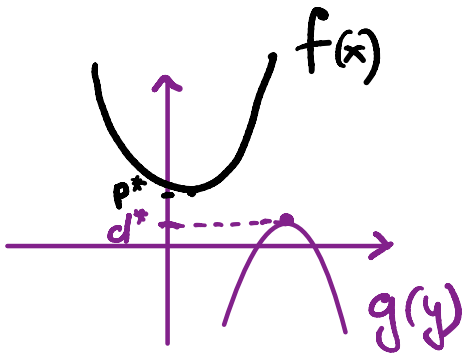
Duality

Duality lets us associate to any constrained optimization problem a concave maximization problem, whose solutions lower bound the optimal value of the original problem. What is interesting is that there are cases, when one can solve the primal problem by first solving the dual one. Now, consider a general constrained optimization problem:

$$\text{Primal: } f(x) \to \min_{x \in S} \qquad \text{Dual: } g(y) \to \max_{y \in \Omega}$$

We'll build $g(y)$, that preserves the uniform bound:

$$\boxed{g(y) \leq f(x)} \qquad \forall x \in S, \forall y \in \Omega$$



As a consequence:

$$\boxed{\max_{y \in \Omega} g(y) \leq \min_{x \in S} f(x)}$$

We'll consider one (of the many) possible way to construct $g(y)$ in case, when we have a general mathematical programming problem with functional constraints:

x^* - решение (P)

$f_0(x^*) = p^*$

Задача (P)

$$\begin{aligned} & f_0(x) \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } f_i(x) \leq 0, \; i = 1, \ldots, m \\ & \quad h_i(x) = 0, \; i = 1, \ldots, p \end{aligned}$$

And the Lagrangian, associated with this problem:

$$\boxed{L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) = f_0(x) + \lambda^\top f(x) + \nu^\top h(x)}$$

We define the Lagrange dual function (or just dual function) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

инфимум

$$\boxed{g(\lambda, \nu) = \inf_{x \in \mathbf{dom} f_0} L(x, \lambda, \nu)} = \inf_{x \in \mathbf{dom} f_0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

двойственная функция

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of

(λ, ν), it is concave, even when the original problem is not convex.

The dual function yields lower bounds on the optimal value p^* of the original problem. For any $\lambda \succeq 0, \nu$:

$$g(\lambda, \nu) \leq p^*$$

Suppose some $\hat{x}$ is a feasible point ($\hat{x} \in S$) for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \lambda^\top f(\hat{x}) + \nu^\top h(\hat{x}) \leq f_0(\hat{x})$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathbf{dom} f_0} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

A natural question is: what is the *best* lower bound that can be obtained from the Lagrange dual function? This leads to the following optimization problem:

$$\begin{aligned} g(\lambda, \nu) &\to \max_{\lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^p} \\ \text{s.t. } \lambda &\succeq 0 \end{aligned}$$

The term "dual feasible", to describe a pair (λ, ν) with $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$, now makes sense. It means, as the name implies, that (λ, ν) is feasible for the dual problem. We refer to (λ^*, ν^*) as dual optimal or optimal Lagrange multipliers if they are optimal for the above problem.

Summary

	Primal	Dual
Function	$f_0(x)$	$g(\lambda, \nu) = \min_{x \in \mathbf{dom} f_0} L(x, \lambda, \nu)$
Variables	$x \in \mathbf{dom} f_0 \subseteq \mathbb{R}^n$	$\lambda \in \mathbb{R}^m_+, \nu \in \mathbb{R}^p$
Constraints	$f_i(x) \leq 0,\ i = 1, \ldots, m$ $h_i(x) = 0,\ i = 1, \ldots, p$	$\lambda_i \geq 0, \forall i \in \overline{1, m}$
Problem	$f_0(x) \to \min_{x \in \mathbb{R}^n}$ s.t. $f_i(x) \leq 0,\ i = 1, \ldots, m$ $h_i(x) = 0,\ i = 1, \ldots, p$	$g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p}$ s.t. $\lambda \succeq 0$

Optimal	x^* if feasible, $p^* = f_0(x^*)$	λ^*, ν^* if max is achieved, $d^* = g(\lambda^*, \nu^*)$

Weak duality

It is common to name this relation between optimals of primal and dual problems as weak duality. For problem, we have:

$$p^* \geq d^*$$

While the difference between them is often called **duality gap:**

$$p^* - d^* \geq 0$$

Note, that we always have weak duality, if we've formulated primal and dual problem. It means, that if we have managed to solve the dual problem (which is always convex, no matter whether the initial problem was or not), then we have some lower bound. Surprisingly, there are some notable cases, when these solutions are equal.

