Subgradient. Optimality conditions

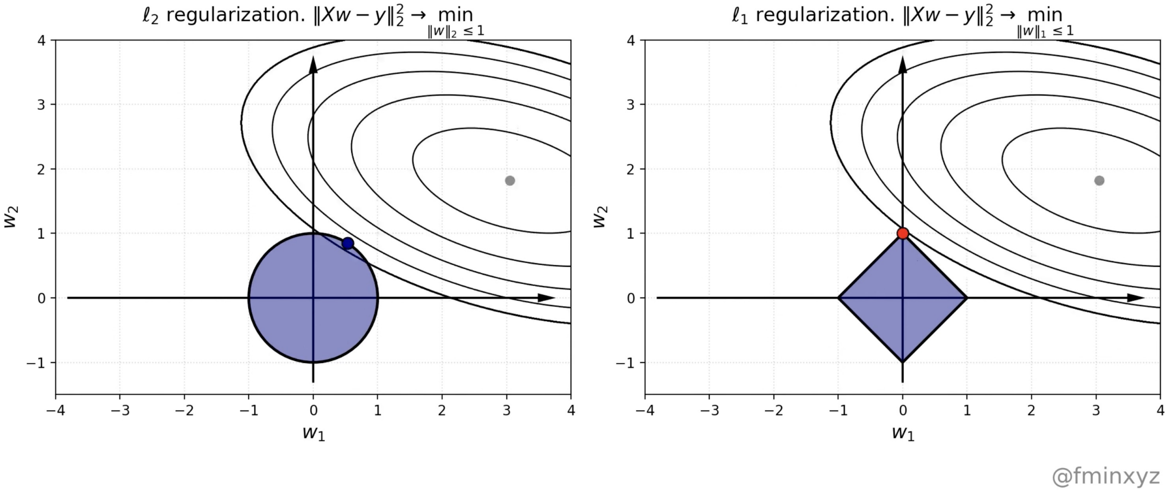
Daniil Merkulov

Optimization methods. MIPT

Subgradient and Subdifferential

ℓ_1-regularized linear least squares

ℓ_1 induces sparsity



Norms are not smooth

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

A classical convex optimization problem is considered. We assume that $f(x)$ is a convex function, but now we do not require smoothness.

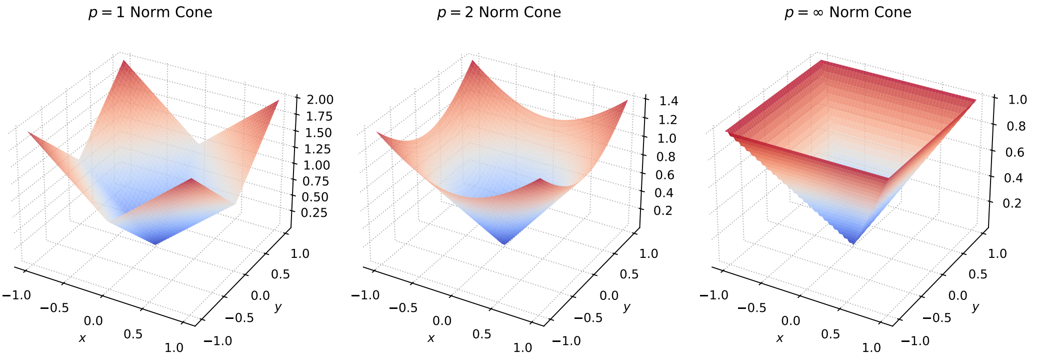


Figure 1: Norm cones for different p - norms are non-smooth

Convex function linear lower bound

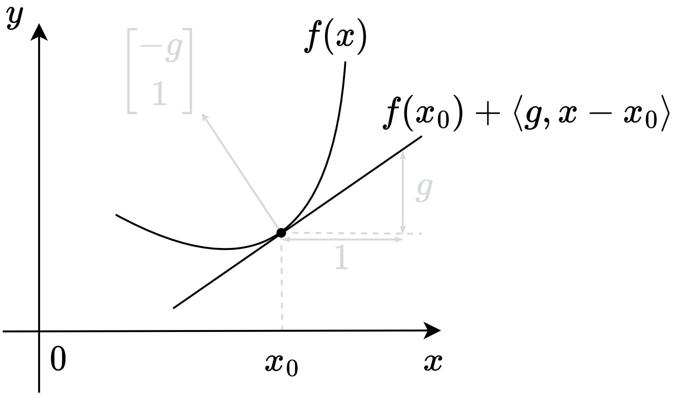


Figure 2: Taylor linear approximation serves as a global lower bound for a convex function

