Continuous Optimisation: Chpt 4 Exercises

October 21, 2016

1. Find the gradient of the Lagrangian function.

Solution: We have

$$
\begin{aligned}
L(\mathbf{x}, \mathbf{y}) &= f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}), \\
\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) &= \nabla_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} y_i \nabla_{\mathbf{x}} g_i(\mathbf{x}), \\
\frac{\partial L}{\partial y_i}(\mathbf{x}, \mathbf{y}) &= g_i(\mathbf{x}) \quad \text{for all } i, \\
\nabla_{\mathbf{y}} L(\mathbf{x}, \mathbf{y}) &= \begin{pmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{pmatrix}, \\
\nabla L(\mathbf{x}, \mathbf{y}) &= \begin{pmatrix} \nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \\ \nabla_{\mathbf{y}} L(\mathbf{x}, \mathbf{y}) \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} y_i \nabla_{\mathbf{x}} g_i(\mathbf{x}) \\ g_1(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{pmatrix}
\end{aligned}
$$

2. Assuming $f, g_1, \ldots, g_m \in C^2$, find the Hessian of the Lagrangian function.

Solution: We have

$$
\begin{aligned}
L(\mathbf{x}, \mathbf{y}) &= f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}), \\
\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) &= \nabla_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} y_i \nabla_{\mathbf{x}} g_i(\mathbf{x}),
\end{aligned}
$$

$$
\begin{aligned}
\nabla^2_{\mathbf{x}} L(\mathbf{x},\mathbf{y}) &= \nabla^2_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} y_i \nabla^2_{\mathbf{x}} g_i(\mathbf{x}),\\
\nabla_{\mathbf{y}}\nabla_{\mathbf{x}} L(\mathbf{x},\mathbf{y}) &= \begin{pmatrix} \nabla_{\mathbf{x}} g_1(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{x}} g_m(\mathbf{x}) \end{pmatrix},\\
\nabla_{\mathbf{y}} L(\mathbf{x},\mathbf{y}) &= \begin{pmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{pmatrix},\\
\nabla^2_{\mathbf{y}} L(\mathbf{x},\mathbf{y}) &= 0,\\
\nabla_{\mathbf{x}}\nabla_{\mathbf{y}} L(\mathbf{x},\mathbf{y}) &= \begin{pmatrix} \nabla_{\mathbf{x}} g_1(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{x}} g_m(\mathbf{x}) \end{pmatrix},\\
\nabla L(\mathbf{x},\mathbf{y}) &= \begin{pmatrix} \nabla_{\mathbf{x}} L(\mathbf{x},\mathbf{y}) \\ \nabla_{\mathbf{y}} L(\mathbf{x},\mathbf{y}) \end{pmatrix},\\
\nabla^2 L(\mathbf{x},\mathbf{y}) &= \begin{pmatrix} \nabla^2_{\mathbf{x}} L(\mathbf{x},\mathbf{y}) & \nabla_{\mathbf{x}}\nabla_{\mathbf{y}} L(\mathbf{x},\mathbf{y}) \\ \nabla_{\mathbf{y}}\nabla_{\mathbf{x}} L(\mathbf{x},\mathbf{y}) & \nabla^2_{\mathbf{y}} L(\mathbf{x},\mathbf{y}) \end{pmatrix}\\
&= \begin{pmatrix} \nabla^2_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} y_i \nabla^2_{\mathbf{x}} g_i(\mathbf{x}) & \nabla_{\mathbf{x}} g_1(\mathbf{x}) & \cdots & \nabla_{\mathbf{x}} g_m(\mathbf{x}) \\ \nabla_{\mathbf{x}} g_1(\mathbf{x})^{\mathsf{T}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \nabla_{\mathbf{x}} g_m(\mathbf{x})^{\mathsf{T}} & 0 & \cdots & 0 \end{pmatrix}
\end{aligned}
$$

3. For $f, g_1, \dots, g_m \in C^2$, give a necessary and sufficient condition for L to be convex function.

Solution: We have that

$$
\begin{aligned}
L \text{ is convex} \quad &\Leftrightarrow \quad \nabla_{\mathbf{x}} g_i(\mathbf{x}) = \mathbf{0} \text{ for all } \mathbf{x} \in \mathbb{R}^n,\ i = 1, \dots, m\\
&\qquad \text{and } 0 \preceq \nabla^2_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} y_i \nabla^2_{\mathbf{x}} g_i(\mathbf{x}) = \nabla^2_{\mathbf{x}} f(\mathbf{x})\\
&\Leftrightarrow \quad g_i \text{ is a constant function for all } i = 1, \dots, m\\
&\qquad \text{and } 0 \preceq \nabla^2_{\mathbf{x}} f(\mathbf{x})\\
&\Leftrightarrow \quad g_i \text{ is a constant function for all } i = 1, \dots, m\\
&\qquad \text{and } f \text{ is a convex function.}
\end{aligned}
$$

Having g_i as constant function for all i means that either $\mathcal{F} = \emptyset$ or $\mathcal{F} = \mathbb{R}^n$. Therefore L is only convex in what are effectively unconstrained convex problems.

