Reading Notes

Alto

Last Compiled: March 29, 2025

Abstract

Reports on papers read. This is a LaTEX file for my own notes taking. It may accelerate the process of writing my thesis for my PhD degree.

This is just pseudo text. This is just pseudo text. This is just pseudo text. This is just pseudo text [1].

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

Chapter 1

The Basics of Optimization Theories

{def:bregman-div} Notations in this chapter are not shared, and they are for this chapter only.

Definition 1.0.1 (Bregman Divergence) *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a differentiable function. Define Bregman Divergence:*

{ass:smooth-add-nonsmooth}

$$D_f : \mathbb{R}^n \times \operatorname{dom} \nabla f \to \overline{\mathbb{R}} := (x, y) \mapsto f(x) - f(y) - \langle \nabla f(y), x - y \rangle.$$

Assumption 1.0.2 (smooth plus nonsmooth) Let $F = f + g$ where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is differentiable and there exists $q \in \mathbb{R}$ such that $g - q/2\| \cdot \|^2$ is convex.

Definition 1.0.3 (proximal gradient operator) *Suppose $F = f+g$ satisfies Assumption 1.0.2. Define the proximal gradient operator by:*

{thm:pg-ineq-wcnvx-generic}

$$\begin{aligned} T_{\beta^{-1}, f, g}(x) &= \operatorname{prox}_{\beta^{-1} g}\left(x - \beta^{-1}\nabla f(x)\right) \\ &= \underset{z}{\operatorname{argmin}} \left\{ g(z) + f(x) + \langle \nabla f(x), z - x\rangle + \frac{L}{2}\|x - z\|^2 \right\}. \end{aligned}$$

Theorem 1.0.4 (weakly convex generic proximal gradient inequality) *Suppose $F = f + g$ satisfies Assumption 1.0.2 with $\beta > 0$ and $q \in \mathbb{R}$. Then for all $x \in \mathbb{R}^n, z \in \mathbb{R}^n$, define $\bar{x} = T_{\beta^{-1}, f, g}(x)$, it has:*

$$\frac{q}{2}\|z - x^+\|^2 \le F(z) - F(\bar{x}) - \langle \beta(x - \bar{x}), z - \bar{x}\rangle + D_f(x, \bar{x}) - D_f(z, x).$$

Proof. Nonsmooth analysis calculus rules has

$$\begin{aligned} \bar{x} &\in \operatorname{argmin} z \left\{ g(z) + \langle \nabla f(x), z\rangle + \frac{\beta}{2}\|z - x\|^2 \right\} \\ \implies \mathbf{0} &\in \partial g(x^+) + \nabla f(x) + \beta(x^+ - x) \\ \iff \partial g(x^+) &\ni -\nabla f(x) - \beta(x^+ - x). \end{aligned}$$

The subgradient inequality for weak convexity has

$$
\begin{aligned}
\frac{q}{2}\|z-\bar{x}\|^2 &\le g(z) - g(\bar{x}) + \langle \nabla f(x) + \beta(\bar{x} - x), z - \bar{x}\rangle \\
&= g(z) - g(\bar{x}) + \langle \nabla f(x), z - \bar{x}\rangle + \langle \beta(\bar{x} - x), z - \bar{x}\rangle \\
&= g(z) - g(\bar{x}) + \langle \nabla f(x), z - x\rangle + \langle \nabla f(x), x - \bar{x}\rangle + \langle \beta(\bar{x} - x), z - \bar{x}\rangle \\
&= g(z) - g(\bar{x}) + (-D_f(z, x) + f(z) - f(x)) \\
&\quad + (D_f(\bar{x}, x) - f(\bar{x}) + f(x)) + \langle \beta(\bar{x} - x), z - \bar{x}\rangle \\
&= F(z) - F(\bar{x}) - D_f(z, x) + D_f(\bar{x}, x) - \langle \beta(x - \bar{x}), z - \bar{x}\rangle.
\end{aligned}
$$