An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \operatorname{dom} f$ the inequality holds:

$$
f(x) \geq f(x_0) + \langle g, x - x_0 \rangle
$$

Convex function linear lower bound

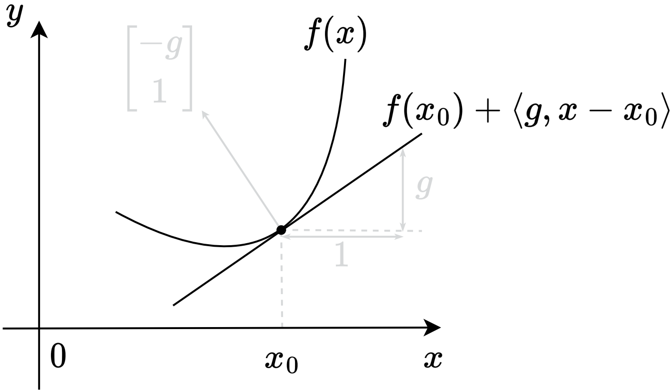


Figure 2: Taylor linear approximation serves as a global lower bound for a convex function

An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g, i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$

Convex function linear lower bound

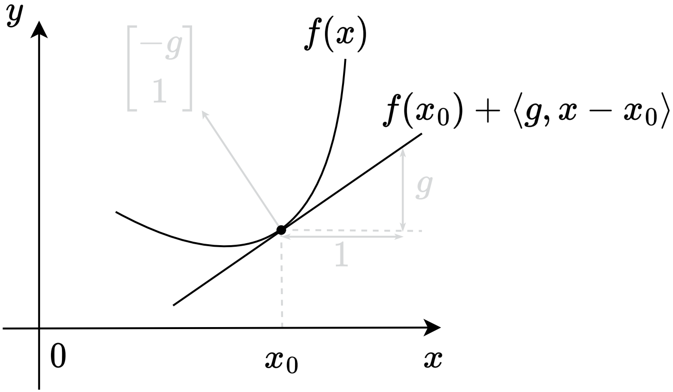


Figure 2: Taylor linear approximation serves as a global lower bound for a convex function

An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g, i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$
- Not all continuous convex functions are differentiable.

Convex function linear lower bound

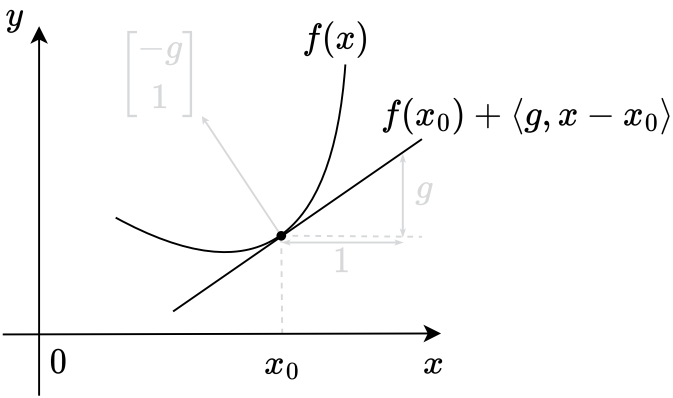


Figure 2: Taylor linear approximation serves as a global lower bound for a convex function

An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g, i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$
- Not all continuous convex functions are differentiable.