Strong duality

Strong duality happens if duality gap is zero:

$$p^* = d^*$$

Notice: both p^* and d^* may be $+\infty$.

- Several sufficient conditions known!
- "Easy" necessary and sufficient conditions: unknown.

Useful features

- **Construction of lower bound on solution of the direct problem.**

  It could be very complicated to solve the initial problem. But if we have the dual problem, we can take an arbitrary $y \in \Omega$ and substitute it in $g(y)$ - we'll immediately obtain some lower bound.

- **Checking for the problem's solvability and attainability of the solution.**

  From the inequality $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ follows: if $\min\limits_{x \in S} f_0(x) = -\infty$, then

$\Omega = \varnothing$ and vice versa.

Sometimes it is easier to solve a dual problem than a primal one.

In this case, if the strong duality holds: $g(y^*) = f_0(x^*)$ we lose nothing.

Obtaining a lower bound on the function's residual.

$f_0(x) - f_0^* \leq f_0(x) - g(y)$ for an arbitrary $y \in \Omega$ (suboptimality certificate).
Moreover, $p^* \in [g(y), f_0(x)], d^* \in [g(y), f_0(x)]$

$f_0^* \geq g(y)$

$-f_0^* \leq -g(y)$

Dual function is always concave

As a pointwise minimum of affine functions.

Examples

Simple projection onto simplex with duality

To find the Euclidean projection of $x \in \mathbb{R}^n$ onto probability simplex $\mathcal{P} = \{z \in \mathbb{R}^n \mid z \succeq 0, \mathbf{1}^\top z = 1\}$, we solve the following problem:

$$\frac{1}{2}\|y - x\|_2^2 \to \min_{y \in \mathbb{R}^n \succeq 0}$$
$$\text{s.t. } \mathbf{1}^\top y = 1$$

Hint: Consider the problem of minimizing $\frac{1}{2}\|y - x\|_2^2$ subject to subject to $y \succeq 0, \mathbf{1}^\top y = 1$. Form the partial Lagrangian

$$L(y, \nu) = \frac{1}{2}\|y - x\|_2^2 + \nu(\mathbf{1}^\top y - 1),$$

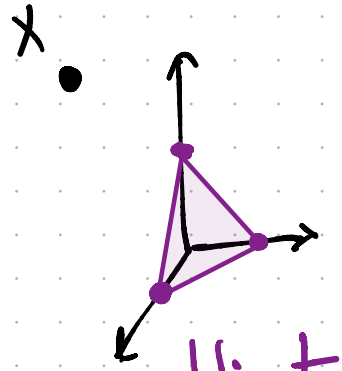
leaving the constraint $y \succeq 0$ implicit. Show that $y = (x - \nu\mathbf{1})_+$ minimizes $L(y, \nu)$ over $y \succeq 0$.

Underdetermined Linear least squares

Problem

Пример: проекция на вероятностный симплекс

$$\frac{1}{2}\|x-y\|_2^2 \to \min_{y \in \mathbb{R}^n}$$
$$1^T y = 1$$
$$y \geq 0$$



Hint

$$f_0(y) = \frac{1}{2}\|x-y\|_2^2 \quad \text{dom} f_0: \; y \geq 0$$

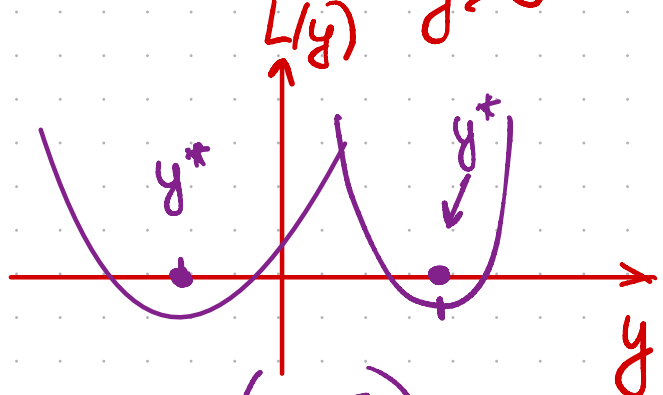
$$L(y,\nu) = \frac{1}{2}\|x-y\|_2^2 + \nu(1^T y - 1) \quad (=)$$

