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 there are cases, when one can solve the primal problem by first solving the dual one. Now, consider 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 uniform bound:

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

As a consequence:

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

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

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

And the Lagrangian, associated with this problem:

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

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 $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

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

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:

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

Hence

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

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:

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

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	$f_0(x)$	$g(\lambda, \nu) = \min\limits_{x \in \mathbf{dom} f_0} L(x, \lambda, \nu)$
Variables	$x \in \mathbf{dom} f_0 \subseteq \mathbb{R}^m$	$\lambda \in \mathbb{R}^m_+, \nu \in \mathbb{R}^m$
Constraints	$f_i(x) \le 0, \ i = 1, \ldots, m$ $h_i(x) = 0, \ i = 1, \ldots, p$	$\lambda_i \ge 0, \forall i \in \overline{1, m}$
Problem	$f_0(x) \to \min\limits_{x \in \mathbb{R}^n}$ $\text{s.t. } f_i(x) \le 0, \ i = 1, \ldots, m$ $h_i(x) = 0, \ i = 1, \ldots, p$	$g(\lambda, \nu) \to \max\limits_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p}$ $\text{s.t. } \lambda \succeq 0$
Optimal	x^* if feasible, $p^* = f(x^*)$	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:

$$p^* \ge d^*$$

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

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

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

Strong duality

Strong duality 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.

- **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 $\max_{y \in \Omega} g(y) \leq \min_{x \in S} f(x)$ follows: if $\min_{x \in S} f(x) = -\infty$, then $\Omega = \varnothing$ and vice versa.

  [handwritten: $\max g \sim \min f$]

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

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

- **Obtaining a lower bound on the function's residual.**

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

- **Dual function is always concave**

  As a pointwise minimum of affine functions.

[handwritten top right:]
$f_0(x) \to \min_x$
$f_i \leq 0$
$Ax = b$
$\inf_x f_0(x) + \lambda^T f(x) + \nu^T(Ax - b) \to \max_{\lambda \geq 0}$
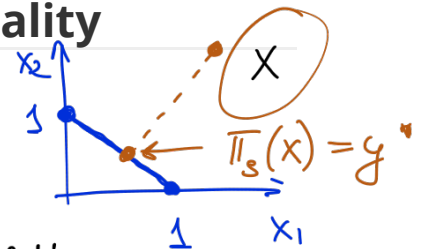ν

Examples

Simple projection onto simplex with duality

[handwritten: Пример 1:]

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

[handwritten: $z_1 + \dots + z_n = 1$; был.п.]

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

[handwritten: dom f]

[handwritten: ищем $y \in P$, δал дальше всего к x]

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

[handwritten: 1)] *[handwritten: Функция Лагранжа]*

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

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

[handwritten:]

2) Найдём $g(\nu) = ?$ $g(\nu) = \inf L(y, \nu) = $ $P_+ = \begin{cases} P_i, & P_i \geq 0 \\ 0, & P_i < 0 \end{cases}$
$\qquad\qquad\qquad\qquad\qquad y \geq 0$
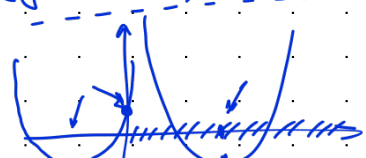
$= \inf_{y \geq 0}\left(\frac{1}{2}\|y - x\|_2^2 + \nu(\mathbf{1}^T y - 1)\right)$

Представим, что у нас деньги $= \inf_{y \geq 0}\left(\frac{1}{2}(y - x)^2 + \nu(y - 1)\right) = \inf_{y \geq 0}\left(\frac{1}{2}y^2 - xy + \frac{1}{2}x^2 + \nu y - \nu\right) =$

$= \inf_{y \geq 0}\left(\frac{1}{2}y^2 - y(x - \nu) + \frac{1}{2}x^2 - \nu\right)$ $y_{opt} = \begin{cases} (x - \nu), & x - \nu \geq 0 \\ 0, & x - \nu < 0 \end{cases}$

