Convex Analysis and Optimization

Chapter 8 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 8: SOLUTION MANUAL

8.1

To obtain a contradiction, assume that q is differentiable at some dual optimal solution $\mu^* \in M$, where $M = \{\mu \in \Re^r \mid \mu \geq 0\}$. Then by the optimality theory of Section 4.7 (cf. Prop. 4.7.2, concave function q), we have

$$\nabla q(\mu^*)(\mu^* - \mu) \geq 0, \qquad \forall\ \mu \geq 0.$$

If $\mu_j^* = 0$, then by letting $\mu = \mu^* + \gamma e_j$ for a scalar $\gamma \geq 0$, and the vector e_j whose jth component is 1 and the other components are 0, from the preceding relation we obtain $\partial q(\mu^*)/\partial \mu_j \leq 0$. Similarly, if $\mu_j^* > 0$, then by letting $\mu = \mu^* + \gamma e_j$ for a sufficiently small scalar γ (small enough so that $\mu^* + \gamma e_j \in M$), from the preceding relation we obtain $\partial q(\mu^*)/\partial \mu_j = 0$. Hence

$$\partial q(\mu^*)/\partial \mu_j \leq 0, \qquad \forall\ j = 1, \ldots, r,$$

$$\mu_j^* \partial q(\mu^*)/\partial \mu_j = 0, \qquad \forall\ j = 1, \ldots, r.$$

Since q is differentiable at μ^*, we have that

$$\nabla q(\mu^*) = g(x^*),$$

for some vector $x^* \in X$ such that $q(\mu^*) = L(x^*, \mu^*)$. This and the preceding two relations imply that x^* and μ^* satisfy the necessary and sufficient optimality conditions for an optimal solution-geometric multiplier pair (cf. Prop. 6.2.5). It follows that there is no duality gap, a contradiction.

8.2 (Sharpness of the Error Tolerance Estimate)

Consider the incremental subgradient method with the stepsize α and the starting point $\overline{x} = (\alpha M C_0, \alpha M C_0)$, and the following component processing order:

M components of the form $|x_1|$ [endpoint is $(0, \alpha M C_0)$],

M components of the form $|x_1 + 1|$ [endpoint is $(-\alpha M C_0, \alpha M C_0)$],

M components of the form $|x_2|$ [endpoint is $(-\alpha M C_0, 0)$],

M components of the form $|x_2 + 1|$ [endpoint is $(-\alpha M C_0, -\alpha M C_0)$],

M components of the form $|x_1|$ [endpoint is $(0, -\alpha M C_0)$],

M components of the form $|x_1 - 1|$ [endpoint is $(\alpha M C_0, -\alpha M C_0)$],

M components of the form $|x_2|$ [endpoint is $(\alpha MC_0, 0)$], and

M components of the form $|x_2 - 1|$ [endpoint is $(\alpha MC_0, \alpha MC_0)$].

With this processing order, the method returns to $\overline{x}$ at the end of a cycle. Furthermore, the smallest function value within the cycle is attained at points $(\pm\alpha MC_0, 0)$ and $(0, \pm\alpha MC_0)$, and is equal to $4MC_0 + 2\alpha M^2 C_0^2$. The optimal function value is $f^* = 4MC_0$, so that

$$\liminf_{k\to\infty} f(\psi_{i,k}) \geq f^* + 2\alpha M^2 C_0^2.$$

Since $m = 8M$ and $mC_0 = C$, we have $M^2 C_0^2 = C^2/64$, implying that

$$2\alpha M^2 C_0^2 = \frac{1}{16}\frac{\alpha C^2}{2},$$

and therefore

$$\liminf_{k\to\infty} f(\psi_{i,k}) \geq f^* + \frac{\beta\alpha C^2}{2},$$

with $\beta = 1/16$.

8.3 (A Variation of the Subgradient Method [CFM75])

At first, by induction, we show that

$$(\mu^* - \mu^k)' d^k \geq (\mu^* - \mu^k)' g^k. \tag{8.0}$$

Since $d^0 = g^0$, the preceding relation obviously holds for $k = 0$. Assume now that this relation holds for $k - 1$. By using the definition of d^k,

$$d^k = g^k + \beta^k d^{k-1},$$

we obtain

$$(\mu^* - \mu^k)' d^k = (\mu^* - \mu^k)' g^k + \beta^k (\mu^* - \mu^k)' d^{k-1}. \tag{8.1}$$

We further have

$$\begin{aligned}(\mu^* - \mu^k)' d^{k-1} &= (\mu^* - \mu^{k-1})' d^{k-1} + (\mu^{k-1} - \mu^k)' d^{k-1} \\ &\geq (\mu^* - \mu^{k-1})' d^{k-1} - \|\mu^{k-1} - \mu^k\| \|d^{k-1}\|.\end{aligned}$$

