Reading Notes

Alto

Last Compiled: April 9, 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 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:*

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

{thm:pg-ineq-wcnvx-generic}

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 $X^+ = \operatorname*{argmin}_{x \in X} f(x) \neq \emptyset$, assume it's a closed set. Assume f has L_f Lipschitz continuous gradient, i.e: for all $x, y \in X$:

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

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

{ineq:pg-opt-cond}
$$x^+ \in X^+ \implies (\forall x \in X)\ \langle \nabla f(x^+), x - x^+ \rangle \ge 0. \tag{2.1.2}$$

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*

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

Then it's not hard to imagine the following natural relaxation of the above conditions.

{def:necoara-weaker-scnvx}

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}

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

{def:necoara-qup}

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

{def:necoara-qgg}

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

{def:necoara-qfg}

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

{def:necoara-peb}

(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 Weaker conditions of strong convexity

{thm:qscnvx-means-qua}

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

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 projection property remains true for nonconvex X^+, however the base case require rethinking.

{thm:qgg-implies-qua}

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.

{thm:qscnvx-implies-qgg}

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.

$$\begin{aligned}
\|x - \Pi_{X^+}x\| &\le \|x - \Pi_{X^+}x^+\| && (2.1.3)\\
&\le \|x - x^+\| + \|x^+ - \Pi_{X^+}x^+\| && (2.1.4)\\
&\le \|x - x^+\| + \beta\|x - \Pi_{X^+}x\| && (2.1.5)\\
\iff 0 &\le \|x - x^+\| - (1-\beta)\|x - \Pi_{X^+}x\|. && (2.1.6)
\end{aligned}$$

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 because it will be used in later proofs. In Assumption 2.1.1, it can be seemed as taking $F = f + g$ in Assumption 1.0.2 with $g = \delta_X$. This makes $\mu_g = 0$ and assuming f is convex we have $\mu_f = 0$. Let $\beta = L_f$, and $x^+ = \Pi_X(x - L_f^{-1}\nabla f(x))$, it has for all $z \in X$:

{ineq:proj-grad}

$$\begin{aligned}
0 &\le f(z) - f(x^+) + \frac{L_f}{2}\|z - x\|^2 - \frac{L_f}{2}\|z - x^+\|^2 \\
&= f(z) - f(x^+) + L_f\langle z - x^+, x^+ - x\rangle + \frac{L_f}{2}\|x - x^+\|^2.
\end{aligned} \tag{2.1.7}$$

Take note that when $z = x$ it has

{ineq:proj-grad2}

$$0 \le f(x) - f(x^+) - \frac{L_f}{2}\|x - x^+\|^2. \tag{2.1.8}$$

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

{thm:qfg-peb-equiv}

Theorem 2.1.13 (equivalence between QFG and PEB) *If f is convex and satisfies Assumption 2.1.1. Then we have:*

$$\begin{aligned}\mathbb{E}(L_f, \kappa_f, X) &\subseteq \mathbb{F}(L_f, \kappa^2/L_f, X),\\ \mathbb{F}(L_f, \kappa_f) &\subseteq \mathbb{E}\left(L_f, \frac{\kappa_f}{\kappa_f/L_f + 1 + \sqrt{\kappa_k/L_f + 1}}, X\right).\end{aligned}$$

Proof. **Firstly, we show that PEB implies QFG.** For any $x \in X$, define the gradient projection steps by $x^+ = \Pi_X(x - L_f^{-1}\nabla f(x))$. Denote $\bar{x}_\Pi^+ = \Pi_{X^+}x^+$. Using (2.1.7) with $z = \bar{x}_\Pi^+$ it yields:

$$\begin{aligned}0 &\geq f(x^+) - f(x_\Pi^+) - L_f\langle x_\Pi^+ - x^+, x^+ - x\rangle - \frac{1}{L_f}\|L_f(x - x^+)\|^2\\ &\geq \frac{\kappa_f}{2}\|x^+ - x_\Pi^+\|^2 - \|L_f(x - x^+)\|\|x_\Pi^+ - x^+\| - \frac{1}{2L_f}\|L_f(x - x^+)\|^2 && \text{QFG, Cauchy}\\ &= \frac{\kappa_f}{2}\|x^+ - x_\Pi^+\|^2 - \frac{1}{2L_f}\left(\|L_f(x - x^+)\|^2 + L_f\|L_f(x - x^+)\|\|x_\Pi^+ - x^+\|\right)\\ &= \frac{\kappa_f + L_f}{2}\|x^+ - x_\Pi^+\|^2 - \frac{1}{2L_f}\left(\|L_f(x - x^+)\| + L_f\|x - x_\Pi^+\|\right)^2\end{aligned}$$

