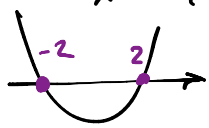
1 Convex optimization problem

DUALITY

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

$x^2 - 4 = 0$



COP

$$
\boxed{\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& Ax = b,
\end{split}}
\tag{COP}
$$

where all the functions $f_0(x), f_1(x), \ldots, f_m(x)$ are convex and all the equality constraints are affine. It sounds a bit strange, but not all convex problems are convex optimization problems.

$$
f_0(x) \to \min_{x \in S},
\tag{CP}
$$

where $f_0(x)$ is a convex function, defined on the convex set S. The necessity of affine equality constraint is essential.

> **Example**
>
> This problem is not a convex optimization problem (but implies minimizing the convex function over the convex set):
>
> $$
> \begin{split}
> & x_1^2 + x_2^2 \to \min_{x \in \mathbb{R}^n} \\
> \text{s.t. } & \frac{x_1}{1 + x_2^2} \leq 0 \\
> & (x_1 + x_2)^2 = 0,
> \end{split}
> \tag{CP}
> $$
>
> НЕ АФИННА
>
> while the following equivalent problem is a convex optimization problem
>
> $$
> \begin{split}
> & x_1^2 + x_2^2 \to \min_{x \in \mathbb{R}^n} \\
> \text{s.t. } & \frac{x_1}{1 + x_2^2} \leq 0 \\
> & x_1 + x_2 = 0,
> \end{split}
> \tag{COP}
> $$

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

2 Materials

- [Convex Optimization — Boyd & Vandenberghe @ Stanford]

Lagrange Duality

1 Motivation

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:

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

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

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

$\max\limits_{y \in \Omega} g(y) \leq p^*$

As a consequence:

$$
d^* = \boxed{\max_{y \in \Omega} g(y) \leq \min_{x \in S} f(x)} = p^*
$$

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

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

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 assume $\boxed{S = \bigcap\limits_{i=0}^m \textbf{dom } f_i \cap \bigcap\limits_{i=1}^p \textbf{dom } h_i}$ is nonempty. 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 $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

конструкт. способ построения $g(y)$

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

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.

двойств. функц $y = \lambda, \nu$

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. 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:

$$
L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{= 0} \leq f_0(\hat{x})
$$

$\hat{x} \in S$

Hence

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

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

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:

Двойственная задача.

$$\boxed{\begin{array}{l} g(\lambda,\nu) \to \max\limits_{\lambda\in\mathbb{R}^m_+,\ \nu\in\mathbb{R}^p} \\ \cancel{\text{s.t. } \lambda \succeq 0} \end{array}}$$

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.

1.1 Summary

ВСЕГДА ВОГНУТАЯ (КАК ПОТОЧ. МИН ВОГН. ФУНК.)

	Primal	Dual
Function	$f_0(x)$	$g(\lambda,\nu) = \min\limits_{x\in S} L(x,\lambda,\nu)$
Variables	$x \in S \subseteq \mathbb{R}^n$	$\lambda \in \mathbb{R}^m_+, \nu \in \mathbb{R}^p$
Constraints	$f_i(x) \le 0,\ i = 1,\dots,m$ $h_i(x) = 0,\ i = 1,\dots,p$	$\lambda_i \ge 0, \forall i \in \overline{1,m}$
Problem	$f_0(x) \to \min\limits_{x\in\mathbb{R}^n}$ s.t. $f_i(x) \le 0,\ i = 1,\dots,m$ $h_i(x) = 0,\ i = 1,\dots,p$	$g(\lambda,\nu) \to \max\limits_{\lambda\in\mathbb{R}^m,\nu\in\mathbb{R}^p}$ s.t. $\lambda \succeq 0$
Optimal	x^* if feasible, $p^* = f_0(x^*)$	λ^*, ν^* if max is achieved, $d^* = g(\lambda^*, \nu^*)$

Least-squares solution of linear equations

We are addressing a problem within a non-empty budget set, defined as follows:

$$\begin{aligned} \min \quad & x^T x \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

with the matrix $A \in \mathbb{R}^{m\times n}$.

