Optimality conditions. KKT. Duality

Daniil Merkulov

Optimization methods. MIPT

> *The reader will find no ~~figures~~ PROOFS in this work. The methods which I set forth do not require either constructions or geometrical or mechanical reasonings: but only algebraic operations, subject to a regular and uniform rule of procedure.*

Preface to Mécanique analytique

Figure 1: Joseph-Louis Lagrange

Optimization with inequality constraints

Example of inequality constraints

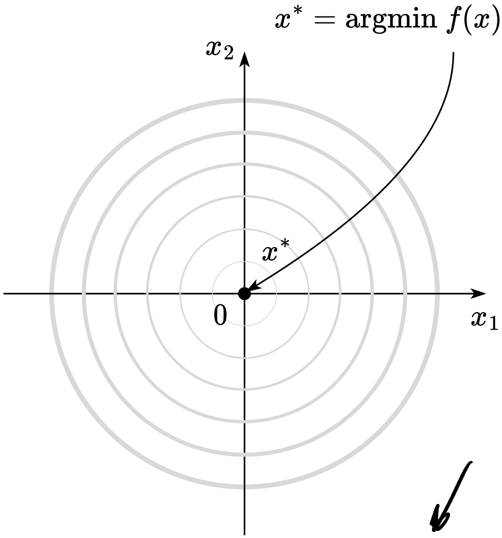
$$f(x) = x_1^2 + x_2^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

$$f(x) \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } g(x) \leq 0$$

Optimization with inequality constraints

линии уровня

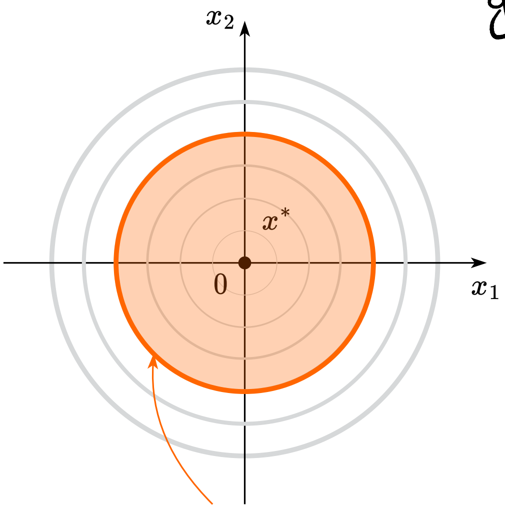


Contour lines of $f(x) = x_1^2 + x_2^2 = C$

Optimization with inequality constraints

$g(x) \leq 0$

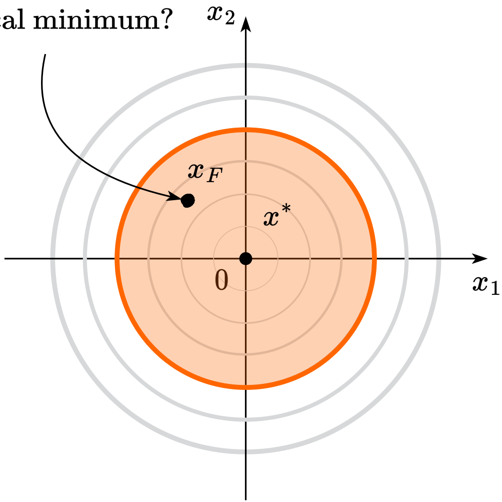
$g(x) = x_1^2 + x_2^2 - 1$



Feasible region $g(x) = x_1^2 + x_2^2 - 1 \leq 0$

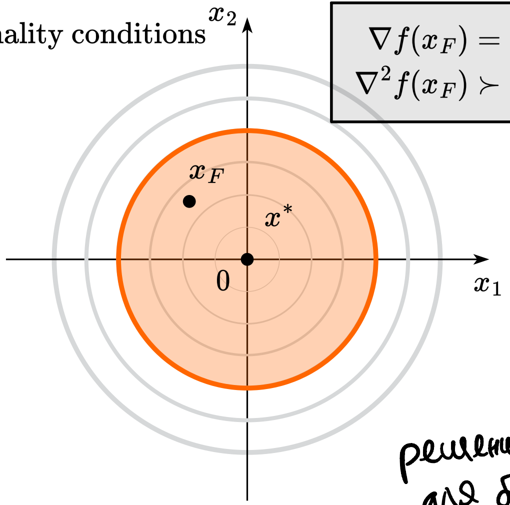
Optimization with inequality constraints

How to recognize that some feasible point is at local minimum?



Optimization with inequality constraints

Easy in this case! Just check unconstrained optimality conditions



$$\nabla f(x_F) = 0$$
$$\nabla^2 f(x_F) \succ 0$$

✓ достаточные условия

$f(x) = x_1^2 + x_2^2$

$\nabla f = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$

$\nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

0 — удовл.
0 — достат. усл.