■

{theorem:pg-ineq}

Theorem 1.0.5 (convex proximal gradient inequality) *Suppose $F = f + g$ satisfies Assumption 1.0.2 such that $q = \mu_g \ge 0$, $\beta \ge L_f$. In addition, suppose that $f : \mathbb{R}^n \to \mathbb{R}$ has L_f Lipschitz continuous gradient, and it's $\mu_f \ge 0$ strongly convex. For all $x \in \mathbb{R}^n, z \in \mathbb{R}^n$, define $\bar{x} = T_{\beta^{-1}, f, g}(x)$ it has*

$$
0 \le F(z) - F(\bar{x}) + \frac{\beta - \mu_f}{2}\|z - x\|^2 - \frac{\beta + \mu_g}{2}\|z - \bar{x}\|^2.
$$

Proof. The Bregman Divergence of f has inequality

$$
(\forall x \in \mathbb{R}^n, y \in \mathbb{R}^n)\ \frac{\mu_f}{2}\|x - y\|^2 \le D_f(x, y) \le \frac{L_f}{2}\|x - y\|^2.
$$

Specializing Theorem 1.0.4, let $x \in \mathbb{R}^n$ and define $\bar{x} = T_{\beta^{-1}, f, g}(x)$ it has $\forall z \in \mathbb{R}^n$:

$$
\begin{aligned}
\frac{\mu_g}{2}\|z - \bar{x}\|^2 &\le F(z) - F(\bar{x}) - D_f(z, x) + D_f(\bar{x}, x) - \langle \beta(x - \bar{x}), z - \bar{x}\rangle \\
&\le F(z) - F(\bar{x}) - \frac{\mu_f}{2}\|z - x\|^2 + \frac{L_f}{2}\|x - \bar{x}\|^2 - \langle \beta(x - \bar{x}), z - x + x - \bar{x}\rangle \\
&= F(z) - F(\bar{x}) - \frac{\mu_f}{2}\|z - x\|^2 + \left(\frac{L_f}{2} - \beta\right)\|x - \bar{x}\|^2 - \langle \beta(x - \bar{x}), z - x\rangle \\
&\le F(z) - F(\bar{x}) - \frac{\mu_f}{2}\|z - x\|^2 - \frac{\beta}{2}\|x - \bar{x}\|^2 - \langle \beta(x - \bar{x}), z - x\rangle \\
&= F(z) - F(\bar{x}) - \frac{\mu_f}{2}\|z - x\|^2 - \frac{\beta}{2}\left(\|x - \bar{x}\|^2 + 2\langle x - \bar{x}, z - x\rangle\right) \\
&= F(z) - F(\bar{x}) + \frac{\beta - \mu_f}{2}\|z - x\|^2 - \frac{\beta}{2}\|z - \bar{x}\|^2.
\end{aligned}
$$

■

Chapter 2

Linear Convergence of First Order Method

In this chapter, we are specifically interested in characterizing linear convergence of well known first order optimization algorithms. In this section, D_f will denote the Bregman Divergence as defined in Definition 1.0.1.

2.1 Necoara's et al's Paper

2.1.1 The Settings

{ass:necoara-2019-settings} The assumption follows give the same setting as Necoara et al. [1].

Assumption 2.1.1 Consider optimization problem:

$$-\infty < f^+ = \min_{x \in X} f(x). \tag{2.1.1}$$

{problem:necoara-2019} $X \subseteq \mathbb{R}^n$ is a closed convex set. Assume projection onto X, denoted by Π_X is easy. Denote $X^+ = \underset{x \in X}{\operatorname{argmin}} f(x) \neq \emptyset$, assume it's a closed set. Assume f has L_f Lipschitz continuous gradient, i.e: for all $x, y \in X$:

$$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\|.$$

Some immediate consequences of Assumption 2.1.1 now follows. The variational inequality characterizing optimal solution has:

$$x^+ \in X^+ \implies (\forall x \in X) \ \langle \nabla f(x^+), x - x^+ \rangle \geq 0.$$

The converse is true if f is convex. The gradient mapping in this case is:

{def:necoara-scnvx}

$$\mathcal{G}_{L_f}x = L_f(x - \Pi_X x).$$

Definition 2.1.2 (Strong convexity) *Suppose f satisfies Assumption 2.1.1. Then $f \in \mathbb{S}(L_f, \kappa_f, X)$ is strongly convex iff*

$$(\forall x, y \in X)\ \kappa_f\|x - y\|^2 \le D_f(x, y) \le L_f\|x - y\|^2.$$

{def:necoara-weaker-scnvx} Then it's not hard to imagine the following natural relaxation of the above conditions.

