My Ideas after Reading Necoara's Papers

Author 1 Name, Author 2 Name *

March 2, 2020

This paper is currently in draft mode. Check source to change options.

Abstract

This is still a note for a draft so no abstract [1]

1 Introduction

Necoara et al. [2] introduced the definition of quasi strongly convex function (Q-SCNVX), Quadratic Under approximations (QUA), Quadratic Gradient Growth (QGG), Proximal Error Bound (PEB) and, Quadratic Function Growth (QFG). These conditions are relaxation of strong convexity which enables linear convergence rate of first order method, including Nesterov's accelerated variants. In this file, we showed a new perspective of their works. Our goal is to relax their definitions and, to extend the linear convergence results, using completely new ideas and perspective.

Notations. Unless specified, our ambient space is $\mathbb{R}^n$ with Euclidean norm $\|\cdot\|$. Let $C \subseteq \mathbb{R}^n$, $\Pi_C(\cdot)$ denotes the projection onto the set C, i.e: the closest point in C to another point in $\mathbb{R}^n$.

The following definitions and assumptions are their.

---

*Subject type, Some Department of Some University, Location of the University, Country. E-mail: `author.name@university.edu`.

Assumption 1.1 (Necoara's linear convergence assumptions)
The following assumptions are about (f, X, X^*, L_f).

(i) $f : \mathbb{R}^n \to \mathbb{R}$ is an L_f Lipschitz smooth function.
(ii) $X \subseteq \mathbb{R}^n$ is a closed convex non-empty set.
(iii) $X^* = \underset{x \in X}{\operatorname{argmin}} f(x) \neq \emptyset$.

Under this assumption, the following definitions are proposed.

Definition 1.2 (Necoara's weaker characterizations of strong convexity)
Suppose that (f, X, X^*, L_f) are given by Assumption 1.1. For all $x \in X$, denote $\bar{x} = \Pi_{X^*} x$. The following definitions are relaxations of strong convexity.

(i) *f is Q-SCNVX if there exists $\kappa_f > 0$ such that $f(\bar{x}) - f(x) - \langle \nabla f(x), \bar{x} - x \rangle \geq \frac{\kappa_f}{2} \|x - \bar{x}\|^2$. Which we denote it by $f \in q\mathcal{S}(f, L_f, \kappa_f)$.*
(ii) *f is QUA if there exists $\kappa_f > 0$ such that $f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq \frac{\kappa_f}{2} \|x - \bar{x}\|^2$. We denote it by $f \in \mathcal{U}(f, L_f, \kappa_f)$.*
(iii) *f is QGG if there exists $\kappa_f > 0$ such that $\langle \nabla f(x) - \nabla f(\bar{x}), x - \bar{x} \rangle \geq \frac{\kappa}{2} \|x - \bar{x}\|^2$. We denote it by $f \in \mathcal{G}(f, L_f, \kappa_f)$.*
(iv) *f is PEB if there exists $\kappa_f > 0$ such that $\|x - L^{-1} \Pi_X (x - L^{-1} \nabla f(x))\| \geq \kappa_f \|x - \bar{x}\|$. We denote it by $f \in \mathcal{E}(f, L_f, \kappa_f)$.*
(v) *f is QFG if there exists $\kappa_f > 0$ such that $f(x) - f(\bar{x}) \geq \frac{\kappa_f}{2} \|x - \bar{x}\|^2$. We denote it by $f \in \mathcal{F}(f, L_f, \kappa_f)$.*

These definitions are the keys which Necoara used to prove the linear convergence of projected gradient, and Nesterov's accelerated gradient method.

2 Our crazy original ideas of extending their definitions

In this section, we will propose our ideas which relaxed Necoara's conditions for strong convexity.

Definition 2.1 (Breman Divergence for differentiable funtion) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$ function. The Bregman divergence is a $\mathbb{R}^n \times \operatorname{dom} \nabla f(x) \to \mathbb{R}$ mapping and, it's defined by:*

$$D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle.$$

Theorem 2.2 (distance squared Bregman Divergence) *Let $C \subseteq \mathbb{R}^n$ be any closed, non-empty and convex set. Let $\varphi = (1/2) \operatorname{dist}(\cdot|C)^2$. Then for all $x, y \in \mathbb{R}^n$, for all $\bar{x} \in \Pi_C x, \bar{y} \in \Pi_C y$, it satisfies:*

$$D_\varphi(x, y) = -\frac{1}{2}\|\bar{x} - \bar{y}\|^2 + \langle x - \bar{x}, \bar{x} - \bar{y}\rangle + \frac{1}{2}\|x - y\|^2.$$

Proof. closed, convex, non-empty C gives unique projection point so Π_C is a convex set.

