Optimization Methods **Fall 2020**

Homework 3

Instructor: Lijun Zhang *Name:* 张逸凯, *StudentId:* 171840708

Notice

- The submission email is: **opt4grad@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: Negative-entropy Regularization

Please show how to compute

$$\operatorname*{argmin}_{x\in\Delta^n} b^\top x + c\cdot\sum_{i=1}^n x_i \ln x_i$$

where $\Delta^n = \{x|\sum_{i=1}^n x_i = 1, x_i \geq 0, i = 1,\cdots,n\}$, $b\in\mathbb{R}^n$ and $c\in\mathbb{R}$.

- **解:**

  该问题是凸问题, 这很容易说明, 其中 $b^\top$ 是仿射的, $c\cdot\sum_{i=1}^n x_i\ln x_i$ 是负熵函数的组合, 并且等式约束也是仿射的, 所以原问题是凸问题, 其KKT条件为:

$$\begin{aligned} x^\top x &= 0 \\ \lambda &\succeq 0 \\ b + c\ln x + (c+\lambda)\cdot\mathbf{1} &= 0 \end{aligned}$$

  其中 $\mathbf{1}\in\mathbb{R}^n$.

  KKT条件直接解是困难的, 我们分析它的对偶函数, 因为 $\ln$ 蕴含了 $x_i > 0$, 所以我们可以消除不等式约束.

$$\begin{aligned} g(\lambda) &= \inf_x b^\top x + c\cdot\sum_{i=1}^n x_i\ln x_i + \lambda\cdot\left(\sum_{i=1}^n x_i - 1\right) \\ &= -c\sum_{i=1}^n e^{\frac{-b_i-c-\lambda}{c}} - \lambda \end{aligned}$$

  求导得到:

$$g(\lambda)' = \sum_{i=1}^n e^{\frac{-b_i-c-\lambda}{c}} - 1$$

  注意此时与直接解KKT的结果是一致的.

  解得 $\lambda^\star = \ln\left(\sum_{i=1}^n e^{\frac{-b_i-c}{c}}\right)^c$, 代入原式即可解得最优解.

Remark 本题这样解出结果是在 $c \geqslant 0$ 并且消除了一个不等式约束的条件下完成的, 请注意, 如果本题的意思是写出计算过程但是不需要结果, 则思路可以归纳为, 原问题直接解不能解:

Slater条件 $\Rightarrow$ 强对偶, 之后解KKT或者对偶问题即可.

Problem 2: One inequality constraint

(1) With $c \neq 0$, express the dual problem of

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & f(x) \leq 0, \end{aligned}$$

in terms of the conjugate $f^\star$.

(2) Explain why the problem you give is convex. We do not assume f is convex.

- 解

  (1)

  先写出拉格朗日函数形式, 由定义得对偶函数:

$$\begin{aligned} \mathcal{L}(x, \lambda) &= c^\top x + \lambda f(x) \\ g(\lambda) &= \inf_x \mathcal{L}(x, \lambda) \\ &= -\sup_x \left(\lambda \left(-\frac{c^\top}{\lambda} x - f(x) \right) \right) \\ &= -\lambda f^\star \left(-\frac{c}{\lambda} \right) \end{aligned}$$

  (2)

  自然地, 对偶问题为:

$$\begin{aligned} & \max g(\lambda) = -\lambda f^\star \left(-\frac{c}{\lambda} \right) \\ & \text{s. t.} \;\; \lambda \succeq 0 \end{aligned}$$

  - 实际上:

$$g(\lambda) = \inf_x c^\top x + \lambda f(x) = c^\top x^\star + \lambda f(x^\star)$$

    其中 $x^\star$ 是使 $\mathcal{L}$ 在取定 λ 时的下界.

    这是关于 λ 的仿射函数, 所以 $g(\lambda)$ 是凸的. 又 $\lambda \succeq 0$ 是凸函数不等式约束, 所以对偶问题是凸问题.