$p^* = x^{*T}x^* = b^T(AA^T)^{-1}b$

$x^* = -\frac{1}{2}A^T\nu = -\frac{1}{2}A^T(-2)(AA^T)^{-1}b = A^T(AA^T)^{-1}b$

Solution

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T(Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

$$\nabla_x L(x, \nu) = 2x + A^T\nu = 0,$$

$\forall \nu \in \mathbb{R}^p \quad \boxed{g(\nu) \le p^*}$

leading to $\boxed{x = -(1/2)A^T\nu}$. As a result, the dual function is articulated as

$$\boxed{g(\nu) = L(-(1/2)A^T\nu, \nu) = -(1/4)\nu^T AA^T\nu - b^T\nu,}$$

emerging as a concave quadratic function within the domain $\mathbb{R}^p$. According to the lower bound property (5.2), for any $\nu \in \mathbb{R}^p$, the following holds true:

$$-(1/4)\nu^T AA^T\nu - b^T\nu \le \inf\{x^T x \mid Ax = b\}.$$

Which is a simple non-trivial lower bound without any problem solving.

Построим двойственную задачу:

$g(\nu) \to \max\limits_{\nu\in\mathbb{R}^p}$

$-\frac{1}{4}\nu^T AA^T\nu - b^T\nu \to \max\limits_{\nu\in\mathbb{R}^p}$

Dual PROBLEM

Two-way partitioning problem

$\frac{1}{4}\nu^T AA^T\nu + b^T\nu \to \min\limits_{\nu\in\mathbb{R}^p}$

$2\frac{1}{4}AA^T\nu + b = 0$

$\nu = -2(AA^T)^{-1}b$

Найдём d^*

$d^* = -\frac{1}{4}(-2)b^T\cancel{(AA^T)^{-1}}\cancel{AA^T}(-2)(AA^T)^{-1}b - b^T(-2)(AA^T)^{-1}b = b^T(AA^T)^{-1}b$

двойственность

$$x^T x \to \min_{x \in \mathbb{R}^n}$$

$$Ax = b \quad \mathbb{R}^n$$

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

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

$$\nabla_x L = 2x + A^T \nu = 0$$

$$x = -\frac{1}{2} A^T \nu$$

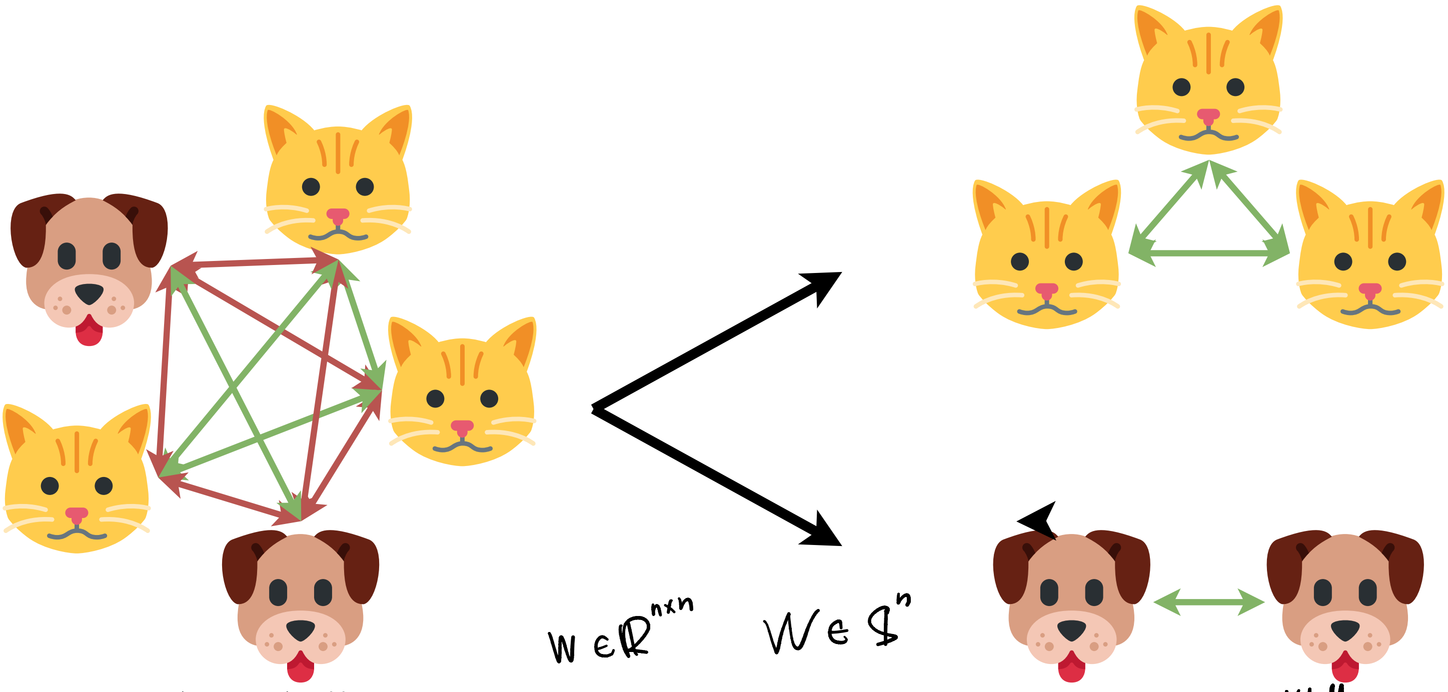
$$Ax = b \qquad A \cdot \left(-\frac{1}{2} A^T \nu\right) = b$$

