ECE 602 – Introduction to Optimization

Solutions to Home Assignment 3

Exercise 1

Note that the last constraint $x \succeq 0$ is implicitly contained in the objective function and thus can be safely ignored. Consequently, we can define the Lagrangian to be

$$L(x, \nu, \eta) = \sum_{k=1}^{n} x_k \log\left(\frac{x_k}{y_k}\right) - \nu^T(Ax - b) - \eta\,(\mathbf{1}^T x - 1).$$

The optimal x of the Lagrangian has k-th component given by

$$x_k \frac{1}{x_k} + \log\left(\frac{x_k}{y_k}\right) - a_k^T \nu - \eta = 0,$$

where a_k is the k-th column of A. The optimal x_k is therefore given by

$$x_k^* = \frac{1}{Z} y_k e^{a_k^T \nu},$$

where Z is a constant in terms of η that allows x to sum to one, i.e.,

$$Z = \sum_k y_k e^{a_k^T \nu}.$$

Plugging this back into the Lagrangian, we obtain

$$g(\nu) = L(x^*, \nu, \eta) = \sum_k \frac{1}{Z} y_k e^{a_k^T \nu} \log\left(\frac{e^{a_k^T \nu}}{Z}\right) - \nu^T \left(A \begin{bmatrix} \frac{1}{Z} y_1 e^{a_1^T \nu} \\ \cdots \\ \frac{1}{Z} y_n e^{a_n^T \nu} \end{bmatrix} - b \right).$$

Simplifying the above equation, we get

$$g(\nu) = -\log Z + b^T \nu = -\log \sum_k y_k e^{a_k^T \nu} + b^T \nu.$$

Exercise 2

The optimization problem has three variables, i.e., $w \in \mathbf{R}^p$, $b \in \mathbf{R}$, and $\zeta \in \mathbf{R}^n$, which can be combined into a single variable $z = [w, b, \zeta] \in \mathbf{R}^d$, where $d = p + n + 1$. We can also define

$$P = \begin{bmatrix} I_{p\times p} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbf{R}^{d\times d}, \quad q = \begin{bmatrix} \mathbf{0}_{p\times 1} \\ 0 \\ \lambda \mathbf{1}_{n\times 1} \end{bmatrix}.$$

Then, the cost function can be rewritten as

$$\frac{1}{2}\|w\|_2^2 + \lambda \sum_{k=1}^{n} \zeta_k = \frac{1}{2} z^T P z + q^T z,$$

which is quadratic.

The first set of inequality constraints can be written as

$$-(y_i x_i)^T w - y_i b - \zeta_i \leq -1, \quad i = 1, 2, \ldots, n.$$

Define $C \in \mathbf{R}^{n\times d}$ as

$$C = \begin{bmatrix} -y_1\, x_1^T & -y_1 & -1 & 0 & 0 & \cdots & 0 \\ -y_2\, x_2^T & -y_2 & 0 & -1 & 0 & \cdots & 0 \\ -y_3\, x_3^T & -y_3 & 0 & 0 & -1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \ddots & \cdots \\ -y_n\, x_n^T & -y_n & 0 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

Then, the constrains become $Cz \preceq -\mathbf{1}_{n\times 1}$. The second inequality constraint, i.e., $\zeta \succeq 0$, can be also defined in terms of z. In particular, let $B = [\mathbf{0}_{n\times p+1},\ -I_{n\times n}]$. Then, the constrain becomes $Bz \preceq \mathbf{0}_{n\times 1}$.

The two inequality constraints can be combined together using matrix $A \in \mathbf{R}^{2n\times d}$ and vector $b \in \mathbf{R}^{2n}$ defined as

$$A = \begin{bmatrix} C \\ B \end{bmatrix}, \quad b = \begin{bmatrix} -\mathbf{1}_{n\times 1} \\ \mathbf{0}_{n\times 1} \end{bmatrix}.$$

In these notations, the original optimization problem can be written as

$$\begin{aligned} \min_z \ & \frac{1}{2} z^T P z + q^T z \\ \text{s.t.} \ & Az \preceq b \end{aligned}$$

which is a QP.

The Lagrangian of this QP is given by

$$L(z, \lambda) = \frac{1}{2} z^T P z + q^T z + \lambda^T (Ax - b) = \frac{1}{2} z^T P z + (q + A^T \lambda)^T z - \lambda^T b.$$

Differentiating (w.r.t. z) and equating to zero results in $z^* = -P^{-1}(q + A^T\lambda)$, which we substitute back into the Lagrangian to compute the dual function. We get

$$g(\lambda) = L(z^*, \lambda) = -(q + A^T\lambda)^T P^{-1}(q + A^T\lambda) - \lambda^T b,$$

which is quadratic. Hence, the dual problem

$$\begin{aligned} \max_{\lambda} \; & -(q+A^T\lambda)^T P^{-1}(q + A^T\lambda) - \lambda^T b \\ \text{s.t.} \; & \lambda \succeq 0 \end{aligned}$$

is a QP as well.

Exercise 3

The Lagrangian is given by

$$L(x, \nu) = \|Ax - b\|_2^2 + \nu^T (Ax - b),$$

with its gradient w.r.t. x equal to

$$\nabla_x L(x, \nu) = 2A^T(Ax - b) + A^T \nu.$$

Consequently, the KKT conditions are:

- stationarity: $2A^T(Ax^* - b) + A^T\nu^* = 0$;
- primal feasibility: $Cx^* = h$;
- dual feasibility: none;
- complementary slackness: none.

From the stationarity, we have

$$x^* = (A^T A)^{-1}(A^T b - (1/2) C^T \nu^*). \tag{1}$$

Then, plugging x^* into the primal feasibility results in

$$C(A^TA)^{-1}A^Tb - (1/2)C(A^TA)^{-1}C^T\nu^* = h,$$

yielding

$$\nu^* = -2\left(C(A^TA)^{-1}C^T\right)^{-1}(h - C(A^TA)^{-1}A^Tb).$$

Finally, the value of ν^* can be used to compute x^* using Eq. (1).