Optimization Methods **Fall 2019**

Homework 3

Instructor: Lijun Zhang *Name:* Renzhe Zhou, *StudentId:* 181220076

Notice

- The submission email is: **njuoptfall2019@163.com**.
- Please use the provided LaTeX file as a template. If you are not familiar with LaTeX, you can also use Word to generate a **PDF** file.

Problem 1: Equality Constrained Least-squares
Consider the equality constrained least-squares problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax-b\|_2^2 \\ \text{subject to} & Gx = h \end{array}$$

where $A \in \mathbf{R}^{m\times n}$ with **rank** $A = n$, and $G \in \mathbf{R}^{p\times n}$ with **rank** $G = p$.

a) Derive the Lagrange dual problem with Lagrange multiplier vector v.

b) Derive expressions for the primal solution $x^\star$ and the dual solution $v^\star$.

Solution.

a) The lagrange function is

$$L(x,v) = \frac{1}{2}\|Ax-b\|_2^2 + v^T(Gx-h) = \frac{1}{2}x^TA^TAx + (v^TG - b^TA)x + \frac{1}{2}b^Tb - v^Th \tag{1}$$

Make $\nabla_x L(x,v) = A^TAx + G^Tv - A^Tb = 0$, we can get minimizer $x = (A^TA)^{-1}(A^Tb - G^Tv)$. So the dual function is

$$g(v) = \inf_x L(x,v) = -\frac{1}{2}(G^Tv - A^Tb)^T(A^TA)^{-1}(G^Tv - A^Tb) + \frac{1}{2}b^Tb - v^Th \tag{2}$$

So the lagrange dual problem is

$$\text{maximize} \quad -\tfrac{1}{2}(G^Tv - A^Tb)^T(A^TA)^{-1}(G^Tv - A^Tb) + \tfrac{1}{2}b^Tb - v^Th$$

b) From KKT conditions we can get $\nabla L(x^*,v^*) = A^TAx^* + G^Tv^* - A^Tb = 0$ and $Gx^* = h$. Make $x^* = (A^TA)^{-1}(A^Tb - G^Tv^*)$ into $Gx^* = h$ we can get

$$v^* = (G(A^TA)^{-1}G^T)^{-1}(G(A^TA)^{-1}A^Tb - h) \tag{3}$$

Therefore we get

$$x^* = (A^TA)^{-1}(A^Tb - G^T(G(A^TA)^{-1}G^T)^{-1}(G(A^TA)^{-1}A^Tb - h)) \tag{4}$$

Done.

□

Problem 2: Support Vector Machines
Consider the following optimization problem

$$\text{minimize} \quad \sum_{i=1}^n \max\left(0, 1 - y_i(w^T x_i + b)\right) + \frac{\lambda}{2}\|w\|_2^2$$

where $x_i \in \mathbf{R}^d, y_i \in \mathbf{R}, i = 1, \cdots, n$ are given, and $w \in \mathbf{R}^d, b \in \mathbf{R}$ are the variables.

a) Derive an equivalent problem by introducing new variables $u_i, i = 1, \cdots, n$ and equality constraints

$$u_i = y_i(w^T x_i + b), i = 1, \cdots, n.$$

b) Derive the Lagrange dual problem of the above equivalent problem.

c) Give the Karush-Kuhn-Tucker conditions.

Hint: Let $\ell(x) = \max(0, 1 - x)$. *Its conjugate function* $\ell^*(y) = \sup_x (yx - \ell(x)) = \begin{cases} y, & -1 \leq y \leq 0 \\ \infty, & \textit{otherwise} \end{cases}$

Solution.