From the last line, by positivity of norm it has

$$\begin{aligned}&\|L_f(x - x^+)\| + L_f\|x^+ - x_\Pi^+\| \geq \sqrt{L_f(\kappa_f + L_f)}\|x^+ - x_\Pi^+\|\\ &\iff \|L_f(x - x^+)\| \geq \left(\sqrt{L_f(\kappa_f + L_f)} - L_f\right)\|x^+ - x_\Pi^+\|.\end{aligned}$$

Using property of projection onto X we have

$$\begin{aligned}\|x - \bar{x}\| &\leq \|x - x_\Pi^+\| \leq \|x - x^+\| + \|x^+ - x_\Pi^+\|\\ &= \frac{1}{L_f}\|L_f(x - x^+)\| + \|x^+ - x_\Pi^+\|\\ \iff \|x - \bar{x}\| - \frac{1}{L_f}\|L_f(x - x^+)\| &\leq \|x^+ - x_\Pi^+\|.\end{aligned}$$

Combining the two above inequalities gives

$$
\begin{aligned}
0 &\le \|L_f(x-x^+)\| - \left(\sqrt{L_f(\kappa_f+L_f)} - L_f\right)\left(\|x-\bar{x}\| - \frac{1}{L_f}\|L_f(x-x^+)\|\right) \\
&= -\left(\sqrt{L_f(\kappa_f+L_f)} - L_f\right)\|x-\bar{x}\| + \left(L_f^{-1}\left(\sqrt{L_f(\kappa_f+L_f)} - L_f\right)+1\right)\|L_f(x-x^+)\| \\
&= -\left(\sqrt{L_f(\kappa_f+L_f)} - L_f\right)\|x-\bar{x}\| + \sqrt{L_f(\kappa_f+L_f)}\|L_f(x-x^+)\| \\
\iff & \frac{\sqrt{L_f(\kappa_f+L_f)} - L_f}{\sqrt{L_f(\kappa_f+L_f)}}\|x-\bar{x}\| \le \|\mathcal{G}_{L_f}x\|.
\end{aligned}
$$

Skipping algebra, the fraction simplifies to

$$
\frac{\kappa_f}{\kappa_f/L_f + 1 + \sqrt{\kappa_k/L_f+1}}.
$$

This gives PEB condition. **We now show QFG implies PEB**. From the error bound condition using κ_f it has

$$
\kappa_f^2\|x-\bar{x}\|^2 \le \|\mathcal{G}_{L_f}(x)\|^2 \underset{(2.1.8)}{\le} 2L_f(f(x)-f(x^+)) \le 2L_f(f(x)-f^+).
$$

■

The following theorem summarizes the hierarchy of the conditions listed in Definition 2.1.3.

{thm:weaker-scnvx-hierarchy}

Theorem 2.1.14 (Hierarchy of weaker S-CNVX conditions) *Let f satisfy Assumption 2.1.1, assuming convexity then the following relations are true:*

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

Proof. $\mathbb{S}' \subseteq \mathbb{G}$ is proved in Theorem 2.1.9 and $\mathbb{G} \subseteq \mathbb{U}$ is proved in 2.1.7. $\mathbb{S} \subseteq \mathbb{S}'$ is obvious and it remains to show $\mathbb{U} \subseteq \mathbb{F}$. Let $f \in \mathbb{U}(\kappa_f, L_f, X)$, it has for all $x \in X$:

$$
\begin{aligned}
0 &\le f(x) - f^+ - \langle \nabla f(\bar{x}), x-\bar{x}\rangle - \frac{\kappa_f}{2}\|x-\bar{x}\|^2 \\
&\underset{(2.1.2)}{\le} f(x) - f^+ - \frac{\kappa_f}{2}\|x-\bar{x}\|^2.
\end{aligned}
$$

■

Remark 2.1.15 It's Theorem 4 in Necoara et al. [1].

