Домашнее задание №2

Бондарь София
Группа Б05-021

Двойственность

Задача №1

Условие:

Выведите двойственную задачу для задачи:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x}-\mathbf{b}\|_2 + \frac{\lambda}{2}\|\mathbf{x}\|_2^2,$$

введя новую переменную $\mathbf{y} \in \mathbb{R}^m$, такую что $\mathbf{y} = A\mathbf{x} - \mathbf{b}$ и соответствующие ограничения. Параметры: $A \in \mathbb{R}^{m\times n}$, $\lambda \geq 0$. Таким образом, мы получим задачу, которая дает нижнюю оценку на решение для задачи безусловной минимизации.

Решение:

Рассмотрим функцию Лагранжа : $L(x, y, \nu) = ||y||_2 + \lambda/2||x||_2^2 + \nu^T(Ax - b - y)$. Тогда при $x \in R^n$ инфинум данной функции будет равен:

$$infL = -\nu^T + inf(||y||_2 + \lambda/2||x||_2^2 + \nu^T(Ax - y))$$

$L'_x = 0 <=> \lambda x^T + \nu^T A = 0$
$L'_y = 0 <=> y/||y|| = \nu$
тогда при $x = -A^T\nu/\lambda$ достигается минимум функции и, если выполнено условие $\nu \leq 1$, то достигается минимум функции в нуле. Теперь рассмотрим снова выражение, в которое подставим полученные значения

$$infL = -\nu^T b + inf(||y||_2 - \nu^T y) - \tfrac{1}{2\lambda}\nu^T(AA^T)\nu$$

$infL = -\infty$ при $\nu \leq 1$, иначе $infL = 0 => max(-\nu^T b - \frac{1}{2\lambda}\nu^T(AA^T)\nu)$ □

Задача №2

Условие:

[8] Рассмотрим некоторую динамическую систему с дискретным временем, которая в каждый момент характеризуется некоторым состоянием $\mathbf{x}_t \in \mathbb{R}^n, t = \overline{0, T}$. Переход из одного состояния системы в другое осуществляется при помощи выбираемого линейного преобразования:

$$\mathbf{x}_{t+1} = A_t \mathbf{x}_t.$$

В данной задаче мы будем считать, что матрицы A выбираются из некоторого конечного множества $\mathcal{A} \in \left\{A^{(1)} \dots A^{(K)}\right\} \subset \mathbb{R}^{n \times n}$. Наша задача выбрать последовательность матриц $\{A_t\}_{t=0}^{T-1}$, которая минимизирует суммарные "потери"за все время $t = \overline{0, T}$, т.е. мы хотим минимизировать функционал вида $\sum\limits_{i=1}^{T} f(\mathbf{x}_t)$ для некоторой заданной функции f. Формально эту задачу можно записать в следующем виде:

$$\begin{aligned} &\min_{\substack{\mathbf{x}_1 \dots \mathbf{x}_T \in \mathbb{R}^n, \\ u_0 \dots u_{T-1} \in \{1 \dots K\}}} \sum_{i=1}^{T} f(\mathbf{x}_t), \\ &\text{s.t.} \mathbf{x}_{t+1} = A^{(u_t)} \mathbf{x}_t, t = \overline{0, T-1}. \end{aligned} \tag{1}$$

Начальное состояние $\mathbf{x}_0$ считаем данным. Функция f для дальнейшего исследования не обязательно должна быть выпуклой, но мы будем считать, что нам известна ее сопряженная функция $f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbb{R}^n} \left(\mathbf{y}^\top \mathbf{x} - f(\mathbf{x})\right)$.

Данная задача является невыпуклой и достаточно тяжело решить её. Вместо этого, мы рассмотрим эвристический подход, основанный на двойственности.

1. [2] Найдите двойственную задачу для задачи (1). Выражение для двойственной задачи может содержать $\mathbf{x}_0$, $A^{(1)} \dots A^{(K)}$, функцию f и её сопряженную f^*.

2. [1] Пусть $\nu_1^* \dots \nu_T^* \in \mathbb{R}^T$ есть оптимальное значение двойственной переменной, соответствующей T ограничениям исходной задачи. По данным $\nu_1^* \dots \nu_T^*$ мы будем искать u_t следующим образом:

$$(\tilde{u}_0 \dots \tilde{u}_{T-1}) \in_{\{u\}_{t=0}^{T-1} \in [K]^T} \inf_{x_1 \dots x_T} L\left(x_1 \dots x_T, u_0 \dots, x_{T-1}, \nu_1^* \dots, \nu_T^*\right).$$

   Опишите алгоритм нахождения $\tilde{u}_0 \dots \tilde{u}_{T-1}$ таким способом. В данном пункте Вы должны описать, какие задачи оптимизации Вы решаете на каждом шаге и что Вы получаете в результате. Задачи оптимизации при построении алгоритма должны быть выписаны настолько просто, насколько возможно, и решены аналитически, если это возможно.

