Quiz 2 Practice Solutions

1. We consider a resource allocation problem of the form

$$\max_{w \in \mathcal{W}} \min_{r \in \mathcal{E}} r^T w$$

where $\mathcal{W} := \{w \in \mathbb{R}^n_+ \ : \ w_1 + \ldots + w_n = 1\}$, and

$$\mathcal{E} := \{\hat{r} + Du \ : \ \|u\|_2 \leq 1\}.$$

Here, $\hat{r} \in \mathbb{R}^n$ and $D = \mathbf{diag}(\sigma_1, \ldots, \sigma_n)$ are given, with $\sigma \in \mathbb{R}^n$ $\sigma > 0$.

The above problem appears when trying to allocate resources to various revenue-generating processes (which could be ads, financial investments, physical sensors, etc). The revenue vector r is unknown but bounded, and the goal is to maximize the worst-case total revenue $\min_{r \in \mathcal{E}} r^T w$.

(a) Describe the shape of $\mathcal{E}$ in simple geometrical terms.

(b) Use the Cauchy-Schwartz inequality to prove that

$$\min_{r \in \mathcal{E}} r^T w = \hat{r}^T w - \|Dw\|_2.$$

(c) Express the problem as an SOCP in standard format, involving the variable w and one extra scalar variable.

Solution:

(a) The set $\mathcal{E}$ is an ellipsoid with axes parallel to the coordinate axes, with semi-axis lengths given by σ_i, $i = 1, \ldots, n$.

(b) For any $w \in \mathcal{W}$, we have

$$\min_{r \in \mathcal{E}} r^T w = \hat{r}^T w + \min_{u \,:\|u\|_2 \leq 1} w^T Du$$

The result follows from the Cauchy-Schwartz inequality. The latter states that for any $u, v \in \mathbb{R}^n$:

$$|u^T v| \leq \|u\|_2 \|v\|_2,$$

with the upper bound attained. This implies that

$$\min_{u \,:\, \|u\|_2 \leq 1} u^T v = -\|v\|_2,$$

which proves the result, upon setting $v = Dw$.

(c) Our problem reads

$$\max_w \hat{r}^T w - \|Dw\|_2 \ : \ w \geq 0, \ \sum_{i=1}^n w_i = 1.$$

We can write the problem in a standard SOCP format:

$$\max_w \hat{r}^T w - t \ : \ t \geq \|Dw\|_2, \ \ w \geq 0, \ \sum_{i=1}^n w_i = 1.$$

2. When a user goes to a website, an advertisement from a set of n ads, labeled $1, \ldots, n$, will be displayed. This is called an *impression* of ad i. We divide a day into T periods, labeled $t = 1, \ldots, T$. Let $N_{it} \geq 0$ denote the number of impressions in period t of ad i. The total number of ad impressions in period t is $I_t > 0$, so we must have $\sum_{i=1}^n N_{it} = I_t$, for $t = 1, \ldots, T$. For simplicity, you can treat all these integer numbers as real. (This is justified since they are typically very large.)

   (a) The revenue for displaying ad i in period t is $R_{it} \geq 0$ per impression. What is the total daily revenue for displaying ads $(1, \ldots, n)$? Given I_t (the number of impressions in period t) and R_{it} (the revenue for displaying ad i in period t), how would you choose N_{it} to maximize revenue for a given day?

   (b) In reality, we also have in place a set of m contracts that require us to display certain numbers of ads, or mixes of ads (say, associated with the products of one company), over certain periods, with a penalty for any shortfalls. For $j = 1, \ldots, m$, contract j is characterized by a set of ads $\mathcal{A}_j \subseteq \{1, \ldots, n\}$, a set of periods $\mathcal{T}_j \subseteq \{1, \ldots, T\}$, a target number of impressions $q_j \geq 0$. The *shortfall* s_j for contract $j = 1, \ldots, m$ is the difference between the target number of impressions of the contract and the total number of ads (single ad or mix of ads, as specified in the contract) which are displayed; if the number of impressions is larger than the target, we count the shortfall as 0. Show that the shortfall for contract j can be written as

   $$s_j = \max(0, q_j - L_j(N)),$$

   where L_j is a linear function of matrix N, which you will determine.

   (c) Contractually, a unit shortfall in contract j results in a given penalty $p_j > 0$, $j = 1, \ldots, m$. What is the total penalty payment? What is the net profit?

   (d) Explain how to find the display numbers N_{it} that maximize net profit via linear programming. Remember that the data in this problem are $R \in \mathbb{R}^{n \times T}$, $I \in \mathbb{R}^T$ (here I is the vector of impressions, not the identity matrix), and the contract data $\mathcal{A}_j$, $\mathcal{T}_j$, q_j and p_j, $j = 1, \ldots, m$. Make sure to state precisely what the objective function, the variables and the constraints are. Explain clearly why your problem is a linear programming problem.