2.1.3 Hoffman error bound and Q-SCNVX

2.1.4 Feasible descent and accelerated feasible descent

This section summarizes results from Necoara et al. on the method of feasible descent, fast feasible descent, and fast feasible descent with restart.

2.1.5 Application, KKT of linear programming

This section extends and ideas in the discussion section of Necoara et al. [1].

Let X_1, X_2, Y be Hilbert spaces. Define linear mapping $E : X_1 \times X_2 \to Y := (x_1, x_2) \mapsto E_1 x_1 + E_2 x_2$ where E_1, E_2 each are mappings of $X_1 \to Y, X_2 \to Y$. Denote the adjoint of linear mapping by $(\cdot)^*$. Let $c = (c_1, c_2) \in X_1 \times X_2$, $b \in Y$. Suppose that $\mathcal{K} \subseteq X_1$ is a simple cone and K^* is its dual cone. We consider the following linear programming problem

{problem:lp-cannon-form}

$$\inf_{x \in X_1 \times X_2} \left\{ \langle -c, x \rangle \mid Ex = b, x \in \mathcal{K} \times X_2 \right\}. \tag{2.1.9}$$

Define linear mapping g, F and indicator function h by the following:

$$\begin{aligned} g : X_1 \times X_2 \to \mathbb{R} &:= x \mapsto \langle -c, x \rangle, \\ F : X_1 \times X_2 \to Y \times X_1 &:= (x_1, x_2) \mapsto (E_1 x_1 + E_2 x_2, x_1), \\ h : Y \times X_1 \to \overline{\mathbb{R}} &:= (y, z) \mapsto \delta_{\{\mathbf{0}\}}(y - b) + \delta_{\mathcal{K}^*}(z). \end{aligned}$$

It's not hard to identify that problem in (2.1.9) has representations

$$\inf_{x \in X_1 \times X_2} \left\{ g(x) + h(Fx) \right\}.$$

The dual problem of the above is given by

$$-\inf_{u \in Y \times X_1} \left\{ h^\star(u) + g^\star(-F^* u) \right\}.$$

Where $h^\star, g^\star$ are the conjugate of h, g and $F^* : Y \times X_1 \to X_1 \times X_2 = (y, z) \mapsto (E_1^* y + z, E_2^* y)$ is the adjoint operator of F. Note that $g^\star(x) = \delta_{\mathbf{0}}(x + c)$ and $h^\star((y, z)) = \langle b, y \rangle + \delta_{\mathcal{K}^*}(z)$. This gives the following dual problem

$$-\inf_{(y,z) \in Y \times \mathcal{K}^*} \left\{ \langle b, y \rangle \mid E_1^* y + z = c_1, E_2^* y = c_2 \right\}.$$

The KKT conditions give the following convex feasibility problem

$$\begin{aligned}
E_1x_1 + E_2x_2 &= b,\\
E_1^*y + z &= c_1,\\
E_2^*y &= c_2,\\
\langle b, y\rangle &= \langle c_1, x_1\rangle + \langle c_2, x_2\rangle,\\
(x_1, x_2) &\in \mathcal{K} \times X_2,\\
(y, z) &\in Y \times \mathcal{K}^*.
\end{aligned}$$

Allow $X_1 = \mathbb{R}^{n_1}, X_2 = \mathbb{R}^{n_2}, Y = \mathbb{R}^m$. Define

$$\begin{aligned}
\mathbf{K} &:= \mathcal{K} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \times \mathcal{K}^*,\\
A &:= \begin{bmatrix} E_1 & E_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E_1^T & I_{n_1} \\ \mathbf{0} & \mathbf{0} & E_2^T & \mathbf{0} \\ c_1^T & c_2^T & -b^T & 0 \end{bmatrix}, v := \begin{bmatrix} x_1 \\ x_2 \\ y \\ z \end{bmatrix} \in \mathbf{K}, d := \begin{bmatrix} b \\ c_1 \\ c_2 \\ 0 \end{bmatrix}.
\end{aligned}$$