3. [1] Рассмотрим случай $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q \mathbf{x}$, где $Q \in \mathbb{S}^n_{++}$. Найдите сопряженную функцию $f^*(\mathbf{y})$ для нее.

Решение:

1. Рассмотрим функцию Лагранжа :
   $L(\{x_i\}_{i=1}^{T}, \{u_j\}_{j=0}^{T-1}, \{\nu_k\}_{k=0}^{T-1}) = \sum_{i=1}^{T} f(x_i) + \sum_{i=0}^{T-1} \nu_i^T (x_{i+1} - A^{u_i} x_i)$

Вытащим из первой суммы последнее слагаемое а из второй суммы первое:
$L(\{x_i\}_{i=1}^T, \{u_j\}_{j=0}^{T-1}, \{\nu_k\}_{k=0}^{T-1}) = \sum_{i=1}^{T-1}(f(x_i) + (\nu_{i-1}^T - \nu_i^T A^{u_i})x_i) + f(x_T) - \nu_0^T A^{u_0} x_0 + \nu_{T-1}^T x_T$.

По условию $f^*(\mathbf{y}) = \sup_{\mathbf{x}\in\mathbb{R}^n}\left(\mathbf{y}^\top\mathbf{x} - f(\mathbf{x})\right)$.
А это значит, что $-f^*(-\nu) = -\sup_x(-\nu^T x - f(x)) = \inf_x(\nu^T x + f(x))$

Возвращаясь к первому равенству, из второго тогда будет следовать:
$\inf_x L(\{x_i\}_{i=1}^T, \{u_j\}_{j=0}^{T-1}, \{\nu_k\}_{k=0}^{T-1}) = -\nu_0^T A^{u_0} x_0 - f^*(\nu_{T-1}) - \sum\limits_{i=1}^{T-1} f^*(A^{u_i}\nu_i - \nu_{i-1})$

Двойственная задача выражается так:
$h(\nu) = -f^*(\nu_{T-1}) - \max_A(\nu_0^T A x_0) - \sum\limits_{i=1}^{T-1} \max_A(f^*(A\nu_i - \nu_{i-1}))$ $\square$

2. Чтоб найти $\tilde{u}_t$:
- Переберем все матрицы, так чтоб максимизировать следующие величины:
- Для $t = 0$ хоттим максимизировать $\nu_0^T A x_0$
- Для остальных t хотим максимизировать $f^*(A\nu_t - \nu_{t-1})$
- Ответом является номер лучшей (для максимизации) матрицы

3. Если $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x}$, то
$f^*(y) = \sup_x(y^T x - f(x)) =>$

$\nabla_x = y - Qx = 0 => x = Q^{-1}y =>$

$f^*(y) = \frac{1}{2}y^T Q^{-1} y$ $\square$

Условия оптимальности

Задача №3

Условие:

Пусть параметры $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n_{++}$ имеют положительные компоненты, при этом компоненты первого вектора отсортированы в порядке убывания $a_n \geq a_k \geq \dots a_1 > 0$, а компоненты второго вектора определены, как $b_k = \frac{1}{a_k}$. Выведите условия ККТ для задачи

$$\min_{\mathbf{x} \in \mathbb{R}^n} - \log(\mathbf{a}^\top \mathbf{x}) - \log(\mathbf{b}^\top \mathbf{x})$$
$$\text{s.t. } \mathbf{x} \geq 0, \quad f1^\top \mathbf{x} = 1$$

и покажите, что вектор $\mathbf{x} = \left(\frac{1}{2}, 0, 0 \dots, 0, 0, \frac{1}{2}\right)^\top$ является решением этой задачи.

[1] Пусть $A \in \mathbb{S}^n_{++}$. Примените результат первой части задачи для вектора $\mathbf{a} = (\lambda_n \dots \lambda_1)^\top$, собственные значения в котором расположены в порядке убывания, чтобы доказать неравенство Канторовича:

$$2\left(\mathbf{u}^\top A \mathbf{u}\right)^{\frac{1}{2}} \left(\mathbf{u}^\top A^{-1} \mathbf{u}\right)^{\frac{1}{2}} \leq \sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}},$$

где вектор $\mathbf{u} \in \mathbb{R}^n, \|\mathbf{u}\|_2 = 1$.

