

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



Рисунок 1: Joseph-Louis Lagrange









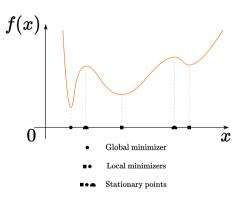


Рисунок 2: Illustration of different stationary (critical) points

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

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

A set S is usually called a **budget set**.

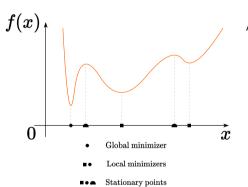


Рисунок 2: Illustration of different stationary (critical) points

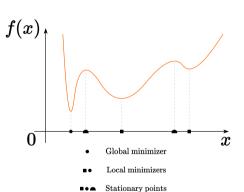


Рисунок 2: Illustration of different stationary (critical) points

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

A set S is usually called a $\mathbf{budget}\ \mathbf{set}.$

We say that the problem has a solution if the budget set is not empty: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

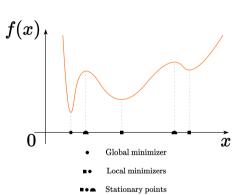


Рисунок 2: Illustration of different stationary (critical) points

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

A set S is usually called a **budget set**.

We say that the problem has a solution if the budget set is not empty: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

• A point x^* is a **global minimizer** if $f(x^*) \leq f(x)$ for all x.

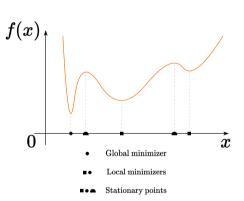


Рисунок 2: Illustration of different stationary (critical) points

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

A set S is usually called a **budget set**.

We say that the problem has a solution if the budget set is **not empty**: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

- A point x^* is a **global minimizer** if $f(x^*) \leq f(x)$ for all x.
- A point x^* is a **local minimizer** if there exists a neighborhood N of x^* such that $f(x^*) \le f(x)$ for all $x \in N$.

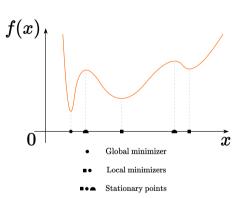


Рисунок 2: Illustration of different stationary (critical) points

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

A set S is usually called a **budget set**.

We say that the problem has a solution if the budget set is **not empty**: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

- A point x^* is a **global minimizer** if $f(x^*) \leq f(x)$ for all x.
- A point x^* is a **local minimizer** if there exists a neighborhood N of x^* such that $f(x^*) \leq f(x)$ for all $x \in N$.
- A point x^* is a **strict local minimizer** (also called a **strong local minimizer**) if there exists a neighborhood N of x^* such that $f(x^*) < f(x)$ for all $x \in N$ with $x \neq x^*$.

⊕ ೧ 0

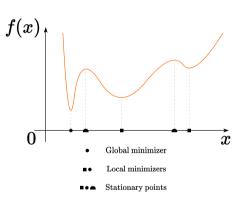


Рисунок 2: Illustration of different stationary (critical) points

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

A set S is usually called a **budget set**.

We say that the problem has a solution if the budget set is not empty: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

- A point x^* is a **global minimizer** if $f(x^*) \leq f(x)$ for all x.
- A point x^* is a **local minimizer** if there exists a neighborhood N of x^* such that $f(x^*) \leq f(x)$ for all $x \in N$.
- A point x^* is a strict local minimizer (also called a strong local **minimizer**) if there exists a neighborhood N of x^* such that $f(x^*) < f(x)$ for all $x \in N$ with $x \neq x^*$.
- We call x^* a stationary point (or critical) if $\nabla f(x^*) = 0$. Any local minimizer of a differentiable function must be a stationary point.

i Theorem

Let $S\subset\mathbb{R}^n$ be a compact set and f(x) a continuous function on S. So, the point of the global minimum of the function f(x) on S exists.



i Theorem

Let $S\subset\mathbb{R}^n$ be a compact set and f(x) a continuous function on S. So, the point of the global minimum of the function f(x) on S exists.



Рисунок 3: A lot of practical problems are theoretically solvable



i Theorem

Let $S\subset\mathbb{R}^n$ be a compact set and f(x) a continuous function on S. So, the point of the global minimum of the function f(x) on S exists.



 ${\sf P}$ исунок 3: A lot of practical problems are theoretically solvable

i Taylor's Theorem

Suppose that $f:\mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and that $p\in \mathbb{R}^n.$ Then we have:

$$f(x+p) = f(x) + \nabla f(x+tp)^T p \quad \text{ for some } t \in (0,1)$$

i Theorem

Let $S\subset\mathbb{R}^n$ be a compact set and f(x) a continuous function on S. So, the point of the global minimum of the function f(x) on S exists.