By the induction hypothesis, we have that

$$(\mu^* - \mu^{k-1})' d^{k-1} \geq (\mu^* - \mu^{k-1})' g^{k-1},$$

while by the subgradient inequality, we have that

$$(\mu^* - \mu^{k-1})' g^{k-1} \geq q(\mu^*) - q(\mu^{k-1}).$$

Combining the preceding three relations, we obtain

$$(\mu^* - \mu^k)' d^{k-1} \geq q(\mu^*) - q(\mu^{k-1}) - \|\mu^{k-1} - \mu^k\| \|d^{k-1}\|.$$

Since

$$\|\mu^{k-1} - \mu^k\| \le s^{k-1}\|d^{k-1}\|,$$

it follows that

$$(\mu^* - \mu^k)'d^{k-1} \ge q(\mu^*) - q(\mu^{k-1}) - s^{k-1}\|d^{k-1}\|^2.$$

Finally, because $0 < s^{k-1} \le \big(q(\mu^*) - q(\mu^{k-1})\big)/\|d^{k-1}\|^2$, we see that

$$(\mu^* - \mu^k)'d^{k-1} \ge 0. \tag{8.2}$$

Since $\beta^k \ge 0$, the preceding relation and equation (8.1) imply that

$$(\mu^* - \mu^k)'d^k \ge (\mu^* - \mu^k)'g^k.$$

Assuming $\mu^k \ne \mu^*$, we next show that

$$\|\mu^* - \mu^{k+1}\| < \|\mu^* - \mu^k\|, \qquad \forall\ k.$$

Similar to the proof of Prop. 8.2.1, it can be seen that this relation holds for $k = 0$. For $k > 0$, by using the nonexpansive property of the projection operation, we obtain

$$\begin{aligned}\|\mu^* - \mu^{k+1}\|^2 &\le \|\mu^* - \mu^k - s^k d^k\|^2 \\ &= \|\mu^* - \mu^k\|^2 - 2s^k(\mu^* - \mu^k)'d^k + (s^k)^2\|d^k\|^2.\end{aligned}$$

By using equation (8.1) and the subgradient inequality,

$$(\mu^* - \mu^k)'g^k \ge q(\mu^*) - q(\mu^k),$$

we further obtain

$$\begin{aligned}\|\mu^* - \mu^{k+1}\|^2 &\le \|\mu^k - \mu^*\|^2 - 2s^k(\mu^* - \mu^k)'g^k + (s^k)^2\|d^k\|^2 \\ &\le \|\mu^k - \mu^*\|^2 - 2s^k\big(q(\mu^*) - q(\mu^k)\big) + (s^k)^2\|d^k\|^2.\end{aligned}$$

Since $0 < s^k \le \big(q(\mu^*) - q(\mu^k)\big)/\|d^k\|^2$, it follows that

$$-2s^k\big(q(\mu^*) - q(\mu^k)\big) + (s^k)^2\|d^k\|^2 \le -s^k\big(q(\mu^*) - q(\mu^k)\big) < 0,$$

implying that

$$\|\mu^* - \mu^{k+1}\|^2 < \|\mu^k - \mu^*\|^2.$$

We next prove that

$$\frac{(\mu^* - \mu^k)'d^k}{\|d^k\|} \ge \frac{(\mu^* - \mu^k)'g^k}{\|g^k\|}.$$

It suffices to show that

$$\|d^k\| \le \|g^k\|,$$

since this inequality and Eq. (8.0) imply the desired relation. If $g^{k'}d^{k-1} \geq 0$, then by the definition of d^k and β^k, we have that $d^k = g^k$, and we are done, so assume that $g^{k'}d^{k-1} < 0$. We then have

$$\|d^k\|^2 = \|g^k\|^2 + 2\beta^k g^{k'}d^{k-1} + (\beta^k)^2\|d^{k-1}\|^2.$$

Since $\beta^k = -\gamma g^{k'}d^{k-1}/\|d^{k-1}\|^2$, it follows that

$$2\beta^k g^{k'}d^{k-1} + (\beta^k)^2\|d^{k-1}\|^2 = 2\beta^k g^{k'}d^{k-1} - \gamma\beta^k g^{k'}d^{k-1} = (2-\gamma)\beta^k g^{k'}d^{k-1}.$$

Furthermore, since $g^{k'}d^{k-1} < 0$, $\beta^k \geq 0$, and $\gamma \in [0,2]$, we see that

$$2\beta^k g^{k'}d^{k-1} + (\beta^k)^2\|d^{k-1}\|^2 \leq 0,$$

implying that

$$\|d^k\|^2 \leq \|g^k\|^2.$$

8.4 (Subgradient Randomization for Stochastic Programming)

The stochastic programming problem of Example 8.2.2 can be written in the following form

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^m \pi_i\big(f_0(x) + f_i(x)\big) \\ &\text{subject to} \quad x \in X, \end{aligned}$$

where

$$f_i(x) = \max_{B_i'\lambda_i \leq d_i} (b_i - Ax)'\lambda_i, \qquad i = 1, \ldots, m,$$

and the outcome i occurs with probability π_i. Assume that for each outcome $i \in \{1, ..., m\}$ and each vector $x \in \Re^n$, the maximum in the expression for $f_i(x)$ is attained at some $\lambda_i(x)$. Then, the vector $A'\lambda_i(x)$ is a subgradient of f_i at x. One possible form of the randomized incremental subgradient method is

$$x_{k+1} = P_X\big(x_k - \alpha_k(g_k + A'\lambda^k_{\omega_k})\big),$$

where g_k is a subgradient of f_0 at x_k, $\lambda^k_{\omega_k} = \lambda_{\omega_k}(x_k)$, and the random variable ω_k takes value i from the set $\{1, \ldots, m\}$ with probability π_i. The convergence analysis of Section 8.2.2 goes through in its entirety for this method, with only some adjustments in various bounding constants.

