ISyE 8813-MON (Winter 2019)
Topics in Convex Analysis

Prof. R. D. C. Monteiro
Georgia Institute of Technology

LaTeXer: W. Kong
http://wwkong.github.io
Last Revision: February 23, 2020

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors.

Acknowledgments:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.

Abstract

The purpose of these notes is to provide the reader with a secondary reference to the material covered in ISyE 8813-MON.

1 Convex Optimization

1.1 Subgradient Method

Consider the problem

$$\min f(x)$$
$$\text{s.t. } x \in X$$

where $X \neq$ is a closed convex set and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex. Observe that:

(1) f is continuous

(2) $\partial f(x) \neq \varnothing$ and compact for $x \in \mathbb{R}^n$

Proposition 1.1. *(a) For every $x, x' \in X$,*

$$|f(x') - f(x)| \leq L\|x' - x\|$$

where $L = \sup\{\|g\| : g \in \partial f(x), x \in X\}$. Observe that $L > -\infty$ since $\partial f(X) \neq \varnothing$.

(b) If X is bounded then $L < \infty$.

Proof. (b) Let $s \in \partial f(x), x \in X$. Then

$$f(x + d) - f(x) \geq \langle s, d \rangle, \quad \forall d.$$

Take $d = s/\|s\|$ and conclude that

$$f(x + d) - f(x) \geq \|s\|.$$

Since

$$\sup\{f(x + d) - f(x) : \|d\| \leq 1, x \in X\}$$

is finite then (b) follows.

(a) Let $s \in \partial f(x), x \in X$. Then $\forall x' \in \mathbb{R}$ we have

$$f(x') \geq f(x) + \langle s, x' - x \rangle$$
$$f(x) \leq f(x') + \langle s, x - x' \rangle$$

$$f(x) \leq f(x') + \langle s, x - x' \rangle \leq f(x') + \|s\|\|x - x'\|$$

and so

$$f(x) - f(x') \leq \|s\|\|x - x'\|.$$

Similarly,

$$f(x') - f(x) \leq \|s'\|\|x - x'\|.$$

□

Subgradient Method

(1) $x_0 \in X$ is given

(2) For $k = 0, 1, \ldots$

$x_{k+1} = P_X(x_k - \alpha_k s_k)$ where α_k and $s_k \in \partial f(x_k)$

Observation: Consider the iteration

$$x^+ = P_X(x - \alpha s), \quad \alpha > 0.$$

Then $f(x^+) < f(x)$ is not necessarily true. For example,

$$a \in (0,1), \quad x_0 = (0,1), \quad f(x_1, x_2) = |x_1| + a|x_2|.$$

We have

$$f(0,1) = a, \quad \partial f(0,1) = [-1,1] \times \{a\}, \quad s = [\eta; a] \in \partial f(0,1)$$

where $\eta \in [0,1]$. Now

$$\begin{aligned} f(x - \alpha s) &= f(-\alpha\eta, 1 - \alpha a) \\ &= \alpha|\eta| + a|1 - \alpha a| \\ &\geq \alpha|\eta| + a(1 - \alpha a) \\ &= a + \alpha(|\eta| - a^2). \end{aligned}$$

Let $|\eta| > a^2$. Then, the subgradient method on this problem is not a descent method.

Fact. For all $x \in \mathbb{R}^n$, $\exists \bar{s} \in \partial f(x)$ such that $f'(x; -\bar{s}) < 0$. In particular, $\bar{s} = \min\{\|s\| : s \in \partial f(x)\}$.

Lemma 1.1. *P_X is nonexpansive, i.e.*

$$\|P_X(x) - P_X(x')\| \leq \|x - x'\| \quad \forall x, x' \in X.$$

Lemma 1.2. *For any $u \in X$ and $k \geq 0$,*

$$\|x_{k+1} - u\|^2 \leq \|x_k - u\|^2 - 2\lambda_k[f(x_k) - f(u)] + \lambda_k^2\|s_k\|^2$$

Proof. We have

$$\begin{aligned} \|x_{k+1} - u\| &= \|P_X(x_k - \lambda s_k) - P_X(u)\|^2 \\ &\leq \|x_k - \lambda s_k - u\|^2 \\ &= \|x_k - u\|^2 + \lambda_k^2\|s_k\|^2 - 2\lambda_k \langle s_k, x_k - u\rangle \\ &\leq \|x_k - u\|^2 + \lambda_k^2\|s_k\|^2 - 2\lambda_k [f(x_k) - f(u)]. \end{aligned}$$

□

Corollary 1.1. *If $f(u) < f(x_k)$ then*

$$\|x_{k+1} - u\| < \|x_k - u\|$$

for any

$$\lambda_k \in \left(0, \frac{2[f(x_k) - f(u)]}{\|s_k\|^2}\right).$$

Corollary 1.2. *If x_k is not optimal, then*

$$\|x_{k+1} - x^*\| < \|x_k - x^*\|$$

for any

$$\lambda_k \in \left(0, \frac{2[f(x_k) - f_*]}{\|s_k\|^2}\right).$$

Stepsize Rules

1. (*Polyak*) Set $\lambda = [f(x_k) - f_*]/\|s_k\|^2$
2. $\lambda_k = \lambda$ for all k
3. $\lambda_k = C/\|s_k\|^2$ for some constant $C \in \mathbb{R}_{++}$
4. diminishing stepsize where $\sum \lambda_i = \infty$ and $(\sum \lambda_i^2)/(\sum \lambda_i) \to 0$

Let us now analyze the complexity of computing $x_k \in X$ such that $f(x_k) - f_* \le \varepsilon$.

Proposition 1.2. *Suppose* $\exists M > 0$ *such that*

$$\|s\| \le M \quad \forall s \in \partial f(X).$$

The subgradient method with constant stepsize rule where $\lambda = \varepsilon/M^2$ *finds* $K \ge 0$ *such that*

$$\theta_K := \min_{k \le K} [f(x_k) - f_*] \le \varepsilon \quad \textit{and} \quad K \le \left\lfloor \frac{d_0^2 M^2}{\varepsilon^2} \right\rfloor$$

where $d_0 = \min_{x^* \in X_*} \|x_0 - x^*\|$.

Proof. By Lemma 1.2, with $u = x^*$ such that $d_0 = \|x^* - x_0\|$, we have

$$2\lambda_k \varepsilon_k \le \|x_k - x^*\|^2 - \|x_{k+1} - x^*\| + \lambda_k^2 \|s_k\|^2$$

where $\varepsilon_k := f(x_k) - f_*$. Hence, summing the above inequality from $k = 0$ to K,

$$\begin{aligned} 2\sum_{k=0}^{K} \varepsilon_k \lambda_k &\le \|x_0 - x^*\|^2 - \|x_{k+1} - x^*\| + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2 \\ &\le d_0^2 - \|x_{k+1} - x^*\| + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2. \end{aligned}$$

It follows that

$$2\theta_K \left(\sum_{k=0}^{K} \lambda_k \right) \le d_0^2 + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2$$

and so

$$\theta_K \le \frac{d_0^2 + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2}{2\sum_{k=0}^{K} \lambda_k} \le \frac{d_0^2 + \sum_{k=0}^{K} \lambda^2 M^2}{2(K+1)\lambda} = \frac{d_0^2 + (K+1)\lambda\varepsilon}{2(K+1)\lambda} = \frac{d_0^2}{2(K+1)\lambda} + \frac{\varepsilon}{2} = \frac{d_0^2 M}{2(K+1)\varepsilon} + \frac{\varepsilon}{2}.$$

If the bound $d_0^2 M^2/[2(K+1)\varepsilon] \le \varepsilon/2$ holds then $\theta_K \le \varepsilon$. The condition $k \le \lfloor d_0^2 M^2/\varepsilon^2 \rfloor$ clearly implies this bound. ☐

Observation. Under the rule $\lambda_k = \varepsilon_k/\|s_k\|^2$ we have

$$\theta_K \le \frac{d_0^2 + \varepsilon \sum_{k=0}^{K} \lambda_k}{2\sum_{k=0}^{K} \lambda_k} = \frac{d_0^2}{2\sum_{k=0}^{K} \lambda_k} + \frac{\varepsilon}{2}.$$

Also,

$$\sum_{k=0}^{K} \lambda_k = \varepsilon \sum_{k=0}^{K} \frac{1}{\|s_k\|^2} \geq \frac{\varepsilon(K+1)}{M^2}$$

and hence

$$\theta_K \leq \frac{d_0^2 M^2}{2\varepsilon(K+1)} + \frac{\varepsilon}{2}.$$

Observation. Under the Polyak rule,

$$\sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2 = \sum_{k=0}^{K} \lambda_k \varepsilon_k \implies \sum_{k=0}^{K} \lambda_k \varepsilon_k \leq d_0^2 - \|x_{k+1} - x^*\|^2. \tag{1.1}$$

Proposition 1.3. *The subgradient method with Polyak rule satisfies*

(a) $\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$ *and hence* $\{x_k\}$ *is bounded*

(b) $\theta_K \leq \varepsilon$ *for every* $K \geq \lfloor d_0^2 M^2 / \varepsilon^2 \rfloor$

(c) $\{x_k\} \to x^*$ *for some* $x^* \in X_*$

Proof. (a) Exercise.

(b) Assume by contradiction that $\theta_K = \min_{k \leq K} \varepsilon_k > \varepsilon$ and $K \geq \lfloor d_0^2 M^2 / \varepsilon^2 \rfloor$. Using (1.1), we have

$$d_0^2 \geq \sum_{k=0}^{K} \lambda_k \varepsilon_k = \sum_{k=0}^{K} \frac{\varepsilon_k^2}{\|s_k\|^2} \geq \theta_K^2 \sum_{k=0}^{K} \frac{1}{\|s_k\|^2} \geq \frac{\theta_K^2 (K+1)}{M^2}.$$

Now the lower bound on K implies

$$K + 1 \geq \left\lfloor \frac{d_0^2 M^2}{\varepsilon^2} \right\rfloor + 1 \geq \frac{d_0^2 M^2}{\varepsilon^2}$$

and hence

$$\theta_K^2 \leq \frac{d_0^2 M^2}{K+1} \leq \varepsilon^2$$

which produces the contradiction.

(c) $\exists \{x_k\}_{k \in K}$ such that $f(x_k) \to f_*$ with $x_k \to \bar{x}$ on $k \in K$. Continuity gives $f(x_k) \to f(\bar{x})$ for $k \in K$. Since $f(\bar{x}) = f_*$ and so $\bar{x} \in X_*$. □

Proximal Problems

Consider the problem

$$\begin{aligned} f_* = \inf \ & f(x) \\ \text{s.t. } & x \in \mathbb{R}^n \end{aligned}$$

where $f \in \overline{\text{Conv}}(\mathbb{R}^n)$. The proximal point method (PPM) is described as follows.

PPM

Given $x_0 \in \text{dom } f$

For $k = 1, 2, \ldots$

choose $\lambda_k > 0$

set $x_k = \text{argmin}_u \left[f(u) + \|u - x_{k-1}\|^2/(2\lambda_k)\right]$

Obs. The optimality condition for the subproblem is $(x_{k-1} - x_k)/\lambda_k \in \partial f(x_k)$.

IPP (Inexact Proximal Point Method)

Given $x_0 \in \text{dom}\, f$

For $k = 1, 2, \ldots$

choose $\lambda_k > 0$

find $(x_k, \varepsilon_k) \in \mathbb{R}^n \times \mathbb{R}$ such that $v_k = (x_{k-1} - x_k)/\lambda_k \in \partial_{\varepsilon_k} f(x_k)$.

Q. What can I say about x_k?

(1) $f(x_k) - f_* \le \varepsilon \iff 0 \in \partial_\varepsilon f(x_k)$

(2) $v_k \in \partial_{\varepsilon_k} f(x_k)$ s.t. $\|v_k\| \le \rho, \varepsilon_k \le \varepsilon \iff 0 \in \partial_{\varepsilon_k} \left(f - \langle v_k \cdot\rangle\right)(x_k)$

If $\text{dom}\, f$ is bounded with diameter D, then for $u \in \text{dom}\, f$ we have

$$\begin{aligned} f(u) &\ge f(x_k) + \langle v_k, u - x_k\rangle - \varepsilon_k \\ &\ge f(x_k) - (\|v_k\|\|u - x_k\| + \varepsilon_k) \\ &\ge f(x_k) - (\|v_k\|D + \varepsilon_k) \\ &= f(x_k) - \tilde{\varepsilon}_k \end{aligned}$$

where $\tilde{\varepsilon}_k = \varepsilon_k + D\|v_k\|$, and hence $0 \in \partial_{\tilde{\varepsilon}_k} f(x_k)$.

1.2 Composite Subgradient Method

Consider the problem

$$\begin{aligned} f_* = \inf\ & f(x) := \varphi(x) + h(x) \\ \text{s.t. } & x \in \mathbb{R}^n \end{aligned}$$

where $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ convex and $h : \mathbb{R}^n \mapsto \bar{\mathbb{R}} \in \overline{\text{Conv}}(\mathbb{R}^n)$.

Special case: $h = \delta_X$ where $\varnothing \ne X$ closed convex set .

Consider the problem

$$\min_u -a^T u + \lambda h(u) + \frac{1}{2}\|u\|^2.$$

This has the optimality condition

$$-a + \lambda\partial h(u) + u \ni 0 \iff u = (I + \lambda\partial h)^{-1}(0)$$

where the right-hand-side is called the **resolvent** of h. Remark that if $h = \delta_X$ then $(I + \lambda\partial\delta_X)^{-1} = P_X$.

Composite Subgradient Method

Given $x_0 \in \text{dom}\, h$

For $k = 1, 2, \ldots$

choose $s_k \in \partial\varphi(x_k)$

set $x_{k+1} = \text{argmin}_u \left[\langle s_k, u\rangle + h(u) + \|u - x_k\|^2/(2\lambda_{k+1})\right]$.

Obs. The optimality condition of the subproblem is

$$s_k + \partial h(x_{k+1}) + \frac{1}{\lambda_{k+1}}(x_{k+1} - x_k) \ni 0.$$

Now note

$$s_k \in \partial\varphi(x_k) \implies s_k \in \partial_{\varepsilon_{k+1}}\varphi(x_{k+1})$$

where

$$\varepsilon_{k+1} = \varphi(x_{k+1}) - \varphi(x_k) - \langle s_k, x_{k+1} - x_k\rangle .$$

Hence,

$$\begin{aligned} v_{k+1} = \frac{1}{\lambda_{k+1}}(x_k - x_{k+1}) &\in s_k + \partial h(x_{k+1}) \\ &\subseteq \partial_{\varepsilon_{k+1}}\varphi(x_{k+1}) + \partial h(x_{k+1}) \\ &\subseteq \partial_{\varepsilon_{k+1}}(\varphi + h)(x_{k+1}) = \partial_{\varepsilon_{k+1}} f(x_{k+1}). \end{aligned}$$

Proposition 1.4. *For every $k \geq 1$, define*

$$\bar{v}_k = \frac{\sum_{i=1}^k \lambda_i v_i}{\Lambda_k}, \quad \bar{\varepsilon}_k = \frac{\sum_{i=1}^k \lambda_i\left[\varepsilon_i + \langle v_i, x_i - \bar{x}_k\rangle\right]}{2\Lambda_k}$$

where $\Lambda_k = \sum_{i=1}^k \lambda_i$ and $\bar{x}_k \in \mathbb{R}^n$ is a point satisfying

$$f(\bar{x}_k) \leq \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i f(x_i), \quad \bar{x}_k \in \textit{conv}\{x_0, ..., x_k\}$$

where the former will be called (*) and the latter to be called (**). Then, (1)

$$f(\bar{x}_k) - f_* \leq \frac{d_0^2 + \sum_{i=1}^k \tau_i}{2\Lambda_k}, \quad \|\bar{x}_k - x^*\|^2 \leq d_0^2 + \sum_{i=1}^k \tau_i$$

and (2)

$$\bar{x}_k \in \partial_{\bar{\varepsilon}_k} f(\bar{x}_k), \quad \|\bar{v}_k\| \leq \frac{\left(d_0 + \sqrt{d_0^2 + \sum_{i=1}^k \tau_i}\right)}{\Lambda_k}, \quad \bar{\varepsilon}_k \leq \frac{3\left(d_0^2 + \sum_{i=1}^k \tau_i\right)}{\Lambda_k}.$$

where

$$\tau_k = 2\lambda_k\varepsilon_k - \|x_k - x_{k-1}\|^2.$$

Obs. Some choices of $\bar{x}_k$ are:

$$\bar{x}_k = \frac{\sum_{i=1}^k \lambda_i x_i}{\Lambda_k} \quad \text{or} \quad \bar{x}_k = \text{argmin}\{f(x_i) : i = 1, 2, ..., k\}.$$

Obs. Note that in the composite subgradient method we have

$$\bar{v}_k = \frac{\sum_{i=1}^k \lambda_i v_i}{\Lambda_k} = \frac{\sum_{i=1}^k (x_{i-1} - x_i)}{\Lambda_k} = \frac{x_0 - x_k}{\Lambda_k}$$

and also

$$\begin{aligned}
\tau_k &= 2\lambda_k\left[\varphi(x_k) - \varphi(x_{k-1}) - \langle s_{k-1}, x_k - x_{k-1}\rangle\right] - \|x_k - x_{k-1}\|^2 \\
&\leq 2\lambda_k\left[\langle s_k, x_k - x_{k-1}\rangle - \langle s_{k-1}, x_k - x_{k-1}\rangle\right] - \|x_k - x_{k-1}\|^2 \\
&\leq 2\lambda_k\left[\|s_k - s_{k-1}\|\|x_k - x_{k-1}\|\right] - \|x_k - x_{k-1}\|^2 \\
&\leq \max_{t\in\mathbb{R}}\left(2\lambda_k\|s_k - s_{k-1}\|t - t^2\right) \\
&= \lambda_k^2\|s_k - s_{k-1}\|^2 \leq \lambda_k^2(2M)^2 = 4M^2\lambda_k^2
\end{aligned}$$

Corollary 1.3. *For all $k \geq 1$ and $x^* \in X^*$,*

$$\|x_k - x^*\|^2 \leq \|x_0 - x^*\|^2 + \sum_{i=1}^k \tau_i.$$

Hence, if $\bar{x}_k \in$ conv$\{x_1, \ldots, x_k\}$ then

$$\|\overline{x}_k - x^*\|^2 \leq \|x_0 - x^*\| + \sum_{i=1}^k \tau_i.$$

Lemma 1.3. *Define $a_k(u) = \varepsilon_k + \langle v_k, x_k - u\rangle$ for all $u \in \mathbb{R}^n$. For every $k \geq 1$:*

(a) $a_k(u) \geq f(x_k) - f(x)$

(b) $a_k(u) = \frac{1}{2\lambda_k}\left(\tau_k + \|u - x_{k-1}\|^2 - \|u - x_k\|^2\right)$

Proof. (a) Follows from the fact that $v_k \in \partial_{\varepsilon_k} f(x_k)$

(b) Have

$$\begin{aligned}
2\lambda_k a_k(u) &= 2\lambda_k\varepsilon_k + 2\lambda_k\langle v_k, x_k - u\rangle \\
&= 2\lambda_k\varepsilon_k + 2\langle x_{k-1} - x_k, x_k - u\rangle \\
&= 2\lambda_k\varepsilon_k + \|u - x_{k-1}\|^2 - \|u - x_k\|^2 - \|x_k - x_{k-1}\|^2 \\
&= \tau_k + \|u - x_{k-1}\|^2 - \|u - x_k\|^2.
\end{aligned}$$

□

Lemma 1.4. *For every $k \geq 1$:*

(a) $\sum_{i=1}^k \lambda_i a_i(u) \geq \sum_{i=1}^k \lambda_i\left[f(x_i) - f(u)\right]$

(b) $\sum_{i=1}^k \lambda_i a_i(u) = \frac{\|u-x_0\|^2 - \|u-x_k\|^2 + \sum_{i=1}^k \tau_i}{2} =: \theta_k(u)$

Proof. Follows immediately from the previous Lemma. □

Proof. (of previous proposition) By the above lemma, with $u = x^*$ we have

$$f(\bar{x}_k) - f_* \leq \frac{\sum_{i=1}^k \lambda_i\left[f(x_i) - f_*\right]}{\Lambda_k} \leq \frac{\|x^* - x_0\| - \|x^* - x_k\|^2 + \sum_{i=1}^k \tau_i}{2\Lambda_k}.$$

□

Lemma 1.5. *Assume $\bar{x}_k$ satisfies (*). Then*

$$f(\bar{x}_k) - f(u) \le \langle \bar{v}_k, \bar{x}_k - u \rangle + (\bar{\varepsilon}_k - \delta_k)$$

or equivalently $\bar{v}_k \in \partial_{\bar{\varepsilon}_k} f(\bar{x}_k)$ where

$$\bar{v}_k = \frac{x_0 - x_k}{\Lambda_k}, \quad \bar{\varepsilon}_k = \frac{\|\bar{x}_k - x_0\|^2 - \|\bar{x}_k - x_k\|^2 + \sum_{i=1}^k \tau_i}{2\Lambda_k}.$$

Proof. Let

$$\delta_k = \frac{\sum_{i=1}^k \lambda_i f(x_i)}{\sum_{i=1}^k \lambda_i} - f(\bar{x}_k) \ge 0.$$

By part (a) of the previous lemma,

$$\begin{aligned} \Lambda_k \left[\delta_k + f(\bar{x}_k) - f(u) \right] &\le \sum_{i=1}^k \lambda_i a_i(u) \\ &= \theta_k(\bar{x}_k) + \langle \nabla \theta_k, u - \bar{x}_k \rangle \\ &= \Lambda_k \bar{\varepsilon}_k + \langle x_k - x_0, u - \bar{x}_k \rangle \\ &= \Lambda_k \bar{\varepsilon}_k + \langle \Lambda_k \bar{v}_k, u - \bar{x}_k \rangle \\ &= \Lambda_k \left[\bar{\varepsilon}_k + \langle \bar{v}_k, u - \bar{x}_k \rangle \right] \end{aligned}$$

and the result follows after some algebraic manipulation. □

Proposition 1.5. *For every $k \ge 1$:*

(a) if $\bar{x}_k$ satisfies (*) then $\bar{v}_k \in \partial_{\bar{\varepsilon}_k} f(\bar{x}_k)$

(b) if $\bar{x}_k$ satisfies (*) and (**) then

$$\|\bar{v}_k\| \le \frac{d_0 + \sqrt{d_0^2 + \sum_{i=1}^k \tau_i}}{\Lambda_k}, \quad \bar{\varepsilon}_k \le \frac{4d_0^2 + 3\sum_{i=1}^k \tau_i}{\Lambda_k}.$$

Proof. (a) Follows from the previous lemma

(b) Let $x^* \in X^*$ be such that $d_0 = \|x_0 - x^*\|$. Then,

$$\begin{aligned} \Lambda_k \|\bar{v}_k\| &= \|x_0 - x_k\| \\ &\le \|x_0 - x^*\| + \|x^* - x_k\| \\ &\le \|x_0 - x^*\| + \sqrt{\|x_0 - x^*\| + \sum_{i=1}^k \tau_i} \\ &= d_0 + \sqrt{d_0^2 + \sum_{i=1}^k \tau_i}. \end{aligned}$$

and also

$$\begin{aligned}
2\Lambda_k\bar{\varepsilon}_k &\le \|\bar{x}_k - x_0\|^2 + \sum_{i=1}^k \tau_i \\
&\le \left(\|\bar{x}_k - x^*\|^2 + \|x_0 - x^*\|^2\right) + \sum_{i=1}^k \tau_i \\
&\le \left(\sqrt{d_0^2 + \sum_{i=1}^k \tau_i} + d_0^2\right) + \sum_{i=1}^k \tau_i \\
&\le 2\left(d_0^2 + \sum_{i=1}^k \tau_i + d_0^2\right) + \sum_{i=1}^k \tau_i \\
&= 4d_0^2 + 3\sum_{i=1}^k \tau_i.
\end{aligned}$$

Composite Subgradient Method (Cont.)

