Algorithms for Convex Optimization

Suggested Solutions to A.4

1. $\omega \in (\alpha+\beta)\Omega \Rightarrow \frac{\omega}{\alpha+\beta} \in \Omega \Rightarrow \alpha\frac{\omega}{\alpha+\beta} + \beta\frac{\omega}{\alpha+\beta} \in \alpha\Omega + \beta\Omega \Rightarrow \omega \in (\alpha+\beta)\Omega \Rightarrow (\alpha+\beta)\Omega \subset \alpha\Omega + \beta\Omega$. $\omega \in \alpha\Omega + \beta\Omega \Rightarrow \exists\, \omega_\alpha, \omega_\beta \in \Omega$ *s.t.* $\omega = \alpha\omega_\alpha + \beta\omega_\beta \Rightarrow \frac{\alpha}{\alpha+\beta}\omega_\alpha + \frac{\beta}{\alpha+\beta}\omega_\beta \in \Omega$. ($\Omega$ is convex) $\Rightarrow \frac{\alpha\omega_\alpha+\beta\omega_\beta}{\alpha+\beta} = \frac{\omega}{\alpha+\beta} \in \Omega \Rightarrow \omega \in (\alpha+\beta)\Omega \Rightarrow \alpha\Omega + \beta\Omega \subset (\alpha+\beta)\Omega$.

2. Hint: f^q is the composite of $h : [0,\infty) \to [0,\infty)$ and f, where $h(x) = x^q$ is convex and increasing.

3. Hint: $f(x+y) = f(2\frac{x+y}{2}) = 2f(\frac{x}{2} + \frac{y}{2}) \le 2(f(\frac{x}{2}) + f(\frac{y}{2})) = f(x) + f(y)$.

4. Hint:

Theorem 5.12 (L-smoothness and boundedness of the Hessian). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function over $\mathbb{R}^n$. Then for a given $L \ge 0$, the following two claims are equivalent:*

(i) *f is L-smooth w.r.t. the l_p-norm ($p \in [1,\infty]$).*

(ii) *$\|\nabla^2 f(\mathbf{x})\|_{p,q} \le L$ for any $\mathbf{x} \in \mathbb{R}^n$, where $q \in [1,\infty]$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$.*

Proof. (ii) $\Rightarrow$ (i). Suppose that $\|\nabla^2 f(\mathbf{x})\|_{p,q} \le L$ for any $\mathbf{x} \in \mathbb{R}^n$. Then by the fundamental theorem of calculus, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\begin{aligned}\nabla f(\mathbf{y}) &= \nabla f(\mathbf{x}) + \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x})dt \\ &= \nabla f(\mathbf{x}) + \left(\int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))dt\right) \cdot (\mathbf{y}-\mathbf{x}).\end{aligned}$$

Then

$$\begin{aligned}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_q &= \left\|\left(\int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))dt\right) \cdot (\mathbf{y}-\mathbf{x})\right\|_q \\ &\le \left\|\int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))dt\right\|_{p,q} \|\mathbf{y}-\mathbf{x}\|_p \\ &\le \left(\int_0^1 \|\nabla^2 f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))\|_{p,q}dt\right) \|\mathbf{y}-\mathbf{x}\|_p \\ &\le L\|\mathbf{y}-\mathbf{x}\|_p,\end{aligned}$$

establishing (i).

(i) $\Rightarrow$ (ii). Suppose now that f is L-smooth w.r.t. the l_p-norm. Then by the fundamental theorem of calculus, for any $\mathbf{d} \in \mathbb{R}^n$ and $\alpha > 0$,

$$\nabla f(\mathbf{x} + \alpha\mathbf{d}) - \nabla f(\mathbf{x}) = \int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{d})\mathbf{d}dt.$$

Thus,

$$\left\| \left(\int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{d}) dt \right) \mathbf{d} \right\|_q = \|\nabla f(\mathbf{x} + \alpha\mathbf{d}) - \nabla f(\mathbf{x})\|_q \leq \alpha L \|\mathbf{d}\|_p.$$

Dividing by α and taking the limit $\alpha \to 0^+$, we obtain

$$\left\| \nabla^2 f(\mathbf{x})\mathbf{d} \right\|_q \leq L\|\mathbf{d}\|_p \text{ for any } \mathbf{d} \in \mathbb{R}^n,$$

implying that $\|\nabla^2 f(\mathbf{x})\|_{p,q} \leq L$. $\square$

A direct consequence is that for twice continuously differentiable convex functions, L-smoothness w.r.t. the l_2-norm is equivalent to the property that the maximum eigenvalue of the Hessian matrix is smaller than or equal to L.

