First Order Nonsmooth Optimization: Catalyst Acceleration and Unifying Nesterov's Acceleration

Hongda Li

University of British Columbia Okanagan

January 24, 2025

Overview

This talk will be based on the content of our draft paper and selected content of the Catalyst Meta Acceleration Framework. Our preprint:

1. X. Wang and H. Li, *A Parameter Free Accelerated Proximal Gradient Method Without Restarting*, preprint, (2025).

Catalyst Meta Acceleration:

1. H. Lin, J. Mairal and Z. Harchaoui, *A universal catalyst for first-order optimization*, in NISP, vol. 28, (2015).
2. ________, *Catalyst acceleration for first-order convex optimization: from theory to practice*, JMLR, 18 (2018), pp. 1–54.

ToC

1. Introduction
   - Notations and preliminaries
2. Content of the draft paper
   - Direction of future works
3. Selected contents from Catalyst Meta Accelerations
   - Direction of future works
4. References

Notations and preliminaries

Throughout this talk, let $\mathbb{R}^n$ be the ambient space equiped with Euclidean inner product and norm. We consider

$$\min_{x \in \mathbb{R}^n} \left\{F(x) := f(x) + g(x)\right\}. \tag{1}$$

Unless specified, assume:

1. $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is L-Lipscthiz smooth $\mu \geq 0$ strongly convex,
2. $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is closed convex proper.

Notations and preliminaries

Definition (Proximal gradient operator)

Define the proximal gradient operator T_L on all $y \in \mathbb{R}^n$:

$$T_L y := \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ g(x) + f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x - y\|^2 \right\}.$$

Definition (Gradient mapping operator)

Define the gradient mapping operator $\mathcal{G}_L$ on all $y \in \mathbb{R}^n$:

$$\mathcal{G}_L(y) := L(y - T_L y).$$

Proximal gradient inequality

Lemma (The proximal gradient inequality)

For all $y \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, it has:

$$(\forall x \in \mathbb{R}^n) \quad F(x) - F(T_L y) - \langle L(y - T_L y), x - y \rangle - \frac{\mu}{2}\|x - y\|^2 - \frac{L}{2}\|y - T_L y\|^2 \geq 0.$$

This lemma is crucial to developing results in our current draft paper.

Nesterov's estimating sequence example

Definition (Nesterov's estimating sequence)

For all $k \geq 0$, let $\phi_k : \mathbb{R}^n \to \mathbb{R}$ be a sequence of functions. We call this sequence of functions a Nesterov's estimating sequence when it satisfies conditions:

1. There exists another sequence $(x_k)_{k \geq 0}$ such that for all $k \geq 0$ it has $F(x_k) \leq \phi_k^* := \min_x \phi_k(x)$.
2. There exists a sequence of $(\alpha_k)_{k \geq 0}$ where $\alpha_k \in (0, 1)\ \forall k \geq 0$ such that for all $x \in \mathbb{R}^n$ it has $\phi_{k+1}(x) - \phi_k(x) \leq -\alpha_k(\phi_k(x) - F(x))$.

The technique is widespread in the literatures and it's used to derive the convergence rate of acceleration on first order method, and the numerical algorithm itself. It is a two birds one stone technique.

Our works on R-WAPG

Here are contributions of our draft paper. Recall the Nesterov's acceleration has momentum extrapolation updates on $y_{k+1} = x_{k+1} + \theta_{k+1}(x_{k+1} - x_k)$. We proposed the idea of R-WAPG, a generic method that:

1. Describe for momentum sequences that doesn't follow Nesterov's rules.
2. Unifies the convergence rate analysis for several Euclidean variants of the FISTA method.
3. A parameter free numerical algorithm: "Free R-WAPG" method that has competitive numerical performance in practical settings without restarting.

Our work is inspired by considering Nesterov's estimating sequence where $F(x_k) + R_k = \phi_k^*$.

Introduction to Catalyst Acceleration

Citation examples

Citation examples [1]

References I

A. Chambolle and C. Dossal, "On the convergence of the iterates of the "Fast iterative shrinkage/thresholding algorithm"," *Journal of Optimization Theory and Applications*, vol. 166, no. 3, pp. 968–982, Sep. 2015. [Online]. Available: https://doi.org/10.1007/s10957-015-0746-4