Optimality conditions. KKT

Daniil Merkulov

Optimization methods. MIPT

> *The reader will find no figures 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

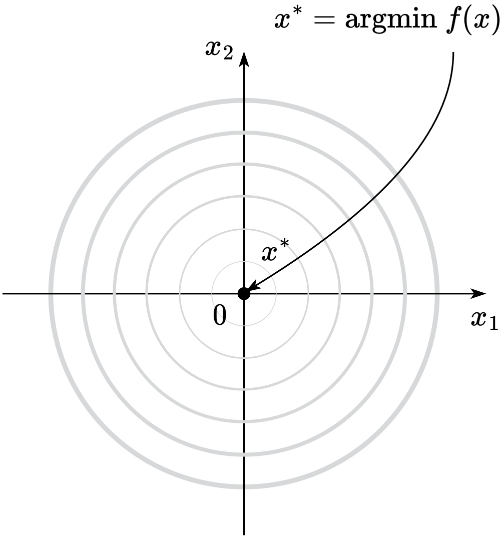
Optimization with inequality constraints

Example of inequality constraints

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

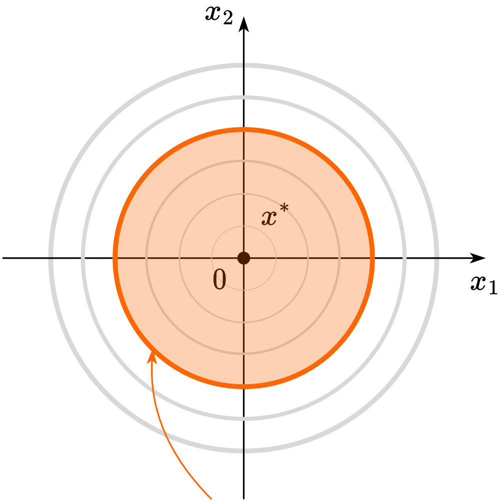
$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } g(x) \leq 0$$

Optimization with inequality constraints



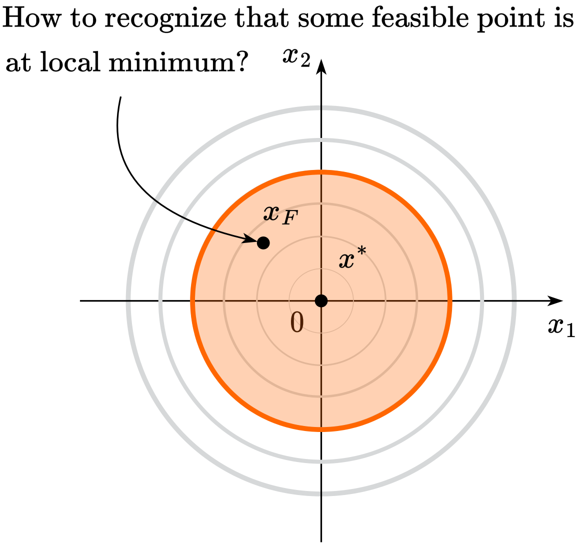
Contour lines of $f(x) = x_1^2 + x_2^2 = C$

Optimization with inequality constraints



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

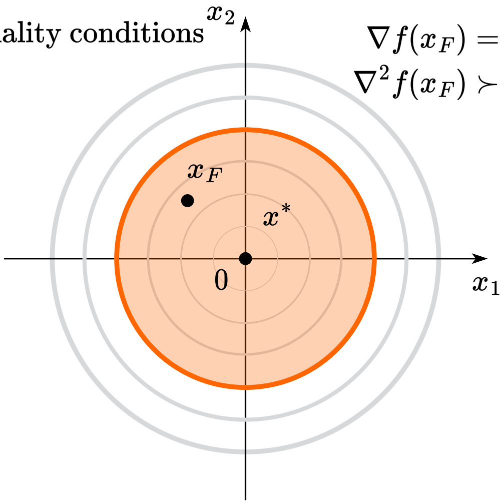
Optimization with inequality constraints



Optimization with inequality constraints

Easy in this case! Just check unconstrained optimality conditions

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



Optimization with inequality constraints

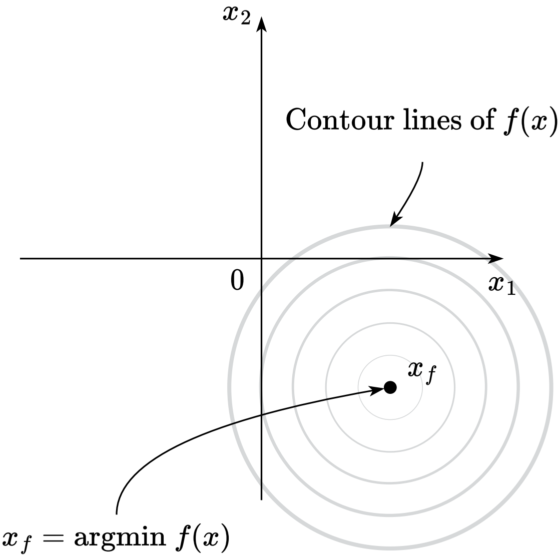
Thus, if the constraints of the type of inequalities are inactive in the constrained problem, then don't worry and write out the solution to the unconstrained problem. However, this is not the whole story. Consider the second childish example