Hint: Если $A\mathbf{v} = \lambda \mathbf{v}$ и A обратимая матрица, то $A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}$.

Решение:

1. a. Задача имеет вид
min $f_0(x) = -log(a^T x) - log(b^T x)$, s.t. $f_i(x) = -x_i \leq 0$, $h(x) = \sum x_i - 1 = 0$

При этом это задача выпуклой оптимизации ($f_i(x)-$ выпуклые, h(x) - афинная -log $(a^T x)$ и - log$(b^T x)$ выпуклые), а значит ККТ - достаточные условия

ККТ:
$f_i(x) \leq 0$
h(x) = 0
$\lambda_i^* \geq 0$
$\lambda_i^* f_i(x) = 0$
$\nabla f_0(x^*) + \sum \lambda_i^* \nabla f_i(x^*) + \nu^* \nabla h(x^*) = 0$

b. Преобразуя последнее уравнение получим (e - единичный вектор, e_i - вектор со всеми нулями и 1 на i месте) $-\frac{a}{a^T x} - \frac{b}{b^T x} - \sum \lambda_i^* + \nu^* = 0$

Тогда $x^* = (1/2, 0...0, 1/2)^T$ удовлетворяет ККТ при $\lambda_2^* = 0, \lambda_i^* = -\frac{a}{a^T x} - \frac{b}{b^T x} + 2$ (i от 2 до n-1)$, \nu^* = 2$

1 и 2 условия выполняются
условие 3: при i = 1 , n выполняется равенство, при других:

$\lambda_i^* = 2 - \frac{2a_i^2 + 2a_1 a_n}{a_i(a_1+a_n)} \geq 0$
$(a_n - a_1)(a_i - a_1) \geq 0$ - выполнено

4 условие: при i от 2 до n-1 $f_i(x^*) = 0$, те 4 тоже выполнено

Подстановкой x^* в уравнение 5 ККТ убеждаемся, что i компонента обнуляется при i от 2 до n-1. При i = 1 , n $b_i = 1/a_i$

$-\frac{2a_i}{a_1+a_n} - \frac{2a_1 a_2}{a_i(a_1+a_2)} + 2 = 0$ равенство выполняется
тогда уравнение 5 выполняется.
чтд

2. $A = O^T \sum O$, $v = Ou => u^T A^{-1} = v^T \sum^{-1} v = \sum v_i^2/\lambda_i$
love sonya
тогда неравенство: $2(v^T a)^{1/2}(v^T b)^{1/2} \leq (\lambda_1 + \lambda_n)^{1/2}(1/\lambda_1 + 1/\lambda_n)^{1/2}$

логарифмируем $ln(v^T a) + \frac{1}{2} ln(v^T b) \leq \frac{1}{2}(ln(\lambda_1 + \lambda_n) + ln(1/\lambda_1 + 1/\lambda_n))$

$-ln(v^T a) - \frac{1}{2} ln(v^T b) \geq -ln(x^{*T} a) - ln(x^{*T} b)$

тогда x^* - минимум f_0 (тк v, a, b удовлетворяет условиям пункта a) $=> f_0(x^*) \leq f_0(v)$
$\square$

Задача №4

Рассмотрим задачу проекции на симплекс:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}-\mathbf{y}\|^2$$
$$\text{s.t. } \sum_{i=1}^{n} x_i = 1, \mathbf{x} \geq 0.$$

1. [2] Выведите условия ККТ, упростите их и сведите задачу к нахождению двойственной переменной $\nu \in \mathbb{R}$, соответствующей единственному ограничению равенству в исходной задаче.
2. [1] Постройте алгоритм для нахождения оптимального ν с учетом зависимости $x(\nu)$, полученной в предыдущем пункте. Оцените сложность данного подхода.

Решение:

1) Для начала запишем необходимые условия ККТ:
a. $x, \lambda \geq 0$
b. $\sum x_i = 1$
c. $\lambda_i x_i = 0$
d. $2x - 2y - \lambda + \nu e = 0$
из этих условий следует, что $x_i = max(0, y_i + \nu e/2)$.
Для таких x выполнены всу условия, кроме условия b. Для его выполнения необходимо выполнение следующего равенства: $\sum max(0, y_i + \nu e/2) = 1$