Definition 2.1.3 (Relaxations of Strong convexity) *Suppose f satisfies Assumption 2.1.1. Let $L_f \ge \kappa_f \ge 0$ such that for all $x \in X$, $\bar{x} = \Pi_{X^+}x$. We define the following:*

{def:neocara-qscnvx}

{def:necoara-qup}

{def:necoara-qgg}

{def:necoara-qfg}

{def:necoara-peb}

(i) *Quasi-strong convexity (Q-SCNVX): $0 \le D_f(\bar{x}, x) - \frac{\kappa_f}{2}\|x - \bar{x}\|^2$. Denoeted by $\mathbb{S}'(L_f, \kappa_f, X)$.*

(ii) *Quadratic under approximation (QUA): $0 \le D_f(x, \bar{x}) - \frac{\kappa_f}{2}\|x - \bar{x}\|^2$. Denoeted by $\mathbb{U}(L_f, \kappa_f, X)$.*

(iii) *Quadratic Gradient Growth (QGG): $0 \le D_f(x, \bar{x}) + D_f(\bar{x}, x) - \kappa_f/2\|x - \bar{x}\|^2$. Denoted by $\mathbb{G}(L_f, \kappa_f, X)$.*

(iv) *Quadratic Function Growth (QFG): $0 \le f(x) - f^* - \kappa_f/2\|x - \bar{x}\|^2$. Denoted by $\mathbb{F}(L_f, \kappa_f, X)$.*

(v) *Proximal Error Bound (PEB): $\|\mathcal{G}_{L_f}x\| \ge \kappa_f\|x - \bar{x}\|$. Denoted by $\mathbb{E}(L_f, \kappa_f, X)$.*

Remark 2.1.4 The error bound condition in Necoara et al. is sometimes referred to as the "Proximal Error Bound".

2.1.2 Major Results in the paper

In Necoara's et al, major results assume convexity of f.

{thm:qscnvx-means-qua} **Theorem 2.1.5** *(Q-SCNVX implies QUA) Let f satisfies Assumption 2.1.1 and assume f is convex:*

$$\mathbb{S}'(L_f, \kappa_f, X) \subseteq \mathbb{U}(L_f, \kappa_f, X).$$

Proof. We proof by induction. Convexity of f makes X^+ convex and $\Pi_{X^+}X$ unique for all $x \in X$. Make inductive hypothesis that there exists $\kappa^{(k)} \geq 0$ such that

$$(\forall x \in X) \quad f(x) \geq f^+ + \langle \nabla f(\Pi_{X^+}x), x - \Pi_{X^+}x\rangle + \kappa^{(k)}/2\|x - \Pi_{X^+}x\|^2.$$

The base case is true by convexity of f with $\kappa_f^{(0)} = 0$. Choose any $x \in X$ define $\bar{x} = \Pi_{X^+}x$. Consider $x_\tau = \bar{x} + \tau(x - \bar{x})$ for $\tau \in [0, 1]$. Calculus rule has

$$\begin{aligned}
f(x) &= f(\bar{x}) + \int_0^1 \langle \nabla f(x_\tau), x - \bar{x}\rangle d\tau \\
&= f(\bar{x}) + \int_0^1 \tau^{-1}\langle \nabla f(x_\tau), \tau(x - \bar{x})\rangle d\tau \\
&= f(\bar{x}) + \int_0^1 \tau^{-1}\langle \nabla f(x_\tau), x_\tau - \bar{x}\rangle d\tau.
\end{aligned}$$

