Error bounds, PL condition, and quadratic growth for weakly convex functions, and linear convergences of proximal point methods

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

WISCONSIN INSTITUTE FOR DISCOVERY
AT THE UNIVERSITY OF WISCONSIN-MADISON

Feng-Yi Liao[1], Lijun Ding[2], Yang Zheng[1]

[1]Department of Electrical and Computer Engineering, University of California, San Diego
[2]Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison

Overview

- (sub)gradient-based methods and their variants are the workhorse algorithms for machine learning applications.
- Well understood for these methods:

  Smoothness + Strong convexity ⇒ Linear convergence

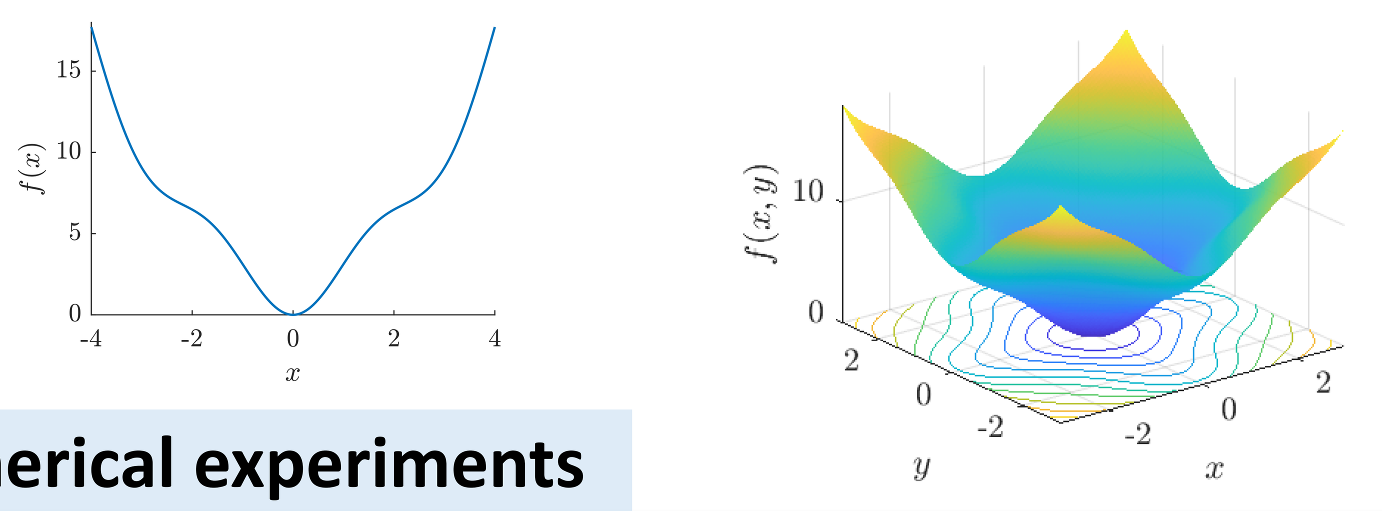
- However, **smoothness** and **strong convexity** are often not satisfied in practice
- Motivate the study of other weaker regularity conditions that also guarantee linear convergence, including RSI, EB, PL, and QG.
- In the class of **weakly convex** functions, common regularity conditions possess certain relationship/equivalence.
- Using the established equivalence in **convex optimization**, we provide
  - **Linear convergence** of the proximal point method (PPM).
  - **Simple** and **clean** proof.
- The linear convergence result extends to **weakly convex** functions.
- Proper control of the inexactness ⇒ Linear convergence of **inexact** PPM.

Preliminary

- Consider a ρ-weakly convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ ($f + \frac{\rho}{2}\|\cdot\|^2$ is convex).
- Define the Fréchet subdifferential

