Optimization Methods **Fall 2019**

Homework 3

Instructor: Lijun Zhang *Name:* 高辰潇, *StudentId:* 181220014

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

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) 解：

该问题的 Lagrange 函数为 $L(x,v) = \frac{1}{2}(Ax-b)^T(Ax-b) + v^T(Gx-h) = \frac{1}{2}x^TA^TAx + (v^TG - b^TA)x + \frac{1}{2}b^Tb - v^Th$，$\nabla_x L(x,v) = A^TAx + G^Tv - A^Tb$.

得到拉格朗日函数的极小值点为 $x = (A^TA)^{-1}(A^Tb - G^Tv)$.

因此 $g(v) = \inf_{x\in D} L(x,v) = \frac{1}{2}(v^TG - b^TA)(A^TA)^{-1}(A^Tb - G^Tv) + \frac{1}{2}b^Tb - v^Th$.

所以对偶问题为

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

b) 解：由于对偶问题是关于变量 v 的无约束优化问题，故 $g(v)$ 对 v 求导得

$\nabla_v g(v) = G(A^TA)^{-1}A^Tb - h - (G(A^TA)^{-1}G^T)v$

因此，$v^\star = (G(A^TA)^{-1}G^T)^{-1}(G(A^TA)^{-1}A^Tb - h)$.

又由 Slater 条件，原问题满足强对偶性。进一步由 KKT 条件，有 $A^TAx^\star + G^Tv^\star - A^Tb = 0$.

代入 $v^\star$ 得到 $x^\star = (A^TA)^{-1}[A^Tb - G^T(G(A^TA)^{-1}G^T)^{-1}(G(A^TA)^{-1}A^b - h)]$

□

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* $\ell^*(y) = \sup_x (yx - \ell(x)) = \begin{cases} y, & -1 \le y \le 0 \\ \infty, & otherwise \end{cases}$

Solution.

a) 解：等价问题为

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

b) 解：拉格朗日函数为 $L(v, u_i, w) = \sum_{i=1}^n \max(0, 1 - u_i) + \frac{\lambda}{2}\|w\|_2^2 + \sum_{i=1}^n v_i[u_i - y_i(w^T x_i + b)]$.

$g(v) = \inf_u\{\sum_{i=1}^n \max(0, 1 - u_i) + \sum_{i=1}^n v_i u_i\} + \inf_w\{\frac{\lambda}{2}\|w\|_2^2 - \sum_{i=1}^n v_i y_i w^T x_i\} + \inf_b\{-v_i y_i b\}$

$\nabla_w L = \lambda w - \sum_{i=1}^n (v_i y_i x_i)$，故极值点为 $w = \frac{1}{\lambda}\sum_{i=1}^n v_i y_i x_i$

由 Hint 知 $\inf_{u_i}\{\max(0, 1 - u_i) + y_i u_i\} = -\sup_{u_i}\{-v_i u_i - \max(0, 1 - u_i)\} = \begin{cases} v_i, & 0 \le v_i \le 1 \\ -\infty, & \text{otherwise} \end{cases}$

而 $\inf_b\{-v_i y_i b\} = \begin{cases} 0, & v_i y_i = 0 \\ -\infty & otherwise \end{cases}$

因此 $g(v) = \sum_{i=1}^n v_i + \frac{1}{2\lambda}|| \sum_{i=1}^n v_i y_i x_i||_2^2 - \sum_{i=1}^n[(\sum_{i=1}^n \frac{1}{\lambda} v_i^2 y_i^2 x_i^T) x_i] = \sum_{i=1}^n v_i - \frac{1}{2\lambda} v_i^2 y_i^2 \sum_{i=1}^n x_i^T \sum_{i=1}^n x_i$，其中 $v_i \in [0, 1], v_i y_i = 0, i \in \{1, ..., n\}$

故对偶问题为：

$$\begin{aligned} &\text{maximize} \quad g(v) = \sum_{i=1}^n v_i - \frac{1}{2\lambda} v_i^2 y_i^2 \sum_{i=1}^n x_i^T \sum_{i=1}^n x_i \\ &\text{subject to} \quad 0 \le v_i \le 1 \\ &\qquad\qquad\quad y_i v_i = 0 \end{aligned} \tag{2}$$

c) 解：KKT 条件中的第五项要求拉格朗日函数对原问题变量可微。然而，拉格朗日函数对原问题中的变量 u_i 不可导, 因此对于变量 $u_i, i \in \{1, 2, ..., n\}$，可回退到拉格朗日对变量 u_i 取极小的过程进行讨论。由于拉格朗日函数为 $L(v, u_i, w) = \sum_{i=1}^n \max(0, 1 - u_i) + \frac{\lambda}{2}\|w\|_2^2 + \sum_{i=1}^n v_i[u_i - y_i(w^T x_i + b)]$，因此拉格朗日对变量 u_i 取极小只需要考虑 $\max(0, 1 - u_i) + v_i u_i$ 取极小的过程。