2) Теперь рассмотрим нужный алгоритм для поиска оптимального значения, для этого введем функцию $l(\nu) = \sum max(0, y_i + \nu e/2) = 1$.
a. рассмотим отрезок от 0 до C (так как при рассмотрении функции нетрудно заметить, что при $nu = 0$ она тоже равна 0, а при $\nu = \infty$ она равна бесконечности и при это она монотонна на этом участке.)
b. выберем мелкость разбиения (M - разбиение) и разобьем наш отрезок
c. теперь будем рассматривать значения в получившихся точках (концы отрезка нашего разбиения). С помощью бинпоиска мы сможем найти решение.
На каждом шаге мы считаем значение функции l за O(n), бинпоиск будет работать за log(|M|) - итого сложность данного подхода **O(nlog(|M|))** □

Задача №5

Условие:

В данной задаче мы докажем следующий геометрический результат:

Пусть C есть многогранник в $\mathbb{R}^n$ вида $C = \{\mathbf{x}| -1 \leq \mathbf{a}_i^\top \mathbf{x} \leq 1, i = \overline{1,p}\}$, такой что $\sum_{i=1}^{p} \mathbf{a}_i\mathbf{a}_i^\top \succeq 0$. Тогда рассмотрим эллипсоид $\mathcal{E} = \{\mathbf{x}|\mathbf{x}^\top Q^{-1}\mathbf{x} \leq 1\}, Q \in \mathbb{S}^n_{++}$ максимального объема, такой что он вписан в C, т.е. $\mathcal{E} \subseteq C$. Тогда $\sqrt{n}\mathcal{E} = \{\mathbf{x}|\mathbf{x}^\top Q^{-1}\mathbf{x} \leq n\}$ содержит C.

[1] Покажите, что $\mathcal{E} \subseteq C$, тогда и только тогда, когда $\mathbf{a}_i^\top Q\mathbf{a}_i \leq 1, i = \overline{1,p}$.

Hint: Может быть полезным представить эллипсоид в виде $\mathcal{E} = \{Q^{1/2}\mathbf{y}| \|\mathbf{y}\|_2 \leq 1\}$.

[2] Объем эллипсоида $\mathcal{E}$ пропорционален $(\det Q)^{1/2}$. Тогда согласно пункту , мы можем определить матрицу Q эллипсоида $\mathcal{E}$ как решение следующей задачи:

$$\begin{aligned} &\min_{Q\in\mathbb{S}^n} -\log\det Q \\ &\text{s.t. } \mathbf{a}_i^\top Q\mathbf{a}_i \leq 1, i = \overline{1,p}. \end{aligned} \tag{2}$$

Пусть Q является оптимальным решением задачи (2). Тогда используя условия ККТ для этой задачи, покажите, что

$$\mathbf{x} \in C \Rightarrow \mathbf{x}^\top Q^{-1}\mathbf{x} \leq n,$$

т.е. $C \in \sqrt{n}\mathcal{E}$.

Решение:

1. $\mathcal{E} = \{Q^{1/2}\mathbf{y}| \|\mathbf{y}\|_2 \leq 1\} \in C$ <=> $||Q^{1/2}a_i||_2 = sup|a_i^T Q^{1/2} y| \leq 1$ при $||y||_2 \leq 1$ для i от 1 до p

2. введем функцию $g(\lambda) = inf_{Q\succeq 0}\ L(Q,\lambda) = inf_{Q\succeq 0}\ (logdetQ^{-1} + tr((\sum \lambda_i a_i a_i^T)Q) - \sum \lambda_i)$

заметим, что inf (log det$X^{-1} + tr(XY)$) = log det Y + n, если $Y \succ 0$, иначе $-\infty$

$-X^{-1} + Y = 0$ => $X = Y^{-1}$ при $Y \succ 0$, иначе существует ненулевое a, такое что $a^T Y a \leq 0$ при $X = I + taa^T$ получим $X = 1 + t||a||_2^2$ и log det X^{-1} + tr(XY) = - log$(1 + ta^T a)$ + trY + tr$a^T Y a$

тогда двойственная функция g(λ) = log det $\sum(\lambda_i a_i a_i^T) - \sum_i + n$ при $\sum(\lambda_i a_i a_i^T) \succ 0$, иначе $-\infty$

итого задача log det $\sum(\lambda_i a_i a_i^T) - \sum_i + n$ -> max, s.t. $\lambda \succeq 0$

ККТ:
$Q \succ 0$
$a_i^T Q a_i \leq 1$
$\lambda \succeq 0$
$\lambda_i(1 - a_i^T Q a_i) = 0$
$Q^{-1} = \sum \lambda_i a_i a_i^T$

тогда $n = \sum \lambda_i tr(Q a_i a_i^T) = \sum \lambda_i$ => $x^T Q^{-1} x = \sum \lambda_i (a_i^T x)^2 \leq \sum \lambda_i = n$ $\square$

Conic Duality

Задача №6

Условие:

Выведите двойственную задачу для задачи:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x}-\mathbf{b}\|_2^2$$
$$\text{s.t. } \|\mathbf{x}\|_1 \le \alpha; \|\mathbf{x}\|_3 \le \beta; \|\mathbf{x}\|_\infty \le \gamma,$$

где $A \in \mathbb{R}^{m\times n}$, $\alpha, \beta, \gamma > 0$ - некоторые положительные константы.