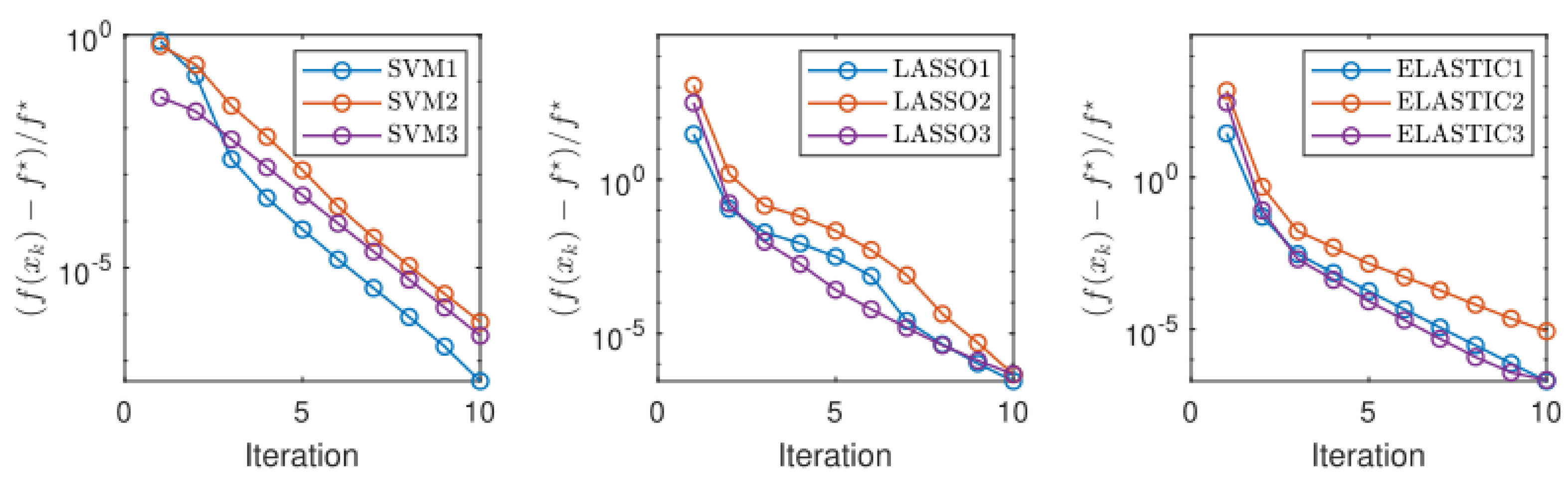
$$\hat{\partial} f(x) = \left\{ s \in \mathbb{R}^n \mid \liminf_{y \to x} \frac{f(y) - f(x) - \langle s, y - x\rangle}{\|y - x\|} \geq 0 \right\}.$$

- Let $S = \operatorname{argmin}_x f(x)$ and $\nu > 0$ and consider the following regularity conditions
  - Local strongly convex **(SC)**: $\exists \mu_{\mathrm{s}} > 0$,
  $$f(x) + \langle g, y - x\rangle + \frac{\mu_{\mathrm{s}}}{2}\|y - x\|^2 \leq f(y),\ \forall x, y \in [f \leq f^\star + \nu], g \in \hat{\partial} f(x).$$
  - Restricted secant inequality **(RSI)**: $\exists \mu_{\mathrm{r}} > 0$,
  $$\mu_{\mathrm{r}} \cdot \operatorname{dist}^2(x, S) \leq \langle g, x - \hat{x}\rangle,\ \forall x \in [f \leq f^\star + \nu], g \in \hat{\partial} f(x), \hat{x} \in \Pi_S(x).$$
  - Polyak-Lojasiewicz **(PL)**: $\exists \mu_{\mathrm{p}} > 0$,
  $$2\mu_{\mathrm{p}} \cdot (f(x) - f^\star) \leq \operatorname{dist}^2(0, \hat{\partial} f(x)),\ \forall x \in [f \leq f^\star + \nu].$$
  - Error bound **(EB)**: $\exists \mu_{\mathrm{e}} > 0$,
  $$\operatorname{dist}(x, S) \leq \mu_{\mathrm{e}} \cdot \operatorname{dist}(0, \hat{\partial} f(x)),\ \forall x \in [f \leq f^\star + \nu].$$
  - Quadratic Growth **(QG)**: $\exists \mu_{\mathrm{q}} > 0$,
  $$\frac{\mu_{\mathrm{q}}}{2} \cdot \operatorname{dist}^2(x, S) \leq f(x) - f^\star,\ \forall x \in [f \leq f^\star + \nu].$$
- **SC**, **RSI**, **EB**, and **PL** imply that all stationary points are global optimal.
- The following two **nonconvex** functions satisfy **RSI**, **EB**, and **PL**



Numerical experiments

- Three machine learning applications



Reference

❒ Feng-Yi Liao, Lijun Ding, and Yang Zheng. Error bounds, PL condition, and quadratic growth for weakly convex functions, and linear convergences of proximal point methods. arXiv preprint arXiv:2307.07651, 2023.

