Journal of Computational Mathematics
Vol.41, No.4, 2023, 683–716.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2110-m2020-0317

A TRUST-REGION METHOD FOR NONSMOOTH NONCONVEX OPTIMIZATION*

Ziang Chen
Department of Mathematics, Duke University, USA
Email: ziang@math.duke.edu
Andre Milzarek
School of Data Science (SDS), The Chinese University of Hong Kong, Shenzhen,
Shenzhen Research Institute for Big Data (SRIBD), China
Email: andremilzarek@cuhk.edu.cn
Zaiwen Wen[1)]
Beijing International Center for Mathematical Research, Peking University, Beijing 100000, China
Email: wenzw@pku.edu.cn

Abstract

We propose a trust-region type method for a class of nonsmooth nonconvex optimization problems where the objective function is a summation of a (probably nonconvex) smooth function and a (probably nonsmooth) convex function. The model function of our trust-region subproblem is always quadratic and the linear term of the model is generated using abstract descent directions. Therefore, the trust-region subproblems can be easily constructed as well as efficiently solved by cheap and standard methods. When the accuracy of the model function at the solution of the subproblem is not sufficient, we add a safeguard on the stepsizes for improving the accuracy. For a class of functions that can be "truncated", an additional truncation step is defined and a stepsize modification strategy is designed. The overall scheme converges globally and we establish fast local convergence under suitable assumptions. In particular, using a connection with a smooth Riemannian trust-region method, we prove local quadratic convergence for partly smooth functions under a strict complementary condition. Preliminary numerical results on a family of ℓ_1-optimization problems are reported and demonstrate the efficiency of our approach.

Mathematics subject classification: 90C30, 65K05, 90C06.
Key words: Trust-region method, Nonsmooth composite programs, Quadratic model function, Global and local convergence.

1. Introduction

We consider the unconstrained nonsmooth nonconvex optimization problems of the composite form:

$$\min_{x\in\mathbb{R}^n} \psi(x) := f(x) + \varphi(x), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable but probably nonconvex function and $\varphi : \mathbb{R}^n \to \mathbb{R}$ is real-valued and convex. The composite program (1.1) is a special form of the general nonsmooth nonconvex optimization problems

$$\min_{x\in\mathbb{R}^n} \psi(x), \tag{1.2}$$

* Received December 17, 2020 / Revised version received July 31, 2021 / Accepted October 8, 2021 / Published online February 25, 2022 /
[1)] Corresponding author

where the objective function $\psi : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous, and has numerous applications, such as ℓ_1-regularized problems [12, 27, 41, 68, 71], group sparse problems [22, 54, 70, 77], penalty approaches [5], dictionary learning [31, 50], and matrix completion [18, 19, 37].

1.1. Related Work

Different types of nonsmooth trust-region methods have already been proposed and analyzed for the general optimization problem (1.2) throughout the last two decades. Several of these nonsmooth trust-region methods utilize abstract model functions on a theoretical level which means that the model function is typically not specified. In [26], a nonsmooth trust-region method is proposed for (1.2) under the assumption that ψ is regular. A nonsmooth trust-region algorithm for general problems is investigated in [64]. In this work, an abstract first-order model is considered that is not necessarily based on subgradient information or directional derivatives. Extending the results in [64], Grohs and Hosseini propose a Riemannian trust-region algorithm, see [34]. Here, the objective function is defined on a complete Riemannian manifold. All mentioned methods derive global convergence under an assumption similar to the concept of a "strict model" stated in [59]. Using this concept, a nonsmooth bundle trust-region algorithm with global convergence is constructed in [4].

In [20], a hybrid approach is presented using simpler and more tractable quadratic model functions. The method switches to a complicated second model if the quadratic model is not accurate enough and if it is strictly necessary. In [3], a quadratic model function is analyzed where the first-order term is derived from a suitable approximation of the steepest descent direction and the second-order term is updated utilizing a BFGS scheme. The authors apply an algorithmic approach proposed in [49] to compute the approximation of the ϵ-subdifferential and steepest descent direction. Another class of methods employs smoothing techniques. In [32], the authors first present a smooth trust-region method without using derivatives, and then, in the nonsmooth case, use this methodology after smoothing the objective function. Furthermore, trust-region algorithms for nonsmooth problems can be developed based on smooth merit functions for the problem. In [66], a nonsmooth convex optimization is investigated and the Moreau envelope is considered as a smooth merit function. A smooth trust-region method is performed on the smooth merit function, where the second-order term of the model function is again updated by the BFGS formula.

Bundle methods are an important and related class of methods for nonsmooth problems [39, 40, 42, 44, 45, 51, 52]. The ubiquitous cutting-plane model in bundle methods is polyhedral, i.e., the supremum of a finite affine family. This model builds approximations of convex functions based on the subgradient inequality. In [67], an efficient bundle technique for convex optimization has been proposed; in [24], a convex bundle method is derived to deal with additional noise, i.e., the case when the objective function and the subgradient can not be evaluated exactly. Different modifications of the bundle ideas for nonconvex problems have been established in [59, 67]. In [35] and [60], the authors consider bundle methods for nonsmooth nonconvex optimization when the function values and the subgradients of ψ can only be evaluated inexactly.

Local convergence properties and rates for nonsmooth problems are typically studied utilizing additional and more subtle structures. In this regard, some fundamental and helpful concepts are the idea of an "active manifold" and the family of "partly smooth" functions introduced by [46]. In particular, the problem (1.1) has been investigated when the nonsmooth term φ is partly smooth relative to a smooth active manifold. The so-called finite activity