Convex function linear lower bound

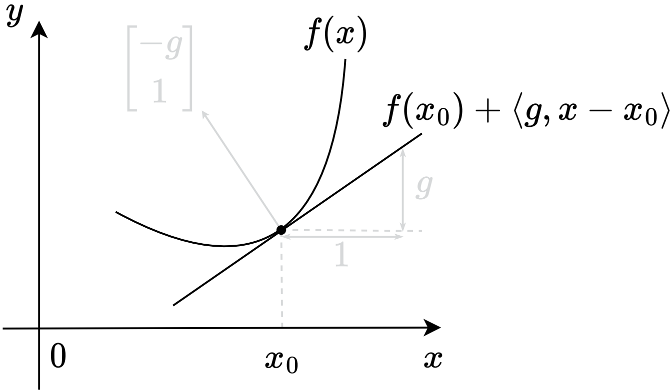


Figure 2: Taylor linear approximation serves as a global lower bound for a convex function

An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g, i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$
- Not all continuous convex functions are differentiable.

We wouldn't want to lose such a nice property.

Subgradient and subdifferential

A vector g is called the **subgradient** of a function $f(x) : S \to \mathbb{R}$ at a point x_0 if $\forall x \in S$:

$$
f(x) \geq f(x_0) + \langle g, x - x_0 \rangle
$$

Subgradient and subdifferential

A vector g is called the **subgradient** of a function $f(x) : S \to \mathbb{R}$ at a point x_0 if $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

The set of all subgradients of a function $f(x)$ at a point x_0 is called the **subdifferential** of f at x_0 and is denoted by $\partial f(x_0)$.

Subgradient and subdifferential

A vector g is called the **subgradient** of a function $f(x) : S \to \mathbb{R}$ at a point x_0 if $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

The set of all subgradients of a function $f(x)$ at a point x_0 is called the **subdifferential** of f at x_0 and is denoted by $\partial f(x_0)$.

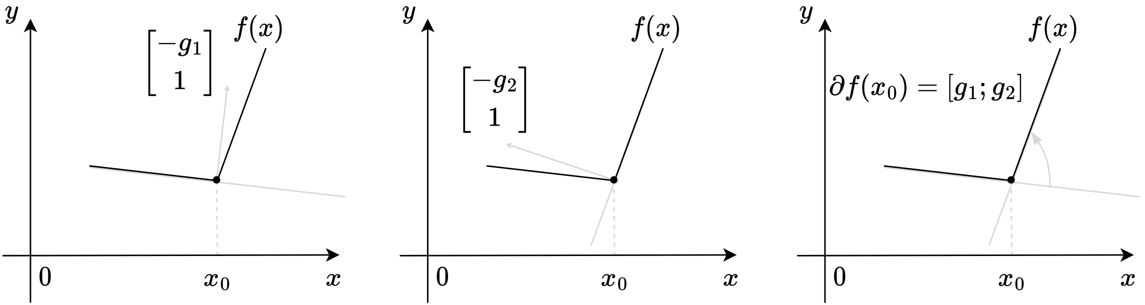


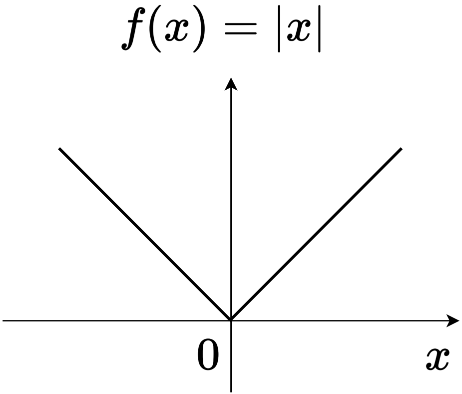
Figure 3: Subdifferential is a set of all possible subgradients

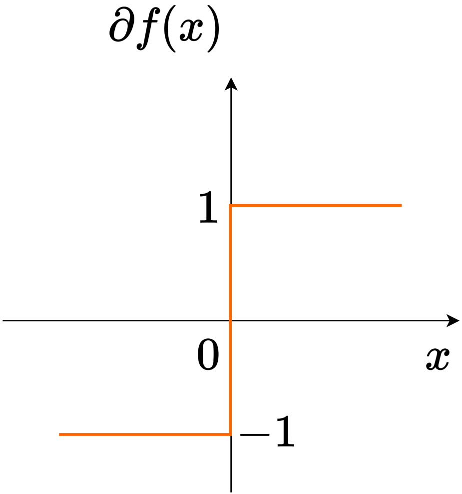
Subgradient and subdifferential

Find $\partial f(x)$, if $f(x) = |x|$

Subgradient and subdifferential

Find $\partial f(x)$, if $f(x) = |x|$





Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S.