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

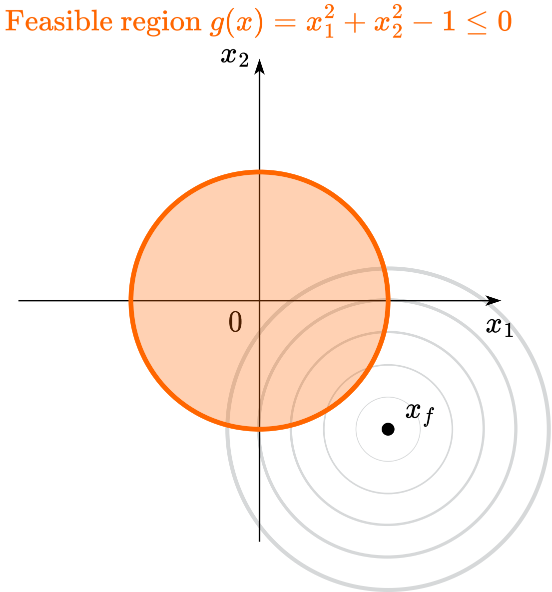
$$\begin{aligned} f(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) &\leq 0 \end{aligned}$$

Optimization with inequality constraints

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

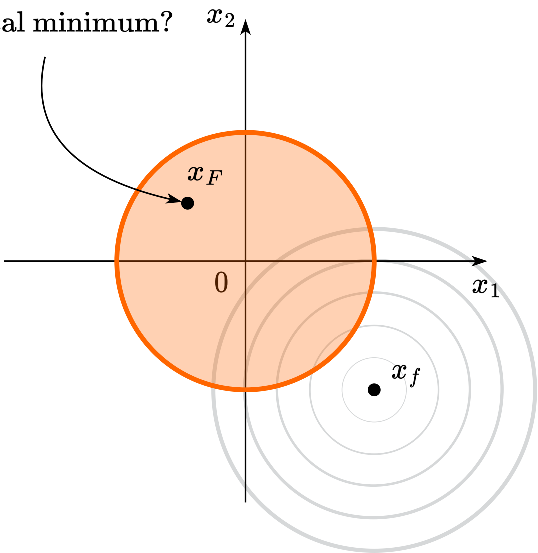


Optimization with inequality constraints



Optimization with inequality constraints

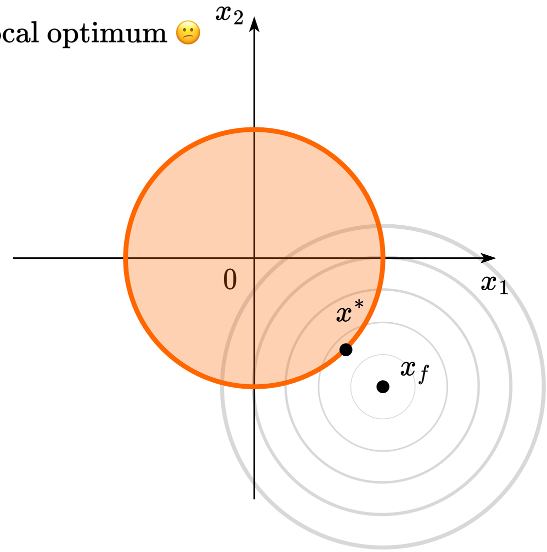
How to recognize that some feasible point is at local minimum?