$$\boxed{\nu = -2 (AA^T)^{-1} b}$$

$$g(\nu) = x^T x + \nu^T (Ax - b) =$$

$$\left(-\frac{1}{2} A^T \nu\right)^T \cdot \left(-\frac{1}{2} A^T \nu\right) + \nu^T \left(A\left(-\frac{1}{2} A^T \nu\right) - b\right)$$

$$=$$



$W \in \mathbb{R}^{n \times n}$ $\qquad W \in \mathbb{S}^n$

We are examining a (nonconvex) problem:

$\boxed{\sum_{ij} x_i \cdot w_{ij} x_j}$ $\quad x_i \cdot x_j = 1 \quad x_i x_j = -1$

$$
\boxed{
\begin{array}{ll}
\text{minimize} & x^T W x \\
\text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n,
\end{array}}
$$

• перебором не решим

• задача не выпуклая

$x_i = \pm 1$ $\qquad 2^n$ точек

Solution

The matrix W belongs to S_n. The constraints stipulate that the values of x_i can only be 1 or -1, rendering this problem analogous to finding a vector, with components ± 1, that minimizes $x^T W x$. The set of feasible solutions is finite, encompassing 2^n points, thereby allowing, in theory, for the resolution of this problem by evaluating the objective value at each feasible point. However, as the count of feasible points escalates exponentially, this approach is viable only for modest-sized problems (for instance, when $n \leq 30$). Generally, and especially when n exceeds 50, the problem poses a formidable challenge to solve.

This problem can be construed as a two-way partitioning problem over a set of n elements, denoted as $\{1, \dots, n\}$: A viable x corresponds to the partition

$$
\{1, \dots, n\} = \{i | x_i = -1\} \cup \{i | x_i = 1\}.
$$

The coefficient W_{ij} in the matrix represents the expense associated with placing elements i and j in the same partition, while $-W_{ij}$ signifies the cost of segregating them. The objective encapsulates the aggregate cost across all pairs of elements, and the challenge posed by problem is to find the partition that minimizes the total cost.

We now derive the dual function for this problem. The Lagrangian is expressed as

$x^T \operatorname{diag}(\nu) x$ $\qquad ax^2 + c$

1) Найдём L

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

$-1^T \nu$

$g(\nu) = \begin{cases} -1^T \nu, \\ -\infty \end{cases}$

By minimizing over x, we procure the Lagrange dual function:

2) Построим $g(\nu)$

$$
\boxed{g(\nu) = \inf_x x^T (W + \operatorname{diag}(\nu)) x - \mathbf{1}^T \nu} = \boxed{\begin{cases} -\mathbf{1}^T \nu & \text{if } W + \operatorname{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}}
$$

exploiting the realization that the infimum of a quadratic form is either zero (when the form is positive semidefinite) or $-\infty$ (when it's not).