4. The following two primal problems are equivalent.

$$\begin{aligned} \min_x \quad & -x^2 \\ \text{s.t.} \quad & -1 \le x \le 1 \\ & x \in \mathbb{R} \end{aligned} \qquad\qquad \begin{aligned} \min_x \quad & -x^2 \\ \text{s.t.} \quad & x^2 \le 1 \\ & x \in \mathbb{R} \end{aligned}$$

What are their optimal values and optimal solutions?

What are the functions ψ for these problems?

What are the optimal values and optimal solutions to their Lagrangian dual problems?

Solution:

1. For the first problem:
   The optimal value is -1 and the optimal solutions are $x^* = \pm 1$.

$$\begin{aligned} f(x) &= -x^2, \qquad g_1(x) = -x - 1, \qquad g_2(x) = x - 1, \\ L(x; y_1, y_2) &= -x^2 + y_1(-x-1) + y_2(x-1) = -x^2 + x(y_1 - y_2) - (y_1 + y_2), \\ \psi(y_1, y_2) &= \inf_x \{L(x; y_1, y_2)\} = -\infty, \\ \sup_{\mathbf{y} \in \mathbb{R}^2_+} \psi(y_1, y_2) &= -\infty. \end{aligned}$$

   Although this is a valid lower bound, it is a rather trivial one.

2. For the second problem:
   The optimal value is -1 and the optimal solutions are $x^* = \pm 1$.

$$\begin{aligned} f(x) &= -x^2, \qquad g(x) = x^2 - 1, \\ L(x; y) &= -x^2 + y(x^2 - 1) = (y-1)x^2 - y, \\ \psi(y_1, y_2) &= \inf_x \{L(x; y_1, y_2)\} = \begin{cases} -y & \text{if } y \ge 1 \\ -\infty & \text{if } y < 1 \end{cases}, \\ \sup_{y \ge 0} \psi(y) &= \sup_{y \ge 1} \{-y\} = -1. \end{aligned}$$

   This lower bound is tight.

5. For $f, g_1, \ldots, g_m$ convex and $\mathcal{C} \subseteq \mathbb{R}^n$ convex and Slater's condition holding, show that

the following set is convex:

$$\mathcal{U} = \left\{ \begin{pmatrix} u_0 \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}^{m+1} : \exists \mathbf{x} \in \mathcal{C} \text{ s. t. } f(\mathbf{x}) < u_0, \begin{array}{c} g_i(\mathbf{x}) \leq u_i \ \forall i \in J_r, \\ g_j(\mathbf{x}) = u_j \ \forall j \in J_s \end{array} \right\}$$

Solution: Consider arbitrary $\begin{pmatrix} u_0 \\ \mathbf{u} \end{pmatrix}$, $\begin{pmatrix} v_0 \\ \mathbf{v} \end{pmatrix} \in \mathcal{U}$ and $\theta \in [0,1]$. We want to show that $\begin{pmatrix} \theta u_0 + (1-\theta)v_0 \\ \theta \mathbf{u} + (1-\theta)\mathbf{v} \end{pmatrix} \in \mathcal{U}$.

There exists $\mathbf{w}, \mathbf{x} \in \mathcal{C}$ such that

$$\begin{array}{lll} f(\mathbf{x}) < u_0, & f(\mathbf{w}) < v_0, & \\ g_i(\mathbf{x}) \leq u_i & g_i(\mathbf{w}) \leq v_i & \forall i \in J_r, \\ g_j(\mathbf{x}) = u_j & g_j(\mathbf{w}) = v_j & \forall j \in J_s. \end{array}$$