a) The problem is:

$$\begin{aligned} &\text{minimize } \sum_{i=1}^n \max(0, 1 - u_i) + \frac{\lambda}{2}\|w\|_2^2 \\ &\text{subject to } u_i = y_i(w^T x_i + b), i = 1, \cdots, n. \end{aligned} \tag{5}$$

b) The Lagrange function is

$$L(w, b, u_1, \cdots, u_n, v_1, \cdots, v_n) = \sum_{i=1}^n \max(0, 1 - u_i) + \frac{\lambda}{2}\|w\|_2^2 + \sum_{i=1}^n v_i(u_i - y_i(w^T x_i + b)) \tag{6}$$

To get the dual function, minimize w, b and u_i. Minimize w, we have $\frac{\partial L}{\partial w} = \lambda w + \sum_{i=1}^n (-v_i y_i x_i) = 0$. So $w^* = \frac{1}{\lambda}\sum_{i=1}^n v_i y_i x_i$. Minimize b, we have $\frac{\partial L}{\partial b} = \sum_{i=1}^n (-v_i y_i) = 0$, if not, the mininum could be $-\infty$. And $\sum_{i=1}^n v_i y_i x_i^T (\sum_{j=1}^n v_j y_j x_j) = \sum_{i=1}^n v_i y_i (\sum_{j=1}^n v_j y_j x_i^T x_j)$. Therefore, in that case, let $l(u_i) = max(0, 1 - u_i)$, we have:

$$\begin{aligned} g(v_1, \cdots, v_n) =& \frac{1}{2|\lambda|}\|\sum_{i=1}^n v_i y_i x_i\|_2^2 - \frac{1}{\lambda}\sum_{i=1}^n v_i y_i (\sum_{j=1}^n v_j y_j x_i^T x_j) + \inf_{u_i}(\sum_{i=1}^n \max(0, 1 - u_i) + v_i u_i) \\ =& \frac{1}{2|\lambda|}\|\sum_{i=1}^n v_i y_i x_i\|_2^2 - \frac{1}{\lambda}\sum_{i=1}^n v_i y_i (\sum_{j=1}^n v_j y_j x_i^T x_j) - \sum_{i=1}^n l^*(-v_i) \end{aligned} \tag{7}$$

To maximize it, we must have $0 \leq v_i \leq 1$. So the Lagrange dual problem is:

$$\begin{aligned} &\text{maximize } \frac{1}{2|\lambda|}\|\sum_{i=1}^n v_i y_i x_i\|_2^2 - \frac{1}{\lambda}\sum_{i=1}^n v_i y_i (\sum_{j=1}^n v_j y_j x_i^T x_j) + \sum_{i=1}^n v_i \\ &\text{subject to } \sum_{i=1}^n v_i y_i = 0, \quad 0 \leq v_i \leq 1, i = 1, \cdots, n. \end{aligned} \tag{8}$$

c) The KKT conditions are:

$$\begin{aligned} &w^* = \frac{1}{\lambda}\sum_{i=1}^n v_i^* y_i x_i \\ &\sum_{i=1}^n v_i^* y_i = 0 \\ &u_i^* = y_i({w^*}^T x_i + b^*), i = 1, \cdots, n. \end{aligned} \tag{9}$$

Done.

$\square$

Problem 3: Euclidean Projection onto the Simplex
Consider the following optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y-x\|_2^2 \\ \text{subject to} & \mathbf{1}^T y = r \\ & y \succeq 0 \end{array}$$

where $r > 0$, $x \in \mathbb{R}^n$ is given, and $y \in \mathbf{R}^n$ is the variable. Give an algorithm to solve this problem and prove the correctness of your algorithm.
Hint: Derive the Lagrangian of this problem and apply the Karush-Kuhn-Tucker conditions. If you need more hints, please read the following paper [1]

Solution.