Рисунок 3: A lot of practical problems are theoretically solvable

Taylor's Theorem

Suppose that $f:\mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and that $p\in \mathbb{R}^n.$ Then we have:

$$f(x+p) = f(x) + \nabla f(x+tp)^T p \quad \text{ for some } t \in (0,1)$$

Moreover, if f is twice continuously differentiable, we have:

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp) p \, dt$$

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p$$

for some $t \in (0,1)$.

mn .u.z Optimality conditions **Unconstrained optimization**





i First-Order Necessary Conditions

If
$$x^*$$
 is a local minimizer and f is continuously differentiable in an open neighborhood, then
$$\nabla f(x^*) = 0$$

i First-Order Necessary Conditions

If x^{st} is a local minimizer and f is continuously differentiable in an open neighborhood, then

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

Proof

Suppose for contradiction that $\nabla f(x^*) \neq 0$. Define the vector $p = -\nabla f(x^*)$ and note that

$$p^T\nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

i First-Order Necessary Conditions

If x^st is a local minimizer and f is continuously differentiable in an open neighborhood, then

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

Proof

Suppose for contradiction that $\nabla f(x^*) \neq 0$. Define the vector $p = -\nabla f(x^*)$ and note that

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Because ∇f is continuous near x^* , there is a scalar T>0 such that

$$p^T \nabla f(x^* + tp) < 0$$
, for all $t \in [0, T]$

i First-Order Necessary Conditions

If x^{st} is a local minimizer and f is continuously differentiable in an open neighborhood, then

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

Proof

Suppose for contradiction that $\nabla f(x^*) \neq 0$. Define the vector $p = -\nabla f(x^*)$ and note that

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Because ∇f is continuous near x^* , there is a scalar T>0 such that

$$p^T\nabla f(x^*+tp)<0, \text{ for all } t\in[0,T]$$

For any $\bar{t} \in (0,T]$, we have by Taylor's theorem that

$$f(x^* + \bar{t}p) = f(x^*) + \bar{t}p^T \nabla f(x^* + tp)$$
, for some $t \in (0, \bar{t})$

i First-Order Necessary Conditions

If x^* is a local minimizer and f is continuously differentiable in an open neighborhood, then

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

Proof

Suppose for contradiction that $\nabla f(x^*) \neq 0$. Define the vector $p = -\nabla f(x^*)$ and note that

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

 $p^T \nabla f(x^* + tp) < 0$, for all $t \in [0, T]$

Because ∇f is continuous near x^* , there is a scalar T>0 such that

For any $\bar{t} \in (0,T]$, we have by Taylor's theorem that

$$f(x^*+\bar{t}p)=f(x^*)+\bar{t}p^T\nabla f(x^*+tp), \ \mbox{for some}\ t\in(0,\bar{t})$$

Therefore, $f(x^* + \bar{t}p) < f(x^*)$ for all $\bar{t} \in (0,T]$. We have found a direction from x^* along which f decreases, so x^* is not a local minimizer, leading to a contradiction.

i Second-Order Sufficient Conditions

Suppose that $\nabla^2 f$ is continuous in an open neighborhood of x^\ast and that

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Then x^* is a strict local minimizer of f.

i Second-Order Sufficient Conditions

Suppose that $\nabla^2 f$ is continuous in an open neighborhood of x^* and that

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Then x^* is a strict local minimizer of f.

Proof

Because the Hessian is continuous and positive definite at x^* , we can choose a radius r>0 such that $\nabla^2 f(x)$ remains positive definite for all x in the open ball $B=\{z\mid \|z-x^*\|< r\}$. Taking any nonzero vector p with $\|p\|< r$, we have $x^*+p\in B$ and so

i Second-Order Sufficient Conditions

Suppose that $\nabla^2 f$ is continuous in an open neighborhood of x^* and that

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Then x^* is a strict local minimizer of f.

Proof

Because the Hessian is continuous and positive definite at x^* , we can choose a radius r>0 such that $\nabla^2 f(x)$ remains positive definite for all x in the open ball $B=\{z\mid \|z-x^*\|< r\}$. Taking any nonzero vector p with $\|p\|< r$, we have $x^*+p\in B$ and so

$$f(x^* + p) = f(x^*) + p^T \nabla f(x^*) + \frac{1}{2} p^T \nabla^2 f(z) p$$

i Second-Order Sufficient Conditions

Suppose that $\nabla^2 f$ is continuous in an open neighborhood of x^* and that

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Then x^* is a strict local minimizer of f.