Corollary 5.13. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable convex function over $\mathbb{R}^n$. Then f is L-smooth w.r.t. the l_2-norm if and only if $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$.*

Proof. Since f is convex, it follows that $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$. Therefore, in this case,

$$\|\nabla^2 f(\mathbf{x})\|_{2,2} = \sqrt{\lambda_{\max}((\nabla^2 f(\mathbf{x}))^2)} = \lambda_{\max}(\nabla^2 f(\mathbf{x})),$$

which, combined with Theorem 5.12, establishes the desired result. $\square$

5. Hint:

[prox of g_5] We will first assume that $\eta < \infty$. Note that $\tilde{u} = \operatorname{prox}_{g_5}(x)$ is the minimizer of

$$w(u) = \frac{1}{2}(u - x)^2$$

over $[0, \eta]$. The minimizer of w over $\mathbb{R}$ is $u = x$. Therefore, if $0 \leq x \leq \eta$, then $\tilde{u} = x$. If $x < 0$, then w is increasing over $[0, \eta]$, and hence $\tilde{u} = 0$. Finally, if $x > \eta$, then w is decreasing over $[0, \eta]$, and thus $\tilde{u} = \eta$. To conclude,

$$\operatorname{prox}_{g_5}(x) = \tilde{u} = \begin{cases} x, & 0 \leq x \leq \eta, \\ 0, & x < 0, \\ \eta, & x > \eta, \end{cases} = \min\{\max\{x, 0\}, \eta\}.$$

For $\eta = \infty$, $g_5(x) = \delta_{[0,\infty)}(x)$, and in this case, g_5 is identical to g_1 with $\mu = 0$, implying that $\operatorname{prox}_{g_5}(x) = [x]_+$, which can also be written as

$$\operatorname{prox}_{g_5}(x) = \min\{\max\{x, 0\}, \infty\}. \quad \square$$

6.

$$(1). \quad \partial f(x) = \begin{cases} -e^{-x} \\ [-1, 1] \\ e^{x} \end{cases}$$

Example 3.41 (subdifferential set of the l_1-norm function—strong result). Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$. Then $f = \sum_{i=1}^n f_i$, where $f_i(\mathbf{x}) \equiv |x_i|$. We have (see also Example 3.34)

$$\partial f_i(\mathbf{x}) = \begin{cases} \{\text{sgn}(x_i)\mathbf{e}_i\}, & x_i \neq 0, \\ [-\mathbf{e}_i, \mathbf{e}_i], & x_i = 0. \end{cases}$$

Thus, by Corollary 3.39,

$$\partial f(\mathbf{x}) = \sum_{i=1}^n \partial f_i(\mathbf{x}) = \sum_{i \in I_{\neq}(\mathbf{x})} \text{sgn}(x_i)\mathbf{e}_i + \sum_{i \in I_0(\mathbf{x})} [-\mathbf{e}_i, \mathbf{e}_i],$$

where

$$I_{\neq}(\mathbf{x}) = \{i : x_i \neq 0\}, \ I_0(\mathbf{x}) = \{i : x_i = 0\},$$

and hence

(2) Hint:

$$\boxed{\partial f(\mathbf{x}) = \{\mathbf{z} \in \mathbb{R}^n : z_i = \text{sgn}(x_i), i \in I_{\neq}(\mathbf{x}), |z_j| \leq 1, j \in I_0(\mathbf{x})\}.}$$

7.

8. .

Proof: (a) For any $y \in R^n$,

$$\begin{aligned} g^*(y) &= \max_x \{\langle y, x\rangle - g(x)\} \\ &= \max_x \left\{\langle y, x\rangle - \alpha f\left(\frac{x}{\alpha}\right)\right\} \\ &= \alpha \max_x \left\{\left\langle y, \frac{x}{\alpha}\right\rangle - f\left(\frac{x}{\alpha}\right)\right\} \\ &\overset{z \leftarrow \frac{x}{\alpha}}{=} \alpha \max_x \{\langle y, z\rangle - f(z)\} \\ &= \alpha f^*(y). \end{aligned}$$

(b) Note that

$$\text{prox}_g(x) = \arg\min_u \left\{g(u) + \frac{1}{2}\|u - x\|^2\right\} = \arg\min_u \left\{\alpha f\left(\frac{u}{\alpha}\right) + \frac{1}{2}\|u - x\|^2\right\}.$$

Making the change of variables $z = \frac{x}{\alpha}$, we can continue to write

$$\begin{aligned} \text{prox}_g(x) &= \alpha \arg\min_z \left\{\alpha f(z) + \frac{1}{2}\|\alpha z - x\|^2\right\} \\ &= \alpha \arg\min_z \left\{\alpha^2\left(\frac{f(z)}{\alpha} + \frac{1}{2}\|z - \frac{x}{\alpha}\|^2\right)\right\} \\ &= \alpha \arg\min_z \left\{\frac{f(z)}{\alpha} + \frac{1}{2}\|z - \frac{x}{\alpha}\|^2\right\} \\ &= \alpha \text{prox}_{f/\alpha}(x/\alpha). \end{aligned}$$

9.