$$g(\nu) = ? \qquad g(\nu) = \inf_{y \in \text{dom} f} L(y,\nu)$$

$$(=) \; \frac{1}{2}\left(x^T x + y^T y - 2x^T y\right) + \nu(1^T y - 1) =$$

$$= \frac{1}{2} y^T y + (\nu 1^T - x^T) y + \frac{1}{2} x^T x - \nu \to \inf_{y \geq 0}$$

$$\boxed{y^* = (x - \nu \cdot 1)_+}$$



$$z_+ = \max(0, z)$$

$$g(\nu) = L(y^*, \nu) = \ldots\ldots$$

$$g(\lambda) \longrightarrow \max_{\lambda \in \mathbb{R}} \qquad (D)$$

В следующей серии мы покажем, что в этой задаче имеется сильная двойственность. (Условие Слейтера)

$$x^\top x \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } Ax = b$$

Lagrangian

$$L(x, \nu) = x^\top x + \nu^\top (Ax - b)$$

Dual function

$$g(\nu) = \min_{x \in \mathbb{R}^n} L(x, \nu)$$

Setting gradient to zero to find minimum 🤔:

$$\nabla_x L(x, \nu) = 2x + \nu^\top A = 2x + A^\top \nu = 0 \;\to\; x = -\frac{1}{2} A^\top \nu$$

$$g(\nu) = \frac{1}{4} \nu^\top A A^\top \nu + \nu^\top (-\frac{1}{2} A A^\top \nu - b) = -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu$$

Here we see lower bound property:

$$p^* \geq -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu, \quad \forall \nu$$

Let's solve the dual problem:

$$d^* = b^\top \left(A A^\top\right)^{-1} b$$

Calculate the primal optimal and check whether this problem has strong duality or not.

LP duality. Standard form

$$c^\top x \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } Ax = b$$
$$x \succeq 0$$

LP duality. Inequality form

$$c^\top x \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } Ax \preceq b$$

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$\begin{aligned} & x^\top Ax + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } x^\top x \le 1 \end{aligned}$$

where $A \in \mathbb{S}^n$, $A \not\succeq 0$ and $b \in \mathbb{R}^n$. Since $A \not\succeq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

Lagrangian and dual function

$$L(x, \lambda) = x^\top Ax + 2b^\top x + \lambda(x^\top x - 1) = x^\top (A + \lambda I)x + 2b^\top x - \lambda$$

$$g(\lambda) = \begin{cases} -b^\top (A + \lambda I)^\dagger b - \lambda & \text{if } A + \lambda I \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Dual problem

$$\begin{aligned} & -b^\top (A + \lambda I)^\dagger b - \lambda \to \max_{\lambda \in \mathbb{R}} \\ & \text{s.t. } A + \lambda I \succeq 0 \end{aligned}$$

$$\begin{aligned} & -\sum_{i=1}^n \frac{(q_i^\top b)^2}{\lambda_i + \lambda} - \lambda \to \max_{\lambda \in \mathbb{R}} \\ & \text{s.t. } \lambda \ge -\lambda_{min}(A) \end{aligned}$$

Connection of Fenchel and Lagrange duality

$$\begin{aligned} & f_0(x) = \sum_{i=1}^n f_i(x_i) \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } a^\top x = b \end{aligned}$$

The dual problem is thus

$$-b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i) \to \max_{\nu \in \mathbb{R}}$$
$$\text{s.t. } \lambda \geq -\lambda_{min}(A)$$

with (scalar) variable $\nu \in \mathbb{R}$. Now suppose we have found an optimal dual variable ν^* (There are several simple methods for solving a convex problem with one scalar variable, such as the bisection method.). It is very easy to recover the optimal value for the primal problem.