Solution:

(a) The revenue of displaying ad i once in period t is given by $R_{it}N_{it}$. The total revenue is given by

$$\sum_{i=1}^{N}\sum_{t=1}^{T} N_{it}R_{it}.$$

In each period t, we only display ad-i with the largest revenue R_{it} such that $I_t = N_{it}$. Then we can achieve the maximum revenue.

(b) The difference between the target number of impressions of contract j and the total number of ads is

$$p_j - \sum_{i\in\mathcal{A}_j}\sum_{t\in\mathcal{T}_j} N_{it}$$

Since the shortfall is set to zero when that number is negative, we have

$$s_j = \max(0, p_j - \sum_{i\in\mathcal{A}_j}\sum_{t\in\mathcal{T}_j} N_{it}).$$

(c) The total penalty payment is given by

$$\sum_{j=1}^{m} p_j s_j = \sum_{j=1}^{m} p_j \max(0, q_j - \sum_{i\in\mathcal{A}_j}\sum_{t\in\mathcal{T}_j} N_{it}).$$

The net profit is given by

$$\sum_{i=1}^{N}\sum_{t=1}^{T} N_{it}R_{it} - \sum_{j=1}^{m} q_j \max(0, p_j - \sum_{i\in\mathcal{A}_j}\sum_{t\in\mathcal{T}_j} N_{it}).$$

(d) The problem is formulated as

$$\begin{aligned}
\max_{s\in\mathbb{R}^m, N\in\mathbb{R}^{n\times p}} \quad & \sum_{i=1}^{N}\sum_{t=1}^{T} N_{it}R_{it} - \sum_{j=1}^{m} q_j s_j \\
\text{s.t.} \quad & \sum_{i=1}^{T} N_{it} = I_t, t = 1, \cdots, T \\
& s_j \geq q_j - \sum_{i\in\mathcal{A}_j}\sum_{t\in\mathcal{T}_j} N_{it}, j = 1, \cdots, m \\
& N \geq 0, s \geq 0.
\end{aligned}$$

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function, with domain the entire space $\mathbb{R}^n$.

(a) Show that we can represent f as a pointwise maximum of affine functions, specifically

$$\forall\, x \in \mathbb{R}^n \;:\; f(x) = \max_{(a,b)\in\mathcal{A}} \; a^T x + b, \tag{1}$$

where $\mathcal{A} \subseteq \mathbb{R}^{n+1}$ is to be determined. *Hint:* use the inequality, valid for any differentiable convex function f, and every $x, y \in \mathbb{R}^n$: $f(x) \geq f(y) + (x-y)^T \nabla f(y)$, and maximize over y for a fixed x.

(b) We define the *perspective* of f as the function $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ with values

$$g(x,t) = \begin{cases} t f(x/t) & \text{if } t > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Prove that g is convex. *Hint:* express g when you rewrite f using (1), leading to a similar expression for g.

(c) Show that the function $h : \mathbb{R}^n \to \mathbb{R}$ with values

$$h(x) = \begin{cases} \dfrac{(a^T x + b)^2}{c^T x + d} & \text{if } c^T x + d > 0, \\ +\infty & \text{otherwise,} \end{cases}$$

is convex. *Hint:* compute the perspective of the function f with values $f(z) = z^2$ on $\mathbb{R}$.

Solution:

(a) Let $x \in \mathbb{R}^n$. We have

$$\forall\, y \in \mathbb{R}^n \;:\; f(x) \geq f(y) + (x-y)^T \nabla f(y).$$

Maximizing over y:

$$f(x) \geq \max_y \; f(y) + (x-y)^T \nabla f(y).$$

Choosing $y = x$ in the expression inside the maximum results in $f(x)$. Hence, the maximum over y is larger than $f(x)$, which leads to the equality:

$$f(x) = \max_y \; f(y) + (x-y)^T \nabla f(y).$$

We thus choose

$$\mathcal{A} = \left\{ (a,b) \in \mathbb{R}^n \times \mathbb{R} \;:\; a = \nabla f(y), \;\; b = f(y) - y^T \nabla f(y) \text{ for some } y \in \mathbb{R}^n \right\}.$$

(b) The domain of g is $\mathbb{R}^n \times \mathbb{R}_{++}$, so it is convex. When $t > 0$:

$$g(x,t) = t \max_{(a,b)\in\mathcal{A}} a^T(x/t) + b = \max_{(a,b)\in\mathcal{A}} a^T x + tb.$$

We observe that g is the point-wise maximum of linear functions of (x,t), hence it is convex.