若 $u_i \geq 1$，则 $\max(0, 1-u_i) + v_i u_i = v_i u_i$，极小值 v_i 在 $u_i = 1$ 时取得。

若 $u_i < 1$，则 $\max(0, 1-u_i) + v_i u_i = 1 - u_i + v_i u_i = 1 + (v_i - 1)u_i \geq 1 + v_i - 1 > v_i$，因此 $u_i < 1$ 时 $\max(0, 1-u_i) + v_i u_i > v_i$

因此当且仅当 $u_i = 1$ 时，拉格朗日函数对该变量取得最小值。故对于变量 u_i，它的 KKT 条件为 $u_i^\star = 1$.

由于拉格朗日函数对其余变量均可微，直接对其余变量求出梯度即可。

综上，KKT 条件为：

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

$u_i^\star = 1, i = 1, ..., n$

$\lambda w^\star - \sum_{i=1}^n v_i^\star y_i x_i = 0$

$y_i v_i^\star = 0, i = 1, ..., n$

并且，$v_i^\star$ 需要满足可行性要求：$v_i^\star \in [0, 1], i = 1, ..., n$.

□

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 [**?**]*

Algorithm 1 Solution

Input: $r > 0, x \in \mathbb{R}^n = (x_1, x_2, ..., x_n)$

```
1: Sort x into u: u_1 ≥ u_2 ≥ ... ≥ u_n
2: j = 1
3: while u_j + (1/j)(r − Σ_{i=1}^{j} u_i) > 0 do
4:     j = j + 1
5: end while
6: ρ = j − 1
7: Define v = −(1/ρ)(r − Σ_{i=1}^{ρ} u_i)
```

Output: optimal $y^\star$s.t.$y_i = \max\{x_i - v, 0\}, i = 1, ..., n$.

Solution. 解：首先给出算法。(见 Solution)

下面证明算法正确性。

该问题的拉格朗日函数为 $L(y, \lambda, v) = \frac{1}{2}\|y - x\|_2^2 - \lambda^T y + v(1^T y - r)$.

故对偶函数 $g(\lambda, v) = \inf_{n \in D} L(y, \lambda, v) = -\frac{1}{2}(x^T + \lambda^T - v \cdot \mathbf{1}^T)(x + \lambda - v \cdot \mathbf{1})$

故原问题的对偶问题为

$$\begin{aligned}&\text{maximize} \;\; g(\lambda, v)\\&\text{subject to } \lambda \succeq 0\end{aligned} \tag{3}$$

由于原问题为凸问题，且满足 Slater 条件，因此原问题的最优解 $y^\star$ 和对偶问题的最优解 $(\lambda^\star, v^\star)$ 应满足 KKT 条件：

$y^\star \succeq 0$

$1^T y^\star = r$

$\lambda^\star \succeq 0$

$\lambda^\star y^\star = 0$

$y^\star - x - \lambda^\star + v^\star \mathbf{1} = 0$

考虑最优解 $y^\star$ 的各分量 y_i，若 $y_i = 0$，则由条件 4 知 $\lambda_i \geq 0$，故 $v^\star - x_i \geq 0$；若 $y_i > 0$，则 $\lambda_i = 0$，$v^\star - x_i = -y_i < 0$.

不失一般性，将给定的向量 x 按照从大到小的顺序重新排列各分量，设排列后得到新向量 $u = (u_1, u_2, ..., u_n)$。

按照原本各维度的对应关系相应地排序最优解，设最优解为 $y^\star = y_1, ..., y_n$。由上述分析知，在 $\lambda_i \geq 0$ 的情况下数值较小的 u_i 一定对应着较小的 y_i. 因此有 $y_1 \geq y_2 \geq ... \geq y_\rho > y_{\rho+1} = ... = y_n = 0$.

又由于 $1^T y^\star = y_1 + y_2 + ... + y_\rho$, 且对于 $y_1, ..., y_m$ 有 $y_i = x_i - v^\star$，因此 $\sum_{i=1}^{\rho}(u_i - v^\star) = r$, $v^\star = -\frac{1}{\rho}(r - \sum_{i=1}^{\rho} u_i)$

为了确定 $v^\star$，还需要确定 ρ 的大小。对于 $j \in \{1, 2, ..., n\}$ 分三种情况考虑：

a) 若 $j < \rho$，则

$$\begin{aligned}&\frac{1}{j}(r - \sum_{i=1}^{j} u_i) + u_j\\&= \frac{1}{j}(r - \sum_{i=1}^{\rho} u_i + \sum_{j+1}^{\rho} u_i + j u_j)\\&= \frac{1}{j}(-\rho v^\star + \sum_{i=j+1}^{\rho} u_i + j u_j)\\&= \frac{1}{j}(j(u_j - v^\star) + \sum_{i=j+1}^{\rho}(u_i - v^\star))\end{aligned} \tag{4}$$

