Journal of Computational Mathematics
Vol.43, No.4, 2025, 1016–1044.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2404-m2023-0128

AN SQP-TYPE PROXIMAL GRADIENT METHOD FOR COMPOSITE OPTIMIZATION PROBLEMS WITH EQUALITY CONSTRAINTS*

Pinzheng Wei[1) and Weihong Yang
School of Mathematical Sciences, Fudan University, Shanghai 200433, China
Emails: pzwei21@m.fudan.edu.cn, whyang@fudan.edu.cn

Abstract

In this paper, we present an SQP-type proximal gradient method (SQP-PG) for composite optimization problems with equality constraints. At each iteration, SQP-PG solves a subproblem to get the search direction, and takes an exact penalty function as the merit function to determine if the trial step is accepted. The global convergence of the SQP-PG method is proved and the iteration complexity for obtaining an ϵ-stationary point is analyzed. We also establish the local linear convergence result of the SQP-PG method under the second-order sufficient condition. Numerical results demonstrate that, compared to the state-of-the-art algorithms, SQP-PG is an effective method for equality constrained composite optimization problems.

Mathematics subject classification: 90C30, 65K05.
Key words: Composite optimization, Proximal gradient method, SQP method, Semismooth Newton method.

1. Introduction

In this paper, we consider the following problem:

$$\begin{aligned} &\min_X \psi(X) := f(X) + h(X) \\ &\text{s.t.} \quad c(X) = 0, \end{aligned} \tag{1.1}$$

where $f : \mathbf{E} \to \mathbb{R}$ is a smooth function, $h : \mathbf{E} \to \mathbb{R}$ is a convex nonsmooth function, and c is a continuously differentiable mapping from $\mathbf{E}$ to $\mathbf{F}$. Here $\mathbf{E}$ and $\mathbf{F}$ are Euclidean spaces. The feasible set for (1.1) is denoted by $\Omega := \{X \in \mathbf{E} : c(X) = 0\}$.

If $\mathbf{E} = \mathbb{R}^{n\times p}$ and $c : \mathbf{E} \to S^p$ is defined by $c(X) = X^\top X - I_p$, where S^p denotes the space of $p \times p$ symmetric matrices, then Ω is just the Stiefel manifold $\mathrm{St}(n,p)$, and (1.1) becomes the composite optimization problem over $\mathrm{St}(n,p)$. Such problems have wide applications in many fields such as machine learning, signal processing and numerical linear algebra. For more details about these applications, we refer the readers to [1, 2, 11] and the references therein.

Recently, many numerical algorithms have been proposed for solving composite optimization problems over the Stiefel manifold. These methods can be classified into four categories: subgradient methods, operator splitting methods, augmented Lagrangian (AL) methods and

* Received June 15, 2023 / Revised version received September 26, 2023 / Accepted April 16, 2024 / Published online June 11, 2024 /
1) Corresponding author

proximal-type methods. For detailed discussions of these methods, we refer the reader to [11,46]. Here we only review them briefly. Grohs and Hosseini [21] propose an ϵ-subgradient method and establish the global converge result. Zhang *et al.* [44] extend the smoothing steepest descent method for nonconvex and non-Lipschitz optimization from Euclidean space to Riemannian manifolds. This method can be applied to solve composite optimization problems. Lai *et al.* [26] propose a splitting method for orthogonality constrained problems. There are several excellent works denoted to AL methods for composite optimization. Gao *et al.* [18] propose a parallelized proximal linearized AL algorithm. Zhou *et al.* [46] present a manifold-based AL method to solve composite optimization problems. The AL methods in [13, 30] can also be used to solve (1.1). For proximal gradient methods, Huang and Wei [25] propose a Riemannian proximal gradient (RPG) method and they analyze the iteration complexity of their method under some assumptions. Chen *et al.* [11] present a proximal gradient method, named ManPG, which can be viewed as an inexact RPG method. To accelerate the ManPG method, Wang and Yang [40] propose a proximal quasi-Newton method.

For manifold-based methods, an important operation is the so-called retraction (see [2, Chapter 4]), which maps a tangent vector to a point on the manifold. For the general equality constraint $c(X) = 0$, if we use manifold-based methods to solve (1.1), we can only use the nearest-point projection as the retraction (see [3, Theorem 15]). Usually, computing the nearest-point is expensive for large-scale problems, which results in that the total computational cost of manifold-based methods is very large.

In this paper, we use a different approach and propose an SQP-type method, named SQP-PG, to solve the problem (1.1). Sequential quadratic programming (SQP) methods were first proposed in 1963 by Wilson [41] and were developed in the 1970s by Garcia-Palomares and Mangasarian [19], Han [22, 23], and Powell [34, 35]. For recent developments in SQP methods, the reader is referred to [7, 8, 10, 15, 20, 28, 29]. We also refer to the monograph [32] and the references therein for detailed discussions on SQP methods.

At the k-th iteration, SQP-PG solves the following subproblem:

$$\begin{aligned} &\min \frac{1}{2}\langle V, \mathcal{B}_k[V]\rangle + \langle \nabla f(X_k), V\rangle + h(X_k + V) \\ &\text{s.t.} \quad c(X_k) + Dc(X_k)[V] = 0, \end{aligned} \tag{1.2}$$

where $\mathcal{B}_k$ is an approximated Hessian operator on $\mathbf{E}$, and $Dc(X_k)$ is the derivative of the mapping c at X_k. Similar to the traditional SQP methods, SQP-PG takes an exact penalty function as the merit function to determine if the trial step $X_{k+1} = X_k + \alpha_k V_k$ is accepted or not, where V_k is a non-zero solution of (1.2). An appealing feature of our method is that, compared to the Riemannian manifold optimization method, it does not involve the computation of retraction. Numerical experiments demonstrate that the SQP-PG method is quite efficient especially when the retraction to $\Omega = \{X \in \mathbf{E} : c(X) = 0\}$ is expensive.

The organization of the paper is shown as follows. In Section 2, we introduce some notations and definitions that will be used throughout the paper. In Section 3, we propose the SQP-PG algorithm in detail. The global convergence of SQP-PG is proved and the iteration complexity for obtaining an ϵ-stationary point is analyzed in Section 4. Under the second-order sufficient condition, we also establish the local linear convergence result of SQP-PG in this section. In Section 5, we compare the SQP-PG method with some state-of-the-art methods in the numerical experiments. The paper ends with some conclusions and a short discussion on possible future works.