Fenchel - Rockafellar problem

Problem

Let $f : E \to \mathbb{R}$ and $g : G \to \mathbb{R}$ — function, defined on the sets E and G in Euclidian Spaces V and W respectively. Let $f^* : E_* \to \mathbb{R}, g^* : G_* \to \mathbb{R}$ be the conjugate functions to the f and g respectively. Let $A : V \to W$ — linear mapping. We call Fenchel - Rockafellar problem the following minimization task:

$$f(x) + g(Ax) \to \min_{x \in E \cap A^{-1}(G)}$$

where $A^{-1}(G) := \{x \in V : Ax \in G\}$ — preimage of G. We'll build the dual problem using variable separation. Let's introduce new variable $y = Ax$. The problem could be rewritten:

$$f(x) + g(y) \to \min_{x \in E,\ y \in G}$$
$$\text{s.t. } Ax = y$$

Lagrangian

$$L(x, y, \lambda) = f(x) + g(y) + \lambda^\top (Ax - y)$$

Dual function

$$\begin{aligned} g_d(\lambda) &= \min_{x \in E,\ y \in G} L(x, y, \lambda) \\ &= \min_{x \in E} \left[f(x) + (A^* \lambda)^\top x \right] + \min_{y \in G} \left[g(y) - \lambda^\top y \right] = \\ &= -\max_{x \in E} \left[(-A^* \lambda)^\top x - f(x) \right] - \max_{y \in G} \left[\lambda^\top y - g(y) \right] \end{aligned}$$

Now, we need to remember the definition of the Conjugate function:

$$\sup_{y \in G} \left[\lambda^\top y - g(y)\right] = \begin{cases} g^*(\lambda), & \text{if } \lambda \in G_* \\ +\infty, & \text{otherwise} \end{cases}$$

$$\sup_{x \in E} \left[(-A^*\lambda)^\top x - f(x)\right] = \begin{cases} f^*(-A^*\lambda), & \text{if } \lambda \in (-A^*)^{-1}(E_*) \\ +\infty, & \text{otherwise} \end{cases}$$

So, we have:

$$g_d(\lambda) = \min_{x \in E, y \in G} L(x, y, \lambda) = \\ = \begin{cases} -g^*(\lambda) - f^*(-A^*\lambda) & \text{if } \lambda \in G_* \cap (-A^*)^{-1}(E_*) \\ -\infty, & \text{otherwise} \end{cases}$$

which allows us to formulate one of the most important theorems, that connects dual problems and conjugate functions:

Fenchel - Rockafellar theorem Let $f : E \to \mathbb{R}$ and $g : G \to \mathbb{R}$ — function, defined on the sets E and G in Euclidian Spaces V and W respectively. Let $f^* : E_* \to \mathbb{R}, g^* : G_* \to \mathbb{R}$ be the conjugate functions to the f and g respectively. Let $A : V \to W$ — linear mapping. Let $p^*, d^* \in [-\infty, +\infty]$ - optimal values of primal and dual problems:

$$p^* = f(x) + g(Ax) \to \min_{x \in E \cap A^{-1}(G)}$$

$$d^* = f^*(-A^*\lambda) + g^*(\lambda) \to \min_{\lambda \in G_* \cap (-A^*)^{-1}(E_*)},$$

Then we have weak duality: $p^* \geq d^*$. Furthermore, if the functions f and g are convex and $A(\mathbf{relint}(E)) \cap \mathbf{relint}(G) \neq \varnothing$, then we have strong duality: $p^* = d^*$. While points $x^* \in E \cap A^{-1}(G)$ and $\lambda^* \in G_* \cap (-A^*)^{-1}(E_*)$ are optimal values for primal and dual problem if and only if:

$$-A^*\lambda^* \in \partial f(x^*) \\ \lambda^* \in \partial g(Ax^*)$$

Convex case is especially important since if we have Fenchel - Rockafellar problem with parameters (f, g, A), than the dual problem has the form $(f^*, g^*, -A^*)$.