решение для безусловной задачи

Optimization with inequality constraints

Thus, if the constraints of the type of inequalities are inactive in the constrained problem, then don't worry and write out the solution to the unconstrained problem. However, this is not the whole story. Consider the second childish example

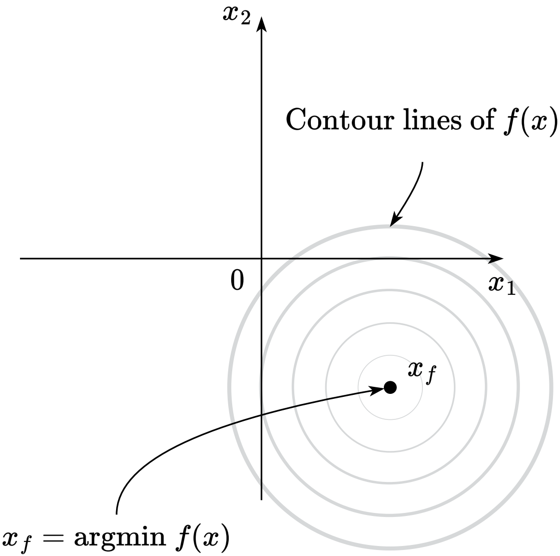
$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

$$\begin{aligned} f(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) \leq 0 \end{aligned}$$

Optimization with inequality constraints

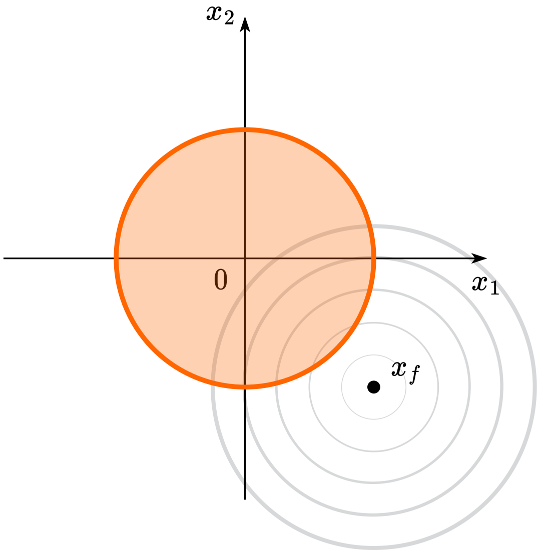
другой случай

$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 = C$$



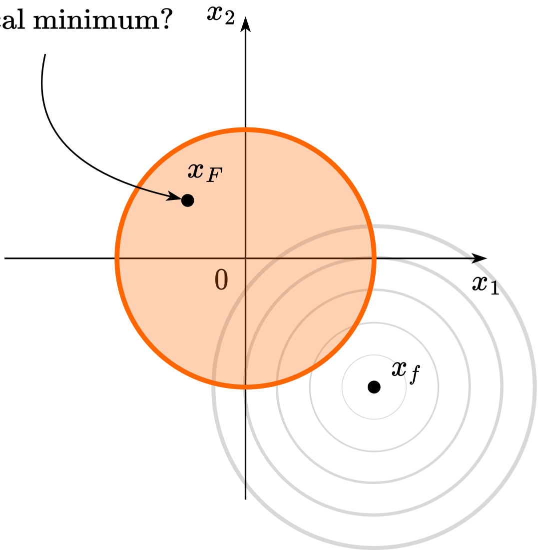
Optimization with inequality constraints

Feasible region $g(x) = x_1^2 + x_2^2 - 1 \leq 0$



Optimization with inequality constraints

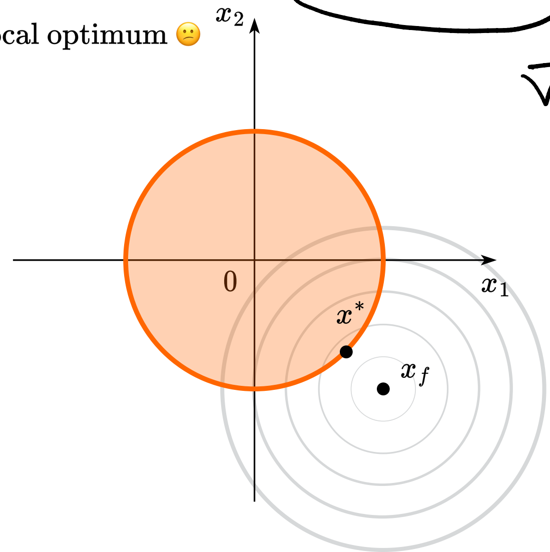
How to recognize that some feasible point is at local minimum?



Optimization with inequality constraints

Not very easy in this case! Even gradient $\neq 0$ at local optimum 😕

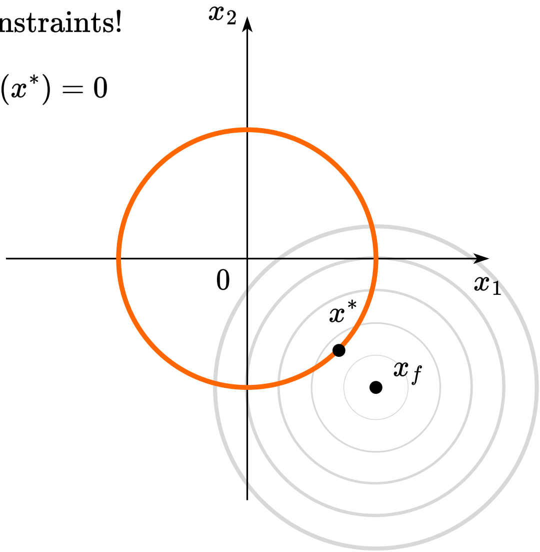
$\nabla f(x^*) \neq 0$



Optimization with inequality constraints

Effectively have a problem with equality constraints!