Optimization with inequality constraints

Not very easy in this case! Even gradient $\neq 0$

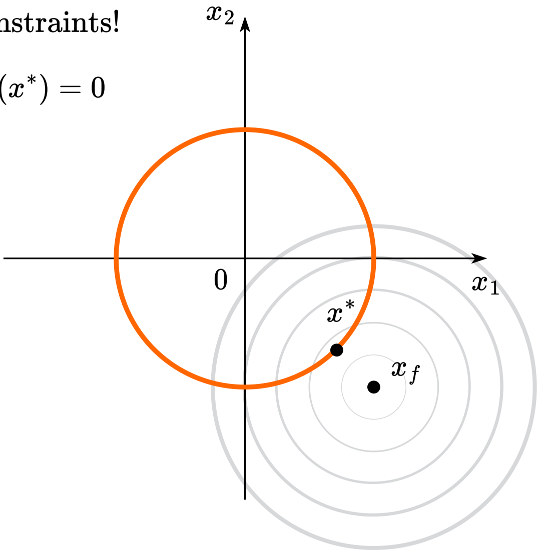
at local optimum 😕



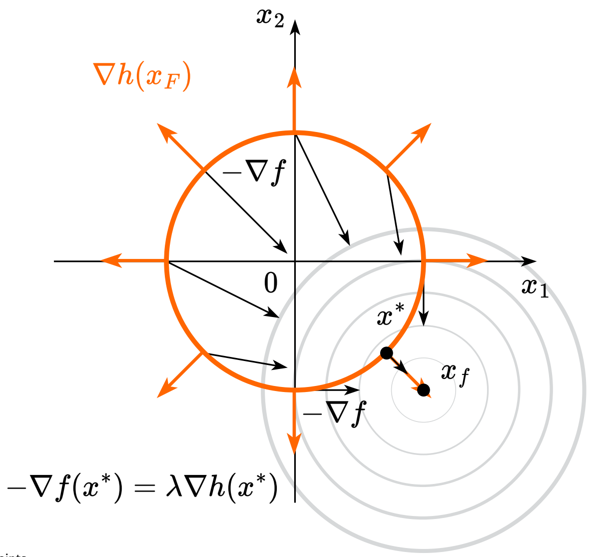
Optimization with inequality constraints

Effectively have a problem with equality constraints!

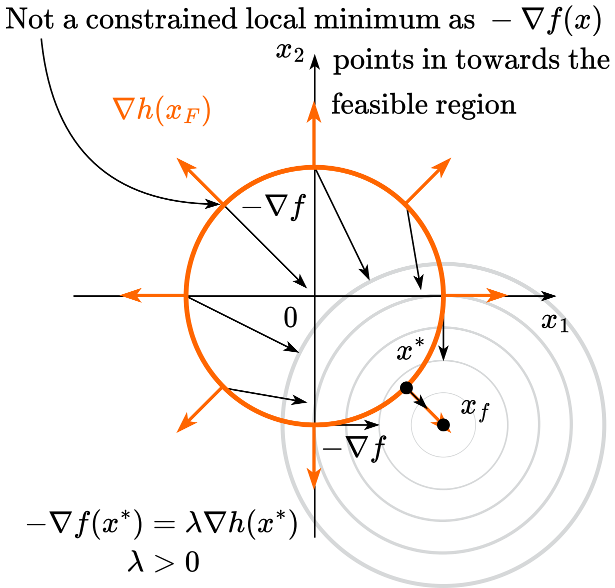
$$g(x^*) = 0$$



Optimization with inequality constraints



Optimization with inequality constraints



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. $g(x^*) < 0$

- $g(x^*) < 0$

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. $g(x^*) < 0$

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

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. $g(x^*) = 0$

- $g(x^*) = 0$

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:

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

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

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

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.

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:

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

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:

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

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:

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

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:

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

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:

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

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:

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

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

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:

$$
\begin{split}
& (1) \; \nabla_x L (x^*, \lambda^*) = 0 \\
& (2) \; \lambda^* \geq 0 \\
& (3) \; \lambda^* g(x^*) = 0 \\
& (4) \; g(x^*) \leq 0\\
& (5) \; \forall y \in C(x^*): \langle y , \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0 \\
& \text{where } C(x^*) = \{y \ \in \mathbb{R}^n | \nabla f(x^*) ^\top y \leq 0 \text{ and } \forall i \in I(x^*): \nabla g_i(x^*)^T y \leq 0 \} \\
& I(x^*) = \{ i \mid g_i(x^*) = 0 \}
\end{split}
$$

KKT

General formulation

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

This formulation is a general problem of mathematical programming.

The solution involves constructing a Lagrange function:

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

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.

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

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

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.

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

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.

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.

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^*.

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

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

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

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 $\lambda_j \geq 0$, $j = 0, 1, \ldots, m$, such that

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

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

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

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.

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

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

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

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

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

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

Example. Projection onto a hyperplane

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

Example. Projection onto a hyperplane

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

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:

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

Example. Projection onto a hyperplane

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

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 $\mathbf{x}$:

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

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

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

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 $\mathbf{x}$:

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

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

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

Example. Projection onto simplex

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

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:

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

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:

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

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:

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

Example. Projection onto simplex

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

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:

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

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:

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

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:

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

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:

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

Question

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

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:

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

Question

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

Question

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

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.