A Framework of Subgradient Methods for Solving Nonsmooth, Non-Lipschitz Convex Optimization Problems

Jun-Kai You

16 December, 2024

Abstract

This paper presents a framework called *radial transformations* to address a general convex optimization problem. Building on this framework, we provide several subgradient methods and show their convergence results, even in the absence of Lipschitz continuity and smoothness assumptions. Additionally, these algorithms eliminate the need for expensive orthogonal projections typically required by standard subgradient methods.

1 Introduction

This paper considers an algorithmic framework first proposed by Renegar's work [Ren16] for solving constrained convex optimization problems, whose objective functions may be non-Lipschitz and nonsmooth. This kind of general convex optimization problem encompasses several applications in many fields, such as portfolio selection in finance [MTZ12], Poisson inverse problems in imaging processing [BBDV09, HW16], quantum state tomography in quantum physics [Hra97, Pv04], and minimization of quantum Rényi divergences in quantum information theory [YCL22]. While their objective functions are convex, they violate the Lipschitz and smoothness assumptions that are common in the literature.

One of the most known approaches to solving these constrained convex optimization problems is a projected subgradient method, which has received much attention recently due to its simplicity and scalability [Nes04, Bub15]. The usual iteration complexity guarantees for subgradient methods are established under two conditions on the objective function: Lipschitz continuity and smoothness. However, these assumptions limit the applicability of subgradient methods, as many objective functions of applications, including those mentioned above, do not meet these two conditions.

To address the absence of Lipschitz continuity and smoothness in the objective function, certain ways to modify classical algorithms have been proposed, including line search methods [LM23], normalized search directions [Gri19], and non-Euclidean metrics [LFN18, BBT17]. The approach we follow in this paper is to consider a

framework using radial projections [Ren16, Gri18, Gri24a]. This *radial transformations* framework provides a foundation for designing projection-free optimization algorithms, significantly reducing the per-iteration computational cost.

2 Mathematical Background

We consider the problem of minimizing a lower semi-continuous convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, where the objective function f is neither assumed to be smooth nor Lipschitz continuous. Any constrained convex optimization can be represented within this setting by assigning an infinite function value to all infeasible points. We start with the radial point and set transformations, which serve as foundational elements for establishing the radial function transformation.

Definition 2.1. *The Radial Point Transformation of a point* $(x, u) \in \mathbb{R}^n \times \mathbb{R}_{++}$ *is*

$$\Gamma(x, u) = \frac{(x, 1)}{u} = (\frac{x}{u}, \frac{1}{u}),$$

where $\mathbb{R}_{++} = \{u \in \mathbb{R} \mid u > 0\}$ *is the set of positive real numbers.*

Applying the radial point transformation elementwise to a set gives *raidal set transformation*. We want to use these two to give transformation between functions lining up what we had for an indicator turning into a gauge.

Definition 2.2. *The Radial Set Transformation of a set* $S \subseteq \mathbb{R}^n \times \mathbb{R}_{++}$ *is*

$$\Gamma S = \{\Gamma(x, u) \mid (x, u) \in S\}.$$

Remark 2.1. *Both the radial point and set transformations are dual since*

$$\Gamma\Gamma(x, u) = \Gamma\frac{(x, 1)}{u} = \frac{(x/u, 1)}{1/u} = (x, u).$$

To extend the radial set transformation to functions, we consider functions $f : \mathbb{R}^n \to \mathbb{R}_{++} \cup \{0, \infty\}$ mapping into the extended positive real numbers. Then we define *the radial function transformation* as follows.

Definition 2.3. *The radial function transformation of* f *is*

$$f^\Gamma(y) = \sup\{v > 0 \mid (y, v) \in \Gamma(\operatorname{epi} f)\},$$

where $\operatorname{epi} f = \{(x, u) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid f(x) \le u\}$ *denotes the epigraph of* f.

The following theorem demonstrates that the duality of the point and set transformations extends to the radial function transformations.

Theorem 2.1. *The radial function transformation of* f *is dual (i.e.,* $f^{\Gamma\Gamma} = f$*) if and only if* f *is convex and upper semi-continuous.*

Remark 2.2. *We find that for a wide range of functions, the duality of the point and set transformations carry over to the function transformations. For example, for any convex set $S \subseteq \mathbb{R}^n$ strictly containing the origin, we define its indicator function as*

$$\hat{\imath}_S(y) = \begin{cases} +\infty & \text{if } y \in S, \\ 0 & \text{if } y \notin S. \end{cases}$$

We have $\hat{\imath}_S^{\Gamma\Gamma}(y) = \hat{\imath}_S(y)$. The radial function transformation of $\hat{\imath}_S(y)$ is the Minkowski gauge of a set S. That is,

$$\hat{\imath}_S^{\Gamma}(y) = \inf\{\lambda \geq 0 \mid y \in \lambda S\}.$$

The following defines the performance metrics to evaluate algorithms in terms of their convergence guarantees.