This dual function furnishes lower bounds on the optimal value of the problem. For instance, we can adopt the particular value of the dual variable

$$
\boxed{\nu = -\lambda_{\min}(W) \mathbf{1}} \in \mathbb{R}^n
$$

which is dual feasible, since

$$
\boxed{W + \operatorname{diag}(\nu) = W - \lambda_{\min}(W) I \succeq 0.}
$$

This renders a simple bound on the optimal value p^*

$$
\boxed{p^* \geq -\mathbf{1}^T \nu = n \lambda_{\min}(W).}
$$

The code for the problem is available here 🧑‍💻

2 Strong duality

слабая двойственная

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

$$\boxed{p^* \geq d^*}$$

зазор двойственности

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

$$\boxed{p^* - d^* \geq 0}$$

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 concave, 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.

сильная двойственность

Strong duality happens if duality gap is zero:

$$\boxed{p^* = d^*}$$

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

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

> **Question**
>
> In the Least-squares solution of linear equations example above calculate the primal optimum p^* and the dual optimum d^* and check whether this problem has strong duality or not.
>
> $p^* = d^*$ (имеется сильная двойств.)

3 Useful features

primal

$p^* \geq g(\lambda, \nu)$ $\forall \lambda, \nu$

- **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 $\boxed{\min\limits_{x \in S} f_0(x) = -\infty}$, then $\boxed{\Omega = \varnothing}$ 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.

  $\forall x, y:$ $\min\limits_{x \in \mathbb{R}^n} f_0(x)$ $g(y)$

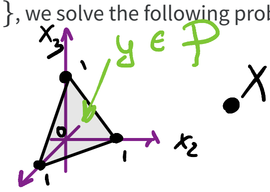
- **Obtaining a lower bound on the function's residual.**

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

- **Dual function is always concave**

  As a pointwise minimum of affine functions.

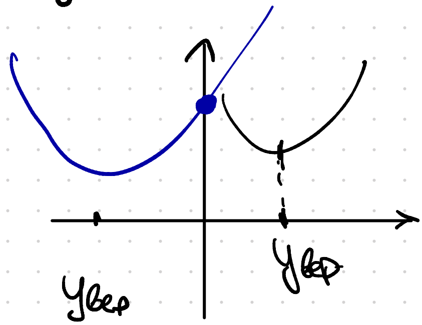
пример

> **Projection onto probability simplex**
>
> 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:
>
> $$\boxed{\begin{aligned} &\frac{1}{2}\|y - x\|_2^2 \to \min_{y \in \mathbb{R}^n, y \succeq 0} \\ &\text{s.t. } \mathbf{1}^\top y = 1 \end{aligned}}$$
>
> 
>
>
> *Hint:* Consider the problem of minimizing $|y - x|_2^2 $ subject to subject to $y \succeq 0, \mathbf{1}^\top y = 1$. Form the partial Lagrangian
>
> 1)
>
> $$\boxed{L(y, \nu) = \frac{1}{2}\|y - x\|_2^2 + \nu(\mathbf{1}^\top y - 1),} \to \min_{y \in \mathbb{R}^n_+}$$
>
> leaving the constraint $y \succeq 0$ implicit. Show that $y = (x - \nu\mathbf{1})_+$ minimizes $L(y, \nu)$ over $y \succeq 0$.

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

2) Dual problem:

$$g(\nu) = \inf_{y \in \mathbb{R}^n_+} L(y,\nu) =$$



$$y^* = \max(0, y_{вер})$$

$$= \frac{1}{2}\|y\|^2 - \cancel{\frac{1}{2}}\, x^T y + \frac{1}{2}\|x\|^2 + \nu 1^T y - \nu$$

$$\frac{1}{2}\|y\|^2 + (\nu \cdot 1 - x)^T y - \nu + \frac{1}{2}\|x\|^2 \to \min_y$$

$$y_{вер} = -(\nu 1 - x) = x - \nu \cdot 1$$

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

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

$$1^T y^* = 1$$

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

$$: \mathbb{R} \to \mathbb{R}$$

имеется

$$\boxed{g(\lambda) \to \max_{\lambda \in \mathbb{R}}}$$

решать очень легко

~~если вдруг~~ сильн. двойств.