References

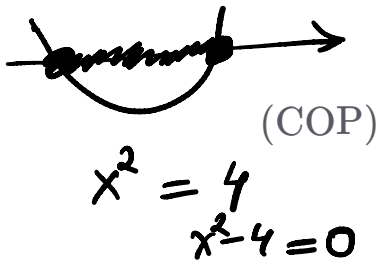
- Convex Optimization — Boyd & Vandenberghe @ Stanford

- Course Notes for EE227C. Lecture 13
- Course Notes for EE227C. Lecture 14
- Optimality conditions. KKT
- Seminar 7 @ CMC MSU
- Seminar 8 @ CMC MSU
- Convex Optimization @ Berkeley - 10th lecture

Convex optimization problem

Note, that there is an agreement in notation of mathematical programming. The problems of the following type are called **Convex optimization problem**:

$$\begin{aligned} & f(x) \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } g_i(x) \leq 0, \ i = 1, \ldots, m \\ & Ax = b, \end{aligned} \tag{COP}$$



where all the functions $f(x), g_1(x), \ldots, g_m(x)$ are convex and all the equality constraints are affine. It sounds a bit strange, but not all convex problems are convex optimization problems.

$$f(x) \to \min_{x \in S}, \tag{CP}$$

where $f(x)$ is a convex function, defined on the convex set S. The necessity of affine equality constraint is essential (see Slater's condition in Duality).

For example, this problem is not a convex optimization problem (but implies minimizing the convex function over the convex set):

$$\begin{aligned} & x_1^2 + x_2^2 \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } \frac{x_1}{1 + x_2^2} \leq 0 \\ & (x_1 + x_2)^2 = 0, \end{aligned} \tag{CP}$$

while the following equivalent problem is a convex optimization problem

$$\begin{aligned} & x_1^2 + x_2^2 \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } \frac{x_1}{1 + x_2^2} \leq 0 \\ & x_1 + x_2 = 0, \end{aligned} \tag{COP}$$

Such confusion in notation is sometimes being avoided by naming problems of type (CP) as *abstract form convex optimization problem*.

Materials

- Convex Optimization — Boyd & Vandenberghe @ Stanford

Two-way partitioning problem:

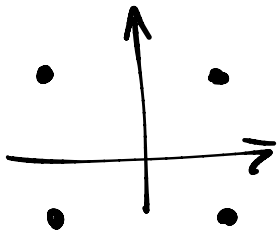
$x^T W x \to \min_{x \in \mathbb{R}^n}$ s.t. $x_i^2 = 1$

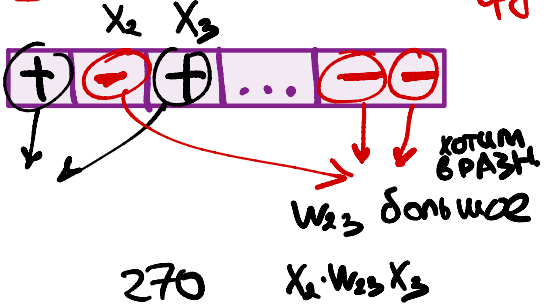
W_{ij} – стоимость нахождения i-ого и j-ого объектов в одной группе

$i = 1, \dots, n$

Брод. МН-во:

2^n точек





$$10^{80} \simeq 10^{81} \simeq (10^3)^{27} \simeq (2^{10})^{27} \simeq 2^{270}$$

Построим двойственную задачу

$g(\lambda, \nu) \to \max_{\lambda, \nu}$, $\lambda \geq 0$

1) $L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) =$

$= x^T W x + x^T \cdot \mathrm{diag}(\nu) x - 1^T \nu =$

$= x^T (W + \mathrm{diag}(\nu)) x - 1^T \nu$

2) $g(\nu) = \inf_{x \in \mathbb{R}^n} L(x, \nu) = \begin{cases} -1^T \nu & W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{иначе} \end{cases}$

3) $p^* \geq g(\nu) \quad \forall \nu$

$p^* \geq -1^T \nu$, ν: $W + \mathrm{diag}(\nu) \succeq 0$

$$p^* \geq n \cdot \lambda_{\min}(W)$$

$$\nu = -\lambda_{\min}(W)$$

вектор из одинак. знач.