Then letting $\mathbf{z} = \theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathcal{C}$ we have the following (recalling that $f, g_1, \ldots, g_m$ are convex, and, as Slater's condition holds, g_i is affine for all $i \in J_s$):

$$\begin{array}{ll} f(\mathbf{z}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{w}) < \theta u_0 + (1-\theta) v_0, & \\ g_i(\mathbf{z}) \leq \theta g_i(\mathbf{x}) + (1-\theta) g_i(\mathbf{w}) \leq \theta u_i + (1-\theta) v_i & \forall i \in J_r, \\ g_j(\mathbf{z}) = \theta g_j(\mathbf{x}) + (1-\theta) g_j(\mathbf{w}) = \theta u_j + (1-\theta) v_j & \forall j \in J_s. \end{array}$$

6. Show that the following systems can not be simultaneously true when (C) is convex, Slater's condition holds and $\gamma \in \mathbb{R}$:

   1. $\exists \mathbf{x} \in \mathcal{C}$ s.t. $f(\mathbf{x}) < \gamma$, $g_i(\mathbf{x}) \leq 0$ for all i.
   2. $\exists \mathbf{y} \in \mathbb{R}^m_+$ s.t. $f(\mathbf{x}) + \sum_{i=1}^m y_i g_i(\mathbf{x}) \geq \gamma$ for all $\mathbf{x} \in \mathcal{C}$.

Solution: If both systems are simultaneously true for some $\mathbf{x} \in \mathcal{C}$ and $\mathbf{y} \in \mathbb{R}^m_+$ then we get the contradiction

$$\gamma \leq \underbrace{f(\mathbf{x})}_{<\gamma} + \sum_{i=1}^m \underbrace{y_i}_{\geq 0} \underbrace{g_i(\mathbf{x})}_{\leq 0} < \gamma$$

7. For $f, g_i : \mathcal{C} \to \mathbb{R}$ for $i \in J_r \cup J_s$, $\mathcal{C} \neq \emptyset$

$$\mathcal{U} := \left\{ (u_0, \mathbf{u}) \in \mathbb{R}^{m+1} : \exists \mathbf{x} \in \mathcal{C} \text{ s. t. } f(\mathbf{x}) < u_0, \begin{array}{c} g_i(\mathbf{x}) \leq u_i \ \forall i \in J_r, \\ g_j(\mathbf{x}) = u_j \ \forall j \in J_s \end{array} \right\},$$

and $(y_0, \mathbf{y}) \in \mathbb{R}^{m+1}$ such that

$$y_0\gamma \leq y_0 u_0 + \sum_{i=1}^{m} y_i u_i \qquad \text{for all } (u_0, \mathbf{u}) \in \mathcal{U}.$$

Show that

I. $y_i \geq 0$ for all $i \in \{0\} \cup J_r$.

II. $y_0\gamma \leq y_0 f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{C}$.

Solution:

I. Consider an arbitrary $\mathbf{x} \in \mathcal{C}$, let $u_0 = f(\mathbf{x}) + 1$ and let $u_i = g_i(\mathbf{x})$ for all i. Also let $\mu = y_0\gamma - y_0 u_0 - \sum_{i=1}^{m} y_i u_i$

For all $\lambda > 0$ we have $(u_0 + \lambda, \mathbf{u}) \in \mathcal{U}$ and thus

$$y_0\gamma \leq y_0(u_0 + \lambda) + \sum_{i=1}^{m} y_i u_i = y_0\lambda + y_0\gamma - \mu,$$

$$\frac{\mu}{\lambda} \leq y_0.$$

Considering the limit as $\lambda \to \infty$, this implies that $y_0 \geq 0$.

Similarly, considering an arbitrary $j \in J_r$, for all $\lambda > 0$ we have $(u_0, \mathbf{u} + \lambda \mathbf{e}_j) \in \mathcal{U}$, where $\mathbf{e}_j \in \mathbb{R}^m$ is the unit vector with jth entry equal to one and all other entries equal to zero. Therefore for all $\lambda > 0$ we have

$$y_0\gamma \leq y_0 u_0 + \sum_{i \neq j} y_i u_i + y_j(u_j + \lambda) = y_j\lambda + y_0\gamma - \mu,$$

$$\frac{\mu}{\lambda} \leq y_j.$$

Considering the limit as $\lambda \to \infty$, this implies that $y_j \geq 0$.

