Mathematical and Numerical Optimization, 3341.454
E. Ryu
Fall 2020

Homework 8

Problem 1: *Further properties of the proximal operator.* Let f be a CCP function on $\mathbb{R}^n$. Show:

(a) $f(\mathrm{Prox}_{\alpha f}(x))$ is a nonincreasing function of $\alpha \in (0, \infty)$ (for a fixed $x \in \mathbb{R}^n$).

(b) $\lim_{\alpha\to\infty} f(\mathrm{Prox}_{\alpha f}(x)) = \inf_x f(x)$ (including the case $\inf_x f(x) = -\infty$).

(c) $f(\mathrm{Prox}_{\alpha f}(x)) \le f(x)$ for any $\alpha > 0$.

(d) $\lim_{\alpha\to 0^+} f(\mathrm{Prox}_{\alpha f}(x)) = f(x)$ for all $x \in \mathrm{dom}\, f$.

Hint. For (a), argue that

$$\alpha f\left(\mathrm{Prox}_{\alpha f}(x)\right) + \frac{1}{2}\left\|\mathrm{Prox}_{\alpha f}(x) - x\right\|^2 \le \alpha f\left(\mathrm{Prox}_{\beta f}(x)\right) + \frac{1}{2}\left\|\mathrm{Prox}_{\beta f}(x) - x\right\|^2$$
$$\beta f\left(\mathrm{Prox}_{\beta f}(x)\right) + \frac{1}{2}\left\|\mathrm{Prox}_{\beta f}(x) - x\right\|^2 \le \beta f\left(\mathrm{Prox}_{\alpha f}(x)\right) + \frac{1}{2}\left\|\mathrm{Prox}_{\alpha f}(x) - x\right\|^2$$

for $\alpha, \beta \in \mathbb{R}$. For (b), let $\varepsilon > 0$ and $M > \inf_x f(x)$. Let $x_{M,\varepsilon}$ be a point such that $f(x_{M,\varepsilon}) < M + \varepsilon/2$. Then

$$f(x_{M,\varepsilon}) + \frac{1}{2\alpha}\|x_{M,\varepsilon} - x\|^2 < M + \varepsilon$$

for large enough α. For (d), show

$$\alpha f(x) \ge \alpha f(\mathrm{Prox}_{\alpha f}(x)) + \frac{1}{2}\|\mathrm{Prox}_{\alpha f}(x) - x\|^2$$

and let $\alpha \to 0$.

Remark. The result of (d) is not necessarily true when $x \notin \mathrm{dom}\, f$. For example, consider $f = \delta_{\{0\}}$ and $x = 1$.

Problem 2: *Consensus + proximable is proximable.* Let r be a CCP function on $\mathbb{R}^n$, C be the consensus set as defined in (2.19), and

$$g(x_1, \dots, x_m) = \delta_C(x_1, \dots, x_m) + \sum_{i=1}^m r(x_i).$$

Show that we can evaluate $\mathrm{Prox}_{\alpha g}$ with

$$\mathrm{Prox}_{\alpha g}(y_1, \dots, y_m) = (x, \dots, x), \quad x = \mathrm{Prox}_{\alpha r}\left(\frac{1}{m}\sum_{i=1}^m y_i\right).$$

Also, what is the proximal operator of $h(x_1, \dots, x_m) = \delta_C(x_1, \dots, x_m) + r(x_1)$?

Problem 3: *Chen–Teboulle is variable metric PPM.* Consider the primal problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) + g(Ax),$$

where f is a CCP function on $\mathbb{R}^n$, g is a CCP function on $\mathbb{R}^m$, and $A \in \mathbb{R}^{m\times n}$, generated by the Lagrangian (convex with respect to $(x, z) \in \mathbb{R}^n \times \mathbb{R}^m$ and concave with respect to $u \in \mathbb{R}^m$)

$$\mathbf{L}(x, z, u) = f(x) + g(z) + \langle u, Ax - z\rangle.$$

Show that the *Chen–Teboulle* method

$$\begin{aligned} p^{k+1} &= y^k + \alpha(Ax^k - z^k) \\ x^{k+1} &= \mathrm{Prox}_{\alpha f}(x^k - \alpha A^\intercal p^{k+1}) \\ z^{k+1} &= \mathrm{Prox}_{\alpha g}(z^k + \alpha p^{k+1}) \\ y^{k+1} &= y^k + \alpha(Ax^{k+1} - z^{k+1}) \end{aligned}$$

is equivalent to an instance of the variable metric proximal point method on $\partial\mathbf{L}$ with

$$M = \begin{bmatrix} (1/\alpha)I & 0 & -A^\intercal \\ 0 & (1/\alpha)I & I \\ -A & I & (1/\alpha)I \end{bmatrix}.$$

Problem 4: *Chen–Teboulle is linearized method of multipliers.* Consider the problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n,\, z\in\mathbb{R}^m}{\text{minimize}} & f(x) + g(z) \\ \text{subject to} & Ax - z = 0, \end{array}$$

where f is a CCP function on $\mathbb{R}^n$, g is a CCP function on $\mathbb{R}^m$, and $A \in \mathbb{R}^{m\times n}$. Show that the *Chen–Teboulle* method of Exercise 3 is equivalent to an instance of the linearized method of multipliers.

Problem 5: *Consensus technique and DYS.* Consider the problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) + \frac{1}{m}\sum_{i=1}^m \left(g_i(x) + h_i(x)\right),$$

where f is CCP, $g_1, \dots, g_m$ are CCP, $h_1, \dots, h_m$ is differentiable and CCP. The consensus techniques yields the equivalent problem

$$\underset{x_1,\dots,x_m\in\mathbb{R}^n}{\text{minimize}} \quad mf(x_1) + \delta_C(x_1, \dots, x_m) + \sum_{i=1}^m \left(g_i(x_i) + h_i(x_i)\right).$$

Show that the FPI with DYS is

$$\begin{aligned} x^{k+1/2} &= \mathrm{Prox}_{\alpha f}\left(\frac{1}{m}\sum_{i=1}^m z_i^k\right) \\ x_i^{k+1} &= \mathrm{Prox}_{\alpha g_i}(2x^{k+1/2} - z_i^k - \alpha\nabla h_i(x^{k+1/2})) \\ z_i^{k+1} &= z_i^k + x_i^{k+1} - x^{k+1/2} \qquad \text{for } i = 1, \dots, m. \end{aligned}$$

Assume computing $\mathrm{Prox}_{\alpha f}$ costs C_f flops and $\mathrm{Prox}_{\alpha g_i}$ costs C_g flops and ∇h_i costs C_h flops for $i = 1, \dots, m$. Assume the cost C_f cannot be further reduced through parallelization. What is the parallel flop count $\mathcal{F}_p\left[\{z_1^k, \dots, z_m^k\} \mapsto \{z_1^{k+1}, \dots, z_m^{k+1}\}\right]$ for $p \le \min\{m, n\}$? For simplicity, you may assume m/p and n/p are an integers.

Hint. Use Exercise 2.

Problem 6: *Parallel SC-FPI.* The SC-FPI can be generalized to the setup where p agents independently and randomly select and update indices. Define $\mathbb{S}_i = (1/\theta)(\mathbb{I} - \mathbb{T}_i)$ for $i = 1, \ldots, m$. *Parallel* SC-FPI is

$$\begin{aligned} i(k,w) &\sim \text{IID Uniform}\{1,\ldots,m\} \qquad \text{for } w = 1,\ldots,p \\ x^{k+1} &= x^k - \sum_{w=1}^{p} \alpha \mathbb{S}_{i(k,w)}(x^k). \end{aligned}$$

The computation of $\mathbb{S}_{i(k,1)}(x^k), \ldots, \mathbb{S}_{i(k,p)}(x^k)$ can be parallelized by p computational agents. We do not require $i(k,1), \ldots, i(k,p)$ to be distinct. Show that

$$\mathbb{E}_k \|x^{k+1} - x^\star\|^2 \leq \|x^k - x^\star\|^2 - (\text{nonnegative term}).$$

Remark. This corresponds to Stage 1 of the proof of Theorem 2. With this martingale inequality established, the convergence of this method can be established by following the same reasoning as that of Stage 2 of Theorem 2.

Problem 7: *Non-uniform selection rules.* Assume $i(k)$ is an IID random variable with $\text{Prob}(i(k) = j) = p_j$, where $p_1, \ldots, p_m > 0$ and $p_1 + \cdots + p_m = 1$. Assume $\mathbb{T}$ is θ-averaged, $\theta \in (0,1)$, and $\text{Fix}\,\mathbb{T} \neq \emptyset$. Define $\mathbb{S}_i = (1/\theta)(\mathbb{I} - \mathbb{T}_i)$ for $i = 1, \ldots, m$. SC-FPI with non-uniform coordinate selection rules is

$$x^{k+1} = x^k - \frac{\eta}{p_{i(k)}} \mathbb{S}_{i(k)}(x^k).$$

converges. Show that under a suitable condition on η,

$$\mathbb{E}_k \|x^{k+1} - x^\star\|^2 \leq \|x^k - x^\star\|^2 - (\text{nonnegative term}).$$

What is the condition on η?

Remark. This corresponds to Stage 1 of the proof of Theorem 2. With this martingale inequality established, the convergence of this method can be established by following the same reasoning as that of Stage 2 of Theorem 2.