$\boxed{y_{opt} = (x - \nu \cdot \mathbf{1})_+}$

$z_+ = \begin{cases} z_i, & z_i \geq 0 \\ 0, & z_i < 0 \end{cases}$

[handwritten right side: x_2, x_1, $\Pi_S(x) = y$, $\mathbf{x}$]

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

$$= \frac{1}{2}\|(x - \nu \cdot 1)_+ - x\|_2^2 + \nu(1^T(x - \nu 1)_+ - 1) =$$

$$\boxed{g(\nu) : \mathbb{R} \to \mathbb{R}}$$

$$(x - \nu \cdot 1)_i \geq 0 \to x - \nu 1$$
$$(x - \nu \cdot 1)_{+_i} < 0 \to 0$$

3) Поставим двойственную задачу:

$$\frac{1}{2}\|(x - \nu \cdot 1)_+ - x\|_2^2 + \nu(1^T(x - \nu 1)_+ - 1) \to \max_\nu$$

двойств. переменная

4) Слейтер ??

зад. вып. опт? ✓

$\exists y^* \in \mathrm{relint}(S)$? ✓

$$y = \frac{1}{n} \cdot 1 \quad \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

$$\Rightarrow \boxed{\text{Сильная двойственность}}$$

$$d^* = p^*$$

Я могу максимизировать $\left(g(\nu) \to \max_\nu\right) \to \nu^*$
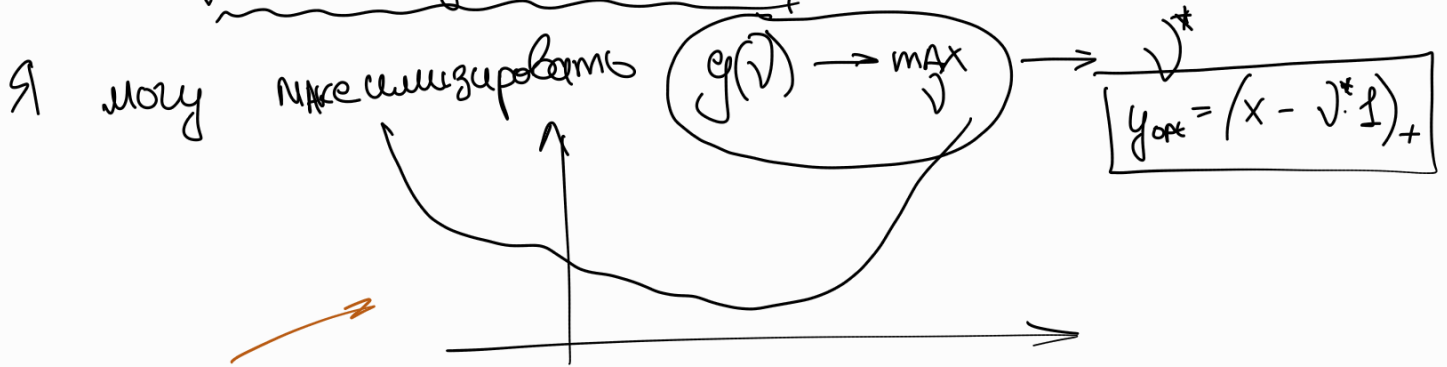
$$\boxed{y_{opt} = (x - \nu^* 1)_+}$$

1. Формулируем двойств задачу

2. Запускаем линейный поиск (метод Брента) $\to \nu^*$ //

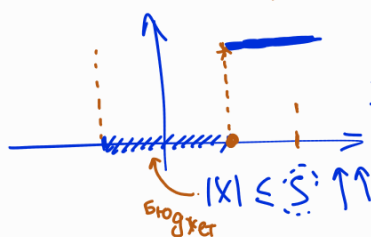3. $y_{opt}^* = (x - \nu^* 1)_+$

$$\min f_0(x)$$
$$f_1(x) \leq S$$
$$Ax = b$$

$$L(x, \lambda, \nu) = f_0(x) + \overset{\geq 0}{\lambda}(f_1(x) - S) + \mu^T(Ax - b)$$

$$\boxed{L(x^*, \lambda^*, \nu^*) = f_0(x^*)} =$$

$$= f_0(x^*) + \overset{\geq 0}{\lambda^*}\underset{=0}{(f_1(x^*) - S)} + \mu^T\overset{=0}{(Ax^* - b)}$$