Consider the problem $f_* = \min f(x) = \varphi(x) + h(x)$ where $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ convex and $h \in \overline{\text{Conv}}(\mathbb{R}^n)$.

Method: $x_0 \in \operatorname{dom} h$ given □

For $k = 1, 2, ...$

choose $s_{k-1} \in \partial\varphi(x_{k=1})$

set $x_k = \operatorname{argmin}_u \left[\langle s_{k-1}, u\rangle + h(u) + \|u - x_{k-1}\|^2/(2\lambda_k)\right]$ for some $\lambda_k > 0$.

Obs. See the previous discussion of the composite subgradient method wherein we have

$$v_k := \frac{x_{k-1} - x_k}{\lambda_k} \in \partial_{\varepsilon_k} f(x_k), \quad s_{k-1} \in \partial\varphi_{\varepsilon_k}(x_k)$$

where

$$\varepsilon_k = \varphi(x_k) - \varphi(x_{k-1}) - \langle s_{k-1}, x_k - x_{k-1}\rangle .$$

Obs. If $\nabla\varphi$ is L-Lipschitz then

$$\begin{aligned}
&\varepsilon_k \le \frac{L}{2}\|x_k - x_{k-1}\|^2, \\
&\tau_k = 2\lambda_k\varepsilon_k - \|x_k - x_{k-1}\|^2 \le (L\lambda_k - 1)\|x_k - x_{k-1}\|^2
\end{aligned}$$

and for $\lambda_k \le 1/L$ we have $\tau_k \le 0$ and $\Lambda_k = k/L$.

1.3 Stochastic Subgradient Method

Consider the problem

$$\begin{aligned}
\min\ & f(x) \\
\text{s.t. } & x \in X
\end{aligned}$$

where $\varnothing \ne X \subseteq \mathbb{R}^n$ is compact convex and $f : \mathbb{R}^n \mapsto \mathbb{R}$. We will assume that we have access to the following oracle:

input: $x \in \mathbb{R}^n$,

output: $s \in \mathbb{R}^n$ s.t. $\hat{s} = E_X(s) \in \partial f(x)$.

A call to the oracle is denoted by $s \in \mathcal{O}(x)$. We will assume that $E_X(\|s\|^2) \leq \tilde{M}^2$.

Stochastic Subgradient Method

Given $x_0 \in X$

For $k = 0, 1, 2, \ldots$

find $s_k \in \mathcal{O}(x_k)$

set $x_{k+1} = P_X(x_k - \lambda_k s_k)$ where $\lambda_k > 0$ is a stepsize

Lemma 1.6. *For every $k > \ell \geq 0$:*

$$2\sum_{i=\ell}^{k} \lambda_i \left(E\left[f(x_i)\right] - f_*\right) \leq E\left(\|x_\ell - x^*\|^2\right) - E\left(\|x_\ell - x^*\|^2\right) + \tilde{M}^2 \sum_{i=\ell}^{k} \lambda_i$$

Proof. For $i \geq 0$, if E_{x_k} is the expectation of x_{k+1} given x_k:

$$\begin{aligned} E_{x_k}\left(\|x_{k+1} - x^*\|^2\right) &= E_{x_k}\left(\|P_X(x_k - \lambda_k s_k) - P_X(x^*)\|^2\right) \\ &\leq E_{x_k}\left(\|x_k - x^* - \lambda_k s_k\|^2\right) \\ &\leq \|x_k - x^*\|^2 - 2\lambda_k \langle \hat{s}_k, x_k - x^* \rangle + \lambda_k^2 E_{x_k}\left(\|s_k\|^2\right) \\ &\leq \|x_k - x^*\|^2 - 2\lambda_k\left[f(x_k) - f_*\right] + \lambda_k^2 \tilde{M}^2 \end{aligned}$$

Taking expectations with respect to x_i,

$$E\left(\|x_{i+1} - x^*\|^2\right) - E\left(\|x_i - x^*\|^2\right) \leq \lambda_i^2 \tilde{M}^2 - 2\lambda_i\left(E\left[f(x_i)\right] - f_*\right)$$

and summing from $i = \ell$ to k yields the result. □

Lemma 1.7. *For $0 \leq \ell \leq k$,*

$$E\left(f_k^{\min}\right) - f_* \leq \frac{E\left(\|x_\ell - x^*\|^2\right) + \tilde{M}^2 \sum_{i=\ell}^{k} \lambda_i^2}{2\sum_{i=\ell}^{k} \lambda_i}$$

where $f_k^{\min} = \min_{1 \leq \ell \leq k} f(x_i)$.

Proof. First observe that $E[f_k^{\min}] \leq \min_{i \leq k} E[f(x_i)]$ and so

$$\begin{aligned} E[f_k^{\min}] - f_* &\leq \min_{1 \leq i \leq k} \left(E[f(x_i)] - f_*\right) \\ &\leq \min_{\ell \leq k} \left(E[f(x_i)] - f_*\right) \\ &\leq \frac{\sum_{i=\ell}^{k} \lambda_i \left(E[f(x_i)] - f\right)}{\sum_{i=\ell}^{k} \lambda_i} \\ &\leq \frac{E\left(\|x_\ell - x^*\|^2\right) + \tilde{M}^2 \sum_{i=\ell}^{k} \lambda_i^2}{2\sum_{i=\ell}^{k} \lambda_i} \end{aligned}$$

where the last inequality follows from applying the previous lemma. □

If $\lambda_i = c/\sqrt{(i+1)}$ then

$$\sum_{i=\lfloor k/2 \rfloor}^{k} \lambda_i^2 = c^2 \sum_{i=\lfloor k/2 \rfloor}^{k} \frac{1}{i+1} = c^2 \log(i+1) \Big]_{\lfloor k/2 \rfloor}^{k} = \mathcal{O}(c^2)$$
$$\sum_{i=\lfloor k/2 \rfloor}^{k} \lambda_i = c \sum_{i=\lfloor k/2 \rfloor}^{k} \frac{1}{\sqrt{i+1}} = c^2 \Omega(\sqrt{k})$$
$$E(f_k^{\min}) - f_* \le \mathcal{O}\left(\frac{D^2 + \tilde{M}^2 c^2}{\sqrt{k}}\right) \overset{(*)}{=} \mathcal{O}\left(\frac{\tilde{M} D}{\sqrt{k}}\right)$$

where $(*)$ is when $c = D/\tilde{M}$.

Application 1

Consider the problem

$$\min f(x) := \sum_{i=1}^{m} f_i(x)$$
$$\text{s.t. } x \in X$$

where $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ is convex. We will assume that for each $i = 1, \dots, m$, $\exists M_i$ such that $\|s\| \le M_i$ for all $s \in \partial f_i(x)$ for all $x \in X$.

Algorithm

Given $x_0 \in X$

For $k = 0, 1, \dots$

- choose $i_k \in \{1, 2, \dots, m\}$ randomly with uniform distribution and then choose $\bar{s}_{i_k} \in \partial f_{i_k}(x_k)$
- set $x_{k+1} = P_X(x_k - \lambda_k s_k)$ where

$$s_k = m\bar{s}_{i_k}, \quad D = \text{diam}(X), \quad \tilde{M} = \left(m \sum_{i=1}^{k} M_i^2\right)^{1/2}, \quad \lambda_k = \frac{D}{\tilde{M}\sqrt{k+1}}.$$

Obs. 1) Given $x_k \in X$ let $\bar{s}_i \in \partial f_i(x_k)$ for $i = 1, \dots, m$. Then, if P_{x_k} denotes the probability given x_k,

$$P_{x_k}(s_k = m\bar{s}_i) = \frac{1}{m}$$

and

$$E_{x_k}(s_k) = \sum_{i=1}^{m} P_{x_k}(s_k = m\bar{s}_i)(m\bar{s}_i)$$
$$= \sum_{i=1}^{m} \bar{s}_i \in \partial f_1(x_k) + \dots + \partial f_m(x_k) = \partial f(x_k)$$

Also,

$$\begin{aligned} E_{x_k}\left(\|s_k\|^2\right) &= \sum_{i=1}^m P_{x_k}\left(s_k = m\bar{s}_i\right)\|m\bar{s}_i\|^2 \\ &= m\sum_{i=1}^m \|\bar{s}_i\|^2 \le m\sum_{i=1}^m M_i^2 = \tilde{M}^2. \end{aligned}$$

The complexity is

$$E(f_k^{\min}) - f^* = \mathcal{O}\left(\frac{DM}{\sqrt{k+1}}\right)$$

where

$$M = \{\|\bar{s}_1 + \ldots + \bar{s}_m\| : \bar{s}_i \in \partial f_i(x), x \in X\} \le M_1 + \ldots + M_m = \|(M_1, \ldots, M_m)\|_1.$$

Application 2

Consider the same framework as before, with

$$f_i(x) = |a_i^T x + b_i|, \quad a_i \in \mathbb{R}^n, \quad b_i \in \mathbb{R}.$$

Then

$$\partial f_i(x) = \begin{cases} a_i, & \text{if } a_i^T x + b_i > 0 \\ -a_i, & \text{if } a_i^T x + b_i < 0 \\ [-1,1]a_i, & \text{if } a_i^T x + b_i = 0. \end{cases}$$

Any $s \in \partial f(X)$ has the form $s = A^T\lambda$, $\lambda \in [-1,1]^m$. Now,

$$\begin{aligned} &\|s\| \le \|A^T\|\|\lambda\| \le \sqrt{m}\|A^T\| \\ &M^2 \le m\|A^T A\| = m\lambda_{\max}(A^T A) \\ &\tilde{M}^2 = m\sum_{i=1}^m \|a_i\|^2 = m\text{tr}(A^T A) = m\sum_{i=1}^m \lambda_i(A^T A) \end{aligned}$$

and hence

$$\frac{1}{m} \le \frac{M^2}{\tilde{M}^2} = \frac{\lambda_{\max}(A^T A)}{\sum_{i=1}^m \lambda_i(A^T A)} \le 1.$$

1.4 Saddle-Point Subgradient Method

Consider the saddle-point problem

$$\min_{x\in X} \underbrace{\max_{y\in Y} \psi(x,y)}_{=p(x)} = \max_{y\in Y} \underbrace{\min_{x\in X} \psi(x,y)}_{=d(y)}$$

where $\psi : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$. We will assume that:

- X, Y are closed convex sets
- ψ is convex-concave differentiable function on $X \times Y$
- $p_* = d_* \in \mathbb{R}$

- $X(y) := \operatorname{argmin}_{x \in X} \psi(x,y) \neq \varnothing$ for all $y \in Y$

Lemma 1.8. *Have:*

(1) $-d \in \overline{Conv}(\mathbb{R}^n)$

(2) $s_y = -\nabla_y \psi(x_y, y) \in \partial(-d)(y)$ *for all* $x_y \in X(y)$ *and* $y \in Y$

Proof. (1) We have $-d = \sup_{x \in X} \{-\psi(x,\cdot) + \delta_Y(\cdot)\}$. So $-d$ is the pointwise supremum of closed convex functions.

(2) Let $y \in Y$ and $x_y \in X(y)$ be given and let $\tilde{y} \in Y$. So,

$$\begin{aligned} -d(\tilde{y}) &= \sup_{x \in X} -\psi(x, \tilde{y}) \\ &\geq -\psi(x_y, \tilde{y}) \\ &\geq -\psi(x_y, y) - \nabla_y [\psi(x_y, y)]^T (\tilde{y} - y) \\ &= -d(y) + s_y^T (\tilde{y} - y) \end{aligned}$$

and thus $s_y \in \partial(-d)(y)$. □

Now let us assume that there exists $M > 0$ such that for all $y \in Y$ and $x_y \in X(y)$ we have $\|s_y\| \leq M$. This leads to the following dual method.

The Dual Projected Subgradient Method

(0) $y_0 \in Y$ given

(1) For $k = 0, 1, \ldots$

compute $x_k \in X(y_k)$

set $y_{k+1} = P_Y(y_k - \lambda_k s_k)$ where

$\lambda_k > 0$ is a stepsize and $s_k = -\nabla \psi(x_k, y_k) \in \partial(-d)(y_k)$.

Obs. For all $y \in Y$ we have

$$\begin{aligned} \|y_{k+1} - y\|^2 &\leq \|y_k - \lambda_k s_k - y\|^2 \\ &= \|y_k - y\|^2 + \lambda_k^2 \|s_k\|^2 - 2\lambda_k \langle s_k, y_k - y \rangle \end{aligned}$$

and so

$$\|y_k - y\|^2 - \|y_{k+1} - y\|^2 + \lambda_k^2 \|s_k\|^2 \geq 2\lambda_k a_k(y)$$

where $a_k(y) = \langle s_k, y_k - y \rangle$.

Lemma 1.9. *For every* $k \geq 0$ *and* $(x, y) \in X \times Y$,

$$\sum_{i=1}^{k} \lambda_i a_i(y) \geq \Lambda_k [\psi(x_k^a, y) - \psi(x, y_k^a)]$$

where

$$\Lambda_k = \sum_{i=0}^{k} \lambda_i, \quad x_k^a = \sum_{i=0}^{k} \lambda_i x_i / \Lambda_k, \quad y_k^a = \sum_{i=0}^{k} \lambda_i y_i / \Lambda_k.$$

Proof. For $i \geq 0$ and $(x, y) \in X \times Y$,

$$\begin{aligned} a_i(y) &= \langle s_i, y_i - y \rangle \\ &= \langle -\nabla_y \psi(x_i, y_i), y_i - y \rangle \\ &\geq -\psi(x_i, y_i) - (-\psi(x_i, y)) \\ &= \psi(x_i, y) - \psi(x_i, y_i) \\ &\geq \psi(x_i, y) - \psi(x, y_i) \end{aligned}$$

since $x_i \in X(y_i)$. So,

$$\begin{aligned} \sum_{i=0}^{k} \lambda_i a_i(y) &\geq \Lambda_k \sum_{i=0}^{k} \frac{\lambda_i}{\Lambda_k} \left[\psi(x_i, y) - \psi(x, y_i) \right] \\ &\geq \Lambda_k \left[\psi(x_i^a, y) - \psi(x, y_i^a) \right]. \end{aligned}$$

□

Lemma 1.10. *For all* $(x, y) \in X \times Y$,

$$\psi(x_k^a, y) - \psi(x, y_k^a) \leq \frac{\|y_0 - y\|^2 - \|y_{k+1} - y\|^2 + \sum_{i=0}^{k} \lambda_i^2 \|s_i\|^2}{2\Lambda_k}$$

or equivalently, for all $y \in Y$,

$$\psi(x_k^a, y) - d(y_k^a) \leq \frac{\|y_0 - y\|^2 - \|y_{k+1} - y\|^2 + \sum_{i=0}^{k} \lambda_i^2 \|s_i\|^2}{2\Lambda_k}.$$

Applications

Consider the problem

$$\begin{aligned} p_* = \min\ & f(x) \\ \text{s.t. } & g(x) \leq 0 \\ & x \in X \end{aligned}$$

where f is convex, $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ convex, X convex. Assume that

(1)$p_* \in \mathbb{R}$,

(2) $\exists \bar{x} \in X$ such that $g(\bar{x}) < 0$.

Define $\psi(x, y) = f(x) + y^T g(x)$, $d(y) = \inf_{x \in X} f(x) + y^T g(x)$, and assume that $X(y) \neq \varnothing$ for all $y \in Y$. Here, $s_y = -g(x_y)$ where $x_y \in X(y)$.

Take $y = g(x_k^a)_+ / \|g(x_k^a)_+\|$ and remark that if $g = g^+ - g^-$ with $g^+, g^- \geq 0$ then

$$\langle y, g^+ - g^- \rangle = \left\langle \frac{g^+}{\|g^+\|}, g^+ - g^- \right\rangle = \|g^+\|.$$

We then have

$$\begin{aligned} f(x_k^a) + y^T g(x_k^a) - d(y_k^a) &= f(x_k^a) - d(y_k^a) + \|g(x_k^a)_+\| \\ &\le \frac{\|y_0 - y\|^2 + \sum_{i=0}^k \lambda_i^2 \|s_i\|^2}{2\Lambda_k} \end{aligned}$$

which will reduce due to the fact that y has a norm of 1.

1.5 Bregman Distance

Consider the problem

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in X \end{aligned}$$

where f convex and X closed convex, under the assumption that there exists $M > 0$ such that $\|s\| \le M$ for all $s \in \partial f(x)$ for all $x \in X$. Consider the iteration scheme given by

$$\begin{aligned} x_{k+1} &= P_X(x_k - \lambda_k s_k) \\ &= \operatorname{argmin}_x \left\{ \ell_f(x; x_k) + \frac{1}{2\lambda_k} \|x - x_k\|^2 : x \in X \right\}. \end{aligned}$$

To be more general, we can consider a σ-strongly convex differentiable function $w : \mathbb{R}^n \mapsto (-\infty, +\infty] \in \overline{\text{Conv}}(\mathbb{R}^n)$ where $X \subseteq \operatorname{dom} w$ and $\operatorname{dom} \partial w \subseteq X$ and define a function $dw : \mathbb{R}^n \times \operatorname{dom} \partial w \mapsto (-\infty, +\infty]$ by

$$dw(x; x_0) \equiv dw_{x_0}(x) = w(x) - w(x_0) - \langle \nabla w(x_0), x - x_0 \rangle .$$

The general method that we will discuss is one of the form

$$x_{k+1} = \operatorname{argmin}_x \left\{ \ell_f(x; x_k) + \frac{1}{2\lambda_k} dw_{x_k}(x) : x \in X \right\}.$$

Mirror Descent Method

Consider the problem

$$\begin{aligned} f_* = &\min f(x) \\ &\text{s.t. } x \in X \end{aligned}$$

where

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ convex
- $X \subseteq \mathbb{R}^n$ closed convex and nonempty
- the set of optimal solutions X_* is nonempty
- $\exists M > 0$ such that $\|s\|_* \le M$ for all $s \in \partial f(x)$ and for all $x \in X$.

Let $w : \mathbb{R}^n \mapsto (-\infty, \infty] \in \overline{\text{Conv}}(\mathbb{R}^n)$ satisfying

- w is differentiable on $\operatorname{dom}(\partial w)$
- $\operatorname{ri} X \subseteq \operatorname{int}(\operatorname{dom} w)$ and $X \subseteq \operatorname{dom} w$
- w is μ-strongly convex on X with respect to $\|\cdot\|$ where $\mu > 0$.

Define $W^0 = \operatorname{int}(\operatorname{dom} w)$ and $W = \operatorname{dom} w$.

Proposition 1.6. $\operatorname{dom}(\partial w) = \mathit{int}(\operatorname{dom} w) \neq \varnothing$.