II. Consider an arbitrary $\mathbf{x} \in \mathcal{C}$ and $\lambda > 0$, let $u_0 = f(\mathbf{x}) + \lambda$ and let $u_i = g_i(\mathbf{x})$ for all i. Then $(u_0, \mathbf{u}) \in \mathcal{U}$ and thus

$$y_0\gamma \leq y_0 u_0 + \sum_{i=1}^{m} y_i u_i = \lambda y_0 + y_0 f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}).$$

Considering $\lambda \to 0$, this implies

$$y_0\gamma \leq y_0 f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}).$$

8. Analyse the following problems, determining whether they are convex, whether Slater's condition holds, what the optimal values to the primal and dual problems are, whether the optimal values are attained and whether there is strong duality:

   1. $\min_{\mathbf{x}\in\mathbb{R}^2}\{x_1^2 + x_2^2 : x_1 + x_2 \geq 1\}$
   2. $\min_{\mathbf{x}\in\mathbb{R}^2}\{x_2 : \exp(x_1) \leq x_2\}$
   3. $\min_{x\in\mathbb{R}}\{x : x^2 \leq 0\}$
   4. $\min_{x\in\mathbb{R}}\{(x-1)^2 : x^3 \geq 4x\}$

Solution:

1. We have

$$\begin{aligned} f(\mathbf{x}) &= x_1^2 + x_2^2, & g(\mathbf{x}) &= 1 - x_1 - x_2, \\ \nabla^2 f(\mathbf{x}) &= 2I \succ 0, & \nabla^2 g(\mathbf{x}) &= 0. \end{aligned}$$

Therefore the problem is convex. Slater's condition holds at $\mathbf{x} = \mathbf{1}$ (and also at any $\mathbf{x} \in \mathbb{R}^2$ such that $x_1 + x_2 \geq 1$ as g is affine). By considering a graph of the problem it can be seen that the optimal value is $\frac{1}{2}$, which is attained at $(x_1^*, x_2^*) = \left(\frac{1}{2}, \frac{1}{2}\right)$. We have

$$\begin{aligned} L(\mathbf{x}, y) &= x_1^2 + x_2^2 + y(1 - x_1 - x_2) = \left(x_1 - \tfrac{1}{2}y\right)^2 + \left(x_2 - \tfrac{1}{2}y\right)^2 + y - \tfrac{1}{2}y^2, \\ \psi(y) &= \inf_{\mathbf{x}\in\mathbb{R}^2}\{L(\mathbf{x}, y)\} = y - \tfrac{1}{2}y^2 = -\tfrac{1}{2}(y-1)^2 + \tfrac{1}{2}, \\ \sup_{y\geq 0}\psi(y) &= \sup_{y\geq 0}\{-\tfrac{1}{2}(y-1)^2 + \tfrac{1}{2}\} = \tfrac{1}{2} \end{aligned}$$

Therefore the dual optimal value is $\frac{1}{2}$. The optimal values to both the primal and dual problems are attained and there is strong duality.

2. We have

$$\begin{aligned} f(\mathbf{x}) &= x_2, & g(\mathbf{x}) &= \exp(x_1) - x_2, \\ \nabla^2 f(\mathbf{x}) &= 0, & \nabla^2 g(\mathbf{x}) &= \begin{pmatrix} \exp(x_1) & 0 \\ 0 & 0 \end{pmatrix} \succeq 0. \end{aligned}$$

Therefore the problem is convex. Slater's condition holds at $\mathbf{x} = 2\mathbf{e}_2$ (and also at any $\mathbf{x} \in \mathbb{R}^2$ such that $\exp(x_1) < x_2$). The optimal value is 0, but this is not attained (it is approached by $(-t, \exp(-t))$ as $t \to \infty$). We have

$$\begin{aligned} L(\mathbf{x}, y) &= x_2 + y(\exp(x_1) - x_2) = x_2(1-y) + y\exp(x_1), \\ \psi(y) &= \inf_{\mathbf{x}\in\mathbb{R}^2}\{L(\mathbf{x}, y)\} = \begin{cases} 0 & \text{if } y = 1 \\ -\infty & \text{if } y \neq 1, \end{cases} \\ \sup_{y\geq 0}\psi(y) &= 0 \end{aligned}$$

Therefore the dual optimal value is 0. The optimal value to the dual problem is attained and there is strong duality.

3. We have

$$f(x) = x, \qquad g(x) = x^2,$$
$$f''(x) = 0, \qquad g''(x) = 1 > 0.$$

Therefore the problem is convex. Slater's condition does not hold. The optimal value is 0, and this is attained at $x^* = 0$. We have

$$L(\mathbf{x}, y) = x + yx^2 = y(x + \tfrac{1}{2y})^2 - \tfrac{1}{4y},$$
$$\psi(y) = \inf_{\mathbf{x} \in \mathbb{R}} \{L(\mathbf{x}, y)\} = \begin{cases} -(4y)^{-1} & \text{if } y > 0 \\ -\infty & \text{if } y \leq 0, \end{cases}$$
$$\sup_{y \geq 0} \psi(y) = \sup_{y > 0} -(4y)^{-1} = 0$$

Therefore the dual optimal value is 0. The optimal value to the dual problem is not attained but there is strong duality.