Proof

Because the Hessian is continuous and positive definite at x^* , we can choose a radius r>0 such that $\nabla^2 f(x)$ remains positive definite for all x in the open ball $B=\{z\mid \|z-x^*\|< r\}$. Taking any nonzero vector p with $\|p\|< r$, we have $x^*+p\in B$ and so

$$f(x^* + p) = f(x^*) + p^T \nabla f(x^*) + \frac{1}{2} p^T \nabla^2 f(z) p$$
$$= f(x^*) + \frac{1}{2} p^T \nabla^2 f(z) p$$

i Second-Order Sufficient Conditions

Suppose that $abla^2 f$ is continuous in an open neighborhood of x^* and that

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Then x^* is a strict local minimizer of f.

Proof

Because the Hessian is continuous and positive definite at x^* , we can choose a radius r>0 such that $\nabla^2 f(x)$ remains positive definite for all x in the open ball $B=\{z\mid \|z-x^*\|< r\}$. Taking any nonzero vector p with $\|p\|< r$, we have $x^*+p\in B$ and so

$$f(x^*+p) = f(x^*) + p^T \nabla f(x^*) + \frac{1}{2} p^T \nabla^2 f(z) p$$

$$= f(x^*) + \frac{1}{2} p^T \nabla^2 f(z) p$$

where $z=x^*+tp$ for some $t\in(0,1)$. Since $z\in B$, we have $p^T\nabla^2 f(z)p>0$, and therefore $f(x^*+p)>f(x^*)$, giving the result.

Unconstrained optimization

\$\Psi\$

Note, that if $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$, i.e. the hessian is positive semidefinite, we cannot be sure if x^* is a local minimum.



Note, that if $\nabla f(x^*)=0, \nabla^2 f(x^*)\succeq 0$, i.e. the hessian is positive semidefinite, we cannot be sure if x^* is a local minimum.

$$f(x,y)=(2x^2-y)(x^2-y)$$

Note, that if $\nabla f(x^*)=0, \nabla^2 f(x^*)\succeq 0$, i.e. the hessian is positive semidefinite, we cannot be sure if x^* is a local minimum.

$$f(x,y) = (2x^2 - y)(x^2 - y) \label{eq:force}$$
 Although the surface does not have a local minimizer

at the origin, its intersection with any vertical plane through the origin (a plane with equation y=mx or x=0) is a curve that has a local minimum at the origin. In other words, if a point starts at the origin (0,0) of the plane, and moves away from the origin along any straight line, the value of $(2x^2-y)(x^2-y)$ will increase at the start of the motion. Nevertheless, (0,0) is not a local minimizer of the function, because moving along a parabola such as $y=\sqrt{2}x^2$ will cause the function value to decrease.

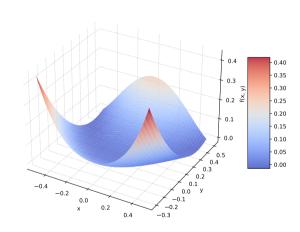


Note, that if $\nabla f(x^*)=0, \nabla^2 f(x^*)\succeq 0$, i.e. the hessian is positive semidefinite, we cannot be sure if x^* is a local minimum.

$$f(x,y) = (2x^2 - y)(x^2 - y)$$

Although the surface does not have a local minimizer at the origin, its intersection with any vertical plane through the origin (a plane with equation u = mx or x=0) is a curve that has a local minimum at the origin. In other words, if a point starts at the origin (0,0) of the plane, and moves away from the origin along any straight line, the value of $(2x^2-y)(x^2-y)$ will increase at the start of the motion. Nevertheless, (0,0) is not a local minimizer of the function, because moving along a parabola such as $y = \sqrt{2}x^2$ will cause the function value to decrease.

Non-convex PL function





Constrained optimization





Direction $d \in \mathbb{R}^n$ is a feasible direction at $x^* \in S \subseteq \mathbb{R}^n$ if small steps along d do not take us outside S.

⊕ ೧ **0**

Direction $d\in\mathbb{R}^n$ is a feasible direction at $x^*\in S\subseteq\mathbb{R}^n$ if small steps along d do

at $x^* \in S \subseteq \mathbb{R}^n$ if small steps along d do not take us outside S.