Proof. $\operatorname{dom}(\partial w) \subseteq \operatorname{int}(\operatorname{dom} w)$. It is also well-known that

$$\operatorname{int}(\operatorname{dom} w) = \operatorname{ri}(\operatorname{dom} w) \subseteq \operatorname{dom}(\partial w).$$

□

Bregman Distance

The **Bregman distance** is a function $dw : \mathbb{R}^n \times W^0 \mapsto (-\infty, +\infty]$ given by

$$dw(x, \bar{x}) \equiv dw_{\bar{x}}(x) := w(x) - \ell_w(x; \bar{x}) = w(x) - w(\bar{x}) - \langle \nabla w(\bar{x}), x - \bar{x} \rangle .$$

Obs: $\operatorname{dom}(dw_{\bar{x}}) = \operatorname{dom} w$

Proposition 1.7. *We have*

(1) $dw_{\bar{x}}(x) \geq 0$ *for all* $x \in W$ *and* $\bar{x} \in W^0$

(2) $dw_{\bar{x}}(x) \geq \mu\|x - \bar{x}\|/2$ *for all* $x \in X$ *and* $\bar{x} \in X \cap W^0$.

Example 1.1. (1) $w(x) = \|\cdot\|_2^2/2$ gives $dw_{\bar{x}}(x) = \|x - \bar{x}\|^2/2$ and $\mu = 1$ with respect to $\|\cdot\|_2$;

(2) $w(x) = \sum_{i=1}^n x_i \log x_i$ gives

$$dw_{\bar{x}}(x) = \sum_{i=1}^n \left(x_i \log \frac{x_i}{\bar{x}_i} + \bar{x}_i - x_i \right)$$

and $\mu = 1$ with respect to $\|\cdot\|_1$;

Mirror Descent Method

(1) Given $x_0 \in X \cap W^0$

(2) For $k = 0, 1, \ldots$

choose $\lambda_k > 0$ and $s_k \in \partial f(x_k)$ and set

$(*)$ $\quad x_{k+1} = \operatorname{argmin}_{x \in X} \ell_f(x; x_k) + dw_{x_k}(x)/\lambda_k$

where $\ell_f(x; x_k) = f(x_k) + \langle s_k, x - x_k \rangle$.

Proposition 1.8. *For every* $k \geq 0$,

(a) x_{k+1} *is well defined*

(b) $x_{k+1} \in X \cap W^0$

Lemma 1.11. *If* $\varphi \in \overline{Conv}(\mathbb{R}^n)$ *is such that* $\operatorname{dom}\varphi = X$ *then*

$$(**) \qquad \inf_{x \in \mathbb{R}^n} \varphi(x) + w(x)$$

has a unique solution $\bar{x} \in X \cap W^0$.

Proof of Proposition. Let $\varphi(x) = \lambda_k \ell_f(x; x_k) + \delta_X(x) - \ell_w(x; x_k)$. So $(*)$ is the same as $(**)$. □

Proof of Lemma. We know that $\varphi + w \in \overline{\text{Conv}}(\mathbb{R}^n)$ and is also strongly convex. So $(**)$ has a unique minimizer $\bar{x}$. Clearly $\bar{x} \in X$. Have $0 \in \partial(\varphi + w)(\bar{x}) = \partial\varphi(\bar{x}) + \partial w(\bar{x})$ which follows from the fact that $W^0 \cap \operatorname{ri} X \neq \varnothing \implies \operatorname{ri}(\operatorname{dom} w) \cap \operatorname{ri}(\operatorname{dom}\varphi) \neq \varnothing$. □

Lemma 1.12. *For every* $k \geq 0$ *we have*

$$\lambda s_k + \nabla w(x_{k+1}) - \nabla w(x_k) + N_X(x_{k+1}) \ni 0$$

or equivalently

$$s_k + \frac{\nabla w(x_{k+1}) - \nabla w(x_k)}{\lambda_k} + n_k = 0, \quad n_k \in N_X(x_{k+1}).$$

Lemma 1.13. *For every* $z_0, z \in W^0$ *and* $u \in W$,

$$dw_{z_0}(u) - dw_{z_0}(z) = \langle \nabla dw_{z_0}(z), u - z \rangle + dw_z(u)$$

Proof. Exercise. □

Lemma 1.14. *For every* $k \geq 0$ *and* $u \in W$,

$$dw_{x_k}(u) - dw_{x_{k+1}}(u) \geq -\frac{\lambda_k^2 M^2}{2\mu} + \lambda_k \langle s_k, x_k - u \rangle.$$

Proof. Let $z_0 = x_k$, $z = x_{k+1}$, $s = s_k$, $n = n_k$ and $\lambda = \lambda_k$. Then, for all $u \in W \cap X \subseteq X$:

$$\begin{aligned}
dw_{x_k}(z) - dw_{x_{k+1}}(u) &= dw_{z_0}(z) - dw_z(u) \\
&= dw_{z_0}(z) + \langle \nabla(dw_{z_0})(z), u - z \rangle \\
&\vdots \\
&\geq \frac{\mu}{2}\|z - z_0\|^2 - \lambda\|s\|_*\|z - z_0\| + \lambda \langle s, x - u \rangle \\
&\geq -\frac{\lambda^2\|s\|_*^2}{2\mu} + \lambda \langle s, x - \mu \rangle
\end{aligned}$$

where $\|s\|_* = \max\{\langle s, x \rangle : \|x\| \leq 1\}$. □

Lemma 1.15. *For every* $k \geq \ell$ *and* $u \in X$,

$$\sum_{i=\ell}^{k} \lambda_i [f(x_i) - f(u)] \leq \sum_{i=\ell}^{k} \lambda_i \langle s_i, x_i - u \rangle \leq dw_{x_\ell}(u) - dw_{x_{k+1}}(u) + \frac{\sum_{i=\ell}^{k} \lambda_i^2 M^2}{2\mu}.$$

Proposition 1.9. *For every $k \geq 0$,*

$$f(\bar{x}_k) - f_* \leq \frac{dw_{x_0}(x^*) + \frac{M^2}{2\mu}\sum_{i=0}^k \lambda_i^2}{\Lambda_k}$$

where $\bar{x}_k$ is any point such that

$$f(\bar{x}_k) \leq \sum_{i=0}^k \lambda_i f(x_i)$$

and $\Lambda_k = \sum_{i=0}^k \lambda_i$.

Constant stepsize scheme

Consider $\lambda = \mu\varepsilon/M^2$ for a given tolerance $\varepsilon > 0$. If $D_0 = \min_{x^*\in X_*}\{dw_{x_0}(x^*)\}$ and $k \geq 2D_0M^2/(\mu\varepsilon^2)$, then

$$f(\bar{x}_k) - f_* \leq \frac{D_0}{\lambda k} + \frac{M^2\lambda}{2\mu} = \frac{D_0M^2}{\mu\varepsilon k} + \frac{\varepsilon}{2} \leq \varepsilon.$$

Application

Consider

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in \Delta_n = \{x \geq 0 : e^T x = 1\} \end{aligned}$$

with $x_0 = e/n$.

In the Euclidean setting, with $w = \|\cdot\|^2/2$, we have

$$dw_{x_0}(x) \leq \frac{1}{2}\left(1 - \frac{1}{n}\right), \quad k \geq \frac{M_2^2}{\varepsilon^2}.$$

In the non-Euclidean setting, with $w(x) = \sum_{i=1}^n x_i \log x_i$, we have

$$dw_{x_0}(x) = \sum_{i=1}^n x_i \log nx_i \leq \log n \implies D_0 = \log n, \quad k \geq \frac{2\log nM_\infty^2}{\varepsilon^2}$$

which could be better or worse, depending on the constants.

1.6 Prox-Subgradient

Consider the problem

$$\begin{aligned} &\min f(x) + h(x) \\ &\text{s.t. } x \in \mathbb{R}^n \end{aligned}$$

where $h \in \overline{\text{Conv}}(\mathbb{R}^n)$, $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex, and for all $x, \tilde{x} \in \mathbb{R}^n$ and $s \in \partial f(x), \tilde{s} \in \partial f(\tilde{x})$ we have

$$\|\tilde{s} - s\| \leq 2M + L\|\tilde{x} - x\|.$$

For example, we could have $f = \varphi + g$ where φ, g are convex, φ is L-smooth and g has M-bounded subgradient.

Proposition 1.10. *For $x, \tilde{x} \in \mathbb{R}^n$ and $s \in \partial f(x)$ we have*

$$f(\tilde{x}) - \ell_f(\tilde{x}; x) \leq 2M\|\tilde{x} - x\| + \frac{L}{2}\|\tilde{x} - x\|^2$$

where

$$\ell_f(\tilde{x}; x) = f(x) + \langle s, \tilde{x} - x\rangle .$$

Proof. We have

$$\begin{aligned}
&f(\tilde{x}) - \ell_f(\tilde{x}; x) \\
&= f(\tilde{x}) - f(x) - \langle s, \tilde{x} - x\rangle \\
&= \int_0^1 \langle s_t, \tilde{x} - x\rangle \, dt - \langle s, \tilde{x} - x\rangle \\
&\leq \int_0^1 \|s_t - s\| \|\tilde{x} - t\| dt \\
&\leq \|\tilde{x} - x\| \int_0^1 (2M + L\|x_t - x\|) dt \\
&= \|\tilde{x} - x\| \left(2M + \frac{L}{2}\|\tilde{x} - x\|\right)
\end{aligned}$$

where $s_t \in \partial f(x_t), x_t = x + t(\tilde{x} - x)$. □

Prox (Composite) Subgradient Method

(1) Given $x_0 \in \mathbb{R}^n$

(2) For $k = 0, 1, \ldots$

choose $\lambda > 0$ and $s_k \in \partial f(x_k)$ and set

$x_{k+1} = \operatorname{argmin}_{x \in X} \ell_f(x; x_k) + \frac{1}{2\lambda}\|x - x_k\|^2 + h(x)$

where $\ell_f(x; x_k) = f(x_k) + \langle s_k, x - x_k\rangle$ and $\lambda = \varepsilon/(4M^2 + \varepsilon L)$.

Optimality Condition

We have

$$s_k + \partial h(x_{k+1}) + \frac{1}{\lambda}(x - x_k) \ni 0, \quad s_k \in \partial f(x_k)$$

and hence

$$\frac{1}{\lambda}(x_k - x_{k-1}) \in \partial_{\varepsilon_{k+1}} f(x_{k+1}) + \partial h(x_{k+1})$$

where

$$\begin{aligned}
\varepsilon_{k+1} &= f(x_{k+1}) - f(x_k) - \langle s_k, x_{k+1} - x_k\rangle , \\
\lambda_{k+1} &= \lambda.
\end{aligned}$$

Note:

$$\begin{aligned}
\varepsilon_{k+1} &\leq 2M\|x_{k+1} - x_0\| + L\|x_{k+1} - x_k\|^2/2 \\
\tau_{k+1} &= 2\lambda_{k+1}\varepsilon_{k+1} - \|x_k - x_{k+1}\|^2 \\
&\leq 4M\lambda_{k+1}\|x_{k+1} - x_k\|^2 + (\lambda L - 1)\|x_{k+1} - x_k\|^2
\end{aligned}$$

Main Result

If $\bar{x}_k \in \mathbb{R}^n$ is such that

$$f(\bar{x}_k) \le \frac{\sum_{i=1}^k \lambda_i f(x_i)}{\Lambda_k} = \frac{\sum_{i=1}^k f(x_i)}{k}, \quad \Lambda_k = \sum_{i=1}^k \lambda_i = \lambda k,$$

then

$$f(\bar{x}_k) - f_* \le \frac{d_0^2 + \sum_{i=1}^k \tau_i}{2\Lambda_k}$$

where $\tau_k = 2\lambda_k \varepsilon_k - \|x_k - x_{k-1}\|^2$.

Lemma 1.16. *For all $k \ge 1$ and $\lambda \in (0, 1/L)$, we have*

$$\tau_k \le \frac{4M^2\lambda^2}{1 - \lambda L}.$$

Proof. Directly,

$$\begin{aligned}
\tau_k &= 2\lambda_k\varepsilon_k - \|x_k - x_{k-1}\|^2 \\
&\le 2\lambda\left[2M \underbrace{\|x_k - x_{k-1}\|}_{\delta_k} + \frac{L}{2}\|x_k - x_{k-1}\|^2\right] - \|x_k - x_{k-1}\|^2 \\
&= 4\lambda M\delta_k - (1 - \lambda L)\delta_k^2 \\
&\le \frac{4\lambda^2 M^2}{1 - \lambda L}.
\end{aligned}$$

□

Lemma 1.17. *For every $k \ge 1$ and $\lambda < 1/L$ we have*

$$f(\bar{x}_k) - f_* \le \frac{d_0^2}{2\lambda k} + \frac{4M^2\lambda}{2(1 - \lambda M)}.$$

Proposition 1.11. *If $\lambda = \varepsilon/(4M^2 + \varepsilon L)$ then*

$$f(\bar{x}_k) - f_* \le \frac{d_0^2}{2k}\left(\frac{4M^2}{\varepsilon} + L\right) + \frac{\varepsilon}{2}.$$

As a consequence, if

$$k \ge \frac{d_0^2}{\varepsilon}\left(\frac{4M^2}{\varepsilon} + L\right)$$

then $f(\bar{x}_k) - f_* \le \varepsilon$.

Lemma 1.18. *For every $k \ge 1$,*

$$f(x_k) - f(u) \le \frac{1}{2\lambda_k}\left(\tau_k + \|u - x_{k-1}\|^2 - \|u - x_k\|^2\right).$$

Corollary 1.4. *For $k \ge 1$,*

$$f(x_{k-1}) - f(x_k) \ge \frac{1}{2\lambda_k}\left(\|x_{k-1} - x_k\|^2 - \tau_k\right).$$

Corollary 1.5. *If $M = 0$ and $\lambda \in (0, 2/L)$ then*

$$\tau_k \leq (\lambda L - 1)\|x_k - x_{k-1}\|^2, \quad f(x_{k-1}) - f(x_k) \geq \frac{2 - \lambda L}{2\lambda}\|x_k - x_{k-1}\|^2.$$

Proof. Consider the case $\lambda > 1/L$. We have

$$\begin{aligned}
\sum_{i=1}^k \tau_i &\leq (\lambda L - 1)\sum_{i=1}^k \|x_i - x_{i-1}\|^2 \\
&\leq (\lambda L - 1)\frac{\lambda}{2 - \lambda L}\sum_{i=1}^k [f(x_{i-1}) - f(x_i)] \\
&= (\lambda L - 1)\frac{2\lambda}{2 - L\lambda}[f(x_0) - f_*]
\end{aligned}$$

and so

$$f(\bar{x}_k) - f_* \leq \frac{d_0^2}{2\lambda k} + \left[\frac{\lambda L - 1}{2 - \lambda L}\right]\left[\frac{f(x_0) - f_*}{k}\right]$$

for $\lambda \in \left[\frac{1}{L}, \frac{2}{L}\right)$. Hence,

$$f(\bar{x}_k) - f_* \leq \frac{d_0^2}{2\lambda k} + \max\left\{0, \frac{\lambda L - 1}{2 - \lambda L}\right\}\left[\frac{f(x_0) - f_*}{k}\right].$$

Extra

Consider the optimality condition

$$\nabla f(x_{k-1}) + \partial h(x_k) + \frac{1}{\lambda}(x_k - x_{k-1}) \ni 0$$

and define

$$v_k = \nabla f(x_k) - \nabla f(x_{k-1}) + \frac{1}{\lambda}(x_{k-1} - x_k).$$

We then have $v_k \in \nabla f(x_k) + \partial h(x_k)$. Now,

$$\|v_k\|^2 \leq \left(L + \frac{1}{\lambda}\right)^2 \|x_k - x_{k-1}\|^2$$

and

$$\begin{aligned}
\min_{i \leq 2k} \|v_i\|^2 \leq \min_{k+1 \leq i \leq 2k} \|v_i\|^2 &\leq \left(L + \frac{1}{\lambda}\right)^2 \min_{k+1 \leq i \leq 2k} \|x_i - x_{i-1}\|^2 \\
&\leq \left(L + \frac{1}{\lambda}\right)^2 \left(\frac{2\lambda}{2 - 2\lambda L}\right) \min_{k+1 \leq i \leq 2k} [f(x_{i-1}) - f(x_i)] \\
&\leq \left(L + \frac{1}{\lambda}\right)^2 \left(\frac{2\lambda}{2 - 2\lambda L}\right)\left(\frac{f(x_k) - f_*}{k}\right) \\
&\overset{(*)}{\leq} \left(L + \frac{1}{\lambda}\right)^2 \left(\frac{2\lambda}{2 - 2\lambda L}\right)\left(\frac{f(\bar{x}_k) - f_*}{k}\right) \\
&\sim \mathcal{O}\left(\frac{1}{k^2}\right)
\end{aligned}$$

where $(*)$ use the fact that $f(x_k) \leq f(x_{k-1})$, i.e. monotonicity. Hence $\min_{i \leq k} \|v_i\| \sim \mathcal{O}(1/k)$. □

1.7 Nesterov's Method

Consider the composite problem

$$\begin{aligned}\varphi_* = \min\ & \varphi(x) = f(x) + h(x) \\ \text{s.t. } & x \in \mathbb{R}^n\end{aligned}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex and differentiable and $h \in \overline{\text{Conv}}(\mathbb{R}^n)$. Assume that:

(1) $\varphi_* \in \mathbb{R}$ and the set of optimal solutions is nonempty;

(2) $\exists L > 0$ such that

$$f(x) \le \ell_f(\tilde{x}; x) + \frac{L}{2}\|\tilde{x} - x\|^2, \quad \forall x, \tilde{x} \in \mathbb{R}^n;$$

(3) φ is μ-strongly convex.

Given $A \ge 0$, $\tau > 0$, and $x, y \in \mathbb{R}^n$, we want to find $A^+ > A$, $\tau^+ > \tau$, and $x^+, y^+ \in \mathbb{R}^n$ such that

$$\begin{aligned}\eta^+ &= A^+\left[\varphi(y^+) - \varphi(u)\right] + \frac{\tau^+}{2}\|u - x^+\|^2 \\ &\le A\left[\varphi(y) - \varphi(u)\right] + \frac{\tau}{2}\|u - x\|^2 = \eta.\end{aligned} \tag{1.2}$$

Considering $\eta = \eta_0$ and $\eta^+ = \eta_k$, we have

$$\varphi(y_k) - \varphi(u) \le \frac{A_0}{A_k}\left[\varphi(y_0) - \varphi(u)\right] + \frac{\tau_0}{2A_k}\|u - x_0\|^2$$

and so $\varphi(y_k) - \varphi_* = \mathcal{O}(d_0^2/A_k)$. Many variants will use $(A_0, \tau_0) = (0, 1)$. Let us remark that (1.2) is equivalent to

$$A^+\varphi(y^+) + \frac{\tau^+}{2}\|u - x^+\|^2 \le A\varphi(y) + a\varphi(u) + \frac{\tau}{2}\|u - x\|^2 \tag{1.3}$$

where $a = A^+ - A$.

Lemma 1.19. *Assume that $\gamma \in \overline{Conv}(\mathbb{R}^n)$ is μ-strongly convex ($\mu \ge 0$), $\gamma \le \varphi$, $y^+ \in \mathbb{R}^n$, and $\gamma \le \varphi$ is such that*

$$A^+\varphi(y^+) \le A\varphi(y) + \min\left\{A\gamma(u) + \frac{\tau}{2}\|u - x\|^2\right\}. \tag{1.4}$$

Then $\tau^+ = \tau + a\mu$ and

$$x^+ = \operatorname{argmin}\left\{a\gamma(u) + \tau\|u - x\|^2/2\right\}$$

satisfies (1.3).

Proof. The equation (1.4) is equivalent to

$$\begin{aligned}A\varphi(y) + a\gamma(u) + \frac{\tau}{2}\|u - x\|^2 &\ge A^+\varphi(y^+) + \frac{a\mu + \tau}{2}\|u - x\|^2 \\ &= A^+\varphi(y^+) + \frac{\tau^+}{2}\|u - x\|^2\end{aligned}$$

and the result follows from the fact that $\gamma \le \varphi$. □

Goal: Given (A, τ, x, y), construct $A^+, y^+, \gamma \le \varphi$ satisfying (1.4). Observe that for all $a > 0$ we have

$$\begin{aligned}
& A\varphi(y) + \min\left\{a\gamma(u) + \frac{\tau}{2}\|u - x\|^2\right\} \\
&= A\varphi(y) + a\gamma(x^+) + \frac{\tau}{2}\|x^+ - x\|^2 \\
&\ge A\gamma(y) + a\gamma(x^+) + \frac{\tau}{2}\|x^+ - x\|^2 \\
&\ge (A + a)\gamma\left(\frac{Ay + ax^+}{A + a}\right) + \frac{\tau}{2}\|x^+ - x\|^2.
\end{aligned}$$

Now let

$$\tilde{y} = \frac{Ay + ax^+}{A + a}, \quad \tilde{x} = \frac{Ay + ax}{A + a}$$

and note that $\|\tilde{y} - \tilde{x}\| = [a/(A + a)]\|x^+ - x\|$. Hence,

$$\begin{aligned}
& A\varphi(y) + \min\left\{a\gamma(u) + \frac{\tau}{2}\|u - x\|^2\right\} \\
&\ge (A + a)\gamma(\tilde{y}) + \frac{\tau}{2}\left(\frac{A + a}{a}\right)^2 \|\tilde{y} - \tilde{x}\|^2 \\
&= (A + a)\left[\gamma(\tilde{y}) + \frac{\tau}{2} \cdot \frac{A + a}{a^2}\|\tilde{y} - \tilde{x}\|^2\right].
\end{aligned}$$

Choose $a > 0$ such that $\tau(A + a)/a^2 = \tilde{L} = L - \mu$ and set $A^+ = A + a$. Then

$$A\varphi(y) + \min\left\{a\gamma(u) + \frac{\tau}{2}\|u - x\|^2\right\} \ge A^+\left[\gamma(\tilde{y}) + \frac{\tilde{L}}{2}\|\tilde{y} - \tilde{x}\|^2\right].$$

Let us now consider two variants under this framework.