For notational simplicity let $\bar{x} = \Pi_C x, \bar{y} = \Pi_C y$, then it has $\nabla\varphi(x) = x - \bar{x}$. With Definition 2.1 and, basic algebra we can show that:

$$\begin{aligned}
D_\varphi(x, y) &= \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x - y\rangle \\
&= \varphi(x) - \varphi(y) - \langle y - \bar{y}, x - y\rangle \\
&= \frac{1}{2}(\|x - \bar{x}\|^2 - \|y - \bar{y}\|^2) - \langle y - \bar{y}, x - y\rangle \\
&= \frac{1}{2}(\|x - \bar{x}\|^2 - \|y - \bar{y}\|^2) - \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\left(\|y - \bar{y}\|^2 + \|x - y\|^2\right) \\
&= \frac{1}{2}\|x - \bar{x}\|^2 - \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\|x - y\|^2 \\
&= \frac{1}{2}\|x - \bar{x}\|^2 - \frac{1}{2}\left(\|x - \bar{x}\|^2 + \|\bar{x} - \bar{y}\|^2 + 2\langle x - \bar{x}, \bar{x} - \bar{y}\rangle\right) + \frac{1}{2}\|x - y\|^2 \\
&= -\frac{1}{2}\|\bar{x} - \bar{y}\|^2 + \langle x - \bar{x}, \bar{x} - \bar{y}\rangle + \frac{1}{2}\|x - y\|^2.
\end{aligned}$$

∎

In the parts that come, we will talk about the set of points that share the same projections onto a set. Let $C \subseteq \mathbb{R}^n$, we define for all $x \in \mathbb{R}^n$ the set:

$$\mathcal{D}(x|C) = \{z \in \mathbb{R}^n : \Pi_C z = \Pi_C x\}. \tag{2.1}$$

The following lemma is a precursor to extend the definition of Q-SCNVX. The abbreviation "DSBD" stands for "distance squared Bregman Divergence".

Lemma 2.3 (conditions when DSBD behaves like norm squared) *Let $C \subseteq \mathbb{R}^n$ be closed and non-empty. Let $\varphi = (1/2) \operatorname{dist}(\cdot|C)^2$. Then, for all $x \in \mathbb{R}^n$ it satisfies*

$$(\forall z \in \mathcal{D}(x|C)) \; \mathcal{D}_\varphi(x, z) = \frac{1}{2}\|x - z\|^2 = D_\varphi(z, x).$$

Proof. The proof is direct. For all $x \in \mathbb{R}^n, z \in \mathcal{D}(x|C)$, the Bregman Divergence simplifies because using Theorem 2.2 it has:

$$D_\varphi(x, z) = -\frac{1}{2}\|\Pi_C x - \Pi_C z\|^2 + \langle x - \Pi_C x, \Pi_C x - \Pi_C z \rangle + \frac{1}{2}\|x - z\|^2$$
$$= 0 + 0 + \frac{1}{2}\|x - z\|^2.$$

This is true because for all $z \in \mathcal{D}(x|C)$, it has $\Pi_C x - \Pi_C z = 0$. Similarly, it has:

$$D_\varphi(x, z) = -\frac{1}{2}\|\Pi_C x - \Pi_C z\|^2 + \langle z - \Pi_C z, \Pi_C z - \Pi_C x \rangle + \frac{1}{2}\|z - x\|^2$$
$$= 0 + 0 + \frac{1}{2}\|z - x\|^2.$$

$\blacksquare$

Now, the definition of $\mathcal{D}(x|C)$ in the above lemma is nothing outrageous. It simply denotes all points that share the same projection onto C as x. Below, we will give some possible scenarios.

(i) When C is convex, the set $\mathcal{D}(x|C) = \Pi_C x + N_C(\Pi_C x)$.

(ii) When C is nonconvex, it has $\mathcal{D}(x|C) \subseteq \Pi_C x + N_C(\Pi_C x)$.

Strong convexity is recovered when C is a singleton. For example when $C = \{\mathbf{0}\}$, and $\varphi = (1/2)\|x\|^2$, this is the usual euclidean Bregman Divergence used to introduce functions that are strongly convex. Furthermore, take note that $\Pi_C x \in D(x|C)$ is always true, hence we have

$$D_\varphi(x, \Pi_C x) = \frac{1}{2}\|x - \Pi_C\|^2 = D_\varphi(\Pi_C x, x). \tag{2.2}$$

{def:pg}   The following theorems will prepare us for linear convergence of first order algorithms.