Definition 2.4. *A point x has an absolute accuracy of ϵ if $f(x) - f^\star < \epsilon$. A point x has a relative accuracy of ϵ if $(f(x) - f^\star) / (f(x_0) - f^\star) < \epsilon$, where x_0 is a point in the interior of the domain of f is known.*

3 Literature Review

The general convex optimization problem in Section 2 is the focus of Renegar's work [Ren16]. Renegar introduces a framework that reformulates such a problem into an equivalent uniformly Lipschitz problem. Using a projection-free subgradient approach on the reformulated problem, he establishes convergence bounds comparable to those of traditional methods that rely on the assumption of Lipschitz continuity. A key distinction in Renegar's approach is that it provides a guarantee of relative accuracy rather than absolute accuracy in the solution.

Grimmer [Gri18] later extends Renegar's work [Ren16] by introducing the concept of a radial transformation. Instead of first converting the convex problem f into conic form, the radial transformation operates directly on the original function. Furthermore, Grimmer presents modifies versions of Renegar's algorithms (referred to as Algorithms A and B in [Ren16]), which not only simplify the conceptual framework but also lead to better convergence rates.

In a subsequent series of papers [Gri24a, Gri24b], Grimmer develops a comprehensive generalization of the ideas from [Ren16] and [Gri18]. The first part of this work [Gri24a] establishes the foundations for radial transformations in the following three ways: (1) The radial transformation is often self-inverse, with a rich structure arising from this property. (2) It creates a novel duality between nonnegative optimization problems. (3) The radial transformation is the unique operation of its kind. The upper transformation described in [Gri24a] is closely related to the transformation used by Grimmer [Gri18] and Renegar [Ren16].

The second part of Grimmer's work [Gri24b] builds upon the theory of radial transformations from [Gri24a] to develop new projection-free algorithms. This work analyzes radial subgradient, smoothing, and accelerated methods for a wide class of non-Lipschitz problems. These algorithms retain the advantages of their predecessors,

avoiding Lipschitz continuity assumptions and costly orthogonal projections. Additionally, this work demonstrates a broad range of applications and algorithms that can be addressed using radial duality theory.

Personal Biography

Jun-Kai You was born in Taichung, Taiwan. He received a master's degree in Electrical Engineering from National Tsing Hua University. He is a first-year PhD student in Industrial and Systems Engineering at Lehigh University, under the supervision of Prof. Tamás Terlaky and Prof. Luis F. Zuluaga. His research interests lie in optimization and machine learning theory, with his previous work focused on the design and analysis of algorithms for solving optimization problems defined with logarithmic losses. From 2023 to 2024, Jun-Kai worked as a Software and Development Engineer (TVC) on the Pixel camera 3A algorithm team at Google in Taiwan. From 2022 to 2023, he was a Research Engineer at MediaTek Research, specializing in chip-placement projects. Between 2019 to 2022, he served as a Research Assistant in Computer Science and Information Engineering at National Taiwan University, where he was advised by Prof. Yen-Huan Li and collaborated with Prof. Hao-Chung Cheng.

References

[BBDV09] M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini. Imagede blurring with poisson data: from cells to galaxies. *Inverse Probl.*, 25, 2009.

[BBT17] Heinz H. Bauschke, Jérôme Bolte, and Marc Teboulle. A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Math. Oper. Res.*, 42(2):330–348, 2017.

[Bub15] Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.*, 8(3-4):231–357, 2015.

[Gri18] Benjamin Grimmer. Radial subgradient method. *SIAM J. Optim.*, 28(1):459–469, 2018.

[Gri19] Benjamin Grimmer. Convergence rates for deterministic and stochastic subgradient methods without Lipschitz continuity. *SIAM J. Optim.*, 29(2):1350–1365, 2019.

[Gri24a] Benjamin Grimmer. Radial duality part i: foundations. *Math. Program.*, 205(1):33–68, 2024.

[Gri24b] Benjamin Grimmer. Radial duality part ii: applications and algorithms. *Math. Program.*, 205(1):69–105, 2024.

[Hra97] Z. Hradil. Quantum-state estimation. *Phys. Rev. A*, 55(3), 1997.

[HW16]   T. Hohage and F. Werner. Inverse problems with poisson data: statistical regularization theory, applications and algorithms. *Inverse Probl.*, 32, 2016.

[LFN18]   Haihao Lu, Robert M. Freund, and Yurii Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM J. Optim.*, 28(1):333–354, 2018.

[LM23]   Zhaosong Lu and Sanyou Mei. Accelerated first-order methods for convex optimization with locally Lipschitz continuous gradient. *SIAM J. Optim.*, 33(3):2275–2310, 2023.

[MTZ12]   Leonard C. MacLean, Edward O. Thorp, and William T. Ziemba. *The Kelly Capital Growth Investment Criterion*. World Scientific, Singapore, 2012.

[Nes04]   Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 2004.

[Pv04]   M. Paris and J. Řeháček. *Quantum State Estimation*. Springer, Berlin, 2004.

[Ren16]   James Renegar. "Efficient" subgradient methods for general convex optimization. *SIAM J. Optim.*, 26(4):2649–2676, 2016.

[YCL22]   Jun-Kai You, Hao-Chung Cheng, and Yen-Huan Li. Minimizing quantum rényi divergences via mirror descent with polyak step size. In *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022.