4. We have

$$f(x) = (x-1)^2, \qquad g(x) = 4x - x^3 = x(2-x)(2+x),$$
$$f''(x) = 2 > 0, \qquad g''(x) = -6x.$$

Therefore the problem is not convex. Slater's condition holds at $x = -1$, but as the problem is nonconvex, that is unimportant. We have $\mathcal{F} = [-2, 0] \cup [2, \infty)$ and thus the optimal value is 1, and this is attained at $x^* \in \{0, 2\}$. We have

$$L(\mathbf{x}, y) = (x-1)^2 + y(4x - x^3),$$
$$\psi(y) = \inf_{\mathbf{x} \in \mathbb{R}} \{L(\mathbf{x}, y)\} = \begin{cases} 0 & \text{if } y = 0 \\ -\infty & \text{if } y \neq 0, \end{cases}$$
$$\sup_{y \geq 0} \psi(y) = 0$$

Therefore the dual optimal value is 0. The optimal value to the dual problem is attained but there is not strong duality.

9. For the following primal problem, find the optimal values and optimal solutions to the

primal problem, its Lagrangian dual and its Wolfe-dual.

$$\begin{aligned} \min_x \quad & \sin x \\ \text{s.t.} \quad & x \le 0 \\ & x \in \mathbb{R} \end{aligned}$$

Solution: The optimal value is -1, which is attained at $x^* = -\pi/2$.

$$\begin{aligned} f(x) &= \sin x, \qquad g(x) = x, \\ L(x,y) &= \sin x + yx, \\ \psi(y) &= \begin{cases} -1 & \text{if } y = 0 \\ -\infty & \text{if } y \ne 0, \end{cases} \\ \sup_{y \ge 0} \psi(y) &= -1 \end{aligned}$$

The Lagrangian dual is -1, and thus there is strong duality (even though the problem is not convex, and thus this would not generally be the case).

For the Wolfe-dual we have

$$\nabla_x L(x,y) = \cos x + y,$$
$$\sup_{x,y} \left\{ L(x,y) : \begin{array}{c} \nabla_x L(x,y) = 0, \\ y \ge 0, \quad x \in \mathbb{R} \end{array} \right\} = \sup_{x,y} \left\{ \sin x + yx : \begin{array}{c} \cos x + y = 0, \\ y \ge 0, \quad x \in \mathbb{R} \end{array} \right\}.$$

For all $t \in \mathbb{N}$ we have that $(\pi + 2\pi t, 1)$ is feasible for this problem, and the value of the objective function at these points is $\sin(\pi + 2\pi t) + (\pi + 2\pi t) = \pi + 2\pi t$. Considering this as $t \to \infty$ implies that the optimal value of the Wolfe-dual is equal to $+\infty$, and thus does not provide a lower bound (as the primal problem is nonconvex).

10. For the parameters $\mathbf{c}, \mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$, give an explicit formulation for the Wolfe-Dual to

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\mathsf{T} \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\mathsf{T} \mathbf{x} \ge b_i \text{ for } i = 1, \dots, m, \\ & \mathbf{x} \in \mathbb{R}^m. \end{aligned}$$

Solution: This is a convex problem, and thus the Wolfe-dual will provide a lower bound. We have

$$f(\mathbf{x}) = \mathbf{c}^\mathsf{T}\mathbf{x}, \qquad g_i(\mathbf{x}) = b_i - \mathbf{a}_i^\mathsf{T}\mathbf{x} \qquad \text{for } i = 1, \ldots, m,$$
$$L(\mathbf{x}, \mathbf{y}) = \mathbf{c}^\mathsf{T}\mathbf{x} + \sum_{i=1}^{m} y_i(b_i - \mathbf{a}_i^\mathsf{T}\mathbf{x}), \qquad \nabla_\mathbf{x} L(\mathbf{x}, \mathbf{y}) = \mathbf{c} - \sum_{i=1}^{m} y_i \mathbf{a}_i.$$