Consider a set $S \subseteq \mathbb{R}^n$ and a function

 $f: \mathbb{R}^n \to \mathbb{R}$. Suppose that $x^* \in S$ is a point of local minimum for f over S, and further assume that f is continuously

differentiable around x^* .

 $f \rightarrow \min$

Direction $d\in\mathbb{R}^n$ is a feasible direction at $x^*\in S\subseteq\mathbb{R}^n$ if small steps along d do

not take us outside S.

Consider a set $S \subseteq \mathbb{R}^n$ and a function

 $f: \mathbb{R}^n \to \mathbb{R}$. Suppose that $x^* \in S$ is a point of local minimum for f over S, and further assume that f is continuously

differentiable around x^* .

1. Then for every feasible direction $d \in \mathbb{R}^n$ at x^* it holds that $\nabla f(x^*)^\top I > 0$

$$\nabla f(x^*)^{\top} d \ge 0.$$

Direction $d \in \mathbb{R}^n$ is a feasible direction at $x^* \in S \subseteq \mathbb{R}^n$ if small steps along d do not take us outside S.

Consider a set $S \subseteq \mathbb{R}^n$ and a function

 $f:\mathbb{R}^n o \mathbb{R}$. Suppose that $x^* \in S$ is a point of local minimum for f over S, and further assume that f is continuously differentiable around x^* .

- 1. Then for every feasible direction $d \in \mathbb{R}^n$ at x^* it holds that $\nabla f(x^*)^\top d \geq 0$.
- 2. If, additionally, S is convex then

$$\nabla f(x^*)^\top (x-x^*) \geq 0, \forall x \in S.$$

Constrained optimization

Direction $d \in \mathbb{R}^n$ is a feasible direction at $x^* \in S \subseteq \mathbb{R}^n$ if small steps along d do not take us outside S.

Consider a set $S \subseteq \mathbb{R}^n$ and a function

 $f:\mathbb{R}^n o \mathbb{R}$. Suppose that $x^* \in S$ is a point of local minimum for f over S, and further assume that f is continuously differentiable around x^* .

1. Then for every feasible direction $d \in \mathbb{R}^n$ at x^* it holds that $\nabla f(x^*)^\top d \geq 0$.

2. If, additionally, S is convex then

$$\nabla f(x^*)^\top (x - x^*) > 0, \forall x \in S.$$

Constrained optimization

General first-order local optimality condition Direction $d \in \mathbb{R}^n$ is a feasible direction

at $x^* \in S \subseteq \mathbb{R}^n$ if small steps along d do not take us outside S.

Consider a set
$$S \subseteq \mathbb{R}^n$$
 and a function $f: \mathbb{R}^n \to \mathbb{R}$. Suppose that $x^* \in S$ is a point of local minimum for f over S , as

point of local minimum for f over S, and further assume that f is continuously differentiable around x^* .

- 1. Then for every feasible direction $d \in \mathbb{R}^n$ at x^* it holds that $\nabla f(x^*)^{\top} d \ge 0.$
- 2. If, additionally, S is convex then

$$\nabla f(x^*)^\top (x - x^*) \ge 0, \forall x \in S.$$

$$f(x)=x_1+x_2 o \min_{x_1,x_2\in \mathbb{R}^2}$$

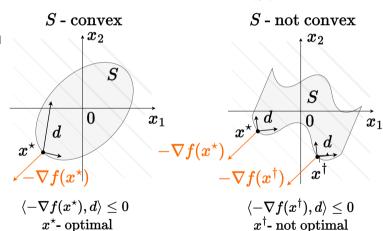


Рисунок 4: General first order local optimality condition

It should be mentioned, that in the **convex** case (i.e., f(x) is convex) necessary condition becomes sufficient.

It should be mentioned, that in the **convex** case (i.e., f(x) is convex) necessary condition becomes sufficient.

One more important result for the convex unconstrained case sounds as follows. If $f(x):S\to\mathbb{R}$ - convex function defined on the convex set S, then:

It should be mentioned, that in the **convex** case (i.e., f(x) is convex) necessary condition becomes sufficient.

One more important result for the convex unconstrained case sounds as follows. If $f(x):S\to\mathbb{R}$ - convex function defined on the convex set S, then:

• Any local minimum is the global one.



It should be mentioned, that in the **convex** case (i.e., f(x) is convex) necessary condition becomes sufficient.

One more important result for the convex unconstrained case sounds as follows. If $f(x):S\to\mathbb{R}$ - convex function defined on the convex set S, then:

- Any local minimum is the global one.
- The set of the local minimizers S^* is convex.