In an alternative method, we could use as components the $m + 1$ functions $f_0, f_1, \ldots, f_m$, with f_0 chosen with probability $1/2$ and each component f_i, $1, \ldots, m$, chosen with probability $\pi_i/2$.

8.5

Consider the cutting plane method applied to the following one-dimensional problem

$$\text{maximize } q(\mu) = -\mu^2,$$
$$\text{subject to } \ \mu \in [0,1].$$

Suppose that the method is started at $\mu^0 = 0$, so that the initial polyhedral approximation is $Q^1(\mu) = 0$ for all μ. Suppose also that in all subsequent iterations, when maximizing $Q^k(\mu)$, $k = 0, 1, ...,$ over $[0,1]$, we choose μ^k to be the largest of all maximizers of $Q^k(\mu)$ over $[0,1]$. We will show by induction that in this case, we have $\mu^k = 1/2^{k-1}$ for $k = 1, 2, ...$

Since $Q^1(\mu) = 0$ for all μ, the set of maximizers of $Q^1(\mu) = 0$ over $[0,1]$ is the entire interval $[0,1]$, so that the largest maximizer is $\mu^1 = 1$. Suppose now that $\mu^i = 1/2^{i-1}$ for $i = 1, ..., k$. Then

$$Q^{k+1}(\mu) = \min\{l_0(\mu), l_1(\mu), ..., l_k(\mu)\},$$

where $l_0(\mu) = 0$ and

$$l_i(\mu) = q(\mu^i) + \nabla q(\mu^i)'(\mu - \mu^i) = -2\mu^i\mu + (\mu^i)^2, \qquad i = 1, ..., k.$$

The maximum value of $Q^{k+1}(\mu)$ over $[0,1]$ is 0 and it is attained at any point in the interval $[0, \mu^k/2]$. By the induction hypothesis, we have $\mu^k = 1/2^{k-1}$, implying that the largest maximizer of $Q^{k+1}(\mu)$ over $[0,1]$ is $\mu^{k+1} = 1/2^k$.

Hence, in this case, the cutting plane method generates an infinite sequence $\{\mu^k\}$ converging to the optimal solution $\mu^* = 0$, thus showing that the method need not terminate finitely even if it starts at an optimal solution.

8.6 (Approximate Subgradient Method)

(a) We have for all $\mu \in \Re^r$

$$\begin{aligned} q(\mu) &= \inf_{x \in X}\big\{f(x) + \mu' g(x)\big\} \\ &\le f(x^k) + \mu' g(x^k) \\ &= f(x^k) + {\mu^k}' g(x^k) + g(x^k)'(\mu - \mu^k) \\ &= q(\mu^k) + \epsilon + g(x^k)'(\mu - \mu^k), \end{aligned}$$

where the last inequality follows from the equation

$$L(x^k, \mu^k) \le \inf_{x \in X} L(x, \mu^k) + \epsilon.$$

Thus $g(x^k)$ is an ϵ-subgradient of q at μ^k.

(b) For all $\mu \in M$, by using the nonexpansive property of the projection, we have

$$\begin{aligned} \|\mu^{k+1} - \mu\|^2 &\le \|\mu^k + s^k g^k - \mu\|^2 \\ &\le \|\mu^k - \mu\|^2 - 2s^k {g^k}'(\mu - \mu^k) + (s^k)^2 \|g^k\|^2, \end{aligned}$$

where

$$s^k = \frac{q(\mu^*) - q(\mu^k)}{\|g^k\|^2},$$

and $g^k \in \partial_\epsilon q(\mu^k)$. From this relation and the definition of an ϵ-subgradient we obtain

$$\|\mu^{k+1} - \mu\|^2 \leq \|\mu^k - \mu\|^2 - 2s^k\big(q(\mu) - q(\mu^k) - \epsilon\big) + (s^k)^2\|g^k\|^2, \qquad \forall\ \mu \in M.$$

Let μ^* be an optimal solution. Substituting the expression for s^k and taking $\mu = \mu^*$ in the above inequality, we have

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^k - \mu^*\|^2 - \frac{q(\mu^*) - q(\mu^k)}{\|g^k\|^2}\big(q(\mu^*) - q(\mu^k) - 2\epsilon\big).$$

Thus, if $q(\mu^k) < q(\mu^*) - 2\epsilon$, we obtain

$$\|\mu^{k+1} - \mu^*\| \leq \|\mu^k - \mu^*\|.$$