The KKT conditions is a convex feasibility problem which can be formulated by best approxmation problem:

{problem:lp-kkt-min}

$$\min_{v\in\mathbf{K}} \frac{1}{2}\|Ax - d\|^2. \tag{2.1.10}$$

It is minimizing a quadratic problem on a simple cone. Solving (2.1.9) can be approached by optimizing (2.1.10). It's necessary to investigate the matrices A, A^T which are essential to solving it numerically. The properties of A^TA will determine the convergence rate of algorithms. The matrix is a block matrix and possibly sparse in practice. Let $v = (x_1, x_2, y, z)$, it admits implicit representation:

$$Av = (E_1x_1 + E_2x_2,\ E_1^Ty + z,\ E_2^Ty,\ c_1^Tx_1 + c_2^Tx_2 - b^Ty).$$

It involves

(i) Two multiplications of E: x_1, x_2 on the right and y on the right,

(ii) inner product using x_1, x_2 and y.

Let $\bar{v} = (\bar{y}, \bar{x}_1, \bar{x}_2, \xi) \in \mathbb{R}^m \times \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}$ then the right mulplication of has:

$$\begin{aligned}
\bar{v}^TA &= (E_1^T\bar{y} + \xi c_1^T,\ E_2^T\bar{y} + \xi c_2^T,\ \bar{x}_1^TE_1^T + \bar{x}_2^TE_2^T - \xi b^T,\ \bar{x}_1^T)\\
&= (E_1^T\bar{y} + \xi c_1,\ E_2^T\bar{y} + \xi c_2,\ E_1\bar{x}_1 + E_2\bar{x}_2 - \xi b,\ \bar{x}_1)^T.
\end{aligned}$$

(i) Two multiplications of E: $\bar{y}$ on the left and for $\bar{x}_1, \bar{x}_2$ on the right,

(ii) one vector addition with $c = (c_1, c_2)$ and b.

Therefore, computing $A^T Av$ has four vector multiplications using E. In practice, a sparse matrix E from the model can speed up computations.

Another key operation would be $A^T Av$. Let $\bar{v} = Av$, then

$$
\begin{aligned}
A^T Av &= \begin{bmatrix} E_1^T(E_1x_1 + E_2x_2) + (c_1^T x_1 + c_2^T x_2 - b^T y)c_1 \\ E_2^T(E_1x_1 + E_2x_2) + (c_1^T x_1 + c_2^T x_2 - b^T y)c_2 \\ E_1(E_1^T y + z) + E_2E_2^T y - (c_1^T x_1 + c_2^T x_2 - b^T y)b \\ E_1^T y + z \end{bmatrix} \\
&= \begin{bmatrix} (E_1^T E_1 + c_1^T)x_1 + (E_1^T E_2 + c_2^T)x_2 - (c_1 b^T)y \\ (E_2^T E_1 + c_1^T)x_1 + (E_2^T E_2 + c_2^T)x_2 - (c_2 b^T)y \\ -(bc_1^T)x_1 - (bc_2^T)x_2 + (E_2E_2^T + E_1E_1^T + bb^T)y + (E_1E_1^T)z \\ E_1^T y + z \end{bmatrix}.
\end{aligned}
$$

In practice, implicitly representing the process of $A^T Av$ is better in computing software. Here we write it out to view, for theoretical interests.

Let $f(v) = (1/2)\|Av - d\|^2$ to be the objective function of optimization problem (2.1.10). Its gradient, objective value are related and if they have:

$$
\begin{aligned}
\nabla f(v) &= A^T Av - A^T d, \\
f(v) &= \frac{1}{2}\langle x, \nabla f(v) - A^T d\rangle + \frac{1}{2}\|d\|^2.
\end{aligned}
$$

The value $\nabla f(v), f(v)$ when evaluated together, require minimal additional computations. This fact is favorable for implementations in practice.

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.