(1) $\gamma = \ell_f(\cdot; \tilde{x}) + h + \mu\| \cdot - \tilde{x}\|^2/2$

If φ is μ-strongly convex then $\gamma \le \varphi$ (exercise)

Iteration: Given $A \ge 0, \tau > 0$, and $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, set $x^+ = \operatorname{argmin}\{a\gamma(u) + \tau\|y - x\|^2/2\}$ and $y^+ = \tilde{y} = (Ay + ax^+)/(A + a)$. Then

$$\begin{aligned}
& A^+\left[\gamma(\tilde{y}) + \frac{\tilde{L}}{2}\|\tilde{y} - \tilde{x}\|^2\right] \\
&= A^+\left[\gamma(\tilde{y}) + \frac{\tilde{L}}{2}\|y^+ - \tilde{x}\|^2\right] \\
&= A^+\left[\ell_f(\tilde{y}; \tilde{x}) + \frac{\tilde{L}}{2}\|y^+ - \tilde{x}\|^2 + h(y^+)\right] \\
&\ge A^+\left[f(y^+) + h(y^+)\right] = A\varphi(y^+).
\end{aligned}$$

(2) [FISTA] $\gamma(u) = \tilde{\gamma}(y^+) + \langle L(\tilde{x} - y^+), u - y^+\rangle + \mu\|u - y^+\|^2/2$ where

$$\tilde{\gamma} = \ell_f(\cdot; \tilde{x}) + h + \mu\| \cdot - \tilde{x}\|^2/2.$$

Define

$$y^+ = \operatorname{argmin}\left\{\tilde{\gamma}(u) + \frac{\tilde{L}}{2}\|u - \tilde{x}\|^2\right\} = \operatorname{argmin}\left\{\ell_f(u;\tilde{x}) + h(u) + \frac{L}{2}\|u - \tilde{x}\|^2\right\}.$$

We have:

(i) $\varphi \geq \tilde{\gamma} \geq \gamma$

Proof. We know that $\tilde{\gamma} \leq \varphi$. Also, by the optimality of y^+ we have $\tilde{L}(\tilde{x} - y^+) \in \partial\tilde{\gamma}(y^+)$. Hence,

$$\varphi(u) \geq \tilde{\gamma}(u) \geq \tilde{\gamma}(y^+) + \left\langle \tilde{L}(\tilde{x} - y^+), u - y^+ \right\rangle + \mu\|u - y^+\|^2/2 = \gamma(u).$$

□

(ii) $\gamma(y^+) = \tilde{\gamma}(y^+)$. (*exercise*)

(iii) the quantity

$$\begin{aligned}
&A^+\left[\gamma(\tilde{y}) + \frac{\tilde{L}}{2}\|\tilde{y} - \tilde{x}\|^2\right] \\
&\geq A^+ \min_u \left[\gamma(u) + \frac{\tilde{L}}{2}\|u - \tilde{x}\|^2\right] \\
&= A^+\left[\gamma(y^+) + \frac{\tilde{L}}{2}\|y^+ - \tilde{x}\|^2\right] \\
&= A^+\left[\tilde{\gamma}(y^+) + \frac{\tilde{L}}{2}\|y^+ - \tilde{x}\|^2\right] \\
&\geq A^+\varphi(y^+).
\end{aligned}$$

<u>Growth of A_k</u>

Let us define

$$\theta := \frac{\tau}{\tilde{L}} = \frac{a^2}{A + a}$$

whereby

$$a = \frac{\lrcorner\theta + \sqrt{\theta^2 + 4A\theta}}{2} \geq \frac{\theta}{2} + \sqrt{A\theta} \geq A + \sqrt{A\theta} + \frac{\theta}{2} \geq \left(\sqrt{A} + \frac{\sqrt{\theta}}{2}\right)^2.$$

Now since $A^+ = A + a$, we have

$$\sqrt{A_{k+1}} \geq \sqrt{A_k} + \frac{\sqrt{\theta_k}}{2} = \sqrt{A_k} + \frac{\sqrt{\tau_k}}{2\sqrt{\tilde{L}}}.$$

If $\mu > 0$ then then improves, by using $\tau_0 = \tau_0 + A_k\mu$, to

$$\sqrt{A_{k+1}} \geq \sqrt{A_k}\left(1 + \frac{\sqrt{\mu}}{2\sqrt{\tilde{L}}}\right)$$

which implies a geometric convergence rate.

<u>FISTA</u>

Recall in FISTA that

$$x^+ = \operatorname{argmin}\left\{\tilde{\gamma}(u) + a\|u - x\|^2\right\}$$

where $\tilde{\gamma}(u) = \gamma(y^+) + \langle(\tilde{x} - y^+)L, u - y^+\rangle$ and $\gamma(u) = \ell_f(u; \tilde{x}) + h(u)$. Also,

$$y^+ = \operatorname{argmin}\left\{\gamma(u) + \frac{L}{2}\|u - \tilde{x}\|^2\right\}, \quad \tilde{x} = \frac{Ay + ax}{A + a}.$$

Lemma 1.20. *We have* $x^+ = (A^+y^+ - Ay)/a$.

Proof. We have

$$\begin{aligned} x^+ &= x - a\nabla\tilde{\gamma} \\ &= x - aL(\tilde{x} - y^+) \\ &= x - aL\left(\frac{Ay + ax}{A + a} - y^+\right) \\ &= x - \frac{A + a}{a}\left(\frac{Ay + ax}{A + a} - y^+\right) \\ &= \frac{A^+y^+ - Ay}{a}. \end{aligned}$$

□

Lemma 1.21. *We have* $\tilde{x}^+ = y^+ + [(t - 1)/(t^+)](y^+ - y)$ *where* $t = A^+/a = aL$.

Proof. We have

$$\begin{aligned} \tilde{x}^+ &= \frac{A^+}{A^{++}}y^+ + \frac{a^+}{A^{++}}x^+ \\ &= \left(y^+ - \frac{a^+}{A^{++}}y^+\right) + \frac{a^+}{A^{++}}\left(\frac{A^+y^+ - Ay}{a}\right) \\ &= y^+ + \frac{a^+}{A^{++}}\left(\frac{A^+}{a} - 1\right)(y^+ - y) \\ &= y^+ + [(t - 1)/t^+](y^+ - y). \end{aligned}$$

□

Lemma 1.22. *We have* $(t^+)^2 - t^+ - t^2 = 0$ *in which*

$$t^+ = \frac{1 + \sqrt{1 + 4t^2}}{2}.$$

1.8 Nesterov's Smooth Approximation Scheme

Consider the problem

$$\varphi_* = \min\{\varphi(x) = f(x) + h(x) : x \in \mathbb{R}^n\}.$$

We previously showed that Nesterov's method has the invariant

$$A_k\left[\varphi(y_k) - \varphi(u)\right] + \frac{\tau_k}{2}\|u - x_k\|^2 \le A\left[\varphi(y_0) - \varphi(u)\right] + \frac{\tau}{2}\|u - x_0\|^2$$

where $\tau_k = \tau_0 + \mu A_k$ and f is assumed to be μ-strongly convex.

Corollary 1.6. *For all $u \in \operatorname{dom} h$ we have*

$$\varphi(y_k) - \varphi(u) \le \frac{1}{A_k}\|x_0 - u\|^2, \quad A_k \ge \frac{k^2}{4L}$$

and also

$$\|x_k - u\| \le \|x_0 - u\| \text{ if } \varphi(u) \le \varphi(y_k).$$

Now consider the problem

$$(P) \quad \varphi_* = \min\{\varphi(x) = f(x) + h(x) + \theta(x) : x \in \mathbb{R}^n\}$$

where:

(1) $h \in \overline{\operatorname{Conv}}(\mathbb{R}^n)$ (easy one)

(2) $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex, differentiable, and ∇f is L-Lipschitz everywhere

(3) $\theta : \mathbb{R}^n \mapsto \mathbb{R}$ is convex(but not easy), , not necessarily differentiable, which is *"smoothable"*

(4) set of optimal solutions is nonempty.

Definition 1.1. θ is (C_1, C_2)-smoothable if for all $\mu > 0$ there exists $\theta_\mu : \mathbb{R}^n \mapsto \mathbb{R}$ convex differentiable such that

$$\theta_\mu \le \theta \le \theta_\mu + C_2\mu, \quad \nabla\theta_\mu \text{ is } \frac{C_1}{\mu}\text{-Lipschitz.}$$

For some $\mu > 0$, we wish to solve the problem

$$(P_\mu) \quad \varphi_{\mu,*} = \min\{\varphi_\mu(x) : x \in \mathbb{R}^n\}$$

where

$$\varphi_\mu(x) = f(x) + h(x) + \theta_\mu(x) = \underbrace{[f + \theta_\mu]}_{=: f_\mu}(x) + h(x).$$

Observe that f_μ is convex, differentiable, and ∇f_μ is $(L + C_1/\mu)$-Lipschitz. Apply FISTA to solve (P_μ):

(0) Let $x_0 \in \operatorname{dom} h, \mu > 0$, where θ_μ is (C_1, C_2)-smoothable;

(1) set $y_0 = \tilde{x}_0, t_0 = 1$;

(2) for $k = 0, 1, 2, \ldots$

$y_{k+1} = \operatorname{argmin}\left\{\ell_{f_\mu}(x; \tilde{x}_k) + h(x) + L\|x - \tilde{x}_k\|^2/2\right\}$

$t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$

$\tilde{x}_{k+1} = y_{k+1} + (t_k - 1)(y_{k+1} - y_k)/t_{k+1}$

Proposition 1.12. *If $\mu = \varepsilon/(2C_2)$ then the above FISTA approach finds y_k such that*

$$\varphi(y_k) - \varphi_* \le \varepsilon$$

in at most

$$\mathcal{O}\left(d_0\left[\sqrt{\frac{L}{\varepsilon}} + \frac{\sqrt{C_1 C_2}}{\varepsilon}\right]\right)$$

iterations.

Proof. We have $\varphi_\mu \le \varphi \le \varphi_\mu + C_2\mu$. Hence, if $x_* \in X_*$ is such that $d_0 = \|x_0 - x_*\|$ then

$$\begin{aligned}\varphi(y_k) - \varphi_* &= \varphi(y_k) - \varphi(x_*)\\ &= \varphi(y_k) - \varphi_\mu(y_k) + \varphi_\mu(y_k) - \varphi_\mu(x_*) + \varphi_\mu(x_*) - \varphi(x_*)\\ &\le C_2\mu + \varphi_\mu(y_k) - \varphi_\mu(x^*) + 0\\ &= \frac{\varepsilon}{2} + \frac{2L_\mu}{k^2}d_0^2 = \frac{\varepsilon}{2} + 2\left(L + \frac{C_1}{\mu}\right)\frac{d_0^2}{k^2}.\end{aligned}$$

We need the last term to be less than ε. This is equivalent to requiring

$$k^2 \ge 4\frac{d_0^2}{\varepsilon}\left(L + \frac{2C_1C_2}{2}\right) \iff k \sim \mathcal{O}\left(d_0\left[\sqrt{\frac{L}{\varepsilon}} + \frac{\sqrt{C_1C_2}}{\varepsilon}\right]\right).$$

<u>Examples</u>

Consider the function

$$\theta(x) = \max_{y\in\mathbb{R}^m} \langle Ax, y\rangle - g(y) = g^*(Ax)$$

where $g \in \overline{\text{Conv}}(\mathbb{R}^n)$ and $\operatorname{dom} g$ is bounded. It is known that $\theta : \mathbb{R}^n \mapsto \mathbb{R}$ is convex. □

Proposition 1.13. *Assume $\tilde{g} \in \overline{Conv}\,\mathbb{R}^m$ is μ-strongly convex. Then,*

$$\tilde{\theta}(z) = (\tilde{g})^*(z) = \sup_y \langle z, y\rangle - \tilde{g}(y)$$

has a unique maximizer $y(z)$, is convex, finite everywhere, differentiable, and its gradient is

$$\nabla\tilde{\theta}(z) = y(z).$$

Moreover, $\nabla\tilde{\theta}$ is $(1/\mu)$-Lipschitz.

Following the above proposition, consider the function

$$\tilde{\theta}_\mu(x) = \max_{y\in\mathbb{R}^m} \langle x, y\rangle - g(y) - \frac{\mu}{2}\|y - y_0\|^2$$

which is $(1/\mu)$-Lipschitz. Take $\theta_\mu(x) = \tilde{\theta}_\mu(x)$ and remark that θ is now (C_1, C_2) smoothable with $C_1 = \|A\|^2$ and

$$C_2 = \max\{\|y - y_0\|^2 : y \in \operatorname{dom} g\}.$$

2 Nonconvex Optimization

Consider the problem

$$\phi_* = \min_{z\in\mathbb{R}^n} [\phi(z) = f(z) + h(z)]$$

where

- $h \in \overline{\text{Conv}}(\mathbb{R}^n)$ such that $\operatorname{dom} h$ is bounded;
- f is differentiable on a a compact convex set $\Omega \supseteq \operatorname{dom} h$ and $\exists L > 0$ such that

$$\|\nabla f(\bar{z}) - \nabla f(z)\| \le L\|\bar{z} - z\|, \quad \forall z, \bar{z} \in \Omega,$$
$$-\frac{m}{2}\|\tilde{z} - z\|^2 \le f(\tilde{z}) - \ell_f(\tilde{z}; z) \le \frac{M}{2}\|\tilde{z} - z\|^2, \quad \forall z, \tilde{z} \in \Omega;$$

- f is nonconvex.

PGM

Given $x \in \operatorname{dom} h$, compute

$$x^+ = \operatorname{argmin} \ell_f(u; x) + \frac{1}{2\lambda}\|u - x\|^2 + h(u).$$

Optimality Condition

This is

$$0 \in \nabla f(x) + \partial h(x^+) + \frac{1}{\lambda}(x^+ - x)$$

or equivalently, for $v^+ = \nabla f(x^+) - \nabla f(x) + (x - x^+)/\lambda$, we have

$$v^+ \in \nabla f(x^+) + \partial h(x^+), \quad \|v^+\| \le \left(L + \frac{1}{\lambda}\right)\|x^+ - x\|.$$

Now, for all $u \in \mathbb{R}^n$:

$$\begin{aligned}
&\ell_f(u; x) + \frac{1}{2\lambda}\|u - x\|^2 - \frac{1}{2\lambda}\|u - x^+\|^2 \\
&\ge \ell_f(x^+; x) + \frac{1}{2\lambda}\|x^+ - x\|^2 \\
&\ge f(x^+) - \frac{M}{2}\|x^+ - x\|^2 + \frac{1}{2\lambda}\|x^+ - x\|^2.
\end{aligned}$$

Taking $u = x$, we have

$$f(x) - f(x^+) \ge \left(\frac{1}{\lambda} - \frac{M}{2}\right)\|x^+ - x\|^2.$$

Assume $\lambda > 0$ satisfying

$$\frac{1}{\lambda} > \frac{M}{2} \text{ or } \lambda < \frac{2}{M}.$$

In terms of the previous inequality, we will have

$$f(x_0) - f(x_k) \ge \left(\frac{1}{\lambda} - \frac{M}{2}\right) k \min_{i \le k}\|x_{i-1} - x_i\|^2 \ge \left(\frac{1}{\lambda} - \frac{M}{2}\right) k \left(L + \frac{1}{\lambda}\right)^{-2} \min_{i \le k}\|v_i\|^2.$$

Thus,

$$\min_{i \le k}\|v_i\|^2 \le \left(L + \frac{1}{\lambda}\right)^2 \left(\frac{2\lambda}{2 - \lambda M}\right)\left(\frac{f(x_0) - f_*}{k}\right).$$

It is possible to get a bound

$$\min_{i \le k}\|v_i\|^2 \le O\left(\frac{mMD^2}{k} + \frac{M^2 d_0^2}{k^2}\right).$$

2.1 George Lan's Method

The accelerated method

(0) Assume M satisfies

$$-\frac{m}{2}\|\tilde{z}-z\|^2 \le f(\tilde{z}) - \ell_f(\tilde{z};z) \le \frac{M}{2}\|\tilde{z}-z\|^2, \quad \forall z, \tilde{z} \in \Omega,$$

and is known; set $\lambda \in (0, 1/M)$. Let $\hat{\rho} > 0$, $y_0 \in \operatorname{dom} h$ be given. Set $A_0 = 0$, $k = 0$, $x_0 = y_0$.

(1) Compute

$$a_k = \frac{1+\sqrt{1+4A_k}}{2}, \quad A_{k+1} = A_k + a_k, \quad \tilde{x}_k = \frac{A_k y_k + a_k x_k}{A_{k+1}}.$$

(2) Compute

$$\begin{aligned} y_{k+1} &= \operatorname{argmin}_u \left\{ \ell_f(u;\tilde{x}_k) + h(u) + \frac{1}{2\lambda}\|u-\tilde{x}_k\|^2 \right\}, \\ x_{k+1} &= P_\Omega\left(x_k - a_k(\tilde{x}_k - y_{k+1})\right). \end{aligned}$$

(3) Set

$$v_{k+1} = \frac{1}{\lambda}(\tilde{x}_k - y_{k+1}) + \nabla f(y_{k+1}) - \nabla f(\tilde{x}_k).$$

If $\|v_{k+1}\| \le \hat{\rho}$ then stop; else set $k \leftarrow k+1$ and go to (1).

Obs

One can show that:

- $a_k^2 = A_{k+1}$ for all $k \ge 0$
- $v_{k+1} \in \nabla f(y_{k+1}) + \partial h(y_{k+1})$ with

$$\|v_{k+1}\| \le \left(L + \frac{1}{\lambda}\right)\|y_{k+1} - \tilde{x}_k\|$$

Lemma 2.1. *Let*

$$\begin{aligned} \tilde{\gamma}_k(u) &= \ell_f(u;\tilde{x}_k) + h(u), \\ \gamma_k(u) &= \tilde{\gamma}_k(y_{k+1}) + \frac{1}{\lambda}\langle \tilde{x}_k - y_{k+1}, u - y_{k+1}\rangle. \end{aligned}$$

Then,

(1) $\gamma_k(y_{k+1}) = \tilde{\gamma}_k(y_{k+1})$, $\gamma_k \le \tilde{\gamma}_k$ *and*

$$y_{k+1} = \operatorname{argmin}\left\{\tilde{\gamma}_k(u) + \frac{1}{2\lambda}\|u-\tilde{x}_k\|^2\right\} = \operatorname{argmin}\left\{\gamma_k(u) + \frac{1}{2\lambda}\|u-\tilde{x}_k\|^2\right\};$$

(2) $\tilde{\gamma}_k(u) - \phi(u) \le \frac{m}{2}\|u-\tilde{x}_k\|^2$;

(3) $x_{k+1} = \operatorname{argmin}_{u\in\Omega}\left\{a_k\gamma_k(u) + \frac{1}{2\lambda}\|u-x_k\|^2\right\}$.

Notation: Let us denote $A^+ = A_{k+1}, A = A_k, a = a_k, y^+ = y_{k+1}, \ldots, \gamma = \gamma_k, \tilde{\gamma} = \tilde{\gamma}_k$.