Problem 3: KKT conditions

Consider the problem

$$\begin{aligned} & \min_{x \in \mathbb{R}^2} x_1^2 + x_2^2 \\ \text{s.t.} \;\; & (x_1 - 1)^2 + (x_2 - 1)^2 \leq 2 \\ & (x_1 - 1)^2 + (x_2 + 1)^2 \leq 2 \end{aligned}$$

where $x = \left[\begin{array}{cc} x_1 & x_2 \end{array} \right]^\top \in \mathbb{R}^2$.
a) Write the Lagrangian for this problem.
b) Does strong duality hold in this problem?
c) Write the KKT conditions for this optimization problem.

- **解:**

  (a)

$$
\begin{aligned}
& \mathcal{L}\left(x_1, x_2, \lambda, v\right) \\
& = x_1^2 + x_2^2 + \lambda\left(\left(x_1-1\right)^2+\left(x_2-1\right)^2-2\right) + v\left(\left(x_1-1\right)^2+\left(x_2+1\right)^2-2\right) \\
& = x_1^2 + x_2^2 + \lambda\left(x_1^2+x_2^2+2\left(x_1+x_2\right)\right) + v\left(x_1^2+x_2^2-2x_1+2x_2\right) \\
& = (1+\lambda+v)x_1^2 + (1+\lambda+v)x_2^2 - 2(\lambda+v)x_1 - 2(\lambda-v)x_2
\end{aligned}
$$

  其中 $\lambda, v \geqslant 0$.

  (b)

  不妨将 $\mathcal{L}$ 写成矩阵形式:

$$
\begin{aligned}
& \text{Let } A = \left(\begin{array}{cc} 1+\lambda+v & 0 \\ 0 & 1+\lambda+v \end{array}\right),\ B = \left(\begin{array}{c} -\lambda-v \\ -\lambda+v \end{array}\right) \\
& \mathcal{L}(x, \lambda, v) = x^\top A x + 2B^\top x
\end{aligned}
$$

  由 $\frac{\partial}{\partial x}\mathcal{L}(x, \lambda, v) = 2Ax + 2B$, $\mathcal{L}$ 在 $\hat{x} = -A^\dagger B$ 时取得最小.
  所以对偶函数为:

$$
g(\lambda, v) = \begin{cases} -B^\top A^\dagger B = -\frac{(\lambda+v)^2+(\lambda-v)^2}{1+\lambda+v} & A \succeq 0 \\ -\infty & \text{otherwise} \end{cases}
$$

  进一步地, 对偶问题为:

$$
\begin{aligned}
& \max -B^\top A^\dagger B = -\tfrac{(\lambda+v)^2+(\lambda-v)^2}{1+\lambda+v} \\
& \text{s.t.} \quad A \succeq 0,\ B \in \mathcal{R}(A),\ \lambda, v \geqslant 0
\end{aligned}
$$

  由Slater条件, 存在 $x = (1,0)^\top \in \text{relint}\mathcal{D}$ 使不等式约束严格成立. 且原问题是凸问题, 所以强对偶性成立.

  - 另一种证法:
    注意到该问题具有可微的目标函数和约束函数, 画图知最优解 $x^\star = (0,0)^\top$, 且满足Slater条件, 将 $x^\star$ 代入到下题的KKT条件中发现满足, 所以强对偶性成立.

  (3)

  自然地, KKT条件为:

$$
\begin{gathered}
(x_1-1)^2+(x_2-1)^2-2\leqslant 0,\quad (x_1-1)^2+(x_2+1)^2-2\leqslant 0\\
\lambda\geqslant 0,\quad v\geqslant 0\\
\lambda\left((x_1-1)^2+(x_2-1)^2-2\right)=v\left((x_1-1)^2+(x_2+1)^2-2\right)=0\\
\begin{pmatrix}2x_1\\2x_2\end{pmatrix}+\lambda\begin{pmatrix}2(x_1-1)\\2(x_2-1)\end{pmatrix}+v\begin{pmatrix}2(x_1-1)\\2(x_2+1)\end{pmatrix}=0
\end{gathered}
$$

从上自下依次为: 原不等式约束, 对偶约束, 互补松弛, 梯度为零 条件.