1. Algorithm:
   Input: $x \in R^n$
   Sort x into u:$u_1 \geq u_2 \geq \cdots \geq u_n$
   Find $\rho = max\{1 \leq j \leq n : u_j + \frac{1}{j}(r - \sum_{i=1}^{j} u_i) > 0\}$
   Define $\lambda = \frac{1}{\rho}(r - \sum_{i=1}^{\rho} u_i)$
   Output:y s.t. $y_i = max\{x_i + \lambda, 0\}, i = 1, \cdots, n.$

2. Proof:
   The lagrange function of this problem is(λ and v is the equality and inequality Lagrange multipliers respectively)
$$L(y, \lambda, v) = \frac{1}{2}\|y-x\|_2^2 - \lambda(\mathbf{1}^T y - r) - v^T y \tag{10}$$
   Assume the optimal solution y then the KKT conditions are following:
$$y_i - x_i - \lambda - v_i = 0, \quad y_i \geq 0, v_i \geq 0, \quad v_i y_i = 0, \quad \sum_{i=1}^{n} y_i = r \tag{11}$$
   From $v_i y_i = 0$ we have
$$\begin{aligned} &if \quad y_i > 0 : v_i = 0, y_i = x_i + \lambda > 0 \\ &if \quad y_i = 0 : v_i \geq 0, y_i = x_i + \lambda + v_i = 0, \text{which is } x_i + \lambda = -v_i < 0 \end{aligned} \tag{12}$$
   So the components of the optimal solution y that are zeros correspond to the smaller components of x. Assume the components of x are sorted and y uses the same ordering:
$$\begin{aligned} &x_1 \geq \cdots \geq x_\rho \geq x_{\rho+1} \geq \cdots \geq x_n \\ &y_1 \geq \cdots \geq y_\rho \geq y_{\rho+1} = \cdots = y_n = 0 \end{aligned} \tag{13}$$
   where ρ is the number of positive components of y. So from $\sum_{i=1}^{n} y_i = r$ we have
$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{\rho} y_i = \sum_{i=1}^{\rho}(x_i + \lambda) = r \tag{14}$$
   Which implies $\lambda = \frac{1}{\rho}(r - \sum_{i=1}^{\rho} x_i)$. So if we can compute ρ then we know λ and then we know the optimal solution y because $y_i = x_i + \lambda$ if $x_i + \lambda > 0$ else $y_i = 0$.
   Therefore, we are going to prove $\rho = max\{1 \leq j \leq n : x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) > 0\}$.
   We have $\lambda\rho = r - \sum_{i=1}^{\rho} x_i$, we are going to show that $x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) > 0$ for $j \leq \rho$ and $x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) \leq 0$ for $j > \rho$.
   1) $j = \rho$, we have $x_\rho + \frac{1}{\rho}(r - \sum_{i=1}^{\rho} x_i) = x_\rho + \lambda = y_\rho > 0$
   2) $j < \rho$,we have
$$\begin{aligned} x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) &= \frac{1}{j}(jx_j + r - \sum_{i=1}^{j} x_i) = \frac{1}{j}(jx_j + \sum_{i=j+1}^{\rho} x_i + r - \sum_{i=1}^{\rho} x_i) \\ &= \frac{1}{j}(jx_j + \sum_{i=j+1}^{\rho} x_i + \rho\lambda + j\lambda - j\lambda) = \frac{1}{j}(j(x_j + \lambda) + \sum_{i=j+1}^{\rho}(x_i + \lambda)) \end{aligned} \tag{15}$$
   Because $x_i + \lambda > 0$ for $i = j, \cdots, \rho$,so $x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) > 0$.

3) $j > \rho$,we have

$$\begin{aligned} x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) =& \frac{1}{j}(jx_j + r - \sum_{i=1}^{j} x_i) = \frac{1}{j}(jx_j + r - \sum_{i=1}^{\rho} x_i - \sum_{i=\rho+1}^{j} x_i) \\ =& \frac{1}{j}(jx_j + \rho\lambda - \sum_{i=\rho+1}^{j} x_i + \rho x_j - \rho x_j) = \frac{1}{j}(\rho(x_j + \lambda) + \sum_{i=\rho+1}^{j}(x_j - x_i)) \end{aligned} \tag{16}$$

Because $x_j + \lambda \le 0$ for $j > \rho$ and $x_j - x_i \le 0$ for $j \ge i$ since x is sorted, we have $x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) < 0$.

