Homework 4

Yujun Kim

IE539 Convex Optimization

Nov 2023

1 Proximal Operator is a Contraction Map

Let $prox_{\eta h}(x) = u, prox_{\eta h}(y) = v$. By lemma, we have

$$x - u \in \eta\partial h(u), \qquad y - x \in \eta\partial h(v)$$

Hence, $x = u + g_u, y = v + g_v$ for some $g_u \in \partial h(u), g_v \in \partial h(v)$. As h is convex,

$$\begin{aligned} h(u) &\geq h(v) + g_v^T(u - v) \\ h(v) &\geq h(u) + g_u^T(v - x) \end{aligned}$$

By adding two inequalities, we obtain $\langle g_u - g_v, u - v\rangle \geq 0$. Gathering all together,

$$\begin{aligned} ||x - y||^2 &= ||(u + g_u) - (v + g_v)||^2 \\ &= ||u - v||^2 + ||g_u - g_v||^2 + 2\langle u - v, g_u - g_v\rangle \\ &\geq ||u - v||^2 + ||g_u - g_v||^2 \\ &\geq ||u - v||^2 \\ &= ||prox_{\eta h}(x) - prox_{\eta h}(y)||^2 \end{aligned}$$

2 Characterization of Minimizer

f as a sum of two convex function is convex. By first order characterization, x^* minimize f if and only if

$$\begin{aligned} &0 \in \partial f(x^*) = \nabla g(x^*) + \partial h(x^*) \\ &\Leftrightarrow -\nabla g(x^*) \in \partial h(x^*) \\ &\Leftrightarrow -\eta\nabla g(x^*) \in \eta\partial h(x^*) \\ &\Leftrightarrow x^* - \eta\nabla g(x^*) \in x^* + \eta\partial h(x^*) \\ &\Leftrightarrow (I - \eta\nabla g)(x^*) \in (I + \eta\partial h)(x^*) \\ &\Leftrightarrow x^* \in (I + \eta\partial h)^{-1}(I - \eta\nabla g)(x^*) \end{aligned}$$

This finishes the proof.

3 Solving Using Lagrangian Dual

(a) $\frac{\partial \mathcal{L}}{\partial x_i} = v_i - \log x_i + 1 - \lambda_i - \mu$. Letting this value equals 0, we obtain

$$\begin{aligned} \log x_i &= \lambda_i + \mu - v_i - 1 \\ x_i &= e^{\lambda_i + \mu - v_i - 1} \\ x_i \log x_i &= (\lambda_i + \mu - v_i - 1)x_i \end{aligned}$$

Thus, $(v_i - \lambda_i - \mu + \log x_i)x_i = -x_i$. Putting these x_i to $\mathcal{L}(x, \lambda, \mu)$ gives

$$\begin{aligned} \mathcal{L}(x, \lambda, \mu) &= \mu + \sum_{i=1}^{d} (v_i - \lambda_i - \mu + \log x_i)x_i \\ &= \mu - \sum_{i=1}^{d} x_i \\ &= \mu - \sum_{i=1}^{d} e^{\lambda_i + \mu - v_i - 1} \end{aligned}$$

Hence, $q(\lambda, \mu) = \mu - \sum_{i=1}^{d} e^{\lambda_i + \mu - v_i - 1}$.

(b) First, $\frac{\partial q}{\partial \lambda_i} = -e^{\lambda_i + \mu - v_i - 1} < 0$ for all $\lambda_i \geq 0$. Hence, among feasible λ_i, $\lambda_i^* = 0$ maximize q with any choice of μ. Thus, $\lambda^* = 0$. Next, with $\lambda^* = 0$,

$$\begin{aligned} \frac{\partial q}{\partial \mu}(\lambda^*, \mu) &= 1 - \sum_{i=1}^{d} e^{\lambda_i^* + \mu - v_i - 1} \\ &= 1 - \sum_{i=1}^{d} e^{\mu - v_i - 1} \\ &= 1 - \left(\sum_{i=1}^{d} e^{-v_i} \right) e^{\mu - 1} \end{aligned}$$

Letting this 0, we obtain

$$e^{\mu^* - 1} = \frac{1}{\sum_{i=1}^{d} e^{-v_i}}$$

(c) Let $f(x)$ be the primal objective, $g_i(x) = x_i \leq 0, h(x) = \sum_{i=1}^{d} x_i - 1 = 0$ be the constraints. The problem clearly satisfies the Slater's condition. By necessity of KKT condition of optimal solution, there exists dual variables $\lambda^* \in \mathbb{R}_+^d, \mu^* \in \mathbb{R}$ such that

$$\nabla f(x^*) + \sum \lambda_i^* \nabla g_i(x^*) + \mu^* \nabla h(x^*) = 0$$

and $\lambda_i^* g_i(x^*) = 0$ for all i. We now show that such λ^*, μ^* is optimal in dual so it satisfies condition obtained from (b). Observe that

$$f(x^*) = \mathcal{L}(x^*, \lambda^*, \mu^*) \leq q(\lambda^*, \mu^*)$$

By weak duality, this shows λ^*, μ^* is optimal value for the dual problem. Considering the i-th component of

$$\nabla f(x^*) + \sum \lambda_j^* \nabla g_j(x^*) + \mu^* \nabla h(x^*) = 0$$

we have

$$\begin{aligned} 0 &= \frac{\partial f}{\partial x_i}(x^*) + \sum_{j=1}^{d} \lambda_j^* \frac{\partial g_j}{\partial x_i}(x^*) + \mu^* \frac{\partial h}{\partial x_i}(x^*) \\ &= v_i + 1 + \log x_i^* - \lambda_i^* - \mu^* \end{aligned}$$

Thus, by results from (b),

$$\begin{aligned} x_i^* &= e^{\lambda_i^* + \mu^* - v_i - 1} \\ &= e^{\mu^* - 1} e^{-v_i} \\ &= \frac{e^{-v_i}}{\sum_{i=1}^{d} e^{-v_i}} \end{aligned}$$

4 Primal-Dual Subgradient Descent

(a) As $\phi(\cdot, \bar{y}), \phi(\bar{x}, \cdot)$ are convex and concave respectively, we have

$$\phi(x, \bar{y}) \geq \phi(\bar{x}, \bar{y}) + g_x^T(x - \bar{x}) \quad (1)$$
$$\phi(\bar{x}, y) \leq \phi(\bar{x}, \bar{y}) + g_y^T(y - \bar{y}) \quad (2)$$

By (2)-(1), we obtain

$$\begin{aligned} \phi(\bar{x}, y) - \phi(x, \bar{y}) &\leq \phi(\bar{x}, \bar{y}) + g_y^T(y - \bar{y}) - (\phi(\bar{x}, \bar{y}) + g_x^T(x - \bar{x})) \\ &= -g_x^T(x - \bar{x}) + g_y^T(y - \bar{y}) \end{aligned}$$

(b) (Step 1) Iteration-wise recursion. Denote $\delta_t = ||(x_t, y_t) - (x, y)||^2$ and $\nabla g = (\nabla g_{x,t}, \nabla g_{y,t})$. Then,

$$\begin{aligned} \delta_{t+1} &= ||x_{t+1} - x||^2 + ||y_{t+1} - y||^2 \\ &= ||proj_X(x_t - \eta_t \nabla g_{x,t}) - x||^2 + ||proj_Y(y_t - \eta_t \nabla g_{y,t}) - y||^2 \\ &\leq ||x_t - x + \eta_t \nabla g_{x,t}||^2 + ||y_t - y + \eta_t \nabla g_{y,t})||^2 \\ &= ||x_t - x||^2 + ||y_t - y||^2 + 2\eta_t(-\nabla g_{x,t}^T(x_t - x) + \nabla g_{y,t}(y_t - y)) + \eta_t^2(||\nabla g_{x,t}||^2 + ||\nabla g_{y,t}||^2) \\ &= ||(x_t, y_t) - (x, y)||^2 - 2\eta_t(-\nabla g_{x,t}^T(x - x_t) + \nabla g_{y,t}(y - y_t)) + \eta_t^2||\nabla g_t||^2 \\ &= \delta_t - 2\eta_t(-\nabla g_{x,t}^T(x - x_t) + \nabla g_{y,t}(y - y_t)) + \eta_t^2||\nabla g_t||^2 \\ &\leq \delta_t - 2\eta_t(\phi(x_t, y) - \phi(x, y_t)) + \eta_t^2||\nabla g_t||^2 \end{aligned}$$

By rearranging the terms,

$$2\eta_t(\phi(x_t, y) - \phi(x, y_t)) \leq \delta_t - \delta_{t+1} + \eta_t^2||\nabla g_t||^2$$

(Step 2) Convergence rate. Sum the obove equation over $t = 1, \cdots, T$ and divide each side by $2\sum_{t=1}^{T} \eta_t$ to obtain

$$\begin{aligned}
\left(\sum_{t=1}^{T} \eta_t\right)^{-1} \sum_{t=1}^{T} \eta_t(\phi(x_t, y) - \phi(x, y_t)) &\leq \frac{1}{2\sum_{t=1}^{T} \eta_t} \left(\delta_1 - \delta_{T+1} + \sum_{t=1}^{T} \eta_t^2 ||\nabla g_t||^2\right) \\
&\leq \frac{1}{2\sum_{t=1}^{T} \eta_t} \left(\delta_1 + \sum_{t=1}^{T} \eta_t^2 ||\nabla g_t||^2\right) \\
&= \frac{1}{2\sum_{t=1}^{T} \eta_t} \left(||(x_1, y_1) - (x, y)||^2 + \sum_{t=1}^{T} \eta_t^2 ||(\nabla g_{x,t}, \nabla g_{y,t})||^2\right)
\end{aligned}$$

As $\phi(\cdot, y), \phi(x, \cdot)$ are convex and concave respectively, we have

$$\begin{aligned}
\phi(\bar{x_t}, y) &\leq \left(\sum_{t=1}^{T} \eta_t\right)^{-1} \sum_{t=1}^{T} \eta_t \phi(x_t, y) \\
\phi(x, \bar{y_t}) &\geq \left(\sum_{t=1}^{T} \eta_t\right)^{-1} \sum_{t=1}^{T} \eta_t \phi(x, y_t)
\end{aligned}$$

Thus, $\phi(\bar{x_t}, y) - \phi(x, \bar{y_t}) \leq \left(\sum_{t=1}^{T} \eta_t\right)^{-1} \sum_{t=1}^{T} \eta_t(\phi(x_t, y) - \phi(x, y_t))$. Hence we conclude

$$\phi(\bar{x_t}, y) - \phi(x, \bar{y_t}) \leq \frac{1}{2\sum_{t=1}^{T} \eta_t} \left(||(x_1, y_1) - (x, y)||^2 + \sum_{t=1}^{T} \eta_t^2 ||(\nabla g_{x,t}, \nabla g_{y,t})||^2\right)$$

5 Moreau Decomposition - General Form

Let $h = \lambda f$. By the given formula,

$$x = prox_h(x) - prox_{h^*}(x)$$

Hence, it suffices to show

$$prox_{h^*}(x) = \lambda prox_{(1/\lambda)f^*}(x/\lambda)$$

(Step 1) Finding h^*.

$$\begin{aligned}
h^*(y) &= \sup_x \{y^T x - h(x)\} \\
&= \sup_x \{y^T x - \lambda f(x)\} \\
&= \lambda \sup_x \{(y/\lambda)^T x - f(x)\} \\
&= \lambda f^*(y/\lambda)
\end{aligned}$$

(Step 2) Finding $prox_{h^*}(x)$. From step 1, it follows that

$$\partial h^*(u) = \partial f^*(u/\lambda)$$

Hence,

$$\begin{aligned}
u = prox_{h^*}(x) &\Leftrightarrow x - u \in \partial h^*(u) \\
&\Leftrightarrow x - u \in \partial f^*(u/\lambda) \\
&\Leftrightarrow \frac{x}{\lambda} - \frac{u}{\lambda} \in \frac{1}{\lambda}\partial f^*(\frac{u}{\lambda}) = \partial \left(\frac{1}{\lambda} f^*\right)(u/\lambda) \\
&\Leftrightarrow \frac{u}{\lambda} = prox_{(1/\lambda)f^*}(x/\lambda) \\
&\Leftrightarrow u = \lambda prox_{(1/\lambda)f^*}(x/\lambda)
\end{aligned}$$

where in the first and fourth equivalence, we used the following lemma

$$u = prox_f(x) \Leftrightarrow x - u \in \partial f(u)$$

Thus we conclude $prox_{h^*}(x) = \lambda prox_{(1/\lambda)f^*}(x/\lambda)$ and so we get the desired equality.