It should be mentioned, that in the **convex** case (i.e., f(x) is convex) necessary condition becomes sufficient.

One more important result for the convex unconstrained case sounds as follows. If $f(x):S\to\mathbb{R}$ - convex function defined on the convex set S, then:

- Any local minimum is the global one.
- The set of the local minimizers S^* is convex.
- If f(x) strictly or strongly convex function, then S^* contains only one single point $S^* = \{x^*\}$.





Things are pretty simple and intuitive in unconstrained problems. In this section, we will add one equality constraint, i.e.



Things are pretty simple and intuitive in unconstrained problems. In this section, we will add one equality constraint, i.e.

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

$$\mathrm{s.t.}\ h(x)=0$$





Things are pretty simple and intuitive in unconstrained problems. In this section, we will add one equality constraint, i.e.

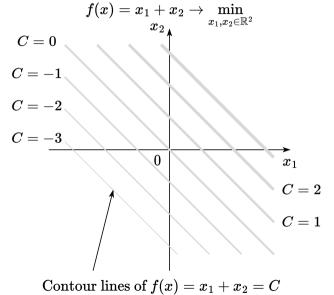
$$f(x)\to \min_{x\in\mathbb{R}^n}$$

$$\mathrm{s.t.}\ h(x)=0$$

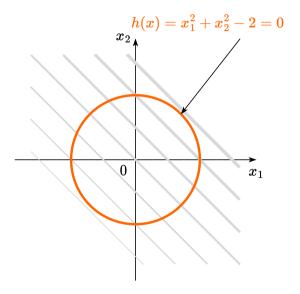
We will try to illustrate an approach to solve this problem through the simple example with $f(x) = x_1 + x_2$ and $h(x) = x_1^2 + x_2^2 - 2$.

 $f \to \min_{x,y,y}$

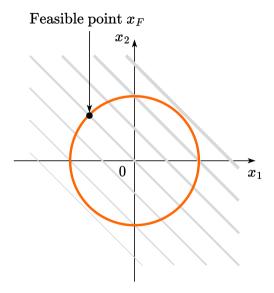
Constrained optimization



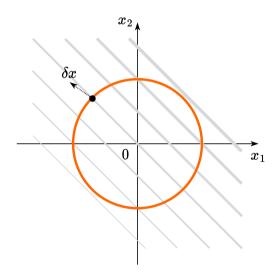




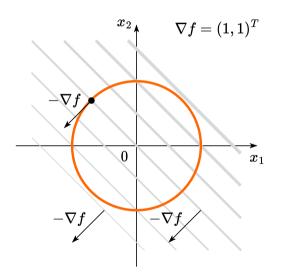




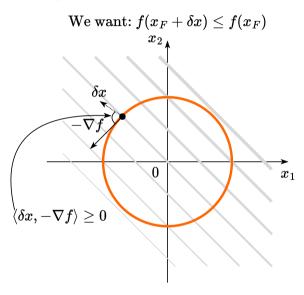




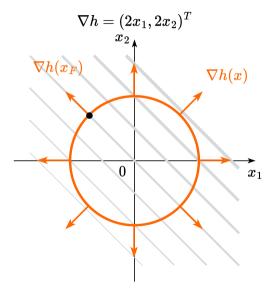




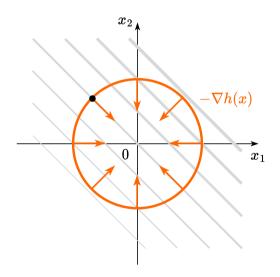
Constrained optimization



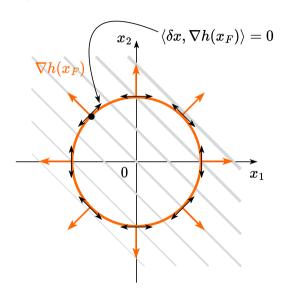














Generally: to move from x_F along the budget set toward decreasing the function, we need to guarantee two conditions:



Generally: to move from x_E along the budget set toward decreasing the function, we need to guarantee two conditions:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$



Generally: to move from x_F along the budget set toward decreasing the function, we need to guarantee two conditions:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$



Generally: to move from x_F along the budget set toward decreasing the function, we need to guarantee two conditions:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Let's assume, that in the process of such a movement, we have come to the point where