Lemma 2.2. *For every $u \in \Omega$:*

$$\begin{aligned} &A^{+}\phi(y^{+}) + \frac{1}{2\lambda}\|u - x^{+}\|^2 + \left(\frac{1-\lambda M}{2\lambda}\right) A^{+}\|y^{+} - \tilde{x}\|^2 \\ &\le [A\gamma(y) + a\gamma(u)] + \frac{1}{2\lambda}\|u - x\|^2. \end{aligned}$$

Proof. Have

$$\phi(y^{+}) \le \tilde{\gamma}(y^{+}) + \frac{M}{2}\|y^{+} - \tilde{x}\|^2$$

and hence we have

$$\begin{aligned} &A^{+}\phi(y^{+}) + \left(\frac{1-\lambda M}{2\lambda}\right) A^{+}\|y^{+} - \tilde{x}\|^2 \\ &\le A^{+}\tilde{\gamma}(y^{+}) + \frac{A^{+}M}{2}\|y^{+} - \tilde{x}\|^2 + \left(\frac{1-\lambda M}{2\lambda}\right) A^{+}\|y^{+} - \tilde{x}\|^2 \\ &= A^{+}\left[\tilde{\gamma}(y^{+}) + \frac{1}{2\lambda}\|y^{+} - \tilde{x}\|^2\right] \\ &\le A^{+}\left[\gamma\left(\frac{Ay + ax^{+}}{A^{+}}\right) + \frac{1}{2\lambda}\left\|\frac{Ay + ax^{+}}{A^{+}} - \frac{Ay + ax}{A^{+}}\right\|^2\right] \\ &= A\gamma(y) + a\gamma(x^{+}) + \frac{1}{2\lambda}\cdot\frac{a^2}{A^{+}}\|x^{+} - x\|^2 \\ &= A\gamma(y) + a\gamma(x^{+}) + \frac{1}{2\lambda}\|x^{+} - x\|^2 \\ &\le A\gamma(y) + a\gamma(u) + \frac{1}{2\lambda}\|u - x\|^2 - \frac{1}{2\lambda}\|u - x^{+}\|^2 \end{aligned}$$

where the last inequality is by the optimality of x^{+}. □

Lemma 2.3. *For all $u \in \Omega$ we have*

$$\left(\frac{1-\lambda M}{2\lambda}\right) A^{+}\|y^{+} - \tilde{x}\|^2 \le \frac{m}{2}\left[A\|y - \tilde{x}\|^2 + a\|u - \tilde{x}\|^2\right] + (\eta - \eta^{+})$$

where

$$\eta = \eta(u) = A\left[\phi(y) - \phi(u)\right] + \frac{1}{2\lambda}\|u - x\|^2.$$

Proof. By a previous result,

$$\begin{aligned} &\eta^{+}(u) + \left(\frac{1-\lambda M}{2\lambda}\right) A^{+}\|y^{+} - \tilde{x}\|^2 \\ &\le A\left[\gamma(y) - \phi(y)\right] + A\left[\phi(y) - \phi(u)\right] + a\left[\gamma(u) - \phi(u)\right] + \frac{1}{2\lambda}\|u - x\|^2 \\ &\le \frac{m}{2}\left[A\|y - \tilde{x}\|^2 + a\|u - \tilde{x}\|^2\right] + \eta(u). \end{aligned}$$

□

Lemma 2.4. *For any $u \in \operatorname{dom} h$,*

$$\begin{aligned} A\|y-\tilde{x}\|^2 + a\|u-\tilde{x}\|^2 &\le 2\|u-x\|^2 + 2(1+a)D_h^2 \\ &\le 2\left(D_\Omega^2 + (1+a)D_h^2\right). \end{aligned}$$

Proof. We have

$$\begin{aligned} &A\|y-\tilde{x}\|^2 + a\|u-\tilde{x}\|^2 \\ &= A\left\|y - \frac{Ay+ax}{A^+}\right\|^2 + a\left\|\frac{A}{A^+}(u-y) + \frac{a}{A^+}(u-x)\right\|^2 \\ &= \frac{Aa^2}{A^+}\|y-x\|^2 + 2a\left(\frac{A^2}{(A^+)^2}\|u-y\|^2 + \frac{A^2}{(A^+)^2}\|u-x\|^2\right) \\ &\le \frac{2A}{A^+}\left(\|y-x\|^2 + \|u-x\|^2\right) + 2a\|u-y\|^2 + \frac{2a}{A^+}\|u-x\|^2 \\ &= 2\|u-x\|^2 + 2(1+a)\|u-y\|^2. \end{aligned}$$

□

Lemma 2.5. *For all $u \in \operatorname{dom} h$ we have*

$$\left(\frac{1-\lambda M}{2\lambda}\right)\sum_{i=0}^{k} A_i\|y_{i+1}-\tilde{x}_i\|^2 \le \eta_0(u) - \eta_{k+1} + \frac{m}{2}\left[2(k+1)(D_\Omega^2 + D_h)^2 + 2D_h^2\sum_{i=0}^{k} a_i\right].$$

Proposition 2.1. *For all $k \ge 0$ we have*

$$\begin{aligned} &\frac{1}{\lambda^2}\min_{i\le k}\|y_{i+1}-\tilde{x}_i\|^2 \\ &\le \frac{2\lambda}{(1-\lambda M)\sum_{i=0}^{k} A_i}\left[\frac{d_0^2}{2\lambda} + m(k+1)\left(D_\Omega^2 + D_h^2\right) + mD_h^2\sum_{i=0}^{k} a_i\right] \\ &\sim \mathcal{O}\left(\frac{d_0^2}{\lambda^2 k^3} + \frac{mD_\Omega^2}{\lambda k^2} + \frac{mD_h^2}{k}\right); \end{aligned}$$

and, furthermore,

$$\|v_i\|^2 \le \left(\frac{1}{\lambda} + M\right)^2\|y_{i+1}-\tilde{x}_i\|^2 = (1+\lambda M)^2\left[\frac{1}{\lambda^2}\|y_{i+1}-\tilde{x}_i\|^2\right].$$

2.2 Inexact Proximal Point Method

Consider the problem

$$\begin{aligned} (*) \quad \phi_* = \inf\ &\Phi(x) := f(x) + h(x) \\ \text{s.t. } &x \in \mathbb{R}^n \end{aligned}$$

where

- $h \in \overline{\operatorname{Conv}}(\mathbb{R}^n)$

- f is differentiable on $\operatorname{dom} h$ and $\exists m, M > 0$ such that

$$\frac{m}{2}\|u-x\|^2 \le f(u) - \ell_f(u;x) \le \frac{M}{2}\|u-x\|^2$$

- $\Phi_* > -\infty$
- $0 < m \ll M$

Obs. If $\operatorname{dom} h$ is bounded, then Nesterov's ACG solves $(*)$ in

$$\mathcal{O}\left(\frac{MmD^2}{\bar{\rho}^2}\right)$$

iterations to obtain $(z, v) \in \operatorname{dom} h \times \mathbb{R}^n$ such that

$$v \in \nabla f(z) + \partial h(z), \quad \|v\| \le \rho$$

where $D = \operatorname{diam}(\operatorname{dom} h)$.

Inexact Proximal Point (IPP) Framework

Input: $\sigma \in (0,1)$, $z_0 \in \operatorname{dom} h$

Steps:

(0) Set $k = 1$

(1) Compute $(\lambda_k, z_k, \tilde{v}_k, \tilde{\varepsilon}_k)$ such that

$$\tilde{v}_k \in \partial_{\tilde{\varepsilon}_k}\left(\lambda_k \phi + \frac{1}{2}\|\cdot - z_{k-1}\|^2\right)(z_k),$$
$$\|\tilde{v}_k\|^2 + 2\tilde{\varepsilon}_k \le \sigma\|z_{k-1} - z_k + \tilde{v}_k\|^2.$$

(2) set $k \leftarrow k+1$ and go to (1)

Analysis

Consider the iterates

$$z_k = \operatorname{argmin}_u \left\{\tilde{\phi}_k(u) := \lambda_k \phi(u) + \frac{1}{2}\|u - z_{k-1}\|^2\right\}$$

where the objective function has curvature pair $(1-\lambda_k m, 1+\lambda_k M)$. So if $\lambda_k \le 1/m$ then $\tilde{\phi}_k$ is convex and $\lambda_k < 1/m$ implies it is strongly convex.

Approximate Solutions

(a) for $\hat{\rho} > 0$, a pair $(\hat{z}, \hat{v}) \in \operatorname{dom} h \times \mathbb{R}^n$ is a $\hat{\rho}$-solution if

$$\hat{v} \in \nabla f(\hat{z}) + \partial h(\hat{z}), \quad \|\hat{v}\| \le \hat{\rho}$$

(b) for $(\bar{\rho}, \bar{\varepsilon})$, a quintuple $(\lambda, z_0, z, v, \varepsilon)$ is a $(\bar{\rho}, \bar{\varepsilon})$-prox solution if

$$v \in \partial_\varepsilon\left(\phi + \frac{1}{2\lambda}\|\cdot - z_0\|^2\right)(z), \quad \|(z - z_0)/\lambda + v\| \le \bar{\rho}, \quad \varepsilon \le \bar{\varepsilon}.$$

Proposition 2.2. *A $(\bar{\rho}, \bar{\varepsilon})$-prox solution yields a ρ-solution where*

$$\rho = 2\left[\bar{\rho} + \sqrt{2\bar{\varepsilon}(M + \lambda^{-1})}\right].$$

Remark. Set $\bar{\rho} = \hat{\rho}/4$ and $\bar{\varepsilon} = \hat{\rho}^2/(32[M + \lambda^{-1}])$. Then $\rho = \hat{\rho}$.

Note in the IPP framework that if $(v_k, \varepsilon_k) = (\tilde{v}_k, \tilde{\varepsilon}_k)/\lambda_k$ then

$$v_k \in \partial_{\varepsilon_k}\left(\phi + \frac{1}{2\lambda_k}\|\cdot - z_{k-1}\|^2\right)(z_k),$$
$$\|\tilde{v}_k\|^2 + 2\frac{\varepsilon_k}{\lambda_k} \le \sigma\theta_k,$$

where

$$\theta_k := \left\|\frac{z_{k-1} - z_k}{\lambda_k} + v_k\right\|^2.$$

The above implies

$$\varepsilon_k \sim O\left(\sigma\lambda_k\theta_k^2\right), \quad \|v_k\|^2 \sim O\left(\sigma\theta_k^2\right)$$

Lemma 2.6. *For all $k \ge 1$,*

$$\lambda_k\theta_k^2 \le \frac{2\left[\phi(z_{k-1}) - \phi(z_k)\right]}{1 - \sigma}.$$

Proof. From the inclusion, we have

$$\lambda_k\phi(z) \ge \frac{1}{2}\|z - z_{k-1}\|^2 \ge \lambda_k\phi(z_{k-1}) + \frac{1}{2}\|z_k - z_{k-1}\|^2 + \langle\tilde{v}_k, z - z_k\rangle - \tilde{\varepsilon}_k,$$

for all z. In particular, using $z = z_{k-1}$ yields

$$\begin{aligned}\lambda_k\left[\phi(z_{k-1}) - \phi(z_k)\right] &\ge \left(\|z_k - z_{k-1}\|^2 + 2\langle\tilde{v}_k, z_{k-1} - z_k\rangle - 2\tilde{\varepsilon}_k\right) \\ &= \frac{1}{2}\left[\|z_{k-1} - z_k + \tilde{v}_k\|^2 - \|\tilde{v}_k\|^2 - 2\tilde{\varepsilon}_k\right] \\ &\ge \frac{1-\sigma}{2}\|z_{k-1} - z_k + \tilde{v}_k\|^2 = \frac{1-\sigma}{2}\lambda_k^2\theta_k^2.\end{aligned}$$

□

Lemma 2.7. *For every $k \ge 1$ we have*

$$\min_{i \le k}\theta_i^2 \le \frac{2\left[\phi(z_0) - \phi_*\right]}{(1-\sigma)\Lambda_k}, \quad \Lambda_k = \sum_{i=1}^{k}\lambda_i.$$

Proof. By the previous lemma,

$$\left(\min_{i \le k}\theta_i^2\right)\Lambda_k \le \sum_{i \le k}\lambda_i\theta_i^2 \le \frac{2\left[\phi(z_0) - \phi_*\right]}{1 - \sigma}.$$

□

Proposition 2.3. *(Refined IPP bound) For every $k \ge 1$ we have*

$$\min_{i \le j}\theta_i^2 \le \frac{2R(\phi; \lambda_1)}{(1-\sigma)^2\lambda_1(\Lambda_k - \lambda_1)}$$

where

$$R(\phi;\lambda_1) = \inf_u \left\{ \lambda(1-\sigma)\left[\phi(u)-\phi_*\right] + \frac{1}{2}\|z_0 - u\|^2 \right\}.$$

In particular,

$$\min_{i\le k} \theta_i^2 \le \frac{\min\left\{2\left[\phi(z_0)-\phi_*\right], \frac{d_0^2}{(1-\sigma)\lambda}\right\}}{\lambda(1-\sigma)(k-1)}.$$

Want: $\theta_k^2 \le \bar{\rho}^2 = O(\hat{\rho}^2)$ and

$$\varepsilon_k \le O\left(\frac{\hat{\rho}^2}{M+\lambda^{-1}}\right) \Longleftarrow \sigma\lambda_k\theta_k^2 \le \frac{\hat{\rho}}{M+\lambda^{-1}} \Longleftrightarrow \theta_k^2 \le O\left(\frac{\rho^2}{\sigma(\lambda M+1)}\right)$$

and so we can choose $\sigma = 1/(\lambda M + 1)$.

Outer Iteration Complexity

Some algebra gives an outer complexity of

$$O\left(\frac{R(\phi;\lambda)}{\lambda^2\hat{\rho}^2}\right).$$

Now consider the composite set-up $\phi = f + h$ and define

$$\phi_s = \lambda f + \frac{1}{2}\|\cdot - z_{k-1}\|^2, \quad \phi = \lambda h$$

Proposition 2.4. *The ACG method, started from z_0, obtains $(z, \tilde{v}, \tilde{\varepsilon})$ such that*

$$\tilde{v} \in \partial_{\tilde{\varepsilon}}(\psi_s + \psi_n)(z)$$
$$\|\tilde{v}\|^2 + 2\tilde{\varepsilon} \le \sigma\|z_0 - z + \tilde{v}\|^2$$

in at most $O(\sqrt{L/\sigma})$ or $O\left(\sqrt{L/\mu}\cdot\log(L/\sigma)\right)$ iterations.

In particular, we are interested in the setup

$$\lambda = \frac{1}{2m}, \quad \mu = 1 - \lambda m = \frac{1}{2}, \quad L = 1 + \lambda M \sim O\left(\frac{M}{m}\right).$$

3 Block Decomposition Methods

Consider the problem

$$\min \varphi(x) = \varphi(x_1, ..., x_p) = f(x_1, ..., x_b) + \sum_{i=1}^{b} h_i(x_i)$$
$$\text{s.t } x \in \mathbb{R}^{n_i}$$

where

- f is convex and differentiable everywhere on $\mathbb{R}^{n_1\times\dots\times n_b}$

- $h_i \in \overline{\text{Conv}}\, \mathbb{R}^{n_i}$
- φ_* is achieved
- $\exists L_i > 0$ such that
$$f(x + U_i(d_i)) - f(x) \le \nabla_i f(x)^T d_i + \frac{L_i}{2}\|d_i\|^2, \quad d_i \in \mathbb{R}^{n_i}$$
where $U_i(d_i) = (0, ..., d_i, ..., 0)$

Method

Let $x^0 = (x_1^0, ..., x_b^0) \in \operatorname{dom} h$.

For $k = 0, 1, ...$

choose $i_k \in \{1, ..., b\}$ randomly

set $x^{k+1} = x^k + U_{i_k}(\hat{x}_i^k - x_i^k)$ where

$$\hat{x}_i^k = \operatorname{argmin}_x \{\langle \nabla_i f(x), u_i - x_i\rangle + h_i(u_i)\} + \frac{L_i}{2}\|u_i - x_i^k\|^2.$$

Theorem 3.1. *If selection is uniform, then for all $k \ge 0$ we have*

$$E\left[\varphi(x^k) - \varphi_*\right] \le \frac{1}{b+k}\left(b\left[\varphi(x^0) - \varphi_*\right] + \frac{1}{2}d_0^2\right)$$

where

$$d_0^2 = \min\left\{\|x^0 - x^*\|_{Lb}^2 : x^* \in X^*\right\},$$
$$\|y\|_\eta^2 = \sum_{i=1}^b \eta_i \|y_i\|^2$$

for $\eta = (\eta_1, ..., \eta_b)$, $y = (y_1, ..., y_b)$, and $Lb = L \circ b$.

Lemma 3.1. *For all $k \ge 0$ and $i = 1, ..., b$ we have*

$$\begin{aligned}&\left\langle \nabla_i f(x_i^k), u_i - x_i^k\right\rangle + h_i(u_i) + \frac{L_i}{2}\|u_i - x_i^k\|^2 \\ &\ge \left\langle \nabla_i f(x_i^k), \hat{x}_i^k - x_i^k\right\rangle + h_i(\hat{x}_i^k) + \frac{L_i}{2}\|\hat{x}_i^k - x_i^k\|^2 + \frac{L_i}{2}\|u_i - \hat{x}_i^k\|^2.\end{aligned}$$

Lemma 3.2. *For all $k \ge 0$ and $i = 1, ..., b$ we have*

$$\begin{aligned}-\frac{L_i}{2}\|\hat{x}_i^k - x_i^k\|^2 &\ge \left\langle \nabla_i f(x_i^k), \hat{x}_i^k - x_i^k\right\rangle + h_i(\hat{x}_i^k) - h_i(x_i^k) + \frac{L_i}{2}\|\hat{x}_i^k - x_i^k\|^2 \\ &\ge \varphi\left(x^k[i]\right) - \varphi(x^k).\end{aligned}$$

Proof. The first inequality follows from the previous lemma with $u_i = x_i^k$. The second one follows from the fact that

$$\begin{aligned}\varphi(x[i]) - \varphi(x) &= [f(x[i]) - f(x)] + [h(x[i]) - h(x)] \\ &\le h_i(x_i^+) - h_i(x_i) + \nabla_i f(x)^T(\hat{x}_i - x_i) + \frac{L_i}{2}\|\hat{x}_i - x_i\|^2.\end{aligned}$$

□

Lemma 3.3. *For all $k \geq 0$ and $i = 1, \ldots, b$ we have*

$$\begin{aligned}&\frac{L_i}{2}\left(\|x_i^k - x_i^*\|^2 - \|\hat{x}_i - x_i^*\|^2\right)\\ &\geq \left\langle \nabla_i f(x^k), x_i^k - x_i^* \right\rangle + h_i(x_i^k) - h_i(x_i^*) + \varphi(x^k[i]) - \varphi(x^k)\end{aligned}$$

Proof. By Lemma 3.1 with $u_i = x_i^*$ we have

$$\begin{aligned}&\left\langle \nabla_i f(x_i^k), x_i^* - x_i^k \right\rangle + h_i(x_i^*) + \frac{L_i}{2}\|x_i^* - x_i^k\|^2\\ &\geq \left\langle \nabla_i f(x_i^k), \hat{x}_i^k - x_i^k \right\rangle + h_i(\hat{x}_i^k) + \frac{L_i}{2}\|\hat{x}_i^k - x_i^k\|^2 + \frac{L_i}{2}\|x_i^* - \hat{x}_i^k\|^2,\end{aligned}$$

and so

$$\begin{aligned}&\frac{L_i}{2}\left(\|x_i^* - x_i^k\|^2 - \|x_i^* - \hat{x}_i^k\|^2\right)\\ &\geq \left\langle \nabla_i f(x_i^k), x_i^k - x_i^* \right\rangle + h_i(x_i^k) - h_i(x_i^*) +\\ &\qquad \left\langle \nabla_i f(x_i^k), \hat{x}_i^k - x_i^k \right\rangle + h_i(\hat{x}_i^k) - h_i(x_i^k) + \frac{L_i}{2}\|\hat{x}_i^k - x_i^k\|^2\\ &\geq \left\langle \nabla_i f(x_i^k), x_i^k - x_i^* \right\rangle + h_i(x_i^k) - h_i(x_i^*) + \left[\varphi(x^k[i]) - \varphi(x^k)\right].\end{aligned}$$

□

Lemma 3.4. *For all $k \geq 0$ we have*

$$\begin{aligned}&\frac{1}{2}\left(\|x^* - x^k\|_L^2 - \|x^* - \hat{x}^k\|_L^2\right)\\ &\geq f(x^k) - \ell_f(x^*; x^k) + h(x^k) - h(x^*) + \sum_{i=1}^{b}\left[\varphi(x^k[i]) - \varphi(x^k)\right]\\ &\geq \varphi(x^k) - \varphi(x^*) + \sum_{i=1}^{b}\left[\varphi(x^k[i]) - \varphi(x^k)\right].\end{aligned}$$

Lemma 3.5. *For all $k \geq 0$, probability vector $p = (p_1, \ldots, p_b)$ used to sample the block, and $\eta = (\eta_1, \ldots, \eta_b) \in \mathbb{R}_{++}^b$:*

$$E_{x^k}\left(\|x^{k+1} - x^*\|_\eta^2 - \|x^k - x^*\|_\eta^2\right) = \|\hat{x}^k - x^*\|_{p\eta}^2 - \|x^k - x^*\|_{p\eta}^2.$$

Proof. We have

$$\begin{aligned}&E_{x^k}\left(\|x^{k+1} - x^*\|_\eta^2\right) - \|x^k - x^*\|^2\\ &= \sum_{i=1}^{b} p_i\left(\|x^k - x^* + U_i(\hat{x}_i - x_i^k)\|_\eta^2 - \|x^k - x^*\|_\eta^2\right)\\ &= \sum_{i=1}^{b} p_i\eta_i\left(\|\hat{x}_i^k - x_i^*\|^2 - \|x_i^k - x_i^*\|^2\right)\\ &= \|\hat{x}^k - x^*\|_{p\eta}^2 - \|x^k - x^*\|_{p\eta}^2.\end{aligned}$$

□

Lemma 3.6. *For $\eta = L/p$ we have*

$$\begin{aligned} &E_{x^k}\left(\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2\right) \\ &\geq \varphi(x^k) - \varphi_* + \sum_{i=1}^{b}\left[\varphi(x^k[i]) - \varphi(x^k)\right] \end{aligned}$$

Lemma 3.7. *Define*

$$\begin{aligned} d_k^2 &= \min\left\{E\left(\|x^k - x^*\|^2\right) : x^* \in X^*\right\}, \\ \theta_k &= E\left(\varphi(x^k) - \varphi_*\right). \end{aligned}$$

For $k \geq 0$ we have

$$\begin{aligned} \frac{1}{2}\left(d_k^2 - d_{k+1}^2\right) &\geq \theta_k + \frac{1}{p_{\min}}\left[E\left(\varphi(x^{k+1}) - \varphi(x^k)\right)\right], \\ &= \theta_k + \frac{1}{p_{\min}}\left[\theta_{k+1} - \theta_k\right]. \end{aligned}$$

Proof. We have

$$E_{x^k}\left[\varphi(x^{k+1}) - \varphi(x^{k+1})\right] \leq \frac{1}{p_{\min}}\sum_{i=1}^{b}\left[\varphi(x^k[i]) - \varphi(x^k)\right]$$

and so

$$\theta_k \leq b\left[\theta_k - \theta_{k+1}\right] + \frac{1}{2}\left(d_k^2 - d_{k+1}^2\right)$$

from which we conclude

$$k\theta_k \leq \sum_{\ell=0}^{k-1}\theta_\ell \leq b[\theta_0 - \theta_k] + \frac{1}{2}\left(d_0^2 - d_k^2\right).$$

Hence, we have

$$\sum_{i=1}^{b}\frac{1}{p_i}p_i\left[\varphi(x^k[i]) - \varphi(x^k)\right] \geq \sum_{i=1}^{b}\frac{1}{p_{\min}}p_i\left[\varphi(x^k[i]) - \varphi(x^k)\right] = \frac{1}{p_{\min}}E\left[\varphi(x^k[i]) - \varphi(x^k)\right]$$

and the complexity

$$\theta_k \leq \left(\frac{\theta_0}{p_{\min}} + \frac{1}{2}d_0^2\right)\Big/\left(k + \frac{1}{p_{\min}}\right).$$

□

3.1 Accelerated Methods

Randomized BC-ACG

Consider the problem

$$\begin{aligned} \phi_* := \min\ & \phi(x) = f(x) + h(x) \\ \text{s.t. } & x = (x_1, \ldots, x_b) \in \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_b} = \mathbb{R}^n \end{aligned}$$

where

- $h(x) = \sum_{i=1}^{b} h_i(x_i)$ and $h_i \in \overline{\text{Conv}}(\mathbb{R}^n)$
- $f : \mathbb{R}^n \mapsto \mathbb{R}$ differentiable, convex, and for every $i = 1, \ldots, b$ there exists $L_i > 0$ such that

$$f(x + U_i d_i) \leq f(x) + \langle \nabla_i f(x), d_i \rangle + \frac{L_i}{2} \|d_i\|^2$$

- $\phi_* \in \mathbb{R}$ and is achieved

The method

(0) Given $x^0 = (x_1^0, \ldots, x_b^0) \in \text{dom}\, h$, set $k = 0, y^0 = x^0, A_0 = 0$

(1) Choose $i_k \in \{1, \ldots, b\}$ randomly. Compute a_k by solving

$$\frac{a_k^2}{A_k + a_k} = \frac{1}{b^2},$$

set

$$A_{k+1} = A_k + a_k, \quad \tilde{x}^k = \frac{A_k y^k + a_k x^k}{A_{k+1}},$$

compute

$$x^{k+1} = x^k[i_k], \quad y^{k+1} = \tilde{x}_k + \frac{1}{b a_k}(x^{k+1} - x^k) = \tilde{x}_k + \frac{b a_k}{A_{k+1}}(x^{k+1} - x^k)$$

where

$$\begin{aligned} x^k[i] &= x^k + U_i(\hat{x}_i^k - x_i^k), \\ \hat{x}_i^k &= \text{argmin}_u \left(a_k \left[\left\langle \nabla_i f(\tilde{x}_i^k), u - \tilde{x}_i^k \right\rangle + h_i(u) \right] + \frac{L_i}{2b} \|u - x_i^k\|^2 \right). \end{aligned}$$

Lemma 3.8. *We have*

$$\hat{x} = \text{argmin}_u \left(a \left[\ell_f(u; \tilde{x}) + h(u) \right] + \frac{1}{2b} \|u - x\|_L^2 \right).$$

Proof. Obvious. □

Lemma 3.9. *Let*

$$\gamma(u) := \frac{1}{ba} \langle x - \hat{x}, u - \hat{x} \rangle + \ell_f(\hat{x}; \tilde{x}) + h(\hat{x}).$$

Then, we have:

(i) γ *is affine;*

(ii) $\gamma \leq \ell_f(\cdot; \tilde{x}) + h(\cdot) \leq \phi$*;*

(iii) $\hat{x} = \text{argmin} \left(a\gamma(u) + \frac{1}{2b} \|u - x\|^2 \right)$.