Scan me

Main Result 1: Relationship/Equivalence

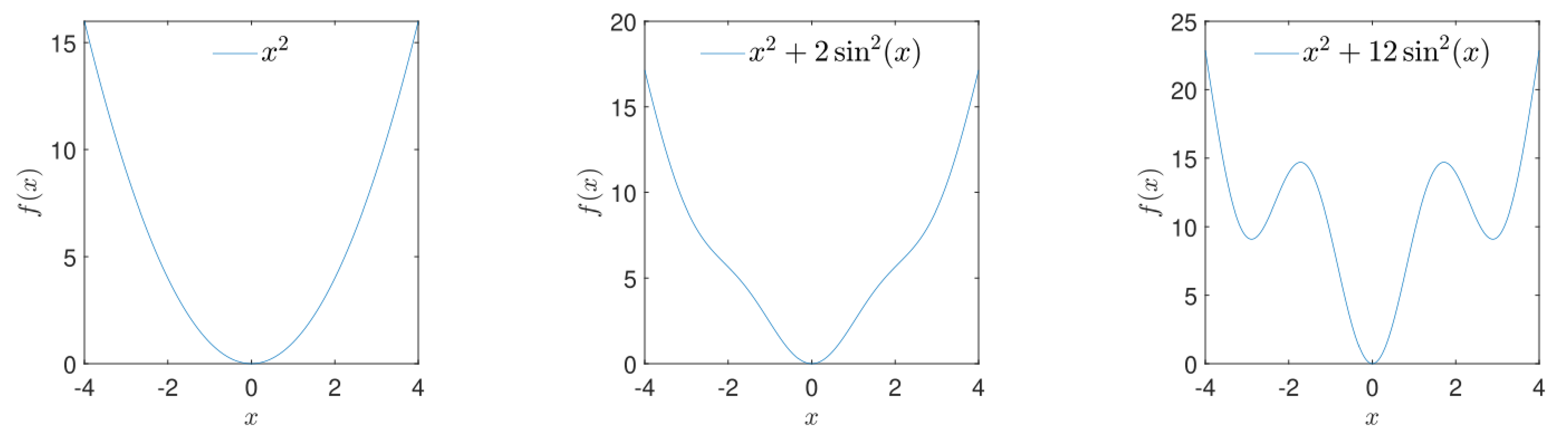
- Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper closed ρ-weakly convex function. Then

$$\textbf{SC} \to \textbf{RSI} \to \textbf{EB} \equiv \textbf{PL} \to \textbf{QG}.$$

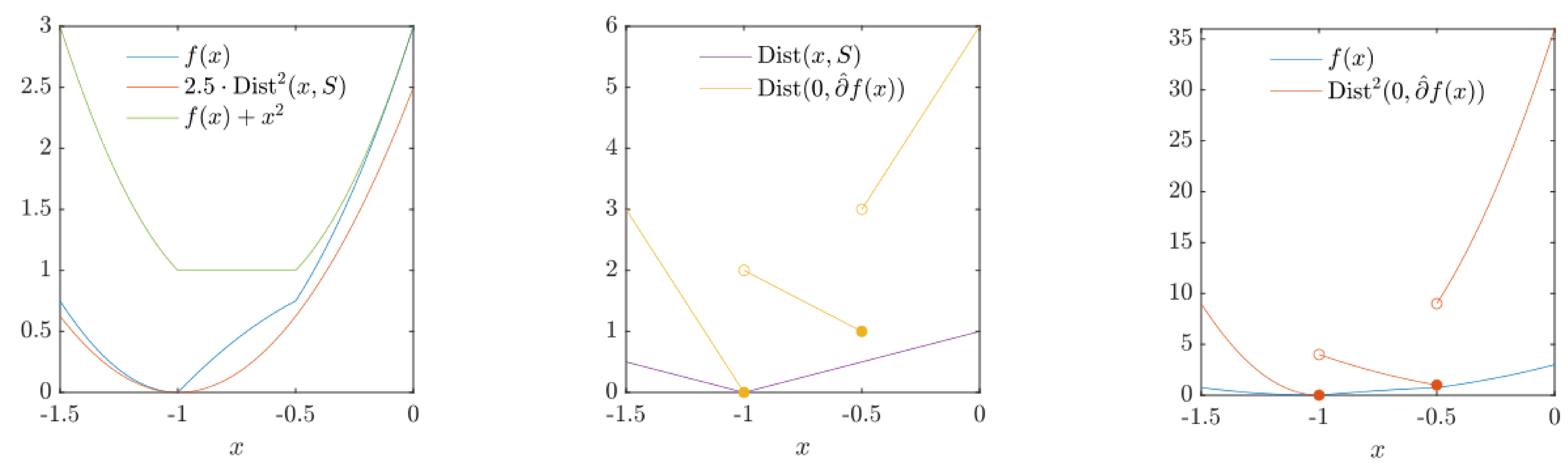
- Furthermore, if the coefficient of **QG** satisfies $\mu_{\mathrm{q}} > \rho$ (including the function f is convex), then the following equivalence holds

$$\textbf{RSI} \equiv \textbf{EB} \equiv \textbf{PL} \equiv \textbf{QG}.$$

- **QG** is the most genral condition.
- **Example 1:** $f(x) = x^2$. f is convex and all properties hold!
- **Example 2:** $f(x) = x^2 + 2\sin^2(x)$. All properties hold but f is not convex!
- **Example 3:** $f(x) = x^2 + 12\sin^2(x)$. All properties fail except **QG**.



- **Example 4:** A **nonsmooth** and **nonconvex** function satisfying $\mu_{\mathrm{q}} > \rho$.



Main result 2: Linear convergence of PPM

- Consider the optimization problem

$$\min_{x \in \mathbb{R}^n} \ f(x)$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper closed **convex** function.

- PPM follows the update formula

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \ f(x) + \frac{1}{2c_k}\|x - x_k\|^2,\ k = 0, 1, 2, \ldots. \tag{1}$$

- Suppose f satisfies **PL** (or **EB**, **RSI**, **QG**) globally. Then, the iterates of PPM with a positive sequence $\{c_k\}_{k\geq 0}$ satisfies

$$f(x_{k+1}) - f^\star \leq \omega_k \cdot (f(x_k) - f^\star), \quad \omega_k < 1,$$
$$\operatorname{dist}(x_{k+1}, S) \leq \theta_k \cdot \operatorname{dist}(x_k, S), \quad \theta_k < 1.$$

- Suppose f is **ρ-weakly convex** and satisfy **QG** with $\mu_{\mathrm{q}} > \rho$ globally. Then, the iterates of PPM with a positive sequence $\{c_k\}_{k\geq 0}$ satisfying $\frac{1}{c_k} > \rho$ enjoy the linear convergence as the convex case.
- Simple proofs:

$$f(x_k) - f(x_{k+1}) \geq \|x_{k+1} - x_k\|^2/(2c_k) \overset{\textbf{O.C.}}{\geq} \frac{c_k}{2}\operatorname{dist}^2(0, \partial f(x_{k+1}))$$
$$\overset{\textbf{PL}}{\geq} c_k\mu_{\mathrm{p}}(f(x_{k+1}) - f^\star)$$
$$\Longrightarrow \frac{1}{1 + c_k\mu_{\mathrm{p}}}(f(x_k) - f^\star) \geq f(x_{k+1}) - f^\star.$$

- If (1) is solved **inexactly**, i.e.,

$$x_{k+1} \approx \operatorname*{argmin}_{x \in \mathbb{R}^n} \ f(x) + \frac{1}{2c_k}\|x - x_k\|^2,\ k = 0, 1, 2, \ldots,$$

with a proper control of inexactness and if f satisfy **QG**, then there exist a nonnegative $\theta < 1$ and a large $\bar{k} > 0$ such that

$$\forall k \geq \bar{k},\ \operatorname{dist}(x_{k+1}, S) \leq \theta_k \operatorname{dist}(x_k, S),\ \theta_k \leq \theta.$$

Acknowledgement

This project is supported by NSF ECCS-2154650 and NSF CMMI-2320697