Mathematical and Numerical Optimization, 3341.454
E. Ryu
Fall 2020

Homework 7

Problem 1: Show that if f is a strictly convex CCP function, then (i) ∂f^* is single-valued and (ii) f^* is differentiable on $\operatorname{int}\operatorname{dom} f^*$.

Remark. Since f^* is CCP, f^* is subdifferentiable on $\operatorname{ri}\operatorname{dom} f^*$ and $\partial f^*(u)$ is a singleton if and only if f^* is differentiable at u.

Note. A function f is *strictly convex* if

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y), \quad \forall\, x, y \in \operatorname{dom} f,\ x \neq y,\ \theta \in (0,1).$$

Previously, there was an error in the slides and the book on the definition strict convexity. It has now been fixed.

Problem 2: *Method of multipliers primal solution convergence.* Show that the method of multipliers converges in the sense of $x^k \to x^\star$ under the stated conditions and strict convexity. Use the following fact: if h is a CCP function that is differentiable on $D \subseteq \mathbb{R}^n$, then $\nabla h\colon D \to \mathbb{R}^n$ is a continuous function, i.e., differentiability and continuous differentiability coincide.

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 $\mathbf{L}_\alpha(x,u) = f(x) + \langle u, Ax-b\rangle + \frac{\alpha}{2}\|Ax-b\|^2$.

Hint. Consider the primal problem

$$\begin{array}{ll} \underset{u\in\mathbb{R}^m,\, v\in\mathbb{R}^n}{\text{minimize}} & f^*(v) + b^\intercal u \\ \text{subject to} & -v - A^\intercal u = 0 \end{array}$$

generated by the Lagrangian $\tilde{\mathbf{L}}(v,u,x) = f^*(v) + b^\intercal u - \langle x, v + A^\intercal u\rangle$, and use Slater's constraint qualification to show that $\mathcal{R}(A^\intercal) \cap \operatorname{ri}\operatorname{dom} f^* \neq \emptyset$ implies strong duality and the existence of a primal solution for the primal-dual problem pair generated by $\mathbf{L}$. Use Exercise 1 to write $x^k = \sigma(u^k)$, where $\sigma\colon \mathbb{R}^m \to \mathbb{R}^n$ is a continuous function.

Remark. The derivation of (2.6) or Exercise 3 of homework 6 establishes $\operatorname{argmin}_x \mathbf{L}_\alpha(x,u^k) \neq \emptyset$, i.e., $x^{k+1} \in \operatorname{argmin}_x \mathbf{L}_\alpha(x,u^k)$ is well-defined for any $u^k \in \mathbb{R}^m$.

Problem 3: In the derivation of PDHG, show that if we instead use

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

(note the sign difference in the off-diagonals) we get an upper triangular system in the inclusion and the method

$$\begin{aligned} u^{k+1} &= \operatorname{Prox}_{\beta g^*}(u^k + \beta A x^k) \\ x^{k+1} &= \operatorname{Prox}_{\alpha f}(x^k - \alpha A^\intercal(2u^{k+1} - u^k)). \end{aligned}$$

Problem 4: *Linearized method of multipliers with BCV.* We used the linearization technique with the proximal method of multipliers to prove convergence of the linearized method of multipliers for $\alpha\beta\lambda_{\max}(A^\intercal A) < 1$. By using the BCV technique, show that in fact $u^k \to u^\star$ for $\alpha\beta\lambda_{\max}(A^\intercal A) \le 1$.

Hint. Apply ADMM to

$$\begin{array}{ll}\underset{x\in\mathbb{R}^p,\,\tilde{y}\in\mathbb{R}^p}{\text{minimize}} & f(x)\\ \text{subject to} & \begin{bmatrix}A\\P\end{bmatrix}x+\begin{bmatrix}0\\I\end{bmatrix}\tilde{y}=\begin{bmatrix}b\\0\end{bmatrix}.\end{array}$$

Problem 5: *PDHG generalizes DRS.* PDHG with $A = I$ and $\beta = 1/\alpha$ is

$$\begin{aligned}x^{k+1} &= \mathrm{Prox}_{\alpha f}(x^k - \alpha u^k)\\ u^{k+1} &= \mathrm{Prox}_{(1/\alpha)g^*}(u^k + (1/\alpha)(2x^{k+1} - x^k)).\end{aligned}$$

DRS with $\mathrm{Prox}_{\alpha f}$ applied first is

$$\begin{aligned}x^{k+1/2} &= \mathrm{Prox}_{\alpha f}(z^k)\\ x^{k+1} &= \mathrm{Prox}_{\alpha g}(2x^{k+1/2} - z^k)\\ z^{k+1} &= z^k + x^{k+1} - x^{k+1/2}.\end{aligned}$$

Show that the two methods are equivalent in the sense that they generate an identical sequence of iterates after a change of variables.

Hint. For PDHG, define $\tilde{z}^k = x^k - \alpha u^k$.

Remark. The BCV technique establishes the converse, that DRS generalizes PDHG.

Problem 6: *PD3O generalizes PAPC/PDFP^{2}O.* PD3O with $f = 0$ is

$$\begin{aligned}x^{k+1} &= x^k - \alpha A^\intercal u^k - \alpha\nabla h(x^k)\\ u^{k+1} &= \mathrm{Prox}_{\beta g^*}\left(u^k + \beta A\left(2x^{k+1} - x^k + \alpha\nabla h(x^k) - \alpha\nabla h(x^{k+1})\right)\right).\end{aligned}$$

PAPC/PDFP^{2}O is

$$\begin{aligned}u^{k+1} &= \mathrm{Prox}_{\beta g^*}\left(u^k + \beta A(x^k - \alpha A^\intercal u^k - \alpha\nabla h(x^k))\right)\\ x^{k+1} &= x^k - \alpha A^\intercal u^{k+1} - \alpha\nabla h(x^k).\end{aligned}$$

Show that the two methods are equivalent in the sense that they generate an identical sequence of iterates after a change of variables.

Note. Earlier versions of the slides had a typo in the PAPC/PDFP^{2}O iteration. It has now been fixed.