f is Q-SCNVX so

$$\begin{aligned}
f^+ - f(x_\tau) &\geq \langle \nabla f(x_\tau), \Pi_{X^+}x_\tau - x_\tau\rangle + \kappa_f/2\|x_\tau - \Pi_{X^+}x_\tau\|^2 \\
&= \langle \nabla f(x_\tau), \bar{x} - x_\tau\rangle + \kappa_f/2\|x_\tau - \bar{x}\|^2 \\
\iff \langle \nabla f(x_\tau), x_\tau - \bar{x}\rangle &\geq f(x_\tau) - f^+ + \kappa_f/2\|x_\tau - \bar{x}\|^2.
\end{aligned}$$

We used $\Pi_{X^+}x_\tau = \bar{x}$ by convexity of f. Therefore:

$$\begin{aligned}
f(x) &\geq f(\bar{x}) + \int_0^1 \tau^{-1}\left(f(x_\tau) - f^+ + \frac{\kappa_f}{2}\|x_\tau - \bar{x}\|^2\right) d\tau \\
&= f(\bar{x}) + \int_0^1 \tau^{-1}\left(f(x_\tau) - f^+\right) + \frac{\tau\kappa_f}{2}\|x - \bar{x}\|^2 d\tau \\
&\geq f(\bar{x}) + \int_0^1 \tau^{-1}\left(\langle \nabla f(\Pi_{X^+}x_\tau), x_\tau - \Pi_{X^+}x_\tau\rangle + \frac{\kappa_f^{(k)}}{2}\|x_\tau - \Pi_{X^+}x_\tau\|^2\right) + \frac{\tau\kappa_f}{2}\|x - \Pi_{X^+}x_\tau\|^2 d\tau \\
&= f(\bar{x}) + \int_0^1 \tau^{-1}\left(\langle \nabla f(\bar{x}), x_\tau - \bar{x}\rangle + \frac{\kappa_f^{(k)}}{2}\|x_\tau - \bar{x}\|^2\right) + \frac{\tau\kappa_f}{2}\|x - \bar{x}\|^2 d\tau \\
&= f(\bar{x}) + \int_0^1 \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \frac{\tau\kappa_f^{(k)}}{2}\|x - \bar{x}\|^2 + \frac{\tau\kappa_f}{2}\|x - \bar{x}\|^2 d\tau \\
&= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \frac{\kappa_f^{(k)} + \kappa_f}{4}\|x - \bar{x}\|^2.
\end{aligned}$$

This is the new inductive hypothesis, and it has $\kappa_f^{(k+1)} = (\kappa_f^{(k)} + \kappa_f)/2$. The induction admits recurrence:

$$\kappa_f^{(n)} = (1/2^n)(\kappa_f^{(0)} + (2^n - 1)\kappa_f).$$

Inductive hypothesis is true for $\kappa_f^{(0)} = 0$ and f being convex is sufficient. It has $\lim_{n\to\infty} \kappa_f^{(n)} = \kappa_f$. ∎

Remark 2.1.6 This is Theorem 1 in the paper. Convexity assumption of f makes X^+ convex, so the projection is unique, and it has $\Pi_{X^+}x_\tau = \bar{x}$ for all $\tau \in [0,1]$. In addition, the inductive hypothesis has $\kappa_f^{(n)} \geq 0$, which is not sufficient for convexity, but necessary. The projectin property remains true for nonconvex X^+, however the base case require rethinking.

Theorem 2.1.7 (QGG implies QUA) *Let f satisfies Assumption 2.1.1, under convexity it has*

$$\mathbb{G}(L_f, \kappa_f, X) \subseteq \mathbb{U}(L_f, \kappa_f, X).$$

Proof. For all $x \in X$, define $\bar{x} = \Pi_{X^+}x$, $x_\tau = \bar{x} + \tau(x - \bar{x})\ \forall \tau \in [0,1]$. Observe that $\frac{d}{d\tau}x_\tau = x - \bar{x}$ and $\Pi_{X^+}x_\tau = \bar{x}\ \forall \tau \in [0,1]$. Using calculus and Theorem 2.1.3 (iii):