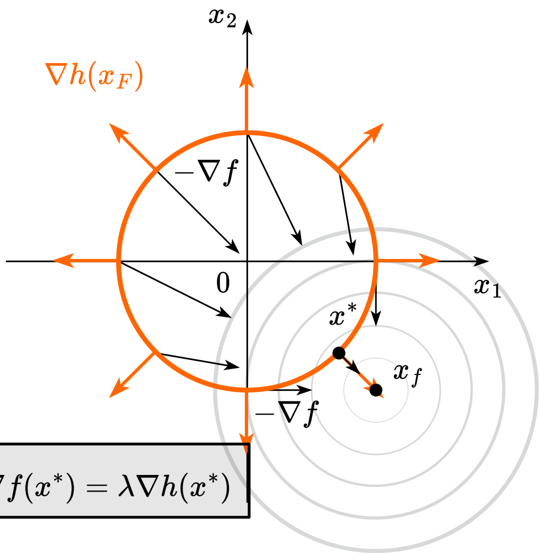
$$g(x^*) = 0$$



оптимальное решение достигается НА ГРАНИЦЕ МНОЖЕСТВА

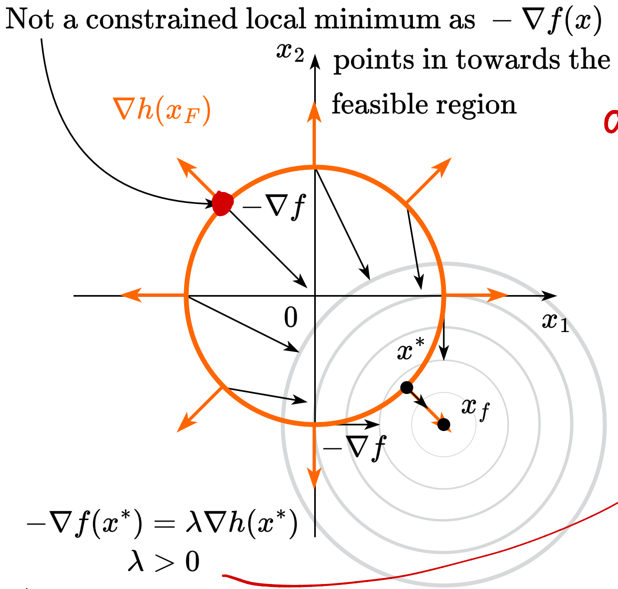
Optimization with inequality constraints

поле антиградиентов



$$\boxed{-\nabla f(x^*) = \lambda \nabla h(x^*)}$$

Optimization with inequality constraints



$$-\nabla f(x^*) = \lambda \nabla h(x^*)$$
$$\lambda > 0$$

антиградиент сонаправлен с $\nabla h(x^*)$

Optimization with inequality constraints

$f(x) \to \min_{x \in \mathbb{R}^n}$

Безусловная задача = БЗ

Условная = УЗ $\min_{x \in S} f(x)$

So, we have a problem:

$x_f = x^*$

оптимум БЗ

= оптимум УЗ

$$
\begin{aligned}
f(x) &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } g(x) &\leq 0
\end{aligned}
$$

Two possible cases:

$g(x) \leq 0$ is inactive. $g(x^*) < 0$

- $g(x^*) < 0$

$x^* \subset S$ строго внутри S

Optimization with inequality constraints

So, we have a problem:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

Two possible cases:

$g(x) \leq 0$ is inactive. $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$

Optimization with inequality constraints

So, we have a problem:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

Two possible cases:

$g(x) \leq 0$ is inactive. $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) > 0$

Optimization with inequality constraints

So, we have a problem:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

Two possible cases:

$g(x) \leq 0$ is inactive. $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) > 0$

63

Optimization with inequality constraints

So, we have a problem:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$

$$
\text{s.t. } g(x) \leq 0
$$

РЕШЕНИЕ лежит на границе S

$x^* \subseteq \partial S$

Two possible cases:

$g(x) \leq 0$ is inactive. $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) > 0$

$g(x) \leq 0$ is active. $g(x^*) = 0$

- $g(x^*) = 0$

Optimization with inequality constraints

So, we have a problem:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

Two possible cases:

$g(x) \leq 0$ is inactive. $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) > 0$

$g(x) \leq 0$ is active. $g(x^*) = 0$

- $g(x^*) = 0$
- Necessary conditions: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$

Optimization with inequality constraints

So, we have a problem:

$$
\begin{aligned}
&f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &g(x) \leq 0
\end{aligned}
$$

задача с ограничениями-
- равенствами

Two possible cases:

$g(x) \leq 0$ is inactive. $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) > 0$

Безусловная
задача

$g(x) \leq 0$ is active. $g(x^*) = 0$

- $g(x^*) = 0$
- Necessary conditions: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$
- Sufficient conditions:
  $\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0, \forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y = 0$

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
\boxed{
\begin{aligned}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{aligned}
}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

If x^* is a local minimum of the problem described above, then there exists a unique Lagrange multiplier λ^* such that:

$$
(1) \; \nabla_x L(x^*, \lambda^*) = 0
$$

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

If x^* is a local minimum of the problem described above, then there exists a unique Lagrange multiplier λ^* such that:

$$
(1)\ \nabla_x L(x^*, \lambda^*) = 0
$$
$$
(2)\ \lambda^* \geq 0
$$

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

If x^* is a local minimum of the problem described above, then there exists a unique Lagrange multiplier λ^* such that:

$$
\begin{split}
& (1) \; \nabla_x L(x^*, \lambda^*) = 0 \\
& (2) \; \lambda^* \geq 0 \\
& (3) \; \lambda^* g(x^*) = 0
\end{split}
$$