Lemma 3.10. *We have, for all* $u \in \mathbb{R}^n$*,*

$$a\gamma(\hat{x}) + \frac{1}{2b} \|\hat{x} - x\|_L^2 \leq a\gamma(u) + \frac{1}{2b} \|u - x\|_L^2 - \frac{1}{2b} \|u - x\|_L^2.$$

Proof. Exercise. □

Lemma 3.11. *Define, for $i = 1, \dots, b$,*

$$y[i] = \tilde{x} + \frac{1}{ba}\left(x[i] - x\right) = \tilde{x} + \frac{ba}{A^+}\left(x[i] - x\right).$$

For all $i = 1, \dots, b$ we have

$$y[i] = \frac{A}{A^+}y + \frac{a}{A^+}x_b[i]$$

where

$$x_b[i] = x + bU_i(\hat{x}_i - x_i) = x + b\left(x[i] - x\right).$$

Proof. We have

$$\begin{aligned} y[i] &= \tilde{x} + \frac{ba}{A^+}\left(x[i] - x\right) \\ &= \frac{Ay + ax}{A^+} + \frac{ba}{A^+}\left(x[i] - x\right) \\ &= \frac{A}{A^+} + \frac{a}{A^+}\left(\underbrace{x + b\left(x[i] - x\right)}_{x_b[i]}\right). \end{aligned}$$

□

Lemma 3.12. *We have $\frac{1}{b}\sum_{i=1}^{b} x_b[i] = \hat{x}$.*

Proof. We have

$$x_b[i] = x + b\left(x[i] - x\right) = x + bU_i\left(\hat{x}_i - x_i\right)$$

and hence

$$\frac{1}{b}\sum_{i=1}^{b} x_b[i] = x + \frac{1}{b}\left[b(\hat{x} - x)\right] = \hat{x}.$$

□

Lemma 3.13. *If $y[i] \in \operatorname{dom} h$ then*

$$A^+\phi(y[i]) \le A\phi(y) + a\tilde{\gamma}(x[i]) + \frac{L_i}{2}\|x[i] - x\|^2$$

where $\tilde{\gamma}(u) = \ell_f(u; \tilde{x}) + h(u)$.

Proof. We have

$$\phi(y[i]) \le \ell_f(x[i]; \tilde{x}) + h(x[i]) + \frac{L_i}{2}\|x[i] - \tilde{x}\|^2 = \tilde{\gamma}(x[i]) + \frac{L_i}{2}\|x[i] - \tilde{x}\|^2.$$

So,

$$\begin{aligned}
A^+\phi(y[i]) &\le A^+\left[\tilde{\gamma}(x[i]) + \frac{L_i}{2}\|x[i]-\tilde{x}\|^2\right] \\
&= A^+\left[\tilde{\gamma}\left(\frac{A}{A^+}y + \frac{a}{A^+}x_b[i]\right) + \frac{L_i b^2 a^2}{2(A^+)^2}\|x[i]-\tilde{x}\|^2\right] \\
&\le A\tilde{\gamma}(y) + a\tilde{\gamma}(x_b[i]) + \frac{L_i}{2}\|x[i]-\tilde{x}\|^2 \\
&\le A\phi(y) + a\tilde{\gamma}(x_b[i]) + \frac{L_i}{2}\|x[i]-\tilde{x}\|^2
\end{aligned}$$

□

Lemma 3.14. *We have*

$$A^+\left(\frac{1}{b}\sum_{i=1}^{b}\phi(y[i])\right) \le A\phi(y) + a\left(\frac{1}{b}\sum_{i=1}^{b}\tilde{\gamma}\left(x_b[i]\right)\right) + \frac{1}{2b}\|\hat{x}-x\|_L^2.$$

Lemma 3.15. *If h is an indicator function, we have*

$$\begin{aligned}
A^+E_{(x,y)}\phi(y^+) &\le A\phi(y) + a\tilde{\gamma}(\hat{x}) + \frac{1}{2b}\|\hat{x}-x\|_L^2 \\
&\le A\phi(y) + a\tilde{\gamma}(u) + \frac{1}{2b}\|u-x\|_L^2 - \frac{1}{2b}\|u-\hat{x}\|_L^2 \\
&\le A\phi(y) + a\phi(u) + \frac{1}{2b}\|u-x\|_L^2 - \frac{1}{2b}\|u-\hat{x}\|_L^2.
\end{aligned}$$

Lemma 3.16. *For all $u \in \mathbb{R}^n$ we have*

$$E_x\left(\|x-u\|_\eta^2 - \|x^+-u\|_\eta^2\right) = \|x-u\|_{\eta p}^2 - \|\hat{x}-u\|_{\eta p}^2.$$

Lemma 3.17. *Let $\eta = L/p$. Then, for all $u \in \mathbb{R}^n$,*

$$A^+E\left[\phi(y^+)-\phi(u)\right] + \frac{1}{2b}\|u-x^+\|_\eta^2 \le AE\left[\phi(y)-\phi(u)\right] + \frac{1}{2b}\|u-x\|_\eta^2.$$

Proposition 3.1. *Let $\eta = L/p$. Then for all $k \ge 0$ we have*

$$A_kE\left[\phi(y_k)-\phi_*\right] + \frac{1}{2b}\|x^*-x^k\|_\eta^2 \le A_0\left[\phi(y_0)-\phi_*\right] + \frac{1}{2b}\|x^*-x_0\|_\eta^2.$$

For $p = 1/b$ and $\eta = bL$ we have

$$E\left[\phi(y_k)-\phi_*\right] \le \frac{1}{2bA_k}\|x^*-x_0\|_{bL}^2 = \frac{1}{2bA_k}\sum_{i=1}^{b}(bL_i)\|x_i^*-x_i^0\|^2.$$

Since $A_k \ge k^2/(4b^2)$ we have

$$E\left[\phi(y_k)-\phi_*\right] \le \frac{2b^2}{k^2}\sum_{i=1}^{b}\|x_i^*-x_i^0\|^2 \le \frac{2b^2L_{\max}}{k^2}\|x^0-x^*\|^2.$$

Recursive Method

(0) Given $x \in \operatorname{dom} h$, set $y = x$ and $A > 0$ (to be determined later)

(1) Choose $i \in \{1, ..., b\}$ randomly. Compute a by solving

$$\frac{a^2}{A + a} = \frac{1}{b^2},$$

set

$$A^+ = A + a, \quad \tilde{x} = \frac{Ay + ax}{A^+},$$

compute

$$x^+ = x[i], \quad y^+ = \tilde{x} + \frac{1}{ba}(x[i] - x) = \tilde{x}_k + \frac{ba}{A^+}(x[i] - x)$$

where

$$\begin{aligned} x[i] &= x + U_i(\hat{x}_i - x_i), \\ \hat{x}_i &= \operatorname{argmin}_u \left(a \left[\langle \nabla_i f(\tilde{x}_i), u - \tilde{x}_i \rangle + h_i(u) \right] + \frac{L_i}{2b} \|u - x_i\|^2 \right), \\ y[i] &= \tilde{x} + \frac{1}{ba} \left(x[i] - x \right) = \tilde{x} + \frac{ba}{A^+} \left(x[i] - x \right). \end{aligned}$$

Obs: We have the relationships / definitions:

$$\begin{aligned} y[i] &= \frac{A}{A^+} y + \frac{a}{A^+} x_b[i], \\ x_b[i] &:= x + b \left(x[i] - x \right), \\ y^+ &= \frac{Ay}{A^+} + \frac{a}{A^+} \left[x + b(x^+ - x) \right], \\ A^+ y^+ - Ay &= bax^+ - (b-1)ax. \end{aligned}$$

Summing up the last relationship, we have

$$\begin{aligned} A_k y_k - A_0 y_0 &= \sum_{\ell=0}^{k-1} \left[ba_\ell x_{\ell+1} - (b-1) a_\ell x_\ell \right] \\ &= ba_{k-1} x_k - (b-1) a_0 x_0 + \sum_{\ell=1}^{k-1} \left[ba_{\ell-1} - (b-1) a_\ell \right] x_0. \end{aligned}$$

Choose $A_0 \geq (b-1)a_0$, and so

$$A_k y_k = ba_{k-1} x_k + \sum_{\ell=1}^{k-1} \left[ba_{\ell-1} - (b-1) a_\ell \right] x_\ell.$$

Lemma 3.18. *For all ℓ we have $ba_{\ell-1} - (b-1)a_\ell \geq 0$.*

Proof. Want to show:

$$ba^- - (b-1)a \geq 0.$$

We have

$$ba^- = \frac{A}{a^- b}, \quad a = \frac{A^+}{b^2 a},$$

so that if we choose $a \geq a_0 \geq 1/b$ we have

$$\begin{aligned}
ba^- - (b-1)a &= \frac{A}{a^- b} - \frac{(b-1)A^+}{b^2 a} \\
&\geq \frac{1}{a}\left[\frac{A}{b} - \frac{(b-a)A^+}{b^2}\right] \\
&= \frac{1}{ab^2}\left[Ab - (b-1)(A+a)\right] \\
&= \frac{1}{ab^2}\left[A - (b-1)a\right] \\
&= \frac{1}{ab^2}\left[A^+ - ba\right] \\
&= \frac{1}{b^2}\left[\frac{A^+}{a} - b\right] \\
&= \frac{1}{b^2}\left[b^2 a - b\right] \\
&= \frac{1}{b}\left[ba - 1\right] \geq 0.
\end{aligned}$$

Def: The sequence $\{\hat{h}(y_k)\}_{k \geq 0}$ is defined by

$$\begin{aligned}
\hat{h}(y_0) &= h(y_0) = h(x_0), \\
\hat{h}(y_{\ell+1}) &= \frac{A_\ell}{A_{\ell+1}}\hat{h}(y_\ell) + \frac{a_\ell}{A_{\ell+1}}\left[h(x_\ell) + b\left[h(x_{\ell+1}) - h(x_\ell)\right]\right], \\
\hat{h}(y_k[i]) &= \frac{A_k}{A_{k+1}}\hat{h}(y_k) + \frac{a_k}{A_{k+1}}\left[h(x_k) + b\left[h(x_k[i]) - h(x_k)\right]\right].
\end{aligned}$$

□

Lemma 3.19. *The following hold:*

(1) for every $\ell \geq 0$ we have $\hat{h}(y_{\ell+1}) \geq h(y_{\ell+1})$;

(2) $\hat{h}(y_k[i]) \geq h(y_k[i])$.

Proof. It is possible to show that

$$\begin{aligned}
&A_k \hat{h}(y_k) \\
&\geq b a_k h(x_k) + \sum_{\ell=1}^{k-1}\left[b a_{\ell-1} - (b-1)a_\ell\right] h(x_\ell) + (A_0 - (b-1)a_0) \\
&\geq A_k h(y_k).
\end{aligned}$$

□

Define $\hat{\phi}(y_\ell) = f(y_\ell) + \hat{h}(y_\ell)$ and remark that

$$\begin{aligned}
&A^+\hat{\phi}(y[i]) \\
&\le A^+\left(f(y[i]) + \hat{h}(y[i])\right) \\
&\le A^+\left(\ell_f(y[i];x) + \hat{h}(y[i]) + \frac{L_i}{2}\|y[i]-\tilde{x}\|^2\right) \\
&= A^+\left[\ell_f\left(\frac{A}{A^+}y + \frac{a}{A^+}x_b[i];\tilde{x}\right) + \frac{L_i b^2 a}{2(A^+)^2}\|x[i]-x\|^2 + \right. \\
&\quad \left. \frac{A}{A^+}\hat{h}(y) + \frac{a}{A}\left[h(x) + b\left(h(x[i]) - h(x)\right)\right]\right] \\
&= A\ell_f(y;\tilde{x}) + A\hat{h}(y) + a\ell_f(x_b[i];\tilde{x}) + \frac{L_i}{2}\|x[i]-x\|^2 + a\left[h(x) + b\left(h(x[i] - h(x)\right)\right] \\
&= A\hat{\phi}(y) + a\ell_f(x_b[i];\tilde{x}) + \frac{L_i}{2}\|x[i]-x\|^2 + a\left[h(x) + b\left(h(x[i] - h(x)\right)\right].
\end{aligned}$$

Multiply by $(1/b)$ and sum to get

$$\begin{aligned}
&A^+E_{(x,y)}\left[\hat{\phi}(y^+)\right] \\
&\le A\hat{\phi}(y) + ah(\hat{x}) + a\left[\ell_f(\hat{x};\tilde{x}) + \frac{1}{2b}\|\hat{x}-x\|_L^2\right] \\
&\le A\hat{\phi}(y) + a\left[\ell_f(\hat{x};\tilde{x}) + h(u)\right] + \frac{1}{2b}\|u-x\|_L^2 - \frac{1}{2b}\|u-\hat{x}\|_L^2 \\
&\le A\hat{\phi}(y) + a\phi(u) + \frac{1}{2b}\|u-x\|_L^2 - \frac{1}{2b}\|u-\hat{x}\|_L^2.
\end{aligned}$$

We are using the fact that

$$\begin{aligned}
\frac{1}{b}\sum_{i=1}^{\ell}\left[h(x) + b\left(h(x[i]) - h(x)\right)\right] &= \left[\sum_{i=1}^{\ell} h(x[i])\right] - (b-1)h(x) \\
&= h(\hat{x}) + (b-1)h(x) - (b-1)h(x).
\end{aligned}$$

To conclude, we have

$$A^+E\left[\hat{\phi}(y_{k+1}) - \phi_*\right] + E\left(\frac{1}{2b}\|x^* - x_{k+1}\|_\eta^2\right) \le A_k E\left[\hat{\phi}(y_k) - \phi_*\right] + E_\eta\left(\frac{1}{2b}\|x^* - x_k\|_\eta^2\right)$$

and hence

$$\begin{aligned}
E\left[\phi(y_k) - \phi_*\right] &\le \frac{A_0\left[\phi(y_0) - \phi_*\right] + \|x_0 - x^*\|_{bL}^2}{A_k} \\
&\le \frac{A_0\left[\phi(y_0) - \phi_*\right] + b^2\max\{L_i\}\|x_0 - x^*\|_{bL}^2}{k^2}
\end{aligned}$$

4 Monotone Inclusion Problems

Consider a point-to-set map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and define

$$\operatorname{gr} T = \{(x, v) : v \in T(x)\}.$$

Definition 4.1. T is monotone if

$$\begin{cases} (x,v) \in \operatorname{gr} T \\ (\tilde{x}, \tilde{v}) \in \operatorname{gr} T \end{cases} \implies \langle \tilde{x} - x, \tilde{v} - v \rangle \geq 0.$$

Definition 4.2. T is maximal monotone if T is monotone and

$$\begin{matrix} \tilde{T} \text{ monotone} \\ \operatorname{gr} \tilde{T} \supseteq \operatorname{gr} T \end{matrix} \implies T = \tilde{T} \quad (\operatorname{gr} T = \operatorname{gr} \tilde{T})$$

Example 4.1. (1) Given $f \in \overline{\text{Conv}}(\mathbb{R}^n)$ and the optimization problem

$$(*) \quad \min\{f(z) : z \in \mathbb{R}^n\},$$

we have $T = \partial f$ is maximal monotone. Also, $\bar{z}$ is an optimal solution of $(*) \iff 0 \in \partial f(\bar{z})$.
(2) $0 \neq C \subseteq \mathbb{R}^n$ closed convex. Then $N_C(\cdot) = \partial \delta_C$ is maximal monotone where

$$N_C(z) = \{n : \langle n, \tilde{z} - z \rangle \leq 0, \forall \tilde{z} \in C\}.$$

(3) $C \subseteq \mathbb{R}^n, D \subseteq \mathbb{R}^m$ nonempty convex sets. The functions $K : C \times D \mapsto \mathbb{R}$ is **closed convex-concave** if $\forall (x, y) \in C \times D$ we have

$$K(\cdot, y) \in \overline{\text{Conv}}\,(C), \quad -K(x, \cdot) \in \overline{\text{Conv}}\,(D),$$

or equivalently

$$K(\cdot, y) - K(x, \cdot) \in \overline{\text{Conv}}\,(C \times D).$$

Proposition 4.1. *Define* $T : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ *as*

$$T(x,y) = \begin{cases} \partial_x K(x,y) \times \partial_y (-K)(x,y), & \textit{if } (x,y) \in C \times D \\ \varnothing, & \textit{otherwise.} \end{cases}$$

where

$$\partial\,[K(\cdot, y)]\,(x) = \partial_x K(x,y), \quad \partial\,[K(x, \cdot)]\,(y) = \partial_y K(x,y).$$

Define the (respective) primal and dual functions

$$\inf_{x \in C} \underbrace{\sup_{y \in D} K(x,y)}_{p(x)} = \inf_{x \in C} p(x) \to \bar{X},$$

$$\sup_{y \in D} \underbrace{\inf_{x \in C} K(x,y)}_{d(y)} = \inf_{y \in D} d(y) \to \bar{Y}.$$

Proposition 4.2. *The following are equivalent:*

(a) $0 \in T(\bar{x}, \bar{y})$,

(b) $\bar{x} \in \operatorname{argmin}_{x \in C} K(x, \bar{y})$ *and* $\bar{y} \in \operatorname{argmax}_{y \in D} K(\bar{x}, y)$,

(c) $K(\bar{x}, y) \le K(\bar{x}, \bar{y}) \le K(x, \bar{y})$ *for all* $(x, y) \in C \times D$,

(d) $p(\bar{x}) = d(\bar{y})$,

(e) $\bar{x} \in \bar{X}, \bar{y} \in \bar{Y}$, *and* $\bar{p} = \bar{d}$.