Условие Слейтера: В задаче выпуклой оптимизации $\exists\, \tilde{x} \in \operatorname{relint}(S)$: $f_i(\tilde{x}) < 0$, $h_j(\tilde{x}) = 0$, то

(COP) $\min\limits_{x \in \mathbb{R}^n} f_0(x)$ ← выпуклы

$f_i(x) \le 0$ ← выпуклы

$h_j(x) = 0$ ← аффинны

имеется нулевой зазор двойственности и ККТ становятся необходимыми и достаточными

$$L(x, \lambda, \nu) = f_0(x) + \lambda^T f(x) + \nu^T h(x)$$

ККТ:

$$\begin{cases} \nabla_x L = 0 \\ \nabla_\nu L = 0 \\ \lambda \ge 0 \\ \lambda_i f_i(x) = 0 \\ f_i(x) \le 0 \end{cases}$$

если $\lambda > 0$, то $f_i(x) = 0$

если $f_i(x) < 0$, то $\lambda = 0$

ККТ • fun fact 1 • У задачи оптимизации (COP) с дифф. f_0, f_i, h_j, в которой присутствует сильная двойственность любой набор оптимальных решений (x^*, λ^*, ν^*) удовл ККТ

KKT fun fact 2

Для любой выпуклой задачи вида (COP) с дифф. f_0, f_i, h_j

ККТ становятся достаточными

BOYD $\forall (x^*, \lambda^*, \nu^*)$ из ККТ $\Rightarrow$ x^*, λ^*, ν^* - оптимальны и имеется сильная двойственность

KKT fun fact 3

Если в выпуклой задаче (COP) с дифф. f_0, f_i, h_j выполнено условие Слейтера, то ККТ становятся необходимыми и достаточными

+ и имеется сильная двойственность

+ и двойственный оптимум достижим

Условие Слейтера - одно из условий регулярности.

Существуют другие

Пример: невыпуклая задача с сильной двойственностью

Сюжет: Связь двойственности Лагранжа и Фенхеля (сопряженных функций)

$$\min_{x \in \mathbb{R}^n} f_0(Ax+b)$$

$$L(x) = f_0(Ax+b)$$

$$g = \inf_{x \in \mathbb{R}^n} L(x) \to \max_{\lambda}$$

$$= \min_{x \in \mathbb{R}^n} f_0(Ax+b)$$

$$\min_{x,z \in \mathbb{R}^n} f_0(z)$$
$$z = Ax+b$$

$$f^*(y) = \sup_x (x^T y - f(x)) =$$
$$= -\inf_x (f(x) - x^T y)$$

$$L(z, \lambda) = f_0(z) + \lambda^T (z - Ax - b)$$

$$g(\lambda) = \inf_{x,z} L(x,z,\lambda) =$$

$$= \inf_{x,z} \left[f_0(z) + \lambda^T z - \lambda^T Ax - \lambda^T b \right]$$

$\lambda^T A = 0$
$A^T \lambda = 0$

$$= \inf_z \left[f_0(z) + \lambda^T z - \lambda^T b \right] =$$

$$= \inf_z \left(f_0(z) + \lambda^T z \right] - \lambda^T b =$$

⊖ $-f_0^*(-\lambda) - \lambda^T b$

$g(\lambda) = -f_0^*(-\lambda) - \lambda^T b$, $A^T\lambda = 0$

Dual Problem:

$$g(\lambda) \to \max_{\substack{\lambda \in \mathbb{R}^n \\ A^T\lambda = 0}}$$

приложение:

$$\log\left(\sum_{i=1}^{m} \exp(a_i^T x + b_i)\right) \to \min_{x \in \mathbb{R}^n}$$

Пусть

$$a_i^T x + b_i = z_i$$

$$\log\left(\sum_{i=} \exp z_i\right) \to \min_{\substack{x, z \in \mathbb{R}^n \\ a_i^T x + b_i = z_i \\ Ax + b = z}}$$

двойственная:

$$g(\lambda) = -f_0^*(-\lambda) - \lambda^T b$$

$$f(z) = \log \sum_{i=1}^{n} \exp z_i$$

$$f_0^*(\lambda) = \sum_{i=1}^{n} \lambda_i \log \lambda_i \,, \quad \lambda \geq 0 \;\; 1^T\lambda = 1$$

$$-\sum_{i=1}^{n}(-\lambda_i)\log -\lambda_i - \lambda^T b \to \max$$

$$\lambda \in \mathbb{R}^n$$

$$-\lambda \geq 0$$

$$1^T(-\lambda) = 1$$

$$A^T\lambda = 0$$

$$-\sum_{i=1}^{n} \alpha_i \log \alpha_i + \alpha^T b \to \max$$

$$\alpha \in \mathbb{R}^n$$

$$\alpha \geq 0$$

$$1^T\alpha = 1$$

$$A^T\alpha = 0$$