CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2021-0016

Vol. **2**, No. 4, pp. 585-651
December 2021

REVIEW ARTICLE

Subspace Methods for Nonlinear Optimization

Xin Liu[1], Zaiwen Wen[2,*] and Ya-Xiang Yuan[3]

[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.*
[2] *Beijing International Center for Mathematical Research, Peking University, China.*
[3] *State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China.*

Received 5 March 2021; Accepted 21 May 2021

Abstract. Subspace techniques such as Krylov subspace methods have been well known and extensively used in numerical linear algebra. They are also ubiquitous and becoming indispensable tools in nonlinear optimization due to their ability to handle large scale problems. There are generally two types of principals: i) the decision variable is updated in a lower dimensional subspace; ii) the objective function or constraints are approximated in a certain smaller subspace of their domain. The key ingredients are the constructions of suitable subspaces and subproblems according to the specific structures of the variables and functions such that either the exact or inexact solutions of subproblems are readily available and the corresponding computational cost is significantly reduced. A few relevant techniques include but not limited to direct combinations, block coordinate descent, active sets, limited-memory, Anderson acceleration, subspace correction, sampling and sketching. This paper gives a comprehensive survey on the subspace methods and their recipes in unconstrained and constrained optimization, nonlinear least squares problem, sparse and low rank optimization, linear and nonlinear eigenvalue computation, semidefinite programming, stochastic optimization and etc. In order to provide helpful guidelines, we emphasize on high level concepts for the development and implementation of practical algorithms from the subspace framework.

AMS subject classifications: 65K05, 90C30

Key words: Nonlinear optimization, subspace techniques, block coordinate descent, active sets, limited memory, Anderson acceleration, subspace correction, subsampling, sketching.

*Corresponding author. *Email addresses:* `liuxin@lsec.cc.ac.cn` (X. Liu), `wenzw@pku.edu.cn` (Z. Wen), `yyx@lsec.cc.ac.cn` (Y.-X. Yuan)

1 Introduction

Large scale optimization problems appear in a wide variety of scientific and engineering domains. In this paper, we consider a general optimization problem

$$\min_x f(x), \quad \text{s.t. } x \in \mathcal{X}, \tag{1.1}$$

where x is the decision variable, $f(x)$ is the objective function and $\mathcal{X}$ is the feasible set. Efficient numerical optimization algorithms have been extensively developed for (1.1) with various types of objective functions and constraints [89,113]. With the rapidly increasing problem scales, subspace techniques are ubiquitous and becoming indispensable tools in nonlinear optimization due to their ability to handle large scale problems. For example, the Krylov subspace methods developed in the numerical linear algebraic community have been widely used for the linear least squares problem and linear eigenvalue problem. The characteristics of the subspaces are clear in many popular optimization algorithms such as the linear and nonlinear conjugate gradient methods, Nesterov's accelerated gradient method, the Quasi-Newton methods and the block coordinate decent (BCD) method. The subspace correction method for convex optimization can be viewed as generalizations of multigrid and domain decomposition methods. The Anderson acceleration or the direct inversion of iterative subspace (DIIS) methods have been successful in computational quantum physics and chemistry. The stochastic gradient type methods usually take a mini-batch from a large collection samples so that the computational cost of each inner iteration is small. The sketching techniques formulate a reduced problem by a multiplication with random matrices with certain properties.

The purpose of this paper is to provide a review of the subspace methods for nonlinear optimization, for their further improvement and for their future usage in even more diverse and emerging fields. The subspaces techniques for (1.1) are generally divided into two categories. The first type is to update the decision variable in a lower dimensional subspace, while the second type is to construct approximations of the objective function or constraints in a certain smaller subspace of functions. Usually, there are three key steps.

- Identify a suitable subspace either for the decision variables or the functions.
- Construct a proper subproblem by various restrictions or approximations.
- Find either an exact or inexact solution of subproblems.

These steps are often mixed together using the specific structures of the problems case by case. The essence is how to reduce the corresponding computational cost significantly. The collection of subspaces techniques is growing ever rich in unconstrained and constrained optimization, nonlinear least squares problem, sparse and low rank optimization, linear and nonlinear eigenvalue computation, semidefinite programming, stochastic optimization, manifold optimization, phase retrieval, variational minimization and