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{aligned}
&f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &g(x) \leq 0
\end{aligned}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

If x^* is a local minimum of the problem described above, then there exists a unique Lagrange multiplier λ^* such that:

Необходимые условия

$$
\begin{aligned}
&(1)\ \nabla_x L(x^*, \lambda^*) = 0 \\
&(2)\ \lambda^* \geq 0 \\
&(3)\ \lambda^* g(x^*) = 0 \\
&(4)\ g(x^*) \leq 0
\end{aligned}
$$

$g(x^*) < 0 \to \lambda^* = 0$ БЗ

$g(x^*) = 0 \to \lambda^* > 0$ Зад. с огр. РАВЕНСТВА

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

If x^* is a local minimum of the problem described above, then there exists a unique Lagrange multiplier λ^* such that:

$$
\begin{split}
& (1) \; \nabla_x L(x^*, \lambda^*) = 0 \\
& (2) \; \lambda^* \geq 0 \\
& (3) \; \lambda^* g(x^*) = 0 \\
& (4) \; g(x^*) \leq 0 \\
& (5) \; \forall y \in C(x^*) : \langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0
\end{split}
$$

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

If x^* is a local minimum of the problem described above, then there exists a unique Lagrange multiplier λ^* such that:

$$
\begin{split}
& (1) \; \nabla_x L (x^*, \lambda^*) = 0 \\
& (2) \; \lambda^* \geq 0 \\
& (3) \; \lambda^* g(x^*) = 0 \\
& (4) \; g(x^*) \leq 0 \\
& (5) \; \forall y \in C(x^*) : \langle y , \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0
\end{split}
$$

where $C(x^*) = \{y \ \in \mathbb{R}^n | \nabla f(x^*) ^\top y \leq 0$ and $\forall i \in I(x^*): \nabla g_i(x^*)^T y \leq 0\}$

Lagrange function for inequality constraints

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{aligned}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{aligned}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

The classical Karush-Kuhn-Tucker first and second-order optimality conditions for a local minimizer x^*, stated under some regularity conditions, can be written as follows.

If x^* is a local minimum of the problem described above, then there exists a unique Lagrange multiplier λ^* such that:

$$
\begin{aligned}
& (1)\ \nabla_x L(x^*, \lambda^*) = 0 \\
& (2)\ \lambda^* \geq 0 \\
& (3)\ \lambda^* g(x^*) = 0 \\
& (4)\ g(x^*) \leq 0 \\
& (5)\ \forall y \in C(x^*) : \langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0
\end{aligned}
$$

where $C(x^*) = \{y \in \mathbb{R}^n | \nabla f(x^*)^\top y \leq 0$ and $\forall i \in I(x^*) : \nabla g_i(x^*)^T y \leq 0\}$

$I(x^*) = \{i \mid g_i(x^*) = 0\}$

Теорема Каруша - Куна - Таккера

KKT

General formulation

общая
задача
мат. программирования

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

This formulation is a general problem of mathematical programming.

The solution involves constructing a Lagrange function:

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

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

Necessary conditions

сейчас читаем как: „если задача достаточно адекватная"

с нулевым зазором двойственности

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*). Let also the functions f_0, f_i, h_i be differentiable.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$

Necessary conditions

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*). Let also the functions f_0, f_i, h_i be differentiable.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$

Necessary conditions

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*). Let also the functions f_0, f_i, h_i be differentiable.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda^*_i \geq 0, i = 1, \ldots, m$

Necessary conditions

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*). Let also the functions f_0, f_i, h_i be differentiable.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda_i^* \geq 0, i = 1, \ldots, m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$

Necessary conditions

$f_0(x) \to \min$

$f_i(x) \le 0$

$0 \in \sum \lambda_i \cdot \partial f_i(x^*)$

$L = f_0(x) + \sum \lambda f_i$

$\nabla f_0(x) + \sum \lambda_i \nabla f_i(x) = 0$

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*). Let also the functions f_0, f_i, h_i be differentiable.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$ (1)
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$ (2)
- $\lambda^*_i \geq 0, i = 1, \ldots, m$ (3)
- $\lambda^*_i f_i(x^*) = 0, i = 1, \ldots, m$ (4)
- $f_i(x^*) \leq 0, i = 1, \ldots, m$ (5)

система ур-ий и нер-в.

ККТ

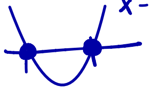
независимо по каждому ограничению

$\lambda_i > 0$

$\lambda_i = 0$

Some regularity conditions

условия „адекватности" задачи



$x^2 - 1 = 0$

These conditions are needed to make KKT solutions the necessary conditions. Some of them even turn necessary conditions into sufficient (for example, Slater's). Moreover, if you have regularity, you can write down necessary second order conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ with *semi-definite* hessian of Lagrangian.

НЕРАВ РАВ

- **Slater's condition.** If for a convex problem (i.e., assuming minimization, f_0, f_i are convex and h_i are affine), there exists a point $\tilde{x}$ such that $h(\tilde{x}) = 0$ and $f_i(\tilde{x}) < 0$ (existence of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.

Some regularity conditions

These conditions are needed to make KKT solutions the necessary conditions. Some of them even turn necessary conditions into sufficient (for example, Slater's). Moreover, if you have regularity, you can write down necessary second order conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ with *semi-definite* hessian of Lagrangian.