(c) The function is the composition of the affine map

$$x \to (z,t) = (a^T x + b, c^T x + d)$$

with the function function $(z,t) \to z^2/t$ on $\mathbb{R} \times \mathbb{R}_{++}$. The latter turns out to be the perspective of the function f with values $f(z) = z^2$ on $\mathbb{R}$. Since f can be represented as a pointwise maximum of affine function, we obtain that its perspective is convex.

4. The log-sump-exp function is defined as the function $f : \mathbb{R}^n \to \mathbb{R}$ with values

$$f(z) = \log\left(\sum_{i=1}^n e^{z_i}\right).$$

Often, the convexity of this function is proven by deriving the Hessian. In this exercise, you will show that f is convex, using another method.

(a) Show that, for any given $s > 0$, we have

$$\log s = -1 + \min_v \; se^v - v.$$

(b) Show that

$$f(z) = -1 + \min_v \sum_{i=1}^n e^{z_i+v} - v.$$

(c) Prove convexity of f based on the above result. *Hint:* use the notion of joint convexity.

Solution:

(a) Since $s > 0$, the differentiable function $v \to se^v - v$ is convex. The optimality condition of the problem of minimizing this function is that its gradient is zero. Differentiating, we obtain that $v^* = -\log s$ is optimal, hence the result.

(b) This follows from applying the result of the previous part to $s = \sum_{i=1}^n e^{z_i}$.

(c) The function

$$(z,v) \in \mathbb{R}^n \times \mathbb{R} \to \sum_{i=1}^n e^{z_i+v} - v$$

is jointly convex in its arguments z, v; hence the minimum over v is convex in z.

5. We consider an investment problem of the form

$$p^* = \max_{x \geq 0} \; r^T x - \frac{1}{2} x^T C x$$

where $r \in \mathbb{R}^n$ is a vector of expected returns, $C = C^T \succ 0$ is a covariance matrix. We assume that C is given as a so-called "single factor model", that is

$$C = D + ff^T,$$

where D is diagonal positive-definite, and $f \in \mathbb{R}^n$.

(a) First show a preliminary result: for every scalars ρ and $\delta > 0$,

$$\max_{\xi \geq 0} \; \rho\xi - \frac{1}{2}\delta\xi^2 = \frac{1}{2\delta} \max(0, \rho)^2.$$

with unique minimizer $\xi^* = \max(0, \rho)/\delta$.

(b) Based on the expression

$$p^* = \max_{x \geq 0} \; r^T x - \frac{1}{2} x^T D x - \frac{1}{2} z^2 \;:\; z = f^T x,$$

show that a dual can be written as

$$\frac{1}{2} \min_{\nu} \; \nu^2 + \sum_{i=1}^{n} \frac{\max(0, r_i - \nu f_i)^2}{D_{ii}}.$$

Hint: for the minimization of the Lagrangian over x, use part(a) .

(c) Does strong duality hold? Justify your answer.

(d) Explain how to recover an optimal primal dual point from an optimal dual variable ν^*.

Solution:

(a) The unconstrained minimizer is unique, and given by

$$\xi^* = \frac{\rho}{\delta}.$$

If this point is feasible (that is, $\rho \geq 0$) then it is optimal; if not, $\xi^* = 0$ is optimal. Hence the maximizer is unique, and given by

$$\xi^* = \frac{\max(0, \rho)}{\delta}.$$

The desired expression follows.

(b) We have

$$p^* = \max_{x \geq 0} \; r^T x - \frac{1}{2} x^T D x - \frac{1}{2} z^2 + \nu(z - f^T x).$$

A dual can be obtained by exchanging the min and max in the above:

$$p^* \leq d^* := \min_{\nu} \; \max_{x \geq 0} \; r^T x - \frac{1}{2} x^T D x - \frac{1}{2} z^2 + \nu(z - f^T x).$$

Maximizing over z leads to the unique solution (for any fixed ν) $z^*(\nu) = \nu$. Maximizing over x leads to a problem of the form seen in part 5a, with $\rho = (r_i - \nu f_i)$, $\delta = D_{ii}$ (here the index i is given). Hence

$$\min_{x_i \geq 0} \, (r_i - \nu f_i) x_i - \frac{1}{2} D_{ii} x_i^2 = \frac{\max(0, r_i - \nu f_i)^2}{D_{ii}}, \quad i = 1, \ldots, n,$$

with unique minimizer

$$x_i^*(\nu) = \frac{\max(0, r_i - \nu f_i)}{D_{ii}}, \quad i = 1, \ldots, n. \tag{2}$$

Plugging in the optimal points $x^*(\nu)$ and $z^*(\nu)$ in the Lagrangian leads to the desired dual.

(c) Strong duality is a direct consequence of Slater's theorem, since the primal is convex, has equality constraints only and is feasible.

(d) We use the fact that for every ν, the minimizer of the Lagrangian over x is unique. This implies that an optimal point x^* can be set as in (2), with ν set to be the optimal dual point.