The Wolfe-dual is thus

$$\begin{aligned}
&\sup_{\mathbf{x},\mathbf{y}} \left\{ L(\mathbf{x}, \mathbf{y}) : \begin{array}{c} \nabla_\mathbf{x} L(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \\ \mathbf{y} \in \mathbb{R}^m_+, \quad \mathbf{x} \in \mathbb{R}^n \end{array} \right\} \\
&= \sup_{\mathbf{x},\mathbf{y}} \left\{ \mathbf{c}^\mathsf{T}\mathbf{x} + \sum_{i=1}^{m} y_i(b_i - \mathbf{a}_i^\mathsf{T}\mathbf{x}) : \begin{array}{c} \mathbf{c} - \sum_{i=1}^{m} y_i \mathbf{a}_i = \mathbf{0}, \\ \mathbf{y} \in \mathbb{R}^m_+, \quad \mathbf{x} \in \mathbb{R}^n \end{array} \right\} \\
&= \sup_{\mathbf{x},\mathbf{y}} \left\{ \left(\mathbf{c} - \sum_{i=1}^{m} y_i \mathbf{a}_i \right)^\mathsf{T} \mathbf{x} + \mathbf{b}^\mathsf{T}\mathbf{y} : \begin{array}{c} \mathbf{c} = \sum_{i=1}^{m} y_i \mathbf{a}_i, \\ \mathbf{y} \in \mathbb{R}^m_+, \quad \mathbf{x} \in \mathbb{R}^n \end{array} \right\} \\
&= \sup_{\mathbf{y}} \left\{ \mathbf{b}^\mathsf{T}\mathbf{y} : \begin{array}{c} \mathbf{c} = \sum_{i=1}^{m} y_i \mathbf{a}_i, \\ \mathbf{y} \in \mathbb{R}^m_+ \end{array} \right\}
\end{aligned}$$

11. For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$, let ν be the optimal value to the following problem:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \quad f(\mathbf{x}) \qquad \text{s.t.} \quad -1 \le x_i \le 1 \text{ for all } i = 1, \ldots, n.$$

Show that the Wolfe-dual to this problem is equivalent to

$$\sup_{\mathbf{x},\mathbf{u},\mathbf{v}} \left\{ f(\mathbf{x}) - \mathbf{x}^\mathsf{T} \nabla f(\mathbf{x}) - \mathbf{1}^\mathsf{T}\mathbf{u} - \mathbf{1}^\mathsf{T}\mathbf{v} : \begin{array}{c} \nabla f(\mathbf{x}) = \mathbf{v} - \mathbf{u}, \\ \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^n_+ \end{array} \right\}.$$

Solution: We can write this problem as

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) \qquad \text{s.t.} \quad g_i(\mathbf{x}) \le 0 \text{ for all } i = 1, \ldots, 2n,$$

where for $i = 1, \ldots, n$ we have

$$g_i(\mathbf{x}) = x_i - 1, \qquad g_{n+i}(\mathbf{x}) = -x_i - 1.$$

We then have the following, where for the sake of simplicity we split $\mathbf{y} \in \mathbb{R}^{2n}$ as $\mathbf{y}^\mathsf{T} = (\mathbf{u}^\mathsf{T},\ \mathbf{v}^\mathsf{T})$ where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n_+$, and we let $\mathbf{1} \in \mathbb{R}^n$ be the all ones vector:

$$\begin{aligned} L(\mathbf{x}, \mathbf{y}) &= f(\mathbf{x}) + \sum_{i=1}^{n} y_i(x_i - 1) + \sum_{i=1}^{n} y_{n+i}(-x_i - 1) \\ &= f(\mathbf{x}) + \mathbf{x}^\mathsf{T}(\mathbf{u} - \mathbf{v}) - \mathbf{1}^\mathsf{T}\mathbf{u} - \mathbf{1}^\mathsf{T}\mathbf{v}, \\ \nabla_\mathbf{x} L(\mathbf{x}, \mathbf{y}) &= \nabla_\mathbf{x} f(\mathbf{x}) + \mathbf{u} - \mathbf{v}. \end{aligned}$$

The Wolfe-dual is then given by the following:

$$\begin{aligned} &\sup_{\mathbf{x},\mathbf{y}} \left\{ L(\mathbf{x}, \mathbf{y}) : \begin{array}{l} \nabla_\mathbf{x} L(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \\ \mathbf{y} \in \mathbb{R}^{2n}_+, \quad \mathbf{x} \in \mathbb{R}^n \end{array} \right\} \\ &= \sup_{\mathbf{x},\mathbf{u},\mathbf{v}} \left\{ f(\mathbf{x}) + \mathbf{x}^\mathsf{T}(\mathbf{u} - \mathbf{v}) - \mathbf{1}^\mathsf{T}\mathbf{u} - \mathbf{1}^\mathsf{T}\mathbf{v} : \begin{array}{l} \nabla f(\mathbf{x}) + \mathbf{u} - \mathbf{v} = \mathbf{0}, \\ \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^n_+ \end{array} \right\} \\ &= \sup_{\mathbf{x},\mathbf{u},\mathbf{v}} \left\{ f(\mathbf{x}) - \mathbf{x}^\mathsf{T}\nabla f(\mathbf{x}) - \mathbf{1}^\mathsf{T}\mathbf{u} - \mathbf{1}^\mathsf{T}\mathbf{v} : \begin{array}{l} \mathbf{v} - \mathbf{u} = \nabla f(\mathbf{x}), \\ \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^n_+ \end{array} \right\}. \end{aligned}$$

12. For a fixed parameter $V > 0$ consider the problem

$$\begin{aligned} \min_{l,b,h} \quad & 2(lb + bh + lh) \\ \text{s. t.} \quad & lbh \geq V \\ & l, b, h \in \mathbb{R}_+. \end{aligned} \tag{1}$$

N.B. This is the problem of minimising the surface area of a box, given that it must have a volume greater than or equal to V.

i) Transform the problem by introducing new variables to obtain:

$$\begin{aligned} \min_{\mathbf{x}} \quad & 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1) \\ \text{s. t.} \quad & x_1 + x_2 + x_3 \geq \ln V \\ & \mathbf{x} \in \mathbb{R}^3 \end{aligned} \tag{2}$$

ii) Prove that problem (2) is convex and satisfies Slater's regularity condition.

iii) Show that the Lagrange dual of problem (2) is:

$$\max_{y \geq 0} \left\{ \left(\tfrac{3}{2} + \ln(V)\right)y - \tfrac{3}{2}y\ln\left(\tfrac{1}{4}y\right) \right\}. \tag{3}$$

iv) Show that the Wolfe dual of problem (2) is equivalent to its Lagrange dual.

v) Use the KKT conditions of problem (2) to show that the cube ($l = b = h = V^{1/3}$) is the optimal solution of problem (1).

vi) Use the dual problem (3) to derive the same result as in part v).

Solution:

i) For any feasible (l, b, h) we have $lbh \geq V$ and $l, b, h \in \mathbb{R}_+$. Therefore $l, b, h > 0$ and there exists $\mathbf{x} \in \mathbb{R}^3$ such that $l = \exp(x_1)$, $b = \exp(x_2)$ and $h = \exp(x_3)$. Making this substitution in problem (1) and simplifying gives us problem (2).

ii) For this problem we have $n = 3$, $m = 1$,

$$f(\mathbf{x}) = 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1),$$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2\exp(x_1 + x_2) + 2\exp(x_3 + x_1) \\ 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) \\ 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1) \end{pmatrix},$$

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2\exp(x_1 + x_2) + 2\exp(x_3 + x_1) & 2\exp(x_1 + x_2) & 2\exp(x_2 + x_3) \\ 2\exp(x_1 + x_2) & 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) & 2\exp(x_2 + x_3) \\ 2\exp(x_3 + x_1) & 2\exp(x_2 + x_3) & 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1) \end{pmatrix}$$

$$= 2\exp(x_1 + x_2) \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}^{\mathsf{T}} + 2\exp(x_2 + x_3) \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}^{\mathsf{T}} + 2\exp(x_3 + x_1) \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^{\mathsf{T}} \succeq 0.$$

$$g_1(\mathbf{x}) = \ln V - x_1 - x_2 - x_3, \quad \nabla g_1(\mathbf{x}) = -\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \nabla^2 g_1(\mathbf{x}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \succeq 0.$$

Therefore the problem is convex. The point $\mathbf{z}_1 = \frac{1}{3}\ln V \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is a Slater point.

The point $\mathbf{z}_2 = \left(1 + \frac{1}{3}\ln V\right) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is an Ideal Slater point.

iii) We have

$$L(\mathbf{x}; y) = 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1) + y(\ln V - x_1 - x_2 - x_3).$$

We want to find $\psi(y) = \inf_{\mathbf{x} \in \mathbb{R}^3} L(\mathbf{x}; y)$. For fixed $y \in \mathbb{R}$ we have that $L(\mathbf{x}; y)$ is a convex function of $\mathbf{x}$. Therefore, if there is a point $\mathbf{x}^y \in \mathbb{R}^3$ such that

$\nabla_{\mathbf{x}} L(\mathbf{x}^y; y) = \mathbf{0}$, then this will be a global minimiser of $L(\mathbf{x}; y)$ with respect to $\mathbf{x}$, and we will have $\psi(y) = L(\mathbf{x}^y; y)$.

