My Ideas after Reading 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]

2010 Mathematics Subject Classification: Primary 47H05, 52A41, 90C25; Secondary 15A09, 26A51, 26B25, 26E60, 47H09, 47A63. **Keywords:**

1 Introduction

Necoara et al. 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.

{ass:necoara-linear} **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) $X^* = \operatorname*{argmin}_{x\in X} f(x) \neq \emptyset$.

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 proposed our ideas which relaxed Necoara's conditions for strong convexity.

{def:bd} **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:*

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

{thm:bd-dist-sq} semi Bregman Divergence is the Bregman Divergence of distance squared to a set.

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

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

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

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

■

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

Lemma 2.3 (conditions when SBD behaves like SCNVX) *Let $C \subseteq \mathbb{R}^n$ be closed and non-empty. Let $\varphi = (1/2)\operatorname{dist}(\cdot|C)^2$. Define the set $\mathcal{D}(x|C) := \{z \in \mathbb{R}^n : \Pi_C z = \Pi_C x\}$. Then, for all $x \in \mathbb{R}^n$ it satisfies*

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

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:

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

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

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

■

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 $\Pi_C x \in D(x|C)$ always so we have

$$D_\varphi(x, \Pi_C x) = \frac{1}{2}\|x - \Pi_C\|^2 = D_\varphi(\Pi_C x, x). \tag{2.1}$$

{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*

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

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

Assumption 2.6 (our assumptions) The following assumption is about (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 l.s.c

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

Definition 2.7 (relatively quasi strongly convex) *Suppose that (f, g, L_f, X^*) satisfies Assumption 2.6. The function $F = f + g$ is relatively quasi strongly convex if there exists $\kappa_f > 0$ such that for all $x \in \operatorname{dom} g$, 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 $\varphi = (1/2)\,\mathrm{dist}(x|X^*)^2$.

The following discussion will give important consequences of our set up.

2.1 examples for relatively quasi strongly convex functions

2.2 theorems useful for convergence analysis of algorithms

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.