First Order Nonsmooth Optimization: Algorithm Design, Analysis, Convergence, and Applications

*Hongda Li

December 22, 2024

Abstract

The research proposal focuses on the theories and practice in solving nonsmooth optimization. The theme of proposal highlight topics of interests that emphasize the computations and applications aspect of algorithms that exhibits both practical and theoretical importance. We summarize our ongoing research in unifying Nesterov type accelerated proximal gradient method and proposes our Free R-WAPG method. We survey literatures under the topic of Catalyst Meta Acceleration framework used in accelerating variance reduced methods in the settings of Data Science and Machine Learning. Furthermore, we present literatures and progress in topics such as Performance Estimation Problem, Inexact Proximal Point, acceleration without convexity. At the end there is a section summarizing a method we developed for tree species classifications using Sentinel-2 satellite remote sensing data using big data analytics by extract spectral signatures of ground vegetation covers.

*Department of Mathematics, I.K. Barber Faculty of Science, The University of British Columbia, Kelowna, BC Canada V1V 1V7. E-mail: alto@mail.ubc.ca.

Contents

1 Introduction

Let $\mathbb{R}^n$ be the ambient space. We consider

$$\min_{x\in\mathbb{R}^n} \left\{F(x) : f(x) + g(x)\right\}. \tag{1.1}$$

Unless specified, assume $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is L-Lipscthiz smooth $\mu \geq 0$ strongly convex and $g : Q \rightarrow \overline{\mathbb{R}}$ is convex. This type of problem is referred to as additive composite problems in the literature.

Our ongoing research concerns accelerated proximal gradient type method for solving (1). In the expository writing by Walkington [1], a variant for of accelerated gradient method for strongly convex function f is discussed. We had two lingering questions after reading it.

(i) Do there exist a unified description for the convergence for both variants of the algorithms?

(ii) Is it possible to attain faster convergence rate without knowledge about the strong convexity of function f?

(iii) Is it possible to describe the convergence of function value for momentum sequences that are much weaker than the Nesterov's rule?

The good news is we have definitive answers for both questions by our own efforts of research. Section 3, 4 are our ongoing research which present the answers to the questions.

In Section 3, we proposed the method of "Relaxed Weak Accelerated Proximal Gradient (R-WAPG)" as the foundation to describe several variants of Accelerated proximal gradient method in the literatures. The convergence theories of R-WAPG allows us to model convergence of accelerated proximal gradient method where the momentum sequence doesn't strictly follow the conditions presented in the literatures. The descriptive power of R-WAPG allows convergence analysis for all the variants using one single theorem.

In Section 4 we propose a practical algorithm that exploits a specific term in the proof of R-WAPG to achieve faster convergence for solving (1) without knowing parameter L, μ in prior. Results of numerical experiments are presented.

Section 5 are results of literatures review in MATH 590. It's based on a series of papers in the topic of Catalyst Meta Acceleration method for First Order Variance Reduced Methods. We will point out potential future direction of research of Catalyst acceleration.

Add citations here.

Section 6, 7, 8 preview literatures in nonsmooth optimization frontier research where progress and impacts can be made.

2 Preliminaries

Assumptions: Additive composite type convex objective until specified.

2.1 Fundamentals in convex analysis

Definitions:

- ☐ Strong convexity of a function.
- ☐ The proximal gradient operator.
- ☐ The proximal mapping operator.
- ☐ The Bregman Divergence of function.

Lemmas:

- ☐ Quadratic growth conditions of a strongly convex function.
- ☐ The proximal gradient envelope.
- ☐ A property of gradient mapping.

Theorems:

- ☐ The proximal gradient inequality, for additive composite objective where the smooth part is $\mu \geq 0$ strongly convex.

2.2 Fundamentals in non-convex analysis

3 Unifying variants of Nesterov's accelerated methods

This section is really about stating the results of the draft paper and no explaining will be done here. Along with the content of the draft paper, we will also explain the origin and inspirations of the ideas.

Definitions:

(i) Method of Nesterov's estimating sequence.
(ii) R-WAPG Sequence.
(iii) R-WAPG algorithm .
(iv) R-WAPG Intermediate form.
(v) R-WAPG Similar triangle form.
(vi) R-WAPG Momentum form.

Theorems:

(i) Convergence of the R-WAPG algorithm.

(ii) R-WAPG First equivalent form.

(iii) R-WAPG Second equivalent form.

(iv) R-WAPG Third equivalent form.

(v) Convergence with constant momentum.

(vi) Convergence with Chambolle, Dossal Sequences.

Lemmas

(i) Inverted FISTA sequence is a R-WAPG sequence.

(ii) Constant R-WAPG sequence.

4 Method Free R-WAPG

Algorithm, and results of numerical experiments with their descriptions.

5 Catalyst accelerations and future works

Literatures review of the topics in Catalyst acceleration method. Here is a list of topics:

(i) The original accelerated PPM.

(ii) The Catalyst with weakly convex objectives.

After the literature reviews of the core literatures, move on and state new research directions and open problems.

A list of relevant literatures:

(i)

6 Performance estimation problems

7 Methods of inexact proximal point

8 Nestrov's acceleration in the non-convex case

9 Using PostGreSQL and big data analytic method for species classification on Sentinel-2 Satellite remote sensing imagery

References

[1] W. Noel, *Nesterov's method for convex optimization*, SIAM Review, 65, pp. 539–562.