We have

$$\nabla_{\mathbf{x}} L(\mathbf{x}; y) = \begin{pmatrix} 2\exp(x_1 + x_2) + 2\exp(x_3 + x_1) - y \\ 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) - y \\ 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1) - y \end{pmatrix}$$

Therefore $\nabla_{\mathbf{x}} L(\mathbf{x}^y; y) = \mathbf{0}$ if and only if $\exp(x_1^y + x_2^y) = \exp(x_2^y + x_3^y) = \exp(x_3^y + x_1^y) = \frac{1}{4}y$, or equivalently $x_1^y = x_2^y = x_3^y = \frac{1}{2}\ln\left(\frac{1}{4}y\right)$ and $y > 0$.

Therefore, for $y > 0$, we have

$$\begin{aligned}\psi(y) &= 6\left(\tfrac{1}{4}y\right) + y\left(\ln V - \tfrac{3}{2}\ln\left(\tfrac{1}{4}y\right)\right) \\ &= \left(\tfrac{3}{2} + \ln(V)\right)y - \tfrac{3}{2}y\ln\left(\tfrac{1}{4}y\right) \\ &= \left(\tfrac{3}{2} + \ln(8V)\right)y - \tfrac{3}{2}y\ln(y).\end{aligned}$$

For $y = 0$ we have $L(\mathbf{x}; 0) = 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1)$ and $\psi(0) = 0$ (which is approached for $\mathbf{x}^y = \mu\mathbf{1}$ with $\mu \to -\infty$). It can then be observed that $\psi(0) = 0 = \lim_{y\to 0^+} \psi(y)$, giving us the required Lagrange dual.

iv) This follows directly from how we calculated the Lagrange dual. In particular, the Wolfe dual is equal to

$$\sup_{\mathbf{x},y} \left\{ L(\mathbf{x}; y) : \begin{array}{c} \nabla_{\mathbf{x}} L(\mathbf{x}; y) = \mathbf{0} \\ \mathbf{x} \in \mathbb{R}^3, \quad y \geq 0 \end{array} \right\} = \sup_{y} \left\{ L(\mathbf{x}^y; y) : y > 0 \right\},$$

which is equivalent to problem (3), where $\mathbf{x}^y$ is as given in the proof of the previous part.

v) The KKT conditions are

$$x_1 + x_2 + x_3 \geq \ln V, \tag{4}$$

$$y \geq 0, \tag{5}$$

$$y(\ln V - x_1 - x_2 - x_3) = 0, \tag{6}$$

$$\begin{pmatrix} 2\exp(x_1 + x_2) + 2\exp(x_3 + x_1) \\ 2\exp(x_1 + x_2) + 2\exp(x_2 + x_3) \\ 2\exp(x_2 + x_3) + 2\exp(x_3 + x_1) \end{pmatrix} = y \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{7}$$

We have that (7) holds if and only if $y > 0$ and $x_1 = x_2 = x_3 = \frac{1}{2}\ln\left(\frac{1}{4}y\right)$. Therefore from (6) we have $\frac{1}{3}\ln V = x_1 = x_2 = x_3$ and $y = 4V^{2/3}$. It can then be checked that equations (4)–(7) do indeed hold for these. As we have a convex problem, this means that $\mathbf{x}$ is an optimal solution to (2), and the optimal solution to (1) is $l = b = h = V^{1/4}$. The optimal values to both these problems are then $6V^{2/3}$.

vi) Consider the points $\mathbf{x}^* = \frac{1}{3}\ln(V)\mathbf{1}$ and $y^* = 4V^{2/3}$. These are feasible for (2) and (3) respectively. Therefore, letting ν_2 and ν_3 be the optimal values to these problems respectively we have

$$\nu_2 \leq f(\mathbf{x}^*) = 6V^{2/3},$$
$$\nu_3 \geq \psi(y^*) = \left(\tfrac{3}{2} + \ln(V)\right) 4V^{2/3} - \tfrac{3}{2}4V^{2/3}\ln\left(\tfrac{1}{4}4V^{2/3}\right) = 6V^{2/3}.$$

Then recalling by weak duality that $\nu_2 \geq \nu_3$ we have that $\nu_2 = \nu_3 = 6V^{2/3}$, with $\mathbf{x}^*$ and y^* being the optimal solutions to (2) and (3) respectively.