10.

11.

12. Choose

$$f_1(x) = f_2(x) = \begin{cases} |x|, & \text{if } x \neq 0. \\ -1, & \text{otherwise .} \end{cases}$$

Then $(f_1 * f_2)(x) = x^2$ if $x \neq 0$ and $(f_1 * f_2)(x) = 1$ otherwise.

13. Just calculate the subdifferential formula. Notice for the interval $\Omega = [a, b]$, we have

$$\partial d(\bar{x}; \Omega) = \begin{cases} \{0\} & \text{if } \bar{x} \in (a, b), \\ [0, 1] & \text{if } \bar{x} = b, \\ \{1\} & \text{if } \bar{x} > b, \\ [-1, 0] & \text{if } \bar{x} = a, \\ \{-1\} & \text{if } \bar{x} < b. \end{cases}$$

Then use the sum rule of subdifferential calculus.

14. .

Proof Suppose $a_{i\max} = \max\{a_1, a_2, \cdots, a_k\}$, on the one hand, if we take $\lambda_{i\max} = 1$, $\lambda_{i \neq i\max} = 0$, then we have

$$\max\{a_1, a_2, \cdots, a_k\} \leq \max_{\lambda \in \Delta_k} \sum_{i=1}^{k} \lambda_i a_i. \tag{1.1}$$

On the other hand, since

$$\max_{\lambda \in \Delta_k} \sum_{i=1}^{k} \lambda_i a_i \leq \{\sum_{i=1}^{k} \lambda_i a_{i\max}, \mid \lambda \in \Delta_k\} = a_{i\max},$$

and thus we have

$$\max\{a_1, a_2, \cdots, a_k\} \geq \max_{\lambda \in \Delta_k} \sum_{i=1}^{k} \lambda_i a_i. \tag{1.2}$$

By inequalities (1.1) and (1.2), we complete the proof. ☐

15.

16.

17.

18.

19. Counterexample: $f(x) = x^2$ with $\Omega = [1, 4]$, then $f^{-1}(\Omega) = [-2, -1] \cup [1, 2]$ is not convex.

20. Hint: For any $\omega \in \Omega$, the function $|\cdot - \omega|$ is convex. Recall that the maximum (or supremum) function is convex.

21. (c). Recall Theorem (convexity under partial minimization).

(d). Recall second prox theorem and its corollary and Theorem (smoothness of the Moreau envelope)

$x^* \in \arg\min f \Leftrightarrow x^* = \mathrm{prox}_f(x^*) \Leftrightarrow x^* = \mathrm{prox}_{\mu f}(x^*) \Leftrightarrow \nabla M_f^\mu(x^*) = 0 \Leftrightarrow x^* \in \arg\min M_f^\mu$.

(e). For any $x^* \in \arg\min M_f^\mu$, recall the alternative expression of M_f^μ:

$$\min_{x\in\mathbb{R}^n} M_f^\mu(x) = M_f^\mu(x^*) = f\left(\mathrm{prox}_{\mu f}(x^*)\right) + \frac{1}{2\mu}\left\|x^* - \mathrm{prox}_{\mu f}(x^*)\right\|^2 = f(x^*) = \min_{x\in\mathbb{R}^n} f(x).$$

22.

23. Hint:

$$\frac{\lambda x + (1-\lambda)y}{\lambda t_x + (1-\lambda)t_y} = \frac{\lambda t_x}{\lambda t_x + (1-\lambda)t_y}\frac{x}{t_x} + \frac{(1-\lambda)t_y}{\lambda t_x + (1-\lambda)t_y}\frac{y}{t_y}$$

24.

$$\begin{aligned} g^*(y,s) &= \sup_{x/t\in\mathrm{dom}(f),t>0}\{\langle y,x\rangle + st - g(x,t)\} \\ &= \sup_{z\in\mathrm{dom}(f),t>0}\{t\langle y,z\rangle + st - tf(z)\} \\ &= \sup_{t>0}\{t\left(f^*(y)+s\right)\} \\ &= \begin{cases} 0, & \text{if } f^*(y)+s \le 0, \\ \infty, & \text{otherwise} \end{cases} \end{aligned}$$