Generally: to move from x_F along the budget set toward decreasing the function, we need to guarantee two conditions:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Let's assume, that in the process of such a movement, we have come to the point where

$$-\nabla f(x) = \nu \nabla h(x)$$



Generally: to move from x_F along the budget set toward decreasing the function, we need to guarantee two conditions:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Let's assume, that in the process of such a movement, we have come to the point where

$$-\nabla f(x) = \nu \nabla h(x)$$

$$\langle \delta x, -\nabla f(x) \rangle = \langle \delta x, \nu \nabla h(x) \rangle = 0$$





Generally: to move from x_E along the budget set toward decreasing the function, we need to guarantee two conditions:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

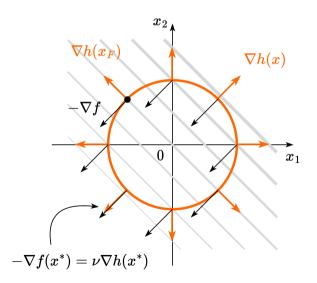
Let's assume, that in the process of such a movement, we have come to the point where

$$-\nabla f(x) = \nu \nabla h(x)$$

$$\langle \delta x, -\nabla f(x) \rangle = \langle \delta x, \nu \nabla h(x) \rangle = 0$$

Then we reached the point of the budget set, moving from which it will not be possible to reduce our function. This is the local minimum in the constrained problem:)







So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$



So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$

Then if the problem is regular (we will define it later) and the point x^* is the local minimum of the problem described above, then there exists ν^* :



So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$

Then if the problem is regular (we will define it later) and the point x^* is the local minimum of the problem described above, then there exists ν^* :

Necessary conditions

So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$

Then if the problem is regular (we will define it later) and the point x^* is the local minimum of the problem described above, then there exists ν^* :

Necessary conditions

$$\nabla_x L(x^*,\nu^*) = 0$$
 that's written above



So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$

Then if the problem is regular (we will define it later) and the point x^* is the local minimum of the problem described above, then there exists ν^* :

Necessary conditions

$$\nabla_x L(x^*,\nu^*) = 0$$
 that's written above

$$\nabla_{\nu}L(x^*,\nu^*)=0$$
 budget constraint



So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$

Then if the problem is regular (we will define it later) and the point x^* is the local minimum of the problem described above, then there exists ν^* :

Necessary conditions

 $\nabla_x L(x^*, \nu^*) = 0$ that's written above

 $\nabla_{\nu}L(x^*,\nu^*)=0$ budget constraint

Sufficient conditions



So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$

Then if the problem is regular (we will define it later) and the point x^* is the local minimum of the problem described above, then there exists ν^* :

Necessary conditions

 $\nabla_x L(x^*, \nu^*) = 0$ that's written above

 $\nabla_{\nu}L(x^*, \nu^*) = 0$ budget constraint Sufficient conditions

$$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$$



So let's define a Lagrange function (just for our convenience):

$$L(x,\nu) = f(x) + \nu h(x)$$

Then if the problem is regular (we will define it later) and the point x^* is the local minimum of the problem described above, then there exists ν^* :

Necessary conditions

 $\nabla_x L(x^*, \nu^*) = 0$ that's written above

$$\nabla_{\nu}L(x^*,\nu^*)=0$$
 budget constraint

Sufficient conditions

$$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$$

$$\forall y \neq 0 \in \mathbb{R}^n : \nabla h(x^*)^\top y = 0$$

We should notice that $L(x^*, \nu^*) = f(x^*)$.

⊕ n €

Equality constrained problem

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
 s.t. $h_i(x) = 0, \ i = 1, \dots, p$

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

Let f(x) and $h_i(x)$ be twice differentiable at the point x^* and continuously differentiable in some neighborhood of x^* . The local minimum conditions for $x \in \mathbb{R}^n, \nu \in \mathbb{R}^p$ are written as

ECP: Necessary conditions

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

$$\nabla_{\nu}L(x^*, \nu^*) = 0$$

ECP: Sufficient conditions

 $\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$

 $\forall u \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top u = 0$

Linear Least Squares

i Example

Pose the optimization problem and solve them for linear system $Ax = b, A \in \mathbb{R}^{m \times n}$ for three cases (assuming the matrix is full rank):

• *m* < *n*

Linear Least Squares

i Example

Pose the optimization problem and solve them for linear system $Ax = b, A \in \mathbb{R}^{m \times n}$ for three cases (assuming the matrix is full rank):

- *m* < *n*
- \bullet m=n

Constrained optimization



Linear Least Squares

i Example

Pose the optimization problem and solve them for linear system $Ax = b, A \in \mathbb{R}^{m \times n}$ for three cases (assuming the matrix is full rank):