Example 4.2. (4) Consider the optimization problem

$$\min\{f(x) : g(x) \le 0, x \in X\}$$

which has the equivalent formulations

$$\min_{x \in X} \max_{y \ge 0} [f(x) + \langle y, g(x) \rangle], \quad \max_{y \ge 0} \min_{x \in X} [f(x) + \langle y, g(x) \rangle].$$

Let $C = X, D = \mathbb{R}^m_+$, and $K : X \times \mathbb{R}^m_+ \mapsto \mathbb{R}$ given by

$$K(x, y) = f(x) + \langle y, g(x) \rangle.$$

Now, if $f_i, g_i \in \overline{\text{Conv}}\,(X)$ then $K(\cdot, y) \in \overline{\text{Conv}}\,(X)$ and $K(x, \cdot) \in \overline{\text{Conv}}\,(\mathbb{R}^m)$, and also

$$\begin{aligned} \partial_x K(x, y) &= \partial f(x) + \sum y_i \partial g_i(x) + N_X(x), \\ \partial_y(-K)(x, y) &= -g(x) + N_{\mathbb{R}^m_+}(y). \end{aligned}$$

(5) Suppose $\varnothing \neq C \subseteq \mathbb{R}^n$ closed convex and $F : C \mapsto \mathbb{R}^n$ continuous monotone. Then $F + N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone. We have

$$\begin{aligned} 0 \in T(\bar{z}) &\iff 0 \in F(\bar{z}) + N_C(\bar{z}) \\ &\iff -F(\bar{z}) \in N_C(\bar{z}) \\ &\iff \langle -F(\bar{z}), z - \bar{z} \rangle \le, \forall z \in C. \end{aligned}$$

If $C = \mathbb{R}^n_+$, then it reduces to

$$F(\bar{x}) \ge 0, \quad \bar{x} \ge 0, \quad \langle \bar{x}, F(\bar{x}) \rangle = 0.$$

If $C = \mathbb{R}^n$ the it reduces to

$$F(\bar{x}) = 0.$$

4.1 Proximal Point Method

Note that given $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ maximal monotone, we have

$$\begin{aligned} 0 \in T(z) &\iff 0 \in \lambda T(z) \\ &\iff z \in z + \lambda T(z) \\ &\iff z \in (I + \lambda T)(z) \\ &\iff z = (I + \lambda T)^{-1}(z) \end{aligned}$$

for $\lambda > 0$, where the existence of the inverse is due to Minty. Here, we are saying that if $S = I + \lambda T$, then for $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ we have $S^{-1} : \mathbb{R}^n \mapsto \mathbb{R}^n$ where

$$S^{-1}(y) = \{x : y \in S(x)\},$$

which follows from the following result.

Lemma 4.1. *Assume that T is monotone, $\lambda > 0$, $z_1 \in (I + \lambda T)(w_1)$, and $z_2 \in (I + \lambda T)(w_2)$. Then,*

$$\langle z_1 - z_2, w_1 - w_2 \rangle \geq \|w_1 - w_2\|^2.$$

As a consequence, $\forall z \in \mathbb{R}^n$, there exists at most one w such that

$$(I + \lambda T)(w) = z.$$

Proposition 4.3. *(Minty) Assume T is monotone and $\lambda > 0$. Then, T is maximal monotone $\iff$ Range $(I + \lambda T) = \mathbb{R}^n \iff \operatorname{dom}(I + \lambda T)^{-1} = \mathbb{R}^n$.*

Corollary 4.1. *Under the above assumptions, $F = (I + \lambda T)^{-1} : \mathbb{R}^n \mapsto \mathbb{R}^n$ satisfies*

$$\langle F(z) - F(\tilde{z}), z - \tilde{z} \rangle \geq \|F(z) - F(\tilde{z})\|^2$$

for every $z, \tilde{z}$. Observe that the above implies

$$\|F(z) - F(\tilde{z})\| \leq \|z - \tilde{z}\|.$$

Proximal Point Method (PPM)

Iterate $z_{k+1} = F(z_k)$ until convergence to z^*.

Lemma 4.2. *For all $k \geq 0$ we have*

$$\langle z_{k+1} - z^*, z_k - z^* \rangle \geq \|z_{k+1} - z^*\|^2$$

or

$$\langle z_{k+1} - z^*, z_k - z_{k+1} \rangle \geq 0.$$

Lemma 4.3. *For all $k \geq 0$ we have*

$$\|z_k - z^*\|^2 \geq \|z_{k+1} - z^*\|^2 + \|z_{k+1} - z_k\|^2$$

and hence

$$\|z_0 - z^*\|^2 \geq \|z_{k+1} - z^*\|^2 + \sum_{i=0}^{k} \|z_{i+1} - z_i\|^2.$$

Proof. We first prove: (1) $\{z_k\}$ is bounded and $\|z_k - z^*\|$ is decreasing.

Assume $z_k \overset{k \in K}{\to} \bar{z}$. Then

$$0 \leftarrow \|F(z_k) - z_k\| = \|z_{k+1} - z_k\| \overset{k \in K}{\to} \|F(\bar{z}) - \bar{z}\| = 0$$

and so $\|z_k - \bar{z}\|$ decreases to 0 along $k \in K$. From a previous lemma,

$$\min_{i \leq k} \|F(x_i) - x_i\| = \|z_{i+1} - z_i\| = \mathcal{O}\left(\frac{d_0}{\sqrt{k}}\right).$$

□

Obs. Note that the PPM, with variable stepsize, is equivalent to

$$x_k \in (I + \lambda_k T)(x_k) \iff \frac{x_{k-1} - x_k}{\lambda_k} \in T(x_k) \iff \frac{x_{k-1} - x_k}{\lambda_k} \in T(x_k), \quad x_k = \tilde{x}_k.$$

Definition 4.3. T^ε denotes the ε–enlargement of T defined as

$$T^\varepsilon(\tilde{x}) = \{\tilde{v} : \langle \tilde{v} - v, \tilde{x} - x\rangle \geq -\varepsilon, \forall (x, v) \in \operatorname{gr} T\}.$$

Properties.

(0) $T^{\varepsilon_1} \subseteq T^{\varepsilon_2}$ if $\varepsilon_1 \subseteq \varepsilon_2$

(1) T monotone $\implies T \subseteq T^0$

(2) T maximal monotone $\iff T = T^0$

(3) If $T = \partial f$, $f \in \overline{\text{Conv}}(\mathbb{R}^n)$, then $\partial_\varepsilon f \subseteq T^\varepsilon$

(4) if $T(x,y) = \partial_x K(x,y) \times \partial_y(-K)(x,y)$ where K is a closed convex-concave function on $C \times D$ then

$$\partial_\varepsilon[K(\cdot, y) - K(x, \cdot)](x,y) \subseteq T^\varepsilon(x,y)$$

Inexact PPM

Given (x_{k-1}, λ_k), compute $(x_k, \tilde{x}_k, \varepsilon_k)$ satisfying

$$\frac{x_{k-1} - x_k}{\lambda_k} \in T^{\varepsilon_k}(\tilde{x}_k), \quad \|x_k - \tilde{x}_k\|^2 + 2\lambda_k\varepsilon_k \leq \sigma^2\|\tilde{x}_k - x_{k-1}\|^2.$$

Lemma 4.4. *For $\tilde{v} \in T^\varepsilon(\tilde{x})$ define $\Gamma(u) = \langle \tilde{v}, u - \tilde{x}\rangle - \varepsilon$ for all u. Then*

$$\Gamma(x^*) \leq 0, \quad \forall x^* \in T^{-1}(0).$$

Proof. We have

$$\tilde{v} \in T^\varepsilon(\tilde{x}), \quad (x^*, 0) \in \operatorname{gr} T \implies \langle \tilde{v} - 0, \tilde{x} - x^*\rangle \geq -\varepsilon \iff \Gamma(x^*) \leq 0.$$

□

Lemma 4.5. *Assume that $(x_0, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++}$ and $(x, \tilde{x}, \varepsilon)$ satisfies $\tilde{v} := \frac{x_0 - x}{\lambda} \in T^\varepsilon(\tilde{x})$ and $\|\tilde{x} - x\|^2 + 2\lambda\varepsilon \leq \sigma^2\|\tilde{x} - x_0\|^2$. Then,*

(a) $x = \operatorname{argmin}\left[\lambda\Gamma(u) + \frac{1}{2}\|u - x_0\|^2\right]$

(b) $\min\left[\lambda\Gamma(u) + \frac{1}{2}\|u - x_0\|^2\right] \geq \frac{1-\sigma^2}{2}\|\tilde{x} - x_0\|^2$

where $\Gamma(u) = \langle \tilde{v}, u - \tilde{x}\rangle - \varepsilon$.

Proof. (a) Obvious.

(b) For all u,

$$\min\left[\lambda\Gamma(u) + \frac{1}{2}\|u - x_0\|^2\right] = \lambda\Gamma(u) + \frac{1}{2}\|u - x_0\|^2 - \frac{1}{2}\|u - x\|^2$$

so take $u = \tilde{x}$ to get

$$\begin{aligned}
&\min\left[\lambda\Gamma(u) + \frac{1}{2}\|u - x_0\|^2\right] \\
&\geq \frac{1}{2}\left(\|\tilde{x} - x_0\|^2 - \|\tilde{x} - x\|^2\right) + \lambda\Gamma(\tilde{x}) \\
&= \frac{1}{2}\left(\|\tilde{x} - x_0\|^2 - \|\tilde{x} - x\|^2 - 2\lambda\varepsilon\right) \\
&\geq \frac{1}{2}\left(\|\tilde{x} - x_0\|^2 - \sigma\|\tilde{x} - x_0\|^2\right).
\end{aligned}$$

□

Lemma 4.6. *Under the assumptions of the previous lemma,*

$$\|u - x_0\|^2 \geq \|u - x\|^2 + (1 - \sigma^2)\|\tilde{x} - x_0\|^2 - 2\lambda\Gamma(u), \quad \forall u \in \mathbb{R}^n.$$

Proof. We have

$$\begin{aligned}
&\lambda\Gamma(u) + \frac{1}{2}\|u - x_0\|^2 \\
&= \lambda\Gamma(x) + \frac{1}{2}\|x - x_0\|^2 + \frac{1}{2}\|u - x\|^2 \\
&\geq \frac{1 - \sigma^2}{2}\|\tilde{x} - x_0\|^2 + \frac{1}{2}\|u - x\|^2.
\end{aligned}$$

□

Lemma 4.7. *For all $x^* \in T^{-1}(0)$ and $k \geq 1$,*

$$\|x^* - x_{k-1}\|^2 \geq \|x^* - x_k\|^2 + (1 - \sigma^2)\|\tilde{x}_k - x_{k-1}\|^2$$

and also

$$\|u - x_{k-1}\|^2 \geq \|u - x_k\|^2 + (1 - \sigma^2)\|\tilde{x}_k - x_{k-1}\|^2 - 2\lambda_k\Gamma_k(u).$$

Proposition 4.4. *We have*

$$\|u - x_0\|^2 \geq \|u - x_k\|^2 + (1 - \sigma^2)\sum_{i=1}^{k}\|\tilde{x}_i - x_{i-1}\|^2 - 2\sum_{i=1}^{k}\lambda_i\Gamma_i(u)$$

and

$$\|x^* - x_0\|^2 \geq \|x^* - x_k\|^2 + (1 - \sigma^2)\sum_{i=1}^{k}\|\tilde{x}_i - x_{i-1}\|^2.$$

<u>Criterion</u>

Given $(\bar{\rho}, \bar{\varepsilon}) > 0$, find $(\tilde{x}, \tilde{v}, \varepsilon)$ such that

$$\tilde{v} \in T^{\varepsilon}(\tilde{x}), \quad \|\tilde{v}\| \leq \bar{\rho}, \quad \varepsilon \leq \bar{\varepsilon}.$$

Have

$$\tilde{v}_k = \frac{x_{k-1} - x_k}{\lambda_k} \in T^{\varepsilon_k}(\tilde{x}_k), \quad (\tilde{x}, \tilde{v}, \varepsilon) = (\tilde{x}_k, \tilde{v}_k, \varepsilon_k).$$

Lemma 4.8. *For all $k \geq 1$,*

$$(1-\sigma)\|\tilde{x}_k - x_{k-1}\|^2 \leq \|\lambda_k \tilde{v}_k\| \leq (1+\sigma)\|\tilde{x}_k - x_{k-1}\| + 2\lambda_k\varepsilon_k \leq \sigma^2\|\tilde{x}_k - x_{k-1}\|^2.$$

Obs. Define

$$\theta_k := \max\left\{\frac{2\lambda_k\varepsilon_k}{\sigma^2}, \frac{\lambda^2\|\tilde{v}_k\|^2}{(1+\sigma)^2}\right\} \leq \|\tilde{x}_k - x_{k-1}\|^2.$$

By a previous proposition,

$$\sum_{i=1}^k \theta_i \leq \frac{\|x_0 - x^*\|^2 - \|x_1 - x^*\|^2}{1-\sigma^2}$$

and also $\sum_{i=1}^k \theta_i \geq k \min_{i\leq k} \theta_i$. Hence,

$$\min_{i\leq k} \theta_i \leq \frac{d_0^2}{k(1-\sigma^2)}$$

and so for every k there exists i such that

$$\varepsilon_i \leq \frac{\sigma^2 d_0^2}{2k(1-\sigma^2)\lambda_i}, \quad \|v_i\|^2 \leq \frac{(1+\sigma)^2 d_0^2}{k(1-\sigma^2)\lambda_i^2}.$$

Properties.

(1) $\{x_k\}$ is bounded.

(2) $\{\tilde{x}_k\}$ is bounded.

(3) there exists a subsequence $\{(\tilde{x}_k, \tilde{v}_k, \varepsilon_k)\} \overset{k\in K}{\to} (\bar{x}, 0, 0)$ such that

$$\tilde{v}_k \in T^{\varepsilon_k}(\tilde{x}_k), \quad 0 \in T^0(\bar{x}) = T(\bar{x})$$

and so $\bar{x} \in T^{-1}(0)$.

(4) $\|x_k - \bar{x}\| \downarrow$ and $\|x_k - \bar{x}\| \overset{k\in K}{\to} 0$.

(5) $x_k \to \bar{x} \implies \tilde{x}_k \to \bar{x}$.

<u>IPP Framework (again)</u>

For $0 \in T(x)$ and T maximal monotone.

(0) Given x_0 and $\sigma \in (0,1)$, set $k = 1$;

(1) choose $\lambda_k > 0$ and compute $(x_k, \tilde{x}_k, \varepsilon_k)$ satisfying

$$\frac{x_{k-1} - x_k}{\lambda_k} \in T^{\varepsilon_k}(\tilde{x}_k), \quad \|x_k - \tilde{x}_k\|^2 + 2\lambda_k\varepsilon_k \leq \sigma^2\|\tilde{x}_k - x_{k-1}\|^2;$$

(2) set $k \leftarrow k+1$ and go to (1).

We showed the following results about $v_k := (x_{k-1} - x_k)/\lambda_k$.

- $\min_{i\leq k} \max\{\|v_i\|^2, \varepsilon_i\} \leq \mathcal{O}(1/k)$
- $x_k \to x^* \in T^{-1}(0)$ and $\tilde{x}_k \to x^* \in T^{-1}(0)$

Proposition 4.5. *For all $k \geq 1$ and $u \in \mathbb{R}^n$ we have*

$$\|u - x_0\|^2 \geq \|u - x_k\|^2 + (1 - \sigma^2) \sum_{i=1}^{k} \|\tilde{x}_i - x_{i-1}\|^2 - 2 \sum_{i=1}^{k} \lambda_i \Gamma_i(u)$$

where

$$\Gamma_i(u) = \varepsilon_i + \langle v_i, u - \tilde{x}_i \rangle .$$

Obs. (1) $\Gamma_i(x^*) \leq$ and $\Gamma_i(\tilde{x}_i) = \varepsilon_i$.

Proposition 4.6. *Let $u_i \in T^{\varepsilon_i}(y_i)$ and $\theta_i \geq 0$ for $i = 1, \ldots, k$ such that $\sum_{i=1}^{k} \theta_i = 1$. Let*

$$u^a = \sum_{i=1}^{k} \theta_i u_i, \quad y^a = \sum_{i=1}^{k} \theta_i y_i,$$

and

$$\varepsilon^a = \sum_{i=1}^{k} \theta_i \left[\varepsilon_i + \langle u_i - u^a, y_i - y^a \rangle \right].$$

Then $u^a \in T^{\varepsilon^a}(y^a)$ and $\varepsilon^a \geq 0$.

Proof. Have for every $i = 1, \ldots, k$:

$$\langle u_i - v, y_i - x \rangle \geq -\varepsilon_i, \quad \forall (x, v) \in \operatorname{gr} T.$$

So, for $(x, v) \in \operatorname{gr} T$, we have

$$\begin{aligned}
&\langle u^a - v, y^a - x \rangle + \varepsilon^a \\
&= \langle u^a - v, y^a - x \rangle + \sum_{i=1}^{k} \theta_i \left[\varepsilon_i + \langle u_i - u^a, y_i - y^a \rangle \right] \\
&= \sum_{i=1}^{k} \theta_i \Big[\langle u_i - v, y^a - x \rangle + \langle u_i - u^a, y_i - y^a \rangle + \varepsilon_i \\
&\qquad \langle u_i - v, y_i - y^a \rangle + \underbrace{\langle v - v^a, y_i - y^a \rangle}_{\text{sum is } 0} \Big] \\
&= \sum_{i=1}^{k} \theta_i \left(\varepsilon_i + \langle u_i - v, y_i - x \rangle \right) \geq 0
\end{aligned}$$

□

Proposition 4.7. *For $\lambda_i \geq 0$ and $\Lambda_k := \sum_{i=1}^{k} \lambda_i$ define*

$$\tilde{x}_k^a = \frac{\sum_{i=1}^{k} \lambda_i \tilde{x}_i}{\Lambda_k}, \quad v_k^a = \frac{\sum_{i=1}^{k} \lambda_i v_i}{\Lambda_k} = \frac{x_0 - x_k}{\Lambda_k},$$

and

$$\varepsilon_k^a = \frac{\sum_{i=1}^{k} \lambda_i \left[\varepsilon_i + \langle v_i, \tilde{x}_i - \tilde{x}_k^a \rangle \right]}{\Lambda_k} = \frac{-\sum_{i=1}^{k} \lambda_i \Gamma_i(\tilde{x}_k^a)}{\Lambda_k}.$$

Then $v_k^a \in T^{\varepsilon_k^a}(\tilde{x}_k^a)$.

Proposition 4.8. $\|v_k^a\| \leq 2d_0/\Lambda_k$ *where* $d_0 = \min\{\|x_0 - x^*\| : x^* \in T^{-1}(0)\}$.

Proof. We have

$$\Lambda_k v_k^a = \sum_{i=1}^k \lambda_i v_i = \sum_{i=1}^k (x_{i-1} - x_i).$$

So, we have

$$\lambda_k \|v_k^a\| \leq \|x_0 - x_k\| \leq \|x_0 - x^*\| + \|x_k - x^*\| \leq 2\|x_0 - x^*\| = 2d_0.$$

□

Proposition 4.9. *For all* $k \geq 0$ *we have*

$$\varepsilon_k^a \leq \frac{\|\tilde{x}_k^a - x_0\|^2}{2\Lambda_k} \leq \frac{\left(2 + \frac{\sigma}{\sqrt{1-\sigma^2}}\right)^2 d_0}{2\Lambda_k}.$$

Proof. We have , from a previous proposition,

$$\Lambda_k \varepsilon_k^a = -\sum_{i=1}^k \lambda_i \Gamma_i(x_k^a) \leq \frac{1}{2}\|x_k^a - x_0\|^2.$$

Now note

$$\begin{aligned} \|x_k^a - x_0\| &\leq \max_{i \leq k} \|\tilde{x}_i - x_0\| \\ &\leq \max_{i \leq k} \left(\|\tilde{x}_i - x_i\| + \|x_i - x_0\|\right) \\ &\leq 2d_0 + \max_{i \leq k} \|\tilde{x}_i - x_i\|. \end{aligned}$$

If $\sigma < 1$ then

$$\max_{i \leq k} \|\tilde{x}_i - x_i\| \leq \sigma \max_{i \leq k} \|\tilde{x}_i - x_{i-1}\| \leq \frac{\sigma d_0}{\sqrt{1 - \sigma^2}}$$

from a previous proposition. Hence

$$\|x_k^a - x_0\| \leq \left(2 + \frac{\sigma}{\sqrt{1 - \sigma^2}}\right) d_0.$$

□

Remark 4.1. We must have $\varepsilon^a \geq 0$ by the following argument.

Assume $\varepsilon^a < 0$. Have $\langle u^a - v, y^a - x\rangle \geq -\varepsilon^a$ for all $(x, v) \in \operatorname{gr} T$. Let $\operatorname{gr} \tilde{T} = \operatorname{gr} T \cup \{(u^a, y^a)\}$. We have $\tilde{T} \neq T$ which contradicts the maximality of T since $\tilde{T}$ is monotone.

Note that this show that if $\operatorname{gr} T^\varepsilon \neq \varnothing$ then $\varepsilon \geq 0$.

Variational Inequalities:

Given $F : X \mapsto \mathbb{R}^n$ monotone continuous satisfying

$$\|F(x) - F(x')\| \leq L\|x - x'\| \quad \forall x, x' \in X,$$

and $\varnothing \neq X \subseteq \mathbb{R}^n$ we want to:

Find $x^* \in X$ such that

$$\begin{aligned} & \langle F(x^*), x - x^* \rangle \geq 0, \quad \forall x \in X \\ \iff & -F(x^*) \in N_X(x^*) = \partial \delta_X(x^*) \\ \iff & 0 \in (\underbrace{F + N_X}_{T})(x^*) \end{aligned}$$

where T is maximally monotone.

Tseng's B-F Splitting Method

(0) Given $x_0 \in \mathbb{R}^n$, choose $\lambda > 0$ and σ such that $\lambda = \sigma/L$.

(1) Solve for $\tilde{x}$ as

$$F(x_0) + \underbrace{N_X(\tilde{x})}_{\tilde{n} \in} + \frac{1}{\lambda}(\tilde{x} - x_0) \ni 0$$

and set $x = x_0 - \lambda[F(\tilde{x}) + \tilde{n}]$ where $\tilde{n} = (\tilde{x} - x_0)/\lambda + F(x_0)$.