- **Slater's condition.** If for a convex 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 $f_i(x) < 0$ (existence of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.
- **Linearity constraint qualification.** If f_i and h_i are affine functions, then no other condition is needed.

KKT являются необх.

Some regularity conditions

These conditions are needed to make KKT solutions the necessary conditions. Some of them even turn necessary conditions into sufficient (for example, Slater's). Moreover, if you have regularity, you can write down necessary second order conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ with *semi-definite* hessian of Lagrangian.

- **Slater's condition.** If for a convex 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 $f_i(x) < 0$ (existence of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.
- **Linearity constraint qualification.** If f_i and h_i are affine functions, then no other condition is needed.
- **Linear independence constraint qualification.** The gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at x^*.

$\nabla h(x^*)$ ЛНЗ $\nabla f_i(x^*)$ $\left(f_i(x^*) = 0 \right.$

активно

Some regularity conditions

These conditions are needed to make KKT solutions the necessary conditions. Some of them even turn necessary conditions into sufficient (for example, Slater's). Moreover, if you have regularity, you can write down necessary second order conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ with *semi-definite* hessian of Lagrangian.

- **Slater's condition.** If for a convex 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 $f_i(x) < 0$ (existence of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.
- **Linearity constraint qualification.** If f_i and h_i are affine functions, then no other condition is needed.
- **Linear independence constraint qualification.** The gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at x^*.
- For other examples, see wiki.

Proof in simple case

$X = 5 \Leftrightarrow \begin{cases} x \le 5 \\ x \ge 5 \end{cases}$

Subdifferential form of KKT

Let X be a linear normed space, and let $f_j : X \to \mathbb{R}$, $j = 0, 1, \ldots, m$, be convex proper (it never takes on the value $-\infty$ and also is not identically equal to ∞) functions. Consider the problem

$$
\boxed{\begin{aligned} &f_0(x) \to \min_{x \in X} \\ \text{s.t. } &f_j(x) \le 0, \; j = 1, \ldots, m \end{aligned}}
$$

Let $x^* \in X$ be a minimum in problem above and the functions f_j, $j = 0, 1, \ldots, m$, be continuous at the point x^*. Then there exist numbers $\lambda_j \ge 0$, $j = 0, 1, \ldots, m$, such that

$\lambda_j = 0, \; f_j < 0$

$\lambda_j > 0, \; f_j = 0$

$$
\sum_{j=0}^{m} \lambda_j = 1,
$$

$$
\lambda_j f_j(x^*) = 0, \quad j = 1, \ldots, m,
$$

СКРЫТ ЛАГРАНЖИАН

$$
0 \in \sum_{j=0}^{m} \lambda_j \partial f_j(x^*).
$$

Proof in simple case (необх. условия)

Proof

1. Consider the function

$$\boxed{f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \ldots, f_m(x)\}.}$$

The point x^* is a global minimum of this function.
Indeed, if at some point $x_e \in X$ the inequality $f(x_e) < 0$ were satisfied, it would imply that $f_0(x_e) < f_0(x^*)$ and $f_j(x_e) < 0$, $j = 1, \ldots, m$, contradicting the minimality of x^* in problem above.
(противоречит тому, что x^* – оптимум исх задачи

Proof in simple case

Proof

1. Consider the function

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \ldots, f_m(x)\}.$$

   The point x^* is a global minimum of this function. Indeed, if at some point $x_e \in X$ the inequality $f(x_e) < 0$ were satisfied, it would imply that $f_0(x_e) < f_0(x^*)$ and $f_j(x_e) < 0$, $j = 1, \ldots, m$, contradicting the minimality of x^* in problem above.
2. Then, from Fermat's theorem in subdifferential form, it follows that

$$\boxed{0 \in \partial f(x^*).}$$

Proof in simple case

Proof

1. Consider the function

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \ldots, f_m(x)\}.$$

   The point x^* is a global minimum of this function. Indeed, if at some point $x_e \in X$ the inequality $f(x_e) < 0$ were satisfied, it would imply that $f_0(x_e) < f_0(x^*)$ and $f_j(x_e) < 0$, $j = 1, \ldots, m$, contradicting the minimality of x^* in problem above.

2. Then, from Fermat's theorem in subdifferential form, it follows that

$$0 \in \partial f(x^*).$$

3. By the Dubovitskii-Milyutin theorem, we have

$$\partial f(x^*) = \text{conv}\left(\bigcup_{j \in I} \partial f_j(x^*)\right),$$

   where $I = \{0\} \cup \{j : \underline{f_j(x^*) = 0}, 1 \leq j \leq m\}$.

[illegible] огран

Proof in simple case

Proof

1. Consider the function

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \ldots, f_m(x)\}.$$

   The point x^* is a global minimum of this function. Indeed, if at some point $x_e \in X$ the inequality $f(x_e) < 0$ were satisfied, it would imply that $f_0(x_e) < f_0(x^*)$ and $f_j(x_e) < 0$, $j = 1, \ldots, m$, contradicting the minimality of x^* in problem above.

2. Then, from Fermat's theorem in subdifferential form, it follows that

$$0 \in \partial f(x^*).$$

3. By the Dubovitskii-Milyutin theorem, we have

$$\partial f(x^*) = \text{conv}\left(\bigcup_{j \in I} \partial f_j(x^*)\right),$$

   where $I = \{0\} \cup \{j : f_j(x^*) = 0, 1 \leq j \leq m\}$.