- *m* < *n*
- \bullet m=n
- m > n

Constrained optimization





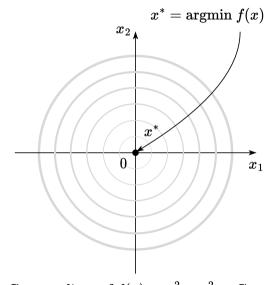
Example of inequality constraints

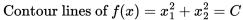
$$f(x) = x_1^2 + x_2^2$$
 $g(x) = x_1^2 + x_2^2 - 1$

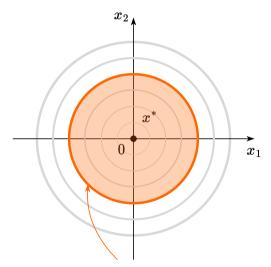
$$f(x)\to \min_{x\in\mathbb{R}^n}$$

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







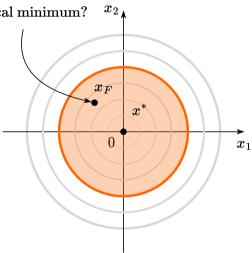


Feasible region $g(x) = x_1^2 + x_2^2 - 1 \le 0$



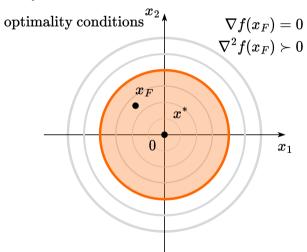


How to recognize that some feasible point is at local minimum?





Easy in this case! Just check unconstrained





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

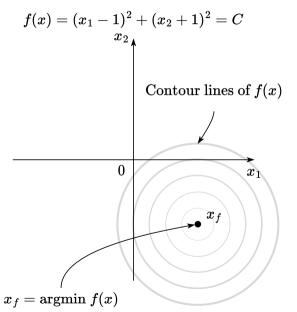
$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

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

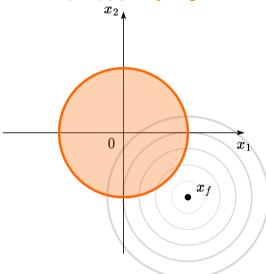
$$\text{s.t. } g(x) \leq 0$$



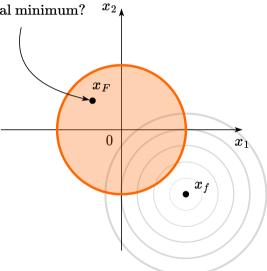




Feasible region $g(x)=x_1^2+x_2^2-1\leq 0$

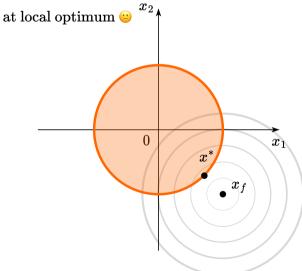


How to recognize that some feasible point is at local minimum? x_{2}



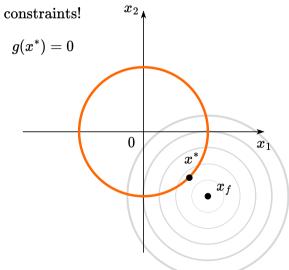


Not very easy in this case! Even gradient $\neq 0$

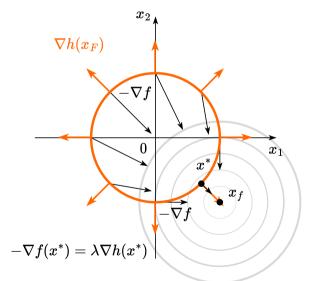




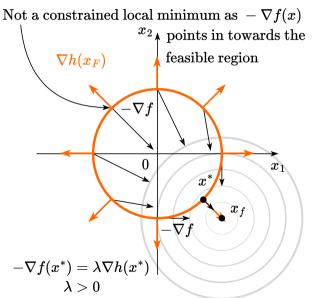
Effectively have a problem with equality











So, we have a problem:

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

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

Two possible cases:

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

• $q(x^*) < 0$



So, we have a problem:

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

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

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

- $q(x^*) < 0$



So, we have a problem:

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

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

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

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



So, we have a problem:

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

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

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

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



So, we have a problem:

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

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

Two possible cases:

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

- $g(x^*) < 0$

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

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

So, we have a problem:

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

$$s.t. \ g(x) \le 0$$

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

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

- q(x) < 0 is active. $q(x^*) = 0$
 - $q(x^*) = 0$
 - Necessary conditions: $-\nabla f(x^*) = \lambda \nabla g(x^*), \ \lambda > 0$