Problem 4: Matrix eigenvalues

We denote by $f(A)$ the sum of the largest r eigenvalues of a symmetric matrix $A\in\mathbb{S}^n$ (with $1\leq r\leq n$), i.e.,

$$f(A)=\sum_{k=1}^{r}\lambda_k(A),$$

where $\lambda_1(A),\ldots,\lambda_n(A)$ are the eigenvalues of A sorted in decreasing order. Show that the optimal value of the optimization problem

$$
\begin{aligned}
\max\quad & \boldsymbol{tr}(AX)\\
\text{s.t.}\quad & \boldsymbol{tr}X=r\\
& 0\preceq X\preceq I,
\end{aligned}
$$

with variable $X\in\mathbb{S}^n$, is equal to $f(A)$.

- **解:**

  该问题的拉格朗日函数为:

  $$\mathcal{L}(X,\lambda,v,u)=-\boldsymbol{tr}(AX)+\lambda\left(\boldsymbol{tr}(X)-r\right)-\boldsymbol{tr}(vX)+\boldsymbol{tr}(u(X-I))$$

  对 X 矩阵求导:

  $$\frac{\partial}{\partial X}\mathcal{L}=-A^\top+\lambda I-v^\top+u^\top$$

  注意到上式是一个常矩阵, 这在矩阵变量意义下是无意义的. 考虑对角化 A,X 的方式:

  $$
  \begin{aligned}
  &\max\boldsymbol{tr}\left(\operatorname{Diag}(\lambda_1,\ldots,\lambda_n)\ (VU^\top)^\top\ \operatorname{Diag}(\gamma_1,\ldots,\gamma_n)VU^\top\right)\\
  &\text{s.t.}\ \ \boldsymbol{tr}X=r\\
  &\qquad 0\preceq X\preceq I
  \end{aligned}
  $$

  其中 U,V 是 A,X 对角化后的单位正交阵, 注意到此时由于特征值矩阵的对角性, 原问题最大化的目标变成关于 $VU^\top$ 的每一个元素的线性组合, 并且权重就是对应的A,X 特征值.

  由上优化目标的线性形式, 我们可以调整特征值使 $\lambda_1>\lambda_2>\cdots>\lambda_n$, 可以得出此时必须满足 $\boldsymbol{tr}X=r$.

Problem 5: Determinant optimization

Derive the dual problem of the following problem

$$\begin{array}{ll} \min & \log\det X^{-1} \\ \text{s.t.} & A_i^T X A_i \preceq B_i,\ i=1,\dots,m \end{array}$$

where $X \in \mathbb{S}^n_{++}$, $A_i \in \mathbb{R}^{n\times k_i}$, $B_i \in \mathbb{S}^{k_i}_{++}$, $k_i \in \mathbb{N}_+$, $i=1,\dots,m$.

- **解:** 如上不等式约束是多个($i\cdot u\cdot v$ 个)仿射的, 不妨令 $A_i^{(u)} \in \mathbb{R}^n$ 是第 u 个列向量, $B_i^{(u,v)}$ 是第u行第v列个元素, 我们有:

  $$A_i^{(u)\top} X A_i^{(v)} = \boldsymbol{tr}\left((A_i^{(v)}A_i^{(u)\top})X\right) \leqslant B_i^{(u,v)}$$

  由书本例3.23我们得到 $\log\det X^{-1}$ 的共轭函数为: $f^\star(Y) = \log\det(-Y)^{-1} - n$

  根据书本结论式(5.11), 上述问题的对偶函数为:

  $$g(\lambda) = \begin{cases} \log\det\left(\sum_{u,v}\sum_{i=1}^m \lambda_i(A_i^{(v)}A_i^{(u)\top})\right) - \sum_{u,v} B^{(u,v)\top}\lambda + n & \sum_{i=1}^m \lambda_i(A_i^{(v)}A_i^{(u)\top}) \succ 0,\ \forall u,v \\ -\infty & \text{otherwise} \end{cases}$$

  这里在满足约束的条件下给出了最优值的下界. 上式也可以直接用矩阵的迹表达拉格朗日函数并求对偶函数.