Решение:

Мы можем расписать условие через скалярные произведения двух векторов, например для $\|\mathbf{x}\|_1 \le \alpha$ - это будет входит в Лагранжиан как $\langle \begin{pmatrix} x \\ \alpha \end{pmatrix}; \begin{pmatrix} \lambda_1 \\ \nu_1 \end{pmatrix} \rangle$, где λ_i- вектор, а ν_i- число.

Тогда $L(x, \lambda, \nu) = \|A\mathbf{x}-\mathbf{b}\|_2^2 - (\lambda_1^T + \lambda_2^T + \lambda_3^T)x - \nu_1\alpha - \nu_2\beta - \nu_3\gamma \ => 0 = \nabla_x = 2A^T(Ax-b) - \lambda_i => \frac{1}{2}\sum_{i=1}^{3}\lambda_i + A^Tb = A^TAx => x = (A^TA)^{-1}(\frac{1}{2}\sum_{i=1}^{3}\lambda_i + A^Tb)$

Двойственная задача получится подставлением найденного x в $L(x, \lambda, \nu)$ с ограничениями на нормы λ_i относительно ν_i $\square$

Задача №7

Условие:

[6] На семинарах мы рассмотрели задачу вида:

$$\min_{\mathbf{x}\in\{-1,1\}^n} \mathbf{x}^\top W\mathbf{x}, \tag{3}$$

где $W \in \mathbb{S}^n$. Мы также вывели для нее двойственную задачу:

$$\begin{aligned} &\max_{\nu\in\mathbb{R}^n} -\mathbf{1}^\top\nu, \\ &\text{s.t. } W + \mathrm{diag}(\nu) \succeq 0. \end{aligned} \tag{4}$$

Данная задача дает оценку снизу на оптимальное значение (3).

1. [2] Покажите, что двойственная задача для задачи (4) имеет вид:

$$\begin{aligned} &\min_{X\in\mathbb{S}^n} \mathrm{tr}(WX), \\ &\text{s.t. } X \succeq 0, \quad X_{ii} = 1, i = \overline{1,n}. \end{aligned} \tag{5}$$

   Покажите, что если оптимальный X в задаче (5) имеет ранг 1, т.е. $\exists \mathbf{x} : X = \mathbf{x}\mathbf{x}^\top$, то $\mathbf{x}$ - есть оптимальное решение задачи (3)

2. [1] Как соотносятся оптимальные значения задач (3), (4) и (5)?

3. [2] Решите задачу (5) при помощи CvxPy.

4. [1] Восстановите приблизительное решение исходной задачи при помощи собственного разложения решения X из пункта 3. Пусть $\mathbf{v}$ есть собственный вектор матрицы X, соотвествующий максимальному собстенному значению. Тогда возьмем в качетсве аппроксимации решения исходной задачи $\hat{\mathbf{x}} = \mathrm{sign}(\mathbf{v})$. Сравните значение исходной задачи для такого вектора и полученное оптимальное значение задачи (5).

Решение

1. Введем функцию $L(\nu, x) = 1^T\nu - tr(XW + Xdiag(\nu)) = \sum \nu_i(1 - X_{ii}) - tr(XW)$,
таким образом мы получаем задачу $min(tr(WX))$ (тк она эквивалентна задаче $max(-tr(WX))$), $s.t. X \succeq 0$, $X_{ii} = 1$, i от 1 до n
Заметим, что выполнено $(xx^T)_{ii} = x_i^2$ и $tr(Wxx^T) = x^TWx$, тогда если оптимальный X в задаче (5) имеет ранг 1, т.е. $\exists \mathbf{x} : X = \mathbf{x}\mathbf{x}^\top$, то $\mathbf{x}$ - есть оптимальное решение задачи (3)

2. Пусть X будет оптимальным решением задачи (7), тогда X (при ранге 1) и оптимален в задаче (5). Нижняя оценка оптимального решения задачи (3) будет совпадать с оптимальными решениями задач (4), (5)