$f_0(x)$

λ^* - велико

$\lambda \geq 0$

> 0

$= 0$ ОГРАН. НЕ АКТИВНО

$|x| \leq \hat{S}$

бюджет

Underdetermined Linear least squares

Problem

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

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 🫠:

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

$$g(\nu) = \frac{1}{4} \nu^\top A A^\top \nu + \nu^\top \left(-\frac{1}{2} A A^\top \nu - b\right) = -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu$$
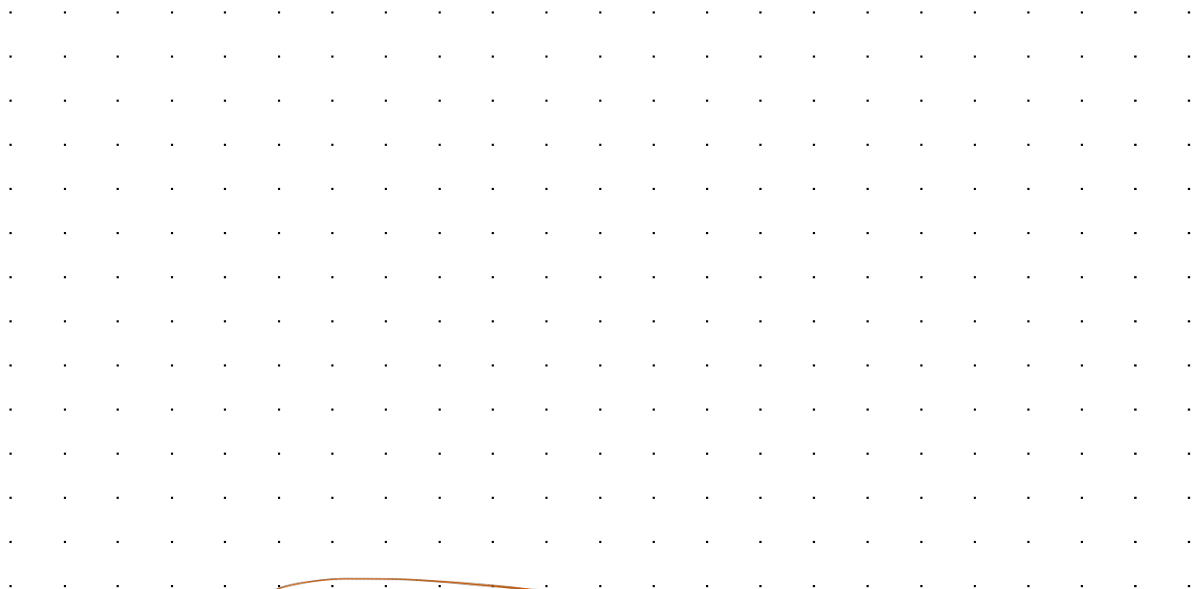
Here we see lower bound property:

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

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

Linear Programming

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

условие ??
слейтера ..