4. Therefore, there exist $g_j \in \partial f_j(x^*)$, $j \in I$, such that

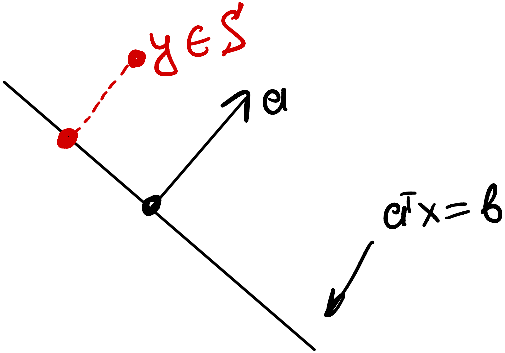
$$\boxed{\sum_{j \in I} \lambda_j g_j = 0,} \quad \sum_{j \in I} \lambda_j = 1, \quad \lambda_j \geq 0, \quad j \in I.$$

no

   It remains to set $\lambda_j = 0$ for $j \notin I$.

Example. Projection onto a hyperplane

$$
\min \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \overbrace{\mathbf{a}^T\mathbf{x} = b}^{S}.
$$



Example. Projection onto a hyperplane

поиск проекции как решение задачи оптимиз.

$$\min \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Solution

Lagrangian:

Example. Projection onto a hyperplane

$$\min \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Solution

Lagrangian:

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

Example. Projection onto a hyperplane

$$
\min \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.
$$

вып. вып.

Solution

Lagrangian:

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

Derivative of L with respect to $\mathbf{x}$:

$$
\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \quad \boxed{\mathbf{x} = \mathbf{y} - \nu\mathbf{a}} = \boxed{y + \frac{b - a^Ty}{a^Ta} a}
$$

$$
\frac{\partial L}{\partial \nu} = 0 \Rightarrow h(x) = 0 \Rightarrow a^Tx = b
$$

$$
a^T(y - \nu a) = b
$$

$$
\Rightarrow a^Ty - \nu a^Ta = b
$$

$$
\Rightarrow \boxed{\nu = \frac{b - a^Ty}{a^Ta}}
$$

Example. Projection onto a hyperplane

$$\min \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Solution

Lagrangian:

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

Derivative of L with respect to $\mathbf{x}$:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}$$

$$\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y} - \nu\mathbf{a}^T\mathbf{a} \qquad \nu = \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}$$

Example. Projection onto a hyperplane

Нужно предъявить $\tilde{x}$:

$h(\tilde{x}) = 0$

$f_i(\tilde{x}) < 0$

ККТ + Слейтер

$$
\min \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.
$$

Solution

Lagrangian:

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

Derivative of L with respect to $\mathbf{x}$:

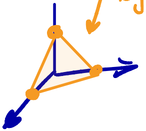
$$
\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}
$$

$$
\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y} - \nu\mathbf{a}^T\mathbf{a} \qquad \nu = \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}
$$

решение зад. опт.

$$
\mathbf{x} = \mathbf{y} - \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}
$$

Example. Projection onto simplex



един. симплекс

$$\min_x \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Example. Projection onto simplex

y - задана (не переменная)

$$
\min_x \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.
$$

$\min\limits_{x\in\mathbb{R}^n} \frac{1}{2}\|x-y\|^2$

$n+1$ множ. Лагр.

$x \geq 0 \qquad -x \leq 0$

$1^T x = 1 \qquad 1^T x - 1 = 0$

KKT Conditions

The Lagrangian is given by:

$$
\boxed{L = \frac{1}{2}\|x-y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)}
$$

Example. Projection onto simplex

$$\min_x \frac{1}{2}\|x - y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

KKT Conditions

The Lagrangian is given by:

$$L = \frac{1}{2}\|x - y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$

Example. Projection onto simplex

$\lambda_i \cdot g_i(x) = 0$

условия дополняющей нежёсткости

$$
\min \frac{1}{2}\|x - y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0. \quad x
$$

KKT Conditions

The Lagrangian is given by:

$$
L = \frac{1}{2}\|x - y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)
$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\underline{\lambda_i x_i = 0}$

Example. Projection onto simplex

$$\min_x \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

KKT Conditions

The Lagrangian is given by:

$$L = \frac{1}{2}\|x-y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$

Example. Projection onto simplex

условие Слейтера

$$
\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0. \quad x
$$

$+ \qquad + \qquad +$

$\tilde{x} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$

$-\tilde{x}_i < 0$ $\qquad \cap S$

$+$

KKT Conditions

The Lagrangian is given by:

$$
L = \frac{1}{2}\|x-y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)
$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $x^\top 1 = 1, \quad x \geq 0$

Необх. + достаточные

Example. Projection onto simplex

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0. \quad x$$

KKT Conditions

The Lagrangian is given by:

$$L = \frac{1}{2}\|x-y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1) =$$

$$= \sum_{i=1}^{n} \left[\frac{1}{2}(x_i - y_i)^2 - \lambda_i x_i + \nu x_i \right] - \nu$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $x^\top 1 = 1, \quad x \geq 0$

$$\boxed{x_i = y_i + \lambda_i - \nu}$$

$\Rightarrow$ можно решать
задачу для
независимо координаты

i:

$$\lambda_i = x_i - y_i + \nu$$

$\lambda_i = 0$

$x_i - y_i + \nu = 0$

$$\boxed{x_i = y_i - \nu}$$

$\lambda_i > 0$

$\nu - y_i > 0$

граница

$x_i = 0$

Example. Projection onto simplex

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0. \quad x$$

KKT Conditions

The Lagrangian is given by:

$$L = \frac{1}{2}\|x-y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $x^\top 1 = 1, \quad x \geq 0$

$$\Rightarrow \begin{cases} x_i = 0, & \lambda_i > 0. \\ x_i = y_i - \nu, & \lambda_i = 0 \end{cases}$$

$$\boxed{x_i = \max(0, y_i - \nu)}$$

$$\Rightarrow \sum_{i=1}^n x_i = 1 \rightarrow \text{выражается } \nu$$

Example. Projection onto simplex

на входе : $y \in \mathbb{R}^n$

$$
\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0. \quad x
$$

KKT Conditions

The Lagrangian is given by:

$$
L = \frac{1}{2}\|x-y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)
$$

Если

$$
x_i = \max(0, y_i - \nu)
$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $x^\top 1 = 1, \quad x \geq 0$

$$
\sum_i x_i = 1
$$

$$
\sum_{\substack{i:\\ x_i > 0\\ y_i > \nu}} (y_i - \nu) = 1
$$

> **Question**
>
> Solve the above conditions in $O(n \log n)$ time.

Example. Projection onto simplex

К сожалению,

$$
\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0. \quad x
$$

нет аналитическ.
решения

KKT Conditions

The Lagrangian is given by:

$$
L = \frac{1}{2}\|x-y\|^2 - \sum_i \lambda_i x_i + \nu(x^\top 1 - 1)
$$

Taking the derivative of L with respect to x_i and writing KKT yields:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $x^\top 1 = 1, \quad x \geq 0$

> **Question**
>
> Solve the above conditions in $O(n \log n)$ time.

> **Question**
>
> Solve the above conditions in $O(n)$ time.

Duality

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:

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:

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

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:

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

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

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

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:

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

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

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

As a consequence:

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

Lagrange duality

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:

Lagrange duality

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:

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

Lagrange duality

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:

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

And the Lagrangian, associated with this problem:

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

Dual function

We assume $\mathcal{D} = \bigcap\limits_{i=0}^{m} \textbf{dom } f_i \cap \bigcap\limits_{i=1}^{p} \textbf{dom } h_i$ 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 $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

Dual function

We assume $\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom } f_i \cap \bigcap_{i=1}^{p} \textbf{dom } h_i$ 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 $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

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

Dual function

We assume $\mathcal{D} = \bigcap\limits_{i=0}^{m} \textbf{dom } f_i \cap \bigcap\limits_{i=1}^{p} \textbf{dom } h_i$ 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 $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

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

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.

Dual function as a lower bound

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 for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

Dual function as a lower bound

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 for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

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

Dual function as a lower bound

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 for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0, \; \lambda \succeq 0$. Then we have:

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

Hence

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

Dual function as a lower bound

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 for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

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

Hence

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

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

Dual function as a lower bound

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 for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0.$ Then we have:

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

Hence

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

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

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:

Dual function as a lower bound

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 for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

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

Hence

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

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

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:

$$g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^p}$$

$$\text{s.t. } \lambda \succeq 0$$

Dual function as a lower bound

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 for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

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

Hence

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

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

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:

$$g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^p}$$

$$\text{s.t. } \lambda \succeq 0$$

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	$f_0(x)$	$g(\lambda, \nu) = \min\limits_{x \in \mathcal{D}} L(x, \lambda, \nu)$
Variables	$x \in S \subseteq \mathbb{R}^{\ltimes}$	$\lambda \in \mathbb{R}^m_+, \nu \in \mathbb{R}^p$
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	$f_0(x) \to \min\limits_{x \in \mathbb{R}^n}$ s.t. $f_i(x) \leq 0,\ i = 1, \ldots, m$ $h_i(x) = 0,\ i = 1, \ldots, p$	$g(\lambda, \nu) \to \max\limits_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p}$ s.t. $\lambda \succeq 0$
Optimal	x^* if feasible, $p^* = f_0(x^*)$	λ^*, ν^* if $\max$ is achieved, $d^* = g(\lambda^*, \nu^*)$

Example. Linear Least Squares

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

Example. Linear Least Squares

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

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

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

Example. Linear Least Squares

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

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

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

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

Example. Linear Least Squares

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

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

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

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

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

Example. Linear Least Squares

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

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

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

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

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

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

Example. Linear Least Squares

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

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

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

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

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

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

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

Example. Linear Least Squares

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

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

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

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

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

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

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

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

Example. Linear Least Squares

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

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

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

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

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

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

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

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

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

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

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

$$
\begin{aligned}
\text{minimize} \quad & x^T W x \\
\text{subject to} \quad & x_i^2 = 1, \quad i = 1, \ldots, n,
\end{aligned}
$$

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

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

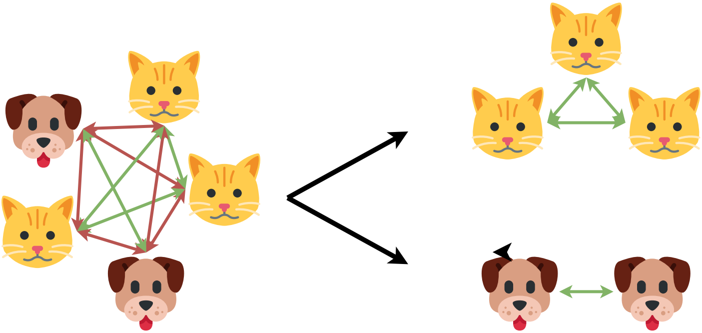


Figure 13: Illustration of two-way partitioning problem

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

$$\begin{aligned} &\text{minimize} \quad x^T W x \\ &\text{subject to} \quad x_i^2 = 1, \quad i = 1, \ldots, n, \end{aligned}$$