Therefore, $\rho = max\{1 \le j \le n : x_j + \frac{1}{j}(r - \sum_{i=1}^{j} x_i) > 0\}$ correctly denotes the number of positive components of y, and we can know λ and we get the optimal solution y by computing $y_i = max\{x_i + \lambda, 0\}$ since x is given.
Above all, the Algorithm can correctly give the optimal solution of the problem.

Done.

□

Problem 4: Optimality Conditions
Consider the problem

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(2X) - \log\det(3X) \\ \text{subject to} & 2Xs = y \end{array}$$

with variable $X \in \mathbf{S}^n$ and domain $\mathbf{S}^n_{++}$. Here, $y \in \mathbf{R}^n$ and $s \in \mathbf{R}^n$ are given, with $s^T y = 1$.

a) Give the Lagrange and then derive the Karush-Kuhn-Tucker conditions.

b) Verify that the optimal solution is given by

$$X^\star = \frac{1}{2}\left(I + yy^T - \frac{ss^T}{s^T s}\right).$$

Solution.

a) The lagrange function is $L(X, v) = \mathrm{tr}(2X) - \log\det(3X) - v^T(2Xs - y)$. And $\nabla_X L(X, v) = 2I - X^{-1} + (vs^T + sv^T)$. So the KKT conditions are:

$$X^* \succ 0, \qquad 2X^* s = y, \qquad 2I - X^{*-1} + (vs^T + sv^T) = 0 \tag{17}$$

b) We have

$$s = \frac{1}{2}X^{*-1}y = (I + \frac{1}{2}(vs^T + sv^T))y = y + \frac{1}{2}(vs^T y + sv^T y) = y + \frac{1}{2}(v + v^T ys) \tag{18}$$

So

$$y^T s = 1 = y^T(y + \frac{1}{2}(v + v^T ys)) = y^T y + \frac{1}{2}(y^T v + y^T v^T ys) = y^T y + \frac{1}{2}(v^T y + (v^T y)y^T s) = y^T y + v^T y \tag{19}$$

Take $v^T y = 1 - y^T y$ into $s = y + \frac{1}{2}(v + (v^T y)s)$ then we have

$$v = (1 + y^T y)s - 2y \tag{20}$$

Therefore,

$$\begin{aligned} X^{*-1} &= 2I + ((1 + y^T y)s - 2y)s^T + s((1 + y^T y)s^T - 2y^T) \\ &= 2I + 2(1 + y^T y)ss^T - 2ys^T - 2sy^T \end{aligned} \tag{21}$$

Multiply it with the condition given above:

$$\begin{aligned}
&[2I + 2(1 + y^T y)ss^T - 2ys^T - 2sy^T][\frac{1}{2}(I + yy^T - \frac{ss^T}{s^T s})] \\
&= (I + yy^T - \frac{ss^T}{s^T s} + (1 + y^T y)(ss^T + sy^T - ss^T) - (ys^T + yy^T - ys^T) - (sy^T + (y^T y)sy^T - \frac{ss^T}{s^T s}) \\
&= I + yy^T - yy^T - (\frac{ss^T}{s^T s}) + (\frac{ss^T}{s^T s}) + (1 + y^T y)sy^T - (1 + y^T y)sy^T \\
&= I
\end{aligned} \tag{22}$$

Therefore, the inverse matrix of $2I + 2(1 + y^T y)ss^T - 2ys^T - 2sy^T$ is $\frac{1}{2}(I + yy^T - \frac{ss^T}{s^T s})$
So $X^* = ({X^*}^{-1})^{-1} = \frac{1}{2}(I + yy^T - \frac{ss^T}{s^T s})$.

Done.

□

References

[1] Weiran Wang, and Miguel Á. Carreira-Peroiñán. Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. *arXiv:1309.1541*, 2013.