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S.

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S.

Subdifferential of a differentiable function

Let $f : S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space $\mathbb{R}^n$. If $x_0 \in \mathbf{ri}(S)$ and f is differentiable at x_0, then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S.

Subdifferential of a differentiable function

Let $f : S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space $\mathbb{R}^n$. If $x_0 \in \mathbf{ri}(S)$ and f is differentiable at x_0, then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S, there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v\rangle$$

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S.

Subdifferential of a differentiable function

Let $f : S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space $\mathbb{R}^n$. If $x_0 \in \mathbf{ri}(S)$ and f is differentiable at x_0, then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S, there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S.

Subdifferential of a differentiable function

Let $f : S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space $\mathbb{R}^n$. If $x_0 \in \mathbf{ri}(S)$ and f is differentiable at x_0, then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S, there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v\rangle$$

which implies:

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle$$

for all $0 < t < \delta$. Taking the limit as t approaches 0 and using the definition of the gradient, we get:

$$\langle \nabla f(x_0), v\rangle = \lim_{t \to 0; 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle$$

2. From this, $\langle s - \nabla f(x_0), v\rangle \geq 0$. Due to the arbitrariness of v, one can set

$$v = -\frac{s - \nabla f(x_0)}{\|s - \nabla f(x_0)\|},$$

leading to $s = \nabla f(x_0)$.

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S.

> **Subdifferential of a differentiable function**
>
> Let $f : S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space $\mathbb{R}^n$. If $x_0 \in \mathbf{ri}(S)$ and f is differentiable at x_0, then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S, there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v\rangle$$

which implies:

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle$$

for all $0 < t < \delta$. Taking the limit as t approaches 0 and using the definition of the gradient, we get:

$$\langle \nabla f(x_0), v\rangle = \lim_{t \to 0; 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle$$

2. From this, $\langle s - \nabla f(x_0), v\rangle \geq 0$. Due to the arbitrariness of v, one can set

$$v = -\frac{s - \nabla f(x_0)}{\| s - \nabla f(x_0)\|},$$

leading to $s = \nabla f(x_0)$.

3. Furthermore, if the function f is convex, then according to the differential condition of convexity $f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0\rangle$ for all $x \in S$. But by definition, this means $\nabla f(x_0) \in \partial f(x_0)$.

Subdifferentiability and convexity

Question

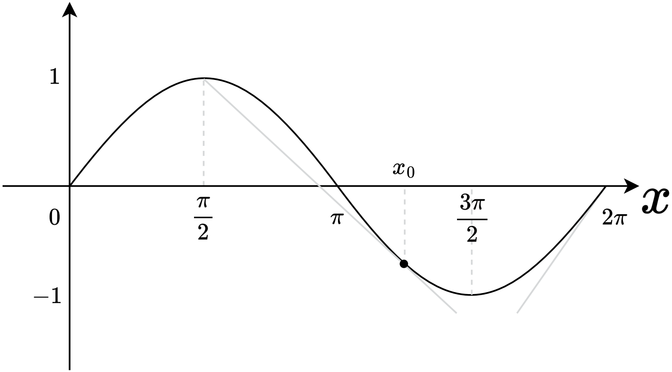
Is it correct, that if the function has a subdifferential at some point, the function is convex?

Subdifferentiability and convexity

Question

Is it correct, that if the function has a subdifferential at some point, the function is convex?

Find $\partial f(x)$, if $f(x) = \sin x, x \in [\pi/2; 2\pi]$



Subdifferentiability and convexity

Question

Is it correct, that if the function is convex, it has a subgradient at any point?

Subdifferentiability and convexity

Question

Is it correct, that if the function is convex, it has a subgradient at any point?

Convexity follows from subdifferentiability at any point. A natural question to ask is whether the converse is true: is every convex function subdifferentiable? It turns out that, generally speaking, the answer to this question is negative.

Let $f : [0, \infty) \to \mathbb{R}$ be the function defined by $f(x) := -\sqrt{x}$. Then, $\partial f(0) = \emptyset$.