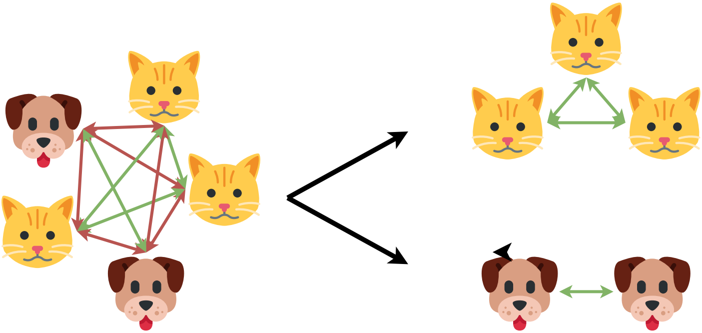


Figure 13: Illustration of two-way partitioning problem

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

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

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

$$\begin{aligned} \text{minimize} \quad & x^T W x \\ \text{subject to} \quad & x_i^2 = 1, \quad i = 1, \ldots, n, \end{aligned}$$

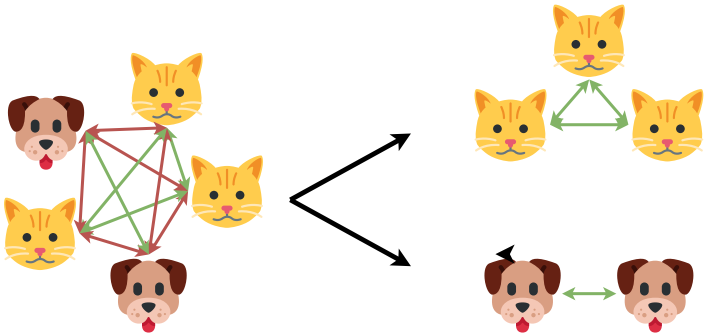


Figure 13: Illustration of two-way partitioning problem

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

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

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.

Example. Two-way partitioning problem

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

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

. . .

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

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

Example. Two-way partitioning problem

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

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

. . .

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

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

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).

Example. Two-way partitioning problem

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

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

. . .

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

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

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

$$\nu = -\lambda_{\min}(W) \mathbf{1}$$

Example. Two-way partitioning problem

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

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

. . .

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

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

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

$$\nu = -\lambda_{\min}(W) \mathbf{1}$$

which is dual feasible, since $W + \mathrm{diag}(\nu) = W - \lambda_{\min}(W) I \succeq 0$.

Example. Two-way partitioning problem

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

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

. . .

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

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

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

$$\nu = -\lambda_{\text{min}}(W) \mathbf{1}$$

which is dual feasible, since $W + \text{diag}(\nu) = W - \lambda_{\text{min}}(W) I \succeq 0$.

This renders a simple bound on the optimal value p^*: $p^* \geq -\mathbf{1}^T \nu = n \lambda_{\text{min}}(W)$.

Example. Two-way partitioning problem

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

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

. . .

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

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

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

$$\nu = -\lambda_{\text{min}}(W)\mathbf{1}$$

which is dual feasible, since $W + \text{diag}(\nu) = W - \lambda_{\text{min}}(W)I \succeq 0$.

This renders a simple bound on the optimal value p^*: $p^* \geq -\mathbf{1}^T\nu = n\lambda_{\text{min}}(W)$.

The code for the problem is available here Open in Colab

Strong duality

Strong duality

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

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

Strong 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:**

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

Strong 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 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.

Strong 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 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.

Strong duality happens if duality gap is zero:

$$p^* = d^*$$

Strong 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 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.

Strong duality happens if duality gap is zero:

$$p^* = d^*$$

Notice: both p^* and d^* may be ∞.

- Several sufficient conditions known!

Strong 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 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.

Strong duality happens if duality gap is zero:

$$
p^* = d^*
$$

Notice: both p^* and d^* may be ∞.

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

Strong duality in linear least squares

Exercise

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.

Useful features of duality

- **Construction of lower bound on solution of the primal 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.

Useful features of duality

* **Construction of lower bound on solution of the primal 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.

Useful features of duality

- **Construction of lower bound on solution of the primal 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.

Useful features of duality

- **Construction of lower bound on solution of the primal 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, $p^* \in [g(y), f_0(x)], d^* \in [g(y), f_0(x)]$

Useful features of duality

- **Construction of lower bound on solution of the primal 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, $p^* \in [g(y), f_0(x)], d^* \in [g(y), f_0(x)]$

- **Dual function is always concave**

  As a pointwise minimum of affine functions.

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

Example

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

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

Example

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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:

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

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

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

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:

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

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

References

- Lecture on KKT conditions (very intuitive explanation) in the course "Elements of Statistical Learning" @ KTH.

References

- Lecture on KKT conditions (very intuitive explanation) in the course "Elements of Statistical Learning" @ KTH.
- One-line proof of KKT

References

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints

References

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
- On Second Order Optimality Conditions in Nonlinear Optimization

References

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
- On Second Order Optimality Conditions in Nonlinear Optimization
- Numerical Optimization by Jorge Nocedal and Stephen J. Wright.