Obs. We have

$$\tilde{x} \leftarrow \min \left\{ \langle F(x_0), u - x_0 \rangle + \frac{1}{2\lambda} \|u - x_0\|^2 : x \in X \right\}$$

and so

$$\tilde{x} = P_X(x_0 - \lambda F(x_0)).$$

IPP Framework Results

Proposition 4.10. *For every $k \geq 0$ we have*

$$\min_{i \leq k} \max\{\|v_i\|^2, \varepsilon_i\} \leq \mathcal{O}(1/k), \quad v_i \in T^{\varepsilon_i}(x_i)$$

and

$$\max\{\|v_k^a\|, \varepsilon_k^a\} \leq \mathcal{O}(1/k), \quad v_k^a \in T^{\varepsilon_k^a}(x_k^a).$$

Composite Korpolevich's Method

We want to get

$$0 \in F(z) + \partial g(z) =: T(z)$$

where $g \in \overline{\text{Conv}}(X)$, $F : X \mapsto \mathbb{R}^n$ is monotone continuous, and F is L-Lipschitz. That is $T(z)$ is maximal monotone.

Obs. $g = \delta_X$ implies the problem of interest is a VIP.

Korpolevich's Method

(0) Given $z_0 \in X$.

(1) Compute $\tilde{z} \in X$ by solving

$$F(z_0) + \partial g(\tilde{z}) + \frac{1}{\lambda}(\tilde{z} - z_0) \ni 0. \quad (a)$$

(2) Compute $z \in X$ by solving

$$F(\tilde{z}) + \partial g(z) + \frac{1}{\lambda}(z - z_0) \ni 0. \quad (b)$$

(3) Set $z_0 \leftarrow z$ and go to (1).

Obs. (a) $\iff$ $\tilde{z} = \operatorname{argmin}_u \left[\langle F(z_0), u\rangle + g(u) + \frac{1}{2\lambda}\|u - z_0\|^2\right]$.

(b) $\iff$ $z = \operatorname{argmin}_u \left[\langle F(\tilde{z}), u\rangle + g(u) + \frac{1}{2\lambda}\|u - z_0\|^2\right]$.

Fact

Assume $p \in \partial g(z)$ and let $\tilde{z} \in \operatorname{dom} g$ and

$$\varepsilon = g(\tilde{z}) - g(z) - \langle p, \tilde{z} - z\rangle .$$

Then, $p \in \partial_\varepsilon(\tilde{z})$.

Proof. Exercise. □

(a) $\iff$ $F(x_0) + \tilde{p} + \frac{1}{\lambda}(\tilde{z} - z_0) = 0, \tilde{p} \in \partial g(\tilde{z})$

(a) $\iff$ $F(x_0) + p + \frac{1}{\lambda}(z - z_0) = 0, p \in \partial g(z)$

So,

$$v = \frac{1}{\lambda}(z_0 - z) = F(\tilde{z}) + p \in F(\tilde{z}) + \partial_\varepsilon g(\tilde{z}) \subseteq (F + \partial g)^\varepsilon(\tilde{z}) = T^\varepsilon(\tilde{z}).$$

We also have

$$\begin{aligned}
\varepsilon &= g(\tilde{z}) - g(z) - \langle p, \tilde{z} - z\rangle \\
&= g(\tilde{z}) - g(z) - \langle \tilde{p}, \tilde{z} - z\rangle + \langle \tilde{p} - p, \tilde{z} - z\rangle \\
&\leq \langle \tilde{p} - p, \tilde{z} - z\rangle .
\end{aligned}$$

Next, note that (a) and (b) imply

$$F(\tilde{z}) - F(z_0) + p - \tilde{p} + \frac{1}{\lambda}(z - \tilde{z}) \ni 0$$

and hence

$$\begin{aligned}
\|\tilde{z} - z\|^2 + 2\lambda\varepsilon &\leq \|\tilde{z} - z\|^2 + 2\lambda \langle \tilde{p} - p, \tilde{z} - z\rangle \\
&\leq \|\tilde{z} - z + \lambda(\tilde{p} - p)\|^2 - \lambda^2\|\tilde{p} - p\|^2 \\
&\leq \|\tilde{z} - z + \lambda(\tilde{p} - p)\|^2 \\
&= \lambda^2\|F(\tilde{z}) - F(z_0)\|^2 \\
&\leq \lambda^2 L^2\|\tilde{z} - z_0\|^2
\end{aligned}$$

and when $\lambda = \sigma/L$ we get

$$\|\tilde{z} - z\|^2 + 2\lambda\varepsilon \leq \sigma^2\|\tilde{z} - z_0\|^2$$

which is the IPP framework.

Complexity

We have

$$v_k = \frac{1}{\lambda}(z_{k-1} - z_k) \in F(\tilde{z}_k) + \partial_{\varepsilon_k} g(\tilde{z}_k)$$

with pointwise convergence

$$\min_{i \leq k} \max\{\|v_i\|^2, \varepsilon_i| \leq \mathcal{O}(1/k).$$

Also note that

$$\tilde{v} = F(\tilde{z}) + \tilde{p} \in (F + \partial g)(\tilde{z})$$

and have

$$\tilde{v} = F(\tilde{z}) - F(z_0) + \frac{1}{\lambda}(z_0 - \tilde{z})$$

so that

$$\|\tilde{v}\| \le \left(L + \frac{1}{\lambda}\right)\|\tilde{z} - z_0\| = (1+\sigma)L\|\tilde{z} - z_0\|.$$

Fact. We have

$$\|\tilde{z} - z_0\| \le \frac{\|z - z_0\|}{1-\sigma} = \frac{\lambda\|v\|}{1-\sigma}.$$

Proof. $\sigma\|\tilde{z} - z_0\| \ge \|\tilde{z} - z\| \ge \|\tilde{z} - z_0\| - \|z_0 - z\|$. □

So

$$\|\tilde{v}\| \le \frac{(1+\sigma)L\lambda\|v\|}{1-\sigma} \implies \min_{i\le k}\|\tilde{v}_i\| \le O\left(\frac{1}{\sqrt{k}}\right).$$

4.2 Saddle-Point Problem

Given $\hat{K} : C \times D \mapsto \mathbb{R}$ closed convex-concave and $C \times D \subseteq \mathbb{R}^n \times \mathbb{R}^m$ convex, define

$$p(x) = \max_{y\in D}\hat{K}(x,y), \quad d(y) = \min_{x\in C}\hat{K}(x,y).$$

Obs: We have $p(x) \ge d(y)$ for all $(x,y) \in C \times D$.

SP → $(x^*, y^*) \in C \times D$ s.t. $p(x^*) = d(y^*)$

ε–SP → $(\bar{x}, \bar{y}) \in C \times D$ s.t. $p(\bar{x}) - d(\bar{y}) = 0$ or equivalently $0 \in \partial_\varepsilon\left[\hat{K}(\cdot,\bar{y}) - \hat{K}(\bar{x},\cdot)\right](\bar{x},\bar{y})$

For $\varepsilon = 0$ this is the problem $0 \in T(z)$ where

$$T(z) = T(x,y) := \partial\left[\hat{K}(\cdot,y) - \hat{K}(x,\cdot)\right](x,y).$$

Smooth Composite Structure

Assume

$$\hat{K}(x,y) = K(x,y) + g_1(x) - g_2(y)$$

where $g_1 \in \overline{\text{Conv}}(C)$, $g_2 \in \overline{\text{Conv}}(C)$, K is a real-valued function which is differentiable on $C \times D$ and ∇K is L–Lipschitz. Here,

$$T(z) = \underbrace{\begin{pmatrix} \nabla_x K(x,y) \\ -\nabla_y K(x,y) \end{pmatrix}}_{=F(z)} + \underbrace{\begin{pmatrix} \partial g_1(x) \\ \partial g_2(x) \end{pmatrix}}_{=\partial g(z)}$$

where $g(z) = g(x,y) = g_1(x) + g_2(y)$. The IPP iteration is

$$\begin{aligned} v_i \in F(z_i) + \partial_{\varepsilon_i} g(z_i) &= \begin{pmatrix} \nabla_x K(\tilde{x}_i, \tilde{y}_i) + \partial_{\varepsilon_i'} g_1(\tilde{x}_i) \\ -\nabla_y K(\tilde{x}_i, \tilde{y}_i) + \partial_{\varepsilon_i'} g_2(\tilde{x}_i) \end{pmatrix} \\ &\subseteq \underbrace{\partial_{\varepsilon_i}\left[\hat{K}(\cdot, \tilde{y}_i) - \hat{K}(\tilde{x}_i, \cdot)\right](\tilde{x}_i, \tilde{y}_i)}_{T[\varepsilon_i](\tilde{x}_i, \tilde{y}_i)} \subseteq T^{\varepsilon_i}(\tilde{x}_i, \tilde{y}_i). \end{aligned}$$

We also have that $v_i \in T[\varepsilon_i](\tilde{x}_i, \tilde{x}_i)$ implies $v_k^a \in T[\varepsilon_k^a](\tilde{x}_k^a, \tilde{y}_k^a)$.

Chambolle-Pock's Algorithm

Consider the problem

$$(P) \quad \min_x \max_y \langle Kx, y\rangle + G(x) - F^*(y)$$

where $G \in \overline{\text{Conv}}(\mathbb{R}^n)$, $F \in \overline{\text{Conv}}(\mathbb{R}^m)$, $K : \mathbb{R}^n \mapsto \mathbb{R}^m$ is linear. The problem (P) is equivalent to

$$\min_x \; F(Kx) + G(x)$$

and has the dual formulation

$$\max_x \min_y \langle Kx, y\rangle + G(x) - F^*(y) = \psi(x,y)$$

or equivalently

$$\max_y -G^*(-K^*y) - F^*(y).$$

Furthermore, let us assume that $\exists (x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$-Kx^* + \partial F^*(y^*) \ni 0, \quad K^*y^* + \partial G(x^*) \ni 0$$

or equivalently

$$(0,0) \in \partial\left[\psi(\cdot, y^*) - \psi(x^*, \cdot)\right](x^*, y^*)$$

Algorithm Description

(0) Choose $\tau_1, \tau_2 > 0$, $\theta = 1$, $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$ and set $\bar{x}^0 = x^0$ and $k = 0$;

(1) compute

$$\begin{aligned} y^{k+1} &= (I + \tau_1 \partial F^*)^{-1}(y^k + \tau_2 K\bar{x}^k), \\ x^{k+1} &= (I + \tau_1 \partial G)^{-1}(x^k + \tau_1 K^* y^{k+1}), \\ \bar{x}^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k). \end{aligned}$$

(2) set $k \leftarrow k+1$ and go to (1).

Facts

We have

$$\begin{aligned} \frac{x^{k+1} - x^k}{\tau_1} + K^* y^{k+1} + \partial G(x^{k+1}) \ni 0, \\ \frac{y^{k+1} - y^k}{\tau_2} + K^* \bar{x}^k + \partial F^*(y^{k+1}) \ni 0. \end{aligned}$$

IPP Framework

The above algorithm is an instance of the following framework.

Proposition 4.11. *The CP algorithm is an instance of the following IPP framework as long as* $\|K\|^2\tau_1\tau_2 \leq \sigma^2$. *Given* (x_k, y_k) *and* $\lambda_{k+1} > 0$, *find* $(x_{k+1}, y_{k+1}), (\tilde{x}_{k+1}, \tilde{y}_{k+1})$ *s.t.*

$$\frac{x_{k+1} - x_k}{\lambda_{k+1}} + \tau_1 \left[K^* \tilde{y}_{k+1} + \partial G(\tilde{x}_{k+1})\right] \ni 0, \tag{a}$$

$$\frac{y_{k+1} - y_k}{\lambda_{k+1}} + \tau_2 \left[-K \tilde{x}_{k+1} + \partial F^*(\tilde{y}_{k+1})\right] \ni 0. \tag{b}$$

We also have, for some $\sigma \in (0,1)$, *the inequality*

$$\frac{1}{\tau_1}\|x_{k+1} - \tilde{x}_{k+1}\|^2 + \frac{1}{\tau_2}\|y_{k+1} - \tilde{y}_{k+1}\|^2 + 2\lambda_{k+1}\varepsilon_{k+1} \leq \sigma \left[\|x_{k+1} - \tilde{x}_k\|^2 + \frac{1}{\tau_2}\|y_{k+1} - \tilde{y}_k\|^2\right]. \tag{c}$$

Proof. Take $\lambda_{k+1} = 1$, $\varepsilon_{k+1} = 0$, and

$$x_{k+1} = \tilde{x}_{k+1} = x^{k+1}, \quad \tilde{y}_{k+1} = y^{k+1},$$
$$y_{k+1} = y^{k+1} + \tau_2 K(\bar{x}^{k+1} - x^{k+1}).$$

The proof of (a) is straightforward. For (b), we have that the right-hand-side of (b) is

$$\begin{aligned}
&y^{k+1} + \tau_2\left[K(\bar{x}^{k+1} - x^{k+1})\right] - y^k - \tau_2\left[K(\bar{x}^k - x^k)\right] \\
&+ \tau_2\left[-Kx^{k+1} + \partial F^*(y^{k+1})\right] \\
&= \tau_2\left[\frac{y_{k+1} - y^k}{\tau_2} + K\left(\bar{x}^{k+1} - x^{k+1} - \bar{x}^k + x^k - x^{k+1}\right) + \partial F^*(y^{k+1})\right] \\
&= \tau_2\left[\frac{y_{k+1} - y^k}{\tau_2} + K\bar{x}^k + \partial F^*(y^{k+1})\right]
\end{aligned}$$

which contains 0 by step 1 of the CP algorithm. We now prove the inequality. We observe that

$$\begin{aligned}
\text{(c)} \iff\ & \frac{1}{\tau_2}\|\tau_2 K(\bar{x}^{k+1} - x^{k+1})\|^2 \leq \sigma\left[\frac{1}{\tau_1}\|x^{k+1} - x^k\|^2 + \frac{1}{\tau_2}\|y^{k+1} - \tilde{y}^k\|^2\right] \\
\iff\ & \|K(\bar{x}^{k+1} - x^{k+1})\|^2 \leq \sigma\left[\frac{1}{\tau_1\tau_2}\|x^{k+1} - x^k\|^2 + \frac{1}{\tau_2^2}\|y^{k+1} - y^k\|^2\right] \\
\Longleftarrow\ & \|K\|^2\|\bar{x}^{k+1} - x^{k+1}\|^2 \leq \frac{\sigma^2}{\tau_1\tau_2}\|x^{k+1} - x^k\|^2 \\
\iff\ & \|K\|^2\|x^{k+1} - x^k\|^2 \leq \frac{\sigma^2}{\tau_1\tau_2}\|x^{k+1} - x^k\|^2 \\
\iff\ & \|K\|^2\tau_1\tau_2 \leq \sigma^2.
\end{aligned}$$

Gauss-Siedel

(0) Given x_0 and y_0.

(1) Find $\tilde{y}$ such that $-Kx_0 + \partial F^*(x) + (\tilde{y} - y_0)/\tau_2 \ni 0$.

(2) Find x such that $-K^*\tilde{y} + \partial G(x) + (x - x_0) \ni 0$.

(3) Get y satisfying $(y - y_0)/\tau_2 - Kx + \partial F^*(\tilde{y}) \ni 0$ □

Claim. CP algorithm is Gauss-Siedel.

Proof? Take $\bar{x}^k = x_0, y^{k+1} = \tilde{y}$,... ???

4.3 ADMM

Consider the problem

$$\begin{aligned} \min f(x) + g(x) \\ \text{s.t. } Ax + By = 0 \in \mathbb{R}^r \end{aligned}$$

where $f \in \overline{\text{Conv}}(\mathbb{R}^n)$ and $g \in \overline{\text{Conv}}(\mathbb{R}^m)$. Define

$$L_p(x, y; \lambda) = f(x) + g(x) + \lambda^T(Ax + By) + \frac{\rho}{2}\|Ax - By\|^2.$$

Augmented Lagrangian Method (ALM)

(0) Given $\lambda_0 \in \mathbb{R}^r$.

(1) Solve $(x, y) \in \operatorname{argmin}_{(x',y')} L_p(x', y'; \lambda_0)$.

(2) Set $\lambda = \lambda_0 + \rho(Ax + By)$.

Optimality Conditions

We have

$$\begin{aligned} \partial f(x) + A^*(\lambda_0 + \rho(Ax + By)) \ni 0, \\ \partial g(y) + B^*(\lambda_0 + \rho(Ax + By)) \ni 0, \\ -Ax - By + \frac{\lambda - \lambda_0}{\rho} = 0, \end{aligned}$$

or equivalently

$$\partial f(x) + A^*\lambda \ni 0, \tag{1}$$

$$\partial g(y) + B^*\lambda \ni 0, \tag{2}$$

$$-Ax - By + \frac{\lambda - \lambda_0}{\rho} = 0. \tag{3}$$

Consider

$$T(z) = T(x, y, \lambda) := \begin{bmatrix} 0 & 0 & A^* \\ 0 & 0 & B^* \\ -A & -B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \partial f(x) \\ \partial g(y) \\ 0 \end{bmatrix}.$$

where T is maximal monotone. The partial prox $(\theta = 1)$ is

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in T(z) + \frac{1}{\theta} \begin{bmatrix} 0 \\ 0 \\ \lambda - \lambda_0 \end{bmatrix}.$$

Also, ALM is a full-prox for

$$0 \in \partial(-d)(\lambda)$$

where

$$d(\lambda) := \inf_{x',y'} L_p(x', y'; \lambda).$$

<u>ADMM</u>

(0) Given (λ_0, y_0)

(1) solve $x \in \operatorname{argmin}_{x'} L_p(x', y_0; \lambda_0)$

(2) solve $y \in \operatorname{argmin}_{y'} L_p(x, y'; \lambda_0)$

(3) set $\lambda = \lambda_0 + \rho(Ax + By)$

<u>Analysis</u>

Define

$$\begin{aligned}\tilde{\lambda} &:= \lambda_0 + \rho(Ax + By_0),\\ (\tilde{x}, \tilde{y}) &:= (x, y),\\ \tilde{z} &:= (\tilde{x}, \tilde{y}, \tilde{\lambda})\\ z &:= (x, y, \lambda).\end{aligned}$$

Exercise. Show that

$$\begin{aligned}(1) &\iff \partial f(\tilde{x}) + A^*\tilde{\lambda} \ni 0\\ (2) &\iff \partial g(\tilde{y}) + B^*\tilde{\lambda} + \rho B^*B(y - y_0) \ni 0\\ (3) &\iff -A\tilde{x} - B\tilde{y} + \frac{\lambda - \lambda_0}{\rho} = 0\end{aligned}$$

and with $\theta = 1, \varepsilon = 0$, we have that the above is equivalent to

$$T^\varepsilon(\tilde{z}) \ni \frac{\nabla w(z_0) - \nabla w(z)}{\theta}$$

where

$$\begin{aligned}w = w(x, y, \lambda) &:= \frac{\rho}{2}\|By\|^2 + \frac{1}{2\rho}\|\lambda\|^2,\\ \nabla w &= \begin{pmatrix} 0 \\ \rho B^*By \\ \frac{1}{\rho}\lambda \end{pmatrix}.\end{aligned}$$

We also have, with $\sigma \in [0, 1]$, the inequality

$$dw_z(\tilde{z}) + \lambda\varepsilon \le \sigma dw_{z_0}(\tilde{z}),$$

where dw_z is the Bregman distance of w, i.e.

$$0 \le dw_z(z') = w(z') - w(z) - \langle \nabla w(z), z' - z\rangle \le \frac{M}{2}\|z' - z\|^2.$$

We have

$$dw_z(\tilde{z}) = \frac{\rho}{2}\|B(y-\tilde{y})\|^2 + \frac{1}{2\rho}\|\lambda - \tilde{\lambda}\|^2 = \frac{1}{2\rho}\|\lambda - \tilde{\lambda}\|^2 = \frac{\rho}{2}\|B(y-y_0)\|^2$$

and

$$dw_{z_0}(\tilde{z}) = \frac{\rho}{2}\|B(y_0-\tilde{y})\|^2 + \frac{1}{2\rho}\|\lambda_0 - \tilde{\lambda}\|^2 = \frac{\rho}{2}\|B(y_0-\tilde{y})\|^2 + \frac{1}{2\rho}\|\lambda_0 - \tilde{\lambda}\|^2.$$

Also,

$$\begin{aligned}\|\tilde{z} - z\| = \|\lambda - \tilde{\lambda}\| &= \rho\|B(y-y_0)\| \\ &= \rho\left(\|B(y-y^*)\| + \|B(y^*-y_0)\|\right),\end{aligned}$$

and since $dw_z(z^*) \le dw_{z_0}(z^*)$ then we have

$$dw_z(\tilde{z}) + \theta\varepsilon \le \sigma dw_{z_0}(\tilde{z})$$

with $\sigma = 1$.

Assumptions on the Bregman

We have $w \in \overline{\text{Conv}}(\mathbb{R}^n)$ and a semi-norm $\|\cdot\|$ for which

$$dw_z(z') \ge \frac{m}{2}\|z'-z\|^2 \quad \forall z', z$$

in the semi-norm. Also,

$$\|\nabla w(z') - \nabla w(z)\|^* \le M\|z'-z\|.$$

Convergence

The IPP inclusion is

$$v^a = \frac{\sum v}{k} \in T^{\varepsilon^a}(z^a)$$

with convergence rates

$$\varepsilon^a \le \left(\frac{3M}{m}\right)\left[\frac{3\left(dw_0 + \sigma p_k\right)}{\theta k}\right], \qquad \|v^a\|^* \le \frac{2\sqrt{2}M(dw_0)^{1/2}}{\sqrt{m}\theta k},$$

where $\rho_k = \max_{i\le k} dw_{z_{i-1}}(\tilde{z}_i)$.