So, we have a problem:

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

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

$$q(x) < 0$$
 is inactive. $q(x^*) < 0$

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

- $q(x) \le 0$ is active. $q(x^*) = 0$
 - $q(x^*) = 0$
 - Necessary conditions: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$
 - Sufficient conditions:

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

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:

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



Combining two possible cases, we can $\text{If } x^* \text{ is a local minimum of the problem described above, then there exists a write down the general conditions for the unique Lagrange multiplier } \lambda^* \text{ such that:}$ problem: $(1) \ \nabla_x L(x^*, \lambda^*) = 0$

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

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

Let's define the Lagrange function:

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



Combining two possible cases, we can $\text{If } x^* \text{ is a local minimum of the problem described above, then there exists a write down the general conditions for the unique Lagrange multiplier } \lambda^* \text{ such that:}$ problem: $(1) \ \nabla_x L(x^*, \lambda^*) = 0$

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

(2)
$$\lambda^* > 0$$

s.t.
$$g(x) \leq 0$$

Let's define the Lagrange function:

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



Combining two possible cases, we can $\text{If } x^* \text{ is a local minimum of the problem described above, then there exists a write down the general conditions for the unique Lagrange multiplier } \lambda^* \text{ such that:}$ problem: $(1) \ \nabla_x L(x^*, \lambda^*) = 0$

 $(2) \lambda^* \ge 0$ $(3) \lambda^* q(x^*) = 0$

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

s.t.
$$q(x) < 0$$

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

Combining two possible cases, we can If x^* is a local minimum of the problem described above, then there exists a write down the general conditions for the unique Lagrange multiplier λ^* such that: problem:

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

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

$$\rightarrow \min_{x \in \mathbb{R}^n}$$

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

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.

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

$$(2) \lambda^* \ge 0$$

$$(3) \lambda^* g(x^*) = 0$$

$$(4) \ g(x^*) \le 0$$

Optimization with inequality constraints

General formulation

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

This formulation is a general problem of mathematical programming.

The solution involves constructing a Lagrange function:

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



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.

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



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.

- $\nabla_{m}L(x^{*},\lambda^{*},\nu^{*})=0$
- $\nabla L(x^*, \lambda^*, \nu^*) = 0$



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.

- $\nabla_{m}L(x^{*},\lambda^{*},\nu^{*})=0$
- $\nabla_{..}L(x^*,\lambda^*,\nu^*)=0$
- $\lambda_i^* \geq 0, i = 1, ..., m$



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.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_{\nu} L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda_i^* \ge 0, i = 1, ..., m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, ..., m$



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.

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

•
$$\nabla_{\nu} L(x^*, \lambda^*, \nu^*) = 0$$

•
$$\lambda^* > 0, i = 1, \dots, m$$

•
$$\lambda_i^* = 0, i = 1, ..., m$$

$$\lambda_i f_i(x^*) = 0, i = 1, \dots, m$$

•
$$f_i(x^*) \leq 0, i = 1, ..., m$$

♥ດ

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.



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

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

Solution

Lagrangian:

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

Solution

Lagrangian:

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

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

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} :

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

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

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} :

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

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

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

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} :

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

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

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

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



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

KKT Conditions

The Lagrangian is given by:

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



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

KKT Conditions

The Lagrangian is given by:

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

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



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

KKT Conditions

The Lagrangian is given by:

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

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



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

KKT Conditions

The Lagrangian is given by:

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

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

•
$$\lambda_i x_i = 0$$

•
$$\lambda_i \geq 0$$



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

KKT Conditions

The Lagrangian is given by:

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

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

$$\lambda_i x_i = 0$$

•
$$\lambda_i \geq 0$$

•
$$x^i = 0$$

• $x^\top 1 = 1$, $x \ge 0$



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

KKT Conditions

The Lagrangian is given by:

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

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

$$\lambda_i x_i = 0$$

•
$$\lambda_i \geq 0$$

•
$$x^{i} = 0$$

• $x^{\top} = 1$, $x \ge 0$



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

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:

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

•
$$\lambda_i \geq 0$$

•
$$x^{\frac{i}{1}} = 1, \quad x \ge 0$$

i Question

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

n Optimization with inequality constraints

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

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:

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

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

$$\lambda_i x_i = 0$$

•
$$\lambda_i \geq 0$$

•
$$x^{\frac{i}{1}} = 1, \quad x \ge 0$$

i Question

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

Solve the above conditions in O(n) time.

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



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





- 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



- 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



- 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.