ИМЕЕТСЯ ИЗ-ЗА условия Слейтера

$d^* = p^*$

опт. реш. двойств. зад.

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

опт. решение прямой

Projection on the Euclidian Ball

Find the projection of a point x on the Euclidian ball

$$
\begin{aligned}
&\frac{1}{2}\|y - x\|_2^2 \to \min_{y \in \mathbb{R}^n} \\
\text{s.t. } &\|y\|_2^2 \leq 1
\end{aligned}
$$

4 Slater's condition

Theorem

If for a convex optimization problem (i.e., assuming minimization, f_0, f_i are convex and h_i are affine), there exists a point x such that $h(x) = 0$ and $f_i(x) < 0$ (existance of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.

An example of convex problem, when Slater's condition does not hold

$$
\min\{f_0(x) = x \mid f_1(x) = \frac{x^2}{2} \leq 0\},
$$

The only point in the budget set is: $x^* = 0$. However, it is impossible to find a non-negative $\lambda^* \geq 0$, such that

$$
\nabla f_0(0) + \lambda^* \nabla f_1(0) = 1 + \lambda^* x = 0.
$$

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

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

where $A \in \mathbb{S}^n$, $A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, 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.

Solution

Lagrangian and dual function

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

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

Dual problem:

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

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

4.1 Reminder of KKT statements:

Suppose we have a general optimization problem (from the chapter)

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

GOP

and convex optimization problem (see corresponding chapter), where all equality constraints are affine: $h_i(x) = a_i^T x - b_i, i \in 1, \dots p$

$$
\begin{aligned}
&f_0(x) \to \min_{x \in \mathbb{R}^n}\\
\text{s.t. } &f_i(x) \leq 0, \; i = 1, \dots, m\\
&Ax = b,
\end{aligned}
\tag{2}
$$

COP

The Lagrangian is

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

The KKT system is:

$$
\begin{aligned}
&\nabla_x L(x^*, \lambda^*, \nu^*) = 0\\
&\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0\\
&\lambda_i^* \geq 0, i = 1, \dots, m\\
&\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m\\
&f_i(x^*) \leq 0, i = 1, \dots, m
\end{aligned}
\tag{3}
$$

KKT

KKT becomes necessary

If x^* is a solution of the original problem Equation 1, then if any of the following regularity conditions is satisfied:

- COP **Strong duality** If $f_1, \dots f_m, h_1, \dots h_p$ are differentiable functions and we have a problem Equation 1 with zero duality gap, then Equation 3 are necessary (i.e. any optimal set x^*, λ^*, ν^* should satisfy Equation 3)
- **LCQ** (Linearity constraint qualification). If $f_1, \dots f_m, h_1, \dots h_p$ are affine functions, then no other condition is needed.
- **LICQ** (Linear independence constraint qualification). The gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at x^*
- COP **SC** (Slater's condition) For a convex optimization problem Equation 2 (i.e., assuming minimization, f_i are convex and h_j is affine), there exists a point x such that $h_j(x) = 0$ and $ g_i(x) < 0$.

Than it should satisfy Equation 3

KKT in convex case

If a convex optimization problem Equation 2 with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality: Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained, so x^* is optimal if and only if there are (λ^*, ν^*) that, together with x^*, satisfy the KKT conditions.

5 Connection between Fenchel duality and Lagrange duality

Example

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

The dual problem is thus

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

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.

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:

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

Lagrangian

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

Dual function

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

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:

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

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((E)) (G) \$, 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:

$$\begin{split}
-A^*\lambda^* &\in \partial f(x^*) \\
\lambda^* &\in \partial g(Ax^*)
\end{split}$$

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

Example

6 References

- Convex Optimization — Boyd & Vandenberghe @ Stanford
- Course Notes for EE227C. Lecture 13
- Course Notes for EE227C. Lecture 14
- Optimality conditions
- Seminar 7 @ CMC MSU
- Seminar 8 @ CMC MSU
- Convex Optimization @ Berkeley - 10th lecture