1) +
2) Если $\exists \; x^*: \boxed{\begin{array}{l} Ax^* = b \\ x^* > 0 \end{array}} \implies p^* = d^*$

$$1)\ L(x,\mu) = c^T x + \mathcal{v}^T(Ax - b) + \mu^T(-x) \qquad \mu \in \mathbb{R}^n_+ \quad -x \le 0$$
$$L(x,\mathcal{v}) = c^T x + \mathcal{v}^T(Ax - b) \qquad x \ge 0 \in \operatorname{dom} f \qquad \mathcal{v} \in \mathbb{R}^m$$

2) Построим $g(\mathcal{v}, \mu)$ $g(\mathcal{v})$:

$$g(\mathcal{v}, \mu) = \inf_x c^T x + \mathcal{v}^T(Ax - b) - \mu^T x = \inf_x \left(c + A^T\mathcal{v} - \mu \right)^T x - \mathcal{v}^T b$$

если $c + A^T\mathcal{v} - \mu = 0 \implies \forall x : \boxed{g(\mathcal{v}, \mu) = -\mathcal{v}^T b}\ \checkmark$

если $c + A^T\mathcal{v} - \mu \ne 0 \implies g(\mathcal{v}, \mu) = -\infty$

$\implies$ двойственная задача

$b^T\mathcal{v} \to \min$

$-b^T\mathcal{v} \to \underset{\mathcal{v} \in \mathbb{R}^m}{\max}$
$-A^T\mathcal{v} \le c$

LP duality. Inequality form

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

$\tilde{c} = b$
$\tilde{A} = -A \qquad \tilde{b} = c$

$$1)\ L(x,\mu) = c^T x + \mu^T(Ax - b)$$

$$2)\ g(\mu) = \inf_x L(x,\mu) = \inf_x \left(c + A^T\mu \right)^T x - b^T\mu$$

$$c + A^T\mu = 0 \implies g(\mu) = -b^T\mu.$$

$c^T x \to \min_{x \in \mathbb{R}^n}$
$Ax = b$
$x \ge 0$ **LP. Standard form**

$b^T\mu \to \min$
$\mu \in \mathbb{R}^m$
$c + A^T\mu = 0$
$\mu \ge 0$

Dual to LP. Standard form имеет форму LP. Inequality form

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

$$x^\top A x + 2 b^\top x \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } x^\top x \le 1$$

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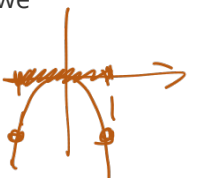
where $A \in \mathbb{S}^n$, $A \not\succeq 0$ and $b \in \mathbb{R}^n$. Since $A \not\succeq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which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

Любая задача миним. квадр функции на квадрат. ограничениях обладает сильной двойственностью

S. Boyd. $\qquad p^* = d^*$
convex optimization

Lagrangian and dual function

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

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

(handwritten annotations:)

$\nabla_x L = 2(A+\lambda I)x^* + 2b = 0$

$(A+\lambda I)x^* = -b$

если $\det(A+\lambda I) = 0,$ то проблемы

$x^* = -(A+\lambda I)^\top b$

$g(\lambda) = \inf_{x \in \mathrm{dom} f} L(x, \lambda)$

Dual problem

$$-b^\top (A + \lambda I)^\dagger b - \lambda \to \max_{\lambda \in \mathbb{R}}$$

$$\text{s.t. } A + \lambda I \succeq 0$$

$$-\sum_{i=1}^{n} \frac{(q_i^\top b)^2}{\lambda_i + \lambda} - \lambda \to \max_{\lambda \in \mathbb{R}}$$

$$\text{s.t. } \lambda \geq -\lambda_{min}(A)$$

(handwritten annotations:)

q_i - собств. вект

λ_i - собств зна

одно и то же

$A + \lambda I \geq 0$

$\lambda_{min}(A + \lambda I) \geq 0$

$\lambda_{min}(A) + \lambda \geq 0$

$\lambda \geq -\lambda_{min}(A)$

Connection of Fenchel and Lagrange duality

$$f_0(x) = \sum_{i=1}^{n} f_i(x_i) \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } a^\top x = b$$

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 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

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

Now, we need to remember the definition of the {% include link.html title='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](#)
- {% include link.html title='Optimality conditions. KKT' %}
- [Seminar 7 @ CMC MSU](#)
- [Seminar 8 @ CMC MSU](#)
- [Convex Optimization @ Berkley - 10th lecture](#)