由于对于 $i \in \{1, 2, ...\rho\}$ 和 j 都有 $u_i - v^\star = y_i > 0$，故 $\frac{1}{j}(1 - \sum_{i=1}^{j} u_i) + u_j > 0$.

b) 若 $j > \rho$，则

$$\begin{aligned}&\frac{1}{j}(r - \sum_{i=1}^{j} u_i) + u_j\\&= \frac{1}{j}(r - \sum_{i=1}^{\rho} u_i - \sum_{\rho+1}^{j} u_i + j u_j)\\&= \frac{1}{j}(-\rho v^\star - \sum_{i=\rho+1}^{j} u_i + j u_j)\\&= \frac{1}{j}(j(u_j - v^\star) - \sum_{i=\rho+1}^{j}(u_i - v^\star))\end{aligned} \tag{5}$$

由于对于 $i\in\{\rho+1,...,n\}$ 和 $j>\rho$ 都有 $u_i-v^\star=<0$，故 $\frac{1}{j}(1-\sum_{i=1}^{j}u_i)+u_j<0$.

c) 若 $j=\rho$，则有

$\frac{1}{\rho}(r-\sum_{i=1}^{\rho}u_i)+u_\rho=\frac{1}{\rho}(\rho(u_\rho-v))>0$.

因此，算法第 3-6 行通过判断 $\frac{1}{j}(1-\sum_{i=1}^{j}u_i)+u_j$ 的正负性，找到使该式为正值的最大的下标，该下标即为 ρ. 再由 $v^\star=-\frac{1}{\rho}(1-\sum_{i=1}^{\rho}u_i)$ 可进一步计算出 $v^\star$（算法的第 7 行）。

而对于原问题最优解 $y^\star$，由 KKT 条件知它的非零分量 $y_i(i\le\rho)$ 均满足 $y_i=x_i-\rho$，且对于它的为零的分量 $y_i(i>\rho)$ 有 $x_i-v^\star\le 0$. 因此，$y^\star$ 的各分量 y_i 可统一表示为 $y_i=\max\{x_i-v^\star,0\}$.

故算法的输出即为原问题的最优解。正确性得证。

□

Problem 4: Optimality Conditions

Consider the problem

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

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 $s^Ty=1$.

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

b) Verify that the optimal solution is given by

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

Solution.

a) 解：Lagrange 函数为 $L(X,v)=\operatorname{tr}(2X)-\log\det(3X)+v^T(2Xs-y)$，$\nabla_X L=2I-X^{-1}+vs^T+sv^T$.

KKT 条件为:

1) $2Xs=y$

2) $2I-X^{-1}+vs^T+sv^T=0$

其中 $X\succ 0$

b) 解：

由 $2Xs=y$ 知 $s=\frac{1}{2}X^{-1}y$

代入 $X^{-1}=2I+vs^T+sv^T$ 得到 $s=\frac{1}{2}(2I+vs^T+sv^T)y=y+\frac{1}{2}(v+(v^Ty)s)$.

$s^Ty=y^Ty+\frac{1}{2}v^Ty+\frac{1}{2}v^Ty=1$，所以有 $y^Ty+v^Ty=1$.

代入 $s=y+\frac{1}{2}(v+(v^Ty)s)$ 可得到 $v=(1+y^Ty)s-2y$.

下面证明，题中所给出的 $X^\star$ 满足 KKT 条件。

1) $2Xs=2\cdot\frac{1}{2}(I+yy^T-\frac{ss^T}{s^Ts})s=s+y-s=y$.

2) 等价于证明 $X^{-1}X = (2I + v^T s + s^T v)X = I$

$$
\begin{aligned}
X^{-1}X &= (2I + (v^\star)^T s + s^T v)X = I \\
&= (2I + 2(1 + y^T y)ss^T - 2ys^T - 2sy^T)(\frac{1}{2}(I + yy^T - \frac{ss^T}{s^T s})) \\
&= I + (1 + y^T y)ss^T - ys^T - sy^T + yy^T + (1 + yy^T)sy^T - yy^T - sy^T yy^T - \frac{ss^T}{s^T s} - (1 + y^T y)ss^T + ys^T +
\end{aligned}
\tag{6}
$$

因此题目所给定的 X 满足 KKT 条件。
又由于 $x^\star = \frac{1}{2}(I + \frac{ys^T}{||s||_2} - \frac{ss^T}{s^T s})(I + \frac{ys^T}{||s||_2} - \frac{ss^T}{s^T s})^T$，因此 $X^\star \succ 0$.
故题目所给定的 X 为原问题的最优解。

□

参考文献

[1] gugu 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.