Assume, that $s \in \partial f(0)$ for some $s \in \mathbb{R}$. Then, by definition, we must have $sx \leq -\sqrt{x}$ for all $x \geq 0$. From this, we can deduce $s \leq -\sqrt{1}$ for all $x > 0$. Taking the limit as x approaches 0 from the right, we get $s \leq -\infty$, which is impossible.

Subdifferential calculus

Moreau - Rockafellar theorem (subdifferential of a linear combination)

Let $f_i(x)$ be convex functions on convex sets $S_i,\ i = \overline{1,n}$. Then if $\bigcap\limits_{i=1}^n \mathbf{ri}(S_i) \neq \emptyset$ then the function $f(x) = \sum\limits_{i=1}^n a_i f_i(x),\ a_i > 0$ has a subdifferential $\partial_S f(x)$ on the set $S = \bigcap\limits_{i=1}^n S_i$ and

$$
\partial_S f(x) = \sum_{i=1}^n a_i \partial_{S_i} f_i(x)
$$

Subdifferential calculus

Moreau - Rockafellar theorem (subdifferential of a linear combination)

Let $f_i(x)$ be convex functions on convex sets $S_i, \ i = \overline{1,n}$. Then if $\bigcap\limits_{i=1}^n \mathbf{ri}(S_i) \neq \emptyset$ then the function $f(x) = \sum\limits_{i=1}^n a_i f_i(x), \ a_i > 0$ has a subdifferential $\partial_S f(x)$ on the set $S = \bigcap\limits_{i=1}^n S_i$ and

$$
\partial_S f(x) = \sum_{i=1}^n a_i \partial_{S_i} f_i(x)
$$

Dubovitsky - Milutin theorem (subdifferential of a point-wise maximum)

Let $f_i(x)$ be convex functions on the open convex set $S \subseteq \mathbb{R}^n, \ x_0 \in S$, and the pointwise maximum is defined as $f(x) = \max\limits_i f_i(x)$. Then:

$$
\partial_S f(x_0) = \mathbf{conv}\left\{ \bigcup_{i \in I(x_0)} \partial_S f_i(x_0) \right\}, \quad I(x) = \{i \in [1
$$

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - convex functions

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - convex functions
- If $g(x) = f(Ax) + b$ then $\partial g(x) = A^T \partial f(Ax + b)$

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - convex functions
- If $g(x) = f(Ax) + b$ then $\partial g(x) = A^T \partial f(Ax + b)$
- $z \in \partial f(x)$ if and only if $x \in \partial f^*(z)$.

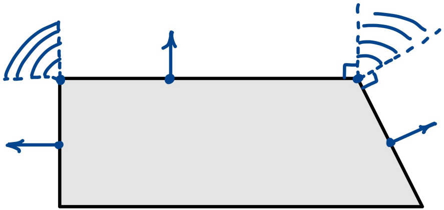
Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - convex functions
- If $g(x) = f(Ax) + b$ then $\partial g(x) = A^T \partial f(Ax + b)$
- $z \in \partial f(x)$ if and only if $x \in \partial f^*(z)$.
- Let $f : E \to \mathbb{R}$ be a convex function and $g : \mathbb{R} \to \mathbb{R}$ be a nondecreasing convex function. Let $x \in E$, and suppose that g is differentiable at the point $f(x)$. Let $h = g \circ f$. Then $\partial h(x) = g'(f(x)) \partial f(x)$.

Connection to convex geometry

Convex set $S \subseteq \mathbb{R}^n$, consider indicator function $I_S : \mathbb{R}^n \to \mathbb{R}$,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$



For $x \in S$, $\partial I_S(x) = \mathcal{N}_S(x)$, the **normal cone** of S at x is, recall

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in S\}$$

Why? By definition of subgradient g,

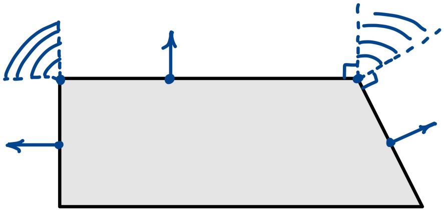
$$I