$$\begin{aligned}
f(x) &= f(\bar{x}) + \int_0^1 \langle \nabla f(x_\tau), x - \bar{x}\rangle d\tau \\
&= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \int_0^1 \langle \nabla f(x_\tau) - \nabla f(\bar{x}), x - \bar{x}\rangle d\tau \\
&= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \int_0^1 \tau^{-1}\langle \nabla f(x_\tau) - \nabla f(\bar{x}), \tau(x - \bar{x})\rangle d\tau \\
&= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \int_0^1 \tau^{-1}\langle \nabla f(x_\tau) - \nabla f(\bar{x}), x_\tau - \bar{x}\rangle d\tau \\
&\geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \int_0^1 \tau^{-1}\kappa_f \|\tau(x - \bar{x})\|^2 d\tau \\
&= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \int_0^1 \tau\kappa_f \|x - \bar{x}\|^2 d\tau \\
&= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + \frac{\kappa}{2}\|x - \bar{x}\|^2.
\end{aligned}$$

■

Remark 2.1.8 This is Theorem 3 in Neocara et al. [1]. There is no immediate use of convexity besides that the projection $\bar{x} = \Pi_{X^+}x$ is a singleton.

Theorem 2.1.9 (Q-SCNVX implies QGG) *Under Assumption 2.1.1 and convexity of f, it has*

$$\mathbb{S}'(L_f, \kappa_f, X) \subseteq \mathbb{G}(L_f, \kappa_f, X).$$

Proof. If $f \in \mathbb{S}'(L_f, \kappa_f, X)$ then Theorem 2.1.5 has $f \in \mathbb{U}(L_f, \kappa_f, X)$. Then, add (ii), (i) in Definition 2.1.3 yield the results. ■

Remark 2.1.10 This is Theorem 2 in the Necoara et al. [1], right after it claims $\mathbb{U}(L_f, \kappa_f, X) \subseteq \mathbb{G}(L_f, \kappa_f/2, X)$ under convexity.

{thm:qfg-suff}

Theorem 2.1.11 (sufficiency of QFG) *Let f satisfies Assumption 2.1.1. For all $0 < \beta < 1$, $x \in X$, let $x^+ = \Pi_X(x - L_f^{-1}\nabla f(x))$. If*

$$\|x^+ - \Pi_{X^+}x^+\| \le \beta\|x - \Pi_{X^+}x\|,$$

then f satisfies the QFG condition with $\kappa_f = L_f(1-\beta)^2$.

Proof. The proof is direct.

$$\begin{aligned}
\|x - \Pi_{X^+}x\| &\le \|x - \Pi_{X^+}x^+\| \qquad &(2.1.2)\\
&\le \|x - x^+\| + \|x^+ - \Pi_{X^+}x^+\| \qquad &(2.1.3)\\
&\le \|x - x^+\| + \beta\|x - \Pi_{X^+}x\| \qquad &(2.1.4)\\
\iff 0 &\le \|x - x^+\| - (1-\beta)\|x - \Pi_{X^+}x\|. \qquad &(2.1.5)
\end{aligned}$$

x^+ has descent lemma hence we have

$$f^+ - f(X) \le f(x^+) - f(x) \le -\frac{L_f}{2}\|x^+ - x\|^2 \le -\frac{L_f}{2}(1-\beta)^2\|x - \Pi_{X^+}\|^2.$$

Hence, it gives the quadratic growth condition. ■

Remark 2.1.12 It's unclear where convexity is used. However, it' still assumed in Necoara et al. paper.

Before we start, we will specialize Theorem 1.0.5 in our context because it will be used in later proofs. In Assumption 2.1.1, it can be seemed as $F = f + g$ in Assumption 1.0.2 with $g = \delta_X$. In this case $\mu_g = 0$ and assuming f is convex we have only $\mu_f = 0$. Therefore the theorem specializes into the following inequality.

The following theorems are about the relation between PEB and QFG.

Theorem 2.1.13 (equivalence between QFG and PEB)

Bibliography

[1] 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.