Definition 2.4 (proximal gradient operator) *Let $F = f + g$, suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is in $\mathcal{C}^1$ and $g : \mathbb{R}^n \to \overline{R}$ is l.s.c. The proximal gradient operator, denoted by $T_{B,f,g}$ is defined as*

$$T_{B,f,g}(x) = \operatorname*{argmin}_{z \in \mathbb{R}^n} \left\{ g(z) + f(x) + \langle \nabla f(x), z - x \rangle + \frac{B}{2}\|x - z\|^2 \right\}$$

Remark 2.5 Usually we can find it in the literature that, $T_{B,f,g} = \operatorname{prox}_{B^{-1}g}(x - B^{-1}\nabla f(x))$.

The following assumption is the weakest and, it doesn't assume convexity at all.

4

Assumption 2.6 The following assumption is about (F, f, g, L_f, X^*).

(i) $f : \mathbb{R}^n \to \mathbb{R}$ is L_f Lipschitz smooth.

(ii) $g : \mathbb{R}^n \to \mathbb{R}$ is closed convex and proper.

(iii) Denote X^* be the fixed point set of $T_{L_f, f, g}$ and assume $X^* \neq \emptyset$.

We take note that, under the above assumption the function $F = f + g$ is a L_f weakly convex function and, $\widehat{\partial} F = \partial F$. The limiting and regular subgradients are the same.

Definition 2.7 (relatively restricted SCNVX) *Suppose that (f, g, L_f, X^*) satisfies Assumption 2.6. The function $F = f + g$ has restricted SCNVX condition if there exists $\kappa_f > 0, D \subseteq \mathrm{dom}\, g$ such that for all $x \in D$, let $\bar{x} = \Pi_{X^*} x$ it satisfies:*

$$(\forall v \in \partial F(\bar{x}))\ \ F(x) - F(\bar{x}) - \langle v, x - \bar{x} \rangle \geq \frac{\kappa_f}{2} \|x - \bar{x}\|^2.$$

Remark 2.8 Take note that the RHS of the inequality $\kappa_f / 2 \|x - \bar{x}\|^2 = D_\varphi(x, \Pi_{X^*} x)$ where $\varphi = (1/2) \mathrm{dist}(x|X^*)^2$. This definition doesn't exclude nonconvex functions. The hope is to have the set $D = \mathrm{dom}\, g$, but we don't think that is always true. Our hope is that the set D is at least somewhat larger than X^*.

Assumption 2.9 (two sets feasibility problem)
The following assumption is about (F, f, g, L_f, X^*). In addition, we assume the following.

(i) Let (f, g, L_f, X^*) satisfies Assumption 2.6. Following that, it has $X^* = \mathrm{Fix}(\Pi_X \Pi_C) \neq \emptyset$.

(ii) $f = (1/2) \mathrm{dist}(x|C)^2$ where $C \subseteq \mathbb{R}^n$ is a closed and convex set.

(iii) $g = \delta_X(x)$ is the indicator function of $X \subseteq \mathbb{R}^n$, where X is a closed and, nonempty set.

The following discussion will give important consequences of this setup.

2.1 examples for restricted SCNVX

In the coming parts we look for conditions for which a function $F = f + g$ satisfying the above assumption also satisfies Definition 2.7.

Theorem 2.10 (characterizing restricted SCNVX) *Suppose that (f, g, L_f, X^*) satisfies Assumption 2.9. Then for all $x \in X, \bar{x} \in \Pi_{X^*}x$ such that $\bar{x} \in \mathcal{D}(x|C)$, it has*

$$D_f(x, \bar{x}) = \frac{1}{2}\|x - \bar{x}\|^2 = D_f(\bar{x}, x).$$

Proof. We use Lemma 2.3, take note that when $z = \Pi_{X^*}x \in \mathcal{D}(x|C)$, then the lemma says: $D_f(x, \Pi_{X^*}x) = \frac{1}{2}\|x - \Pi_{X^*}x\|^2 = D_f(\Pi_{X^*}x, x)$. ∎

Remark 2.11 The conditions, $\Pi_{X^*}x \in \mathcal{D}(x|C)$ is not always a given and, it would require some careful thoughts.

2.2 theorems useful for convergence analysis of algorithms

The proximal gradient inequality and, the Jensen inequality are the most important for deriving convergence rate of first order accelerated method.

3 Linear convergence of first order methods

References

[1] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer International Publishing, Cham, 2017.

[2] I. NECOARA, Y. NESTEROV, AND F. GLINEUR, *Linear convergence of first order methods for non-strongly convex optimization*, Mathematical Programming, 175 (2019), pp. 69–107.