Mathematical and Numerical Optimization, 3341.454
E. Ryu
Fall 2020

Homework 11

Problem 1: Let $f(x,y)$ with $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$ be a CCP function. (f is jointly convex in x and y.) Define

$$\phi(y) = \inf_{x\in\mathbb{R}^p} f(x,y), \qquad \psi(v) = \sup_{\substack{x\in\mathbb{R}^p\\ y\in\mathbb{R}^q}} \{-f(x,y) + \langle v, y\rangle\}.$$

Show:

(a) For any $v \in \mathbb{R}^q$, $(x,y) \in \operatorname{argmax}_{\substack{x\in\mathbb{R}^p\\ y\in\mathbb{R}^q}} \{-f(x,y) + \langle v,y\rangle\}$ implies $(x,y) \in \partial f^*(0,v)$, which in turn implies $y \in \partial\psi(v)$.

(b) $\psi = \phi^*$.

(c) For any $y \in \mathbb{R}^q$, $x \in \operatorname{argmin}_{x\in\mathbb{R}^p} f(x,y)$ and $(0,v) \in \partial f(x,y)$ implies $(x,y) \in \operatorname{argmax}_{\substack{x\in\mathbb{R}^p\\ y\in\mathbb{R}^q}} \{-f(x,y) + \langle v,y\rangle\}$, which in turn implies $v \in \partial\phi(y)$.

Remark. For any $y \in \mathbb{R}^q$, if $x \in \operatorname{argmin}_{x\in\mathbb{R}^p} f(x,y)$ and $(u,v) \in \partial f(x,y)$, then $(0,v) \in \partial f(x,y)$, i.e., we can always find a subgradient such that its first p components are 0.

Remark. Results (a) and (c) provide us with a method to obtain subgradients in $\partial\phi$ and $\partial\psi$ for the primal and dual decomposition setups.

Remark. Under the stated assumptions, ϕ and ψ are convex, but they may not be CCP. While results (a) through (c) hold regardless of whether ϕ and ψ are closed and proper, you may assume ϕ and ψ are CCP for the sake of simplicity.

Problem 2: Consider the primal decomposition formulation

$$\underset{z\in\mathbb{R}^q}{\text{minimize}} \quad \sum_{i=1}^n \phi_i(z)$$

with $\phi_i(z) = \inf_{x\in\mathbb{R}^p} f_i(x,z)$. Show that DRS applied to the consensus formulation

$$\underset{z_1,\dots,z_n\in\mathbb{R}^q}{\text{minimize}} \quad \sum_{i=1}^n \phi_i(z_i) + \delta_C(z_1,\dots,z_n)$$

is equivalent to distributed ADMM:

$$\begin{aligned} z_i^{k+1} &= \operatorname{Prox}_{(1/\alpha)\phi_i}\left(\bar{z}^k - (1/\alpha)u_i^k\right) \\ u_i^{k+1} &= u_i^k + \alpha(z_i^{k+1} - \bar{z}^{k+1}) \end{aligned}$$

for $i = 1,\dots,n$, where $\bar{z}^k = (1/n)(z_1^k + \cdots + z_n^k)$. For simplicity, assume all convex functions are CCP.

Remark. Note that $\operatorname{Prox}_{(1/\alpha)\phi_i}(z_0)$ can be computed by minimizing $f(x,z) + \frac{\alpha}{2}\|z - z_0\|^2$ with respect to x, z and returning z.

Problem 3: Consider the dual decomposition formulation

$$\underset{v_1,\dots,v_n\in\mathbb{R}^q}{\text{minimize}} \quad \sum_{i=1}^{n} \psi_i(v_i) + \delta_{C^*}(v_1,\dots,v_n)$$

with $\psi_i(v_i) = \sup_{\substack{x_i\in\mathbb{R}^p \\ z_i\in\mathbb{R}^q}} \{-f_i(x_i,z_i) + \langle v_i, z_i\rangle\}$ and $C^* = \{v_1+\cdots+v_n = 0\}$. Show that DRS applied to this formulation is equivalent to the method of Exercise 2. For simplicity, assume all convex functions are CCP.

Hint. Show that $(\delta_C)^* = \delta_{C^*}$. Then use part (b) of Exercise 1 and the self-dual property of DRS discussed in §9.3. (Do not directly work out the application of DRS.)

Remark. In the language of convex analysis, the consensus set C is a convex cone (closed, not solid, not pointed) and C^* is in fact the dual cone of C.

Problem 4: *Another ADMM-based decentralized method.* Show that the formulation

$$\begin{array}{ll} \underset{\{x_i,y_i\}_{i\in V}}{\text{minimize}} & \sum_{i\in V} f_i(x_i) \\ \text{subject to} & x_i = y_i \quad \forall\, i\in V \\ & \left\{\begin{array}{c} x_i - y_j = 0 \\ x_j - y_i = 0 \end{array}\right\} \ \forall\,\{i,j\}\in E. \end{array}$$

is equivalent to the formulation (11.1) of the book. Apply ADMM to derive:

$$\begin{aligned} x_i^{k+1} &= \mathrm{Prox}_{(\alpha(|N_i|+1))^{-1}f_i}\left(\frac{1}{|N_i|+1}\sum_{j\in N_i\cup\{i\}} y_j^k - \frac{1}{\alpha}v_i^k\right) \\ y_i^{k+1} &= \frac{1}{|N_i|+1}\sum_{j\in N_i\cup\{i\}} x_j^{k+1} \\ v_i^{k+1} &= v_i^k + \alpha x_i^{k+1} - \frac{\alpha}{|N_i|+1}\sum_{j\in N_i\cup\{i\}} y_j^{k+1} \end{aligned}$$

for all $i\in V$. Also, explain why the method is decentralized.

Problem 5: *Equivalence of consensus conditions.* Consider $\mathbf{x} = \mathrm{stack}(x_1,\dots,x_n)$ and a mixing matrix $W\in\mathbb{R}^{n\times n}$ satisfying $\mathcal{N}(I-W) = \mathrm{span}(\mathbf{1})$. Show that the following conditions are equivalent:

(i) $x_1 = \cdots = x_n$,

(ii) $(I-W)\mathbf{x} = 0$,

(iii) $\|\mathbf{x}\|_{I-W} = 0$ when $W = W^\intercal$ and $\lambda_1(W)\le 1$, and

(iv) $U\mathbf{x} = 0$ when $W = W^\intercal$, $\lambda_1(W)\le 1$, and $U^2 = \frac{1}{2}(I-W)$.

Problem 6: Let $W\in\mathbb{R}^{n\times n}$ and consider the decentralized averaging method

$$\mathbf{x}^{k+1} = W\mathbf{x}^k.$$

Then $\mathbf{x}^k\to\mathbf{x}^\star$ for all starting points $\mathbf{x}^0\in\mathbb{R}^{n\times p}$ if and only if

(i) $W\mathbf{1} = \mathbf{1}$,

(ii) $\mathbf{1}^\intercal W = \mathbf{1}^\intercal$,

(iii) $\lambda_1(W) = 1 > |\lambda_2(W)| \geq \cdots \geq |\lambda_n(W)|$.

Also show that when $W = W^\intercal$ is symmetric, assumptions (i)–(iii) are equivalent to assumptions $\mathcal{N}(I - W) = \mathrm{span}(\mathbf{1})$ and $1 = |\lambda_1| > \max\{|\lambda_2|, \ldots, |\lambda_n|\}$.

Problem 7: Let $W \in \mathbb{R}^{n\times n}$ and consider the decentralized averaging method

$$\mathbf{x}^{k+1} = W\mathbf{x}^k.$$

Since

$$\|\mathbf{x}^k - \mathbf{x}^\star\| \sim (\rho(W - \mathbf{1}\mathbf{1}^\intercal/n))^k \, \|\mathbf{x}^0 - \mathbf{x}^\star\|, \tag{1}$$

where ρ denotes the spectral radius, we interpret $\rho(W - \mathbf{1}\mathbf{1}^\intercal/n)$ as the asymptotic convergence rate. Next, consider a graph $G = (V, E)$ and assume W is decentralized with respect to G. Consider the problem of finding a decentralized mixing matrix with the fastest asymptotic convergence rate:

$$\begin{array}{ll} \underset{W\in\mathbb{R}^{n\times n}}{\text{minimize}} & \rho(W - \mathbf{1}\mathbf{1}^\intercal/n) \\ \text{subject to} & \mathbf{1}^T W = \mathbf{1}^T W,\ W\mathbf{1} = \mathbf{1} \\ & W_{ij} = 0,\ \{i,j\} \notin E,\ i \neq j. \end{array}$$

However, optimizing the spectral radius of non-symmetric matrices is a difficult problem. So we further assume W is symmetric:

$$\begin{array}{ll} \underset{W\in\mathbb{R}^{n\times n}}{\text{minimize}} & \sigma_{\max}(W - \mathbf{1}\mathbf{1}^\intercal/n) \\ \text{subject to} & W = W^\intercal,\ W\mathbf{1} = \mathbf{1} \\ & W_{ij} = 0,\ \{i,j\} \notin E,\ i \neq j, \end{array}$$

where $\sigma_{\max}$ denotes the maximum singular value. This problem is equivalent to

$$\begin{array}{ll} \underset{s\in\mathbb{R},\, W\in\mathbb{R}^{n\times n}}{\text{minimize}} & s \\ \text{subject to} & -sI \preceq W - \mathbf{1}\mathbf{1}^\intercal/n \preceq sI \\ & W = W^\intercal,\ W\mathbf{1} = \mathbf{1} \\ & W_{ij} = 0,\ \{i,j\} \notin E,\ i \neq j, \end{array}$$

where $\preceq$ denotes the partial order in the sense of positive semidefinite matrices.

(a) Show (1).

(b) Numerically solve the problem instance depicted in Figure 1 and establish that the depicted solution is indeed optimal. (The solution is not unique.)

(c) The optimal mixing matrix of part (b) contains negative weights. Show that the negative weights are necessary to obtain the optimal mixing matrix by solving the optimization problem with the added constraint $W_{ij} \geq 0$ for all $i, j \in \{1, \ldots, n\}$.

Remark. For optimization problems with somewhat complicated constraints, such as this one, it is often simpler to solve small problem instances with libraries such as YALMIP, CVX, or CVXPY. For large problem instances, it becomes necessary to use efficient splitting methods.

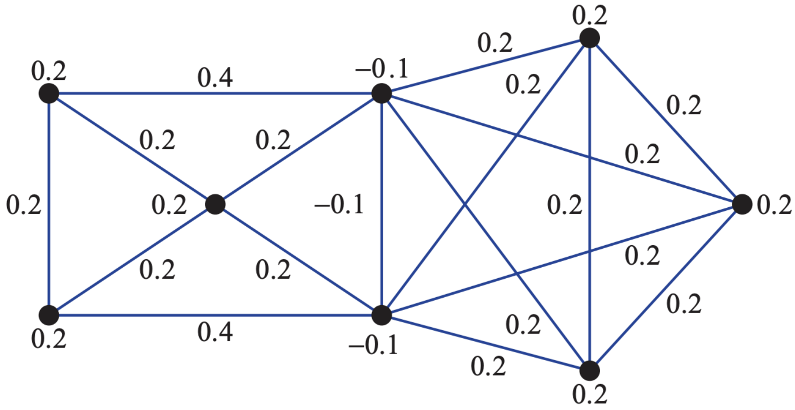

Figure 1: A small graph with 8 nodes and 17 edges. Each edge and node is labeled with the optimal symmetric weights, which give the minimum asymptotic convergence factor.

Problem 8: *Closed graph theorem for maximal monotone operators.* Let $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone. Show $\mathbb{A}$ is *upper hemicontinuituous*, i.e., show that if $x^k \to x^\infty$, $u^k \to u^\infty$, and $u^k \in \mathbb{A}x^k$, then $u^\infty \in \mathbb{A}x^\infty$. (Upper hemicontinuity of $\mathbb{A}$ is equivalent to $\operatorname{Gra}\mathbb{A} \subset \mathbb{R}^n \times \mathbb{R}^n$ being a closed set.)

Hint. The proof can be done in one line using $\mathbf{F}_{\mathbb{A}}(x^\infty, u^\infty) \le \liminf_{k\to\infty} \mathbf{F}_{\mathbb{A}}(x^k, u^k)$ and Lemma 3.

Problem 9: *Method of multipliers primal solution convergence without strict convexity.* Consider the method of multipliers under the stated conditions. Show that any accumulation point of $x^0, x^1, \dots$ is a primal solution.

Hint. Use Exercise 8.

Remark. The stated conditions are: f is CCP, $\mathcal{R}(A^\intercal) \cap \operatorname{ri}\operatorname{dom} f^* \neq \emptyset$, a dual solution exists, $\alpha > 0$, and $\mathbf{L}_\alpha(x,u) = f(x) + \langle u, Ax-b\rangle + \frac{\alpha}{2}\|Ax-b\|^2$.
then $u^k \to u^\star$. If we further assume f is strictly convex, we can show $x^\star \to x^\star$.

Problem 10: *Nonexpansiveness and monotonicity.* Show that if $\mathbb{T}$ is maximal nonexpansive (i.e., nonexpansive and $\operatorname{dom}\mathbb{T} = \mathbb{R}^n$ per Theorem 11) then $\mathbb{A} = \left(\frac{1}{2}\mathbb{T} + \frac{1}{2}\mathbb{I}\right)^{-1} - \mathbb{I}$ is maximal monotone.

Remark. Conversely, if $\mathbb{A}$ is a maximal monotone operator, then $2\mathbb{J}_{\mathbb{A}} - \mathbb{I}$ is maximal nonexpansive. Therefore, the transformation $\mathbb{A} \mapsto 2\mathbb{J}_{\mathbb{A}} - \mathbb{I}$ and its inverse $\mathbb{T} \mapsto \left(\frac{1}{2}\mathbb{T} + \frac{1}{2}\mathbb{I}\right)^{-1} - \mathbb{I}$ provide a one-to-one correspondence between maximal monotone operators and maximal nonexpansive operators.