 8 June 2020; Accepted 21 October 2020

Abstract. Optimization on a unit sphere finds crucial applications in science and engineering. However, derivatives of the objective function may be difficult to compute or corrupted by noises, or even not available in many applications. Hence, we propose a Derivative-Free Geometric Algorithm (DFGA) which, to the best of our knowledge, is the first derivative-free algorithm that takes trust region framework and explores the spherical geometry to solve the optimization problem with a spherical constraint. Nice geometry of the spherical surface allows us to pursue the optimization at each iteration in a local tangent space of the sphere. Particularly, by applying Householder and Cayley transformations, DFGA builds a quadratic trust region model on the local tangent space such that the local optimization can essentially be treated as an unconstrained optimization. Under mild assumptions, we show that there exists a subsequence of the iterates generated by DFGA converging to a stationary point of this spherical optimization. Furthermore, under the Łojasiewicz property, we show that all the iterates generated by DFGA will converge with at least a linear or sublinear convergence rate. Our numerical experiments on solving the spherical location problems, subspace clustering and image segmentation problems resulted from hypergraph partitioning, indicate DFGA is very robust and efficient for solving optimization on a sphere without using derivatives.

AMS subject classifications: 65K05, 90C30, 90C56

Key words: Derivative-free optimization, spherical optimization, geometry, trust region method, Łojasiewicz property, global convergence, convergence rate, hypergraph partitioning.

---

*Corresponding author. *Email addresses:* `ynchen@scnu.edu.cn` (Y. Chen), `mxi@gdufs.edu.cn` (M. Xi), `hozhang@math.lsu.edu` (H. Zhang)

1 Introduction

In this paper, we consider the following spherical optimization problem

$$\min f(\mathbf{x}) \qquad \text{s.t. } \mathbf{x}\in\mathcal{S}^{n-1}, \tag{1.1}$$

where $\mathcal{S}^{n-1}:=\{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}\|=1\}$ is a unit sphere under the Euclidean norm $\|\cdot\|$ and $f:\mathcal{S}^{n-1}\to\mathbb{R}$ is continuously differentiable with Lipschitz continuous gradient. However, we assume that the derivatives of f are unavailable during algorithm designment. The spherical optimization problem (1.1) has extensive applications in science and engineering. For example, the classical Weber problem is to find the best location on a three dimensional sphere which minimizes the weighted sum of the distances to several destination points on the sphere [37,55]. In geophysics, climate modelling and global navigation, various nonlinear optimization problems on a sphere need to be solved for dealing with massive signals on the surface of the earth [13,18]. Finding the largest and smallest Z-eigenvalues of an even order symmetric tensor [26,48] is equivalent to calculate the maximum and minimum values of a homogeneous polynomial associated with a tensor on a unit sphere, respectively. The best rank-one approximation of a symmetric tensor could be also formulated as a spherical optimization [58]. Other spherical optimization problems which have nonsmooth objectives include the robust subspace detection [29], the sparse principal component analysis (PCA) [2,54], and the sparse blind deconvolution [17,34] etc. In addition, for some practical applications, the data may come from simulations or experiments, for which the analytic derivatives of the objective function are unavailable or prohibitively expensive to compute. For example, the objective functions proposed in [27,39] depend on random variables whose distributions are unknown. In particular, for studying precision medicine, it is proposed to maximize the hypervolume under the manifold (HUM) [27], which can be interpreted as the probability of disease detection. Other examples include [28,49], where the evaluation of objective function needs solution of differential equations, and therefore, it is expensive or impossible to compute derivatives of the objective functions at each iteration. Hence, developing derivative-free algorithms for solving (1.1), which only uses the function values of f, has great importance in both theory and applications.

Recently, derivative-free optimization (DFO) has become an important research topic in nonlinear optimization since derivatives of the objective function may be difficult to compute or corrupted by noises, or even not available in many real applications. Hence, DFO methods need to be developed to solve these optimization problems without using derivatives. Currently, the DFO methods can be generally divided into three classes. The first class of methods approximate derivatives by finite-differences and then derivative-based methods can be applied using the approximate derivatives [12,14,32]. For instance, Nocedal et al. [14] combines the classical BFGS updating and an adaptive finite-difference technique for minimization without derivatives. The second class of methods are direct search methods [38], which for example include pattern search methods [33,53], Nelder-Mead simplex method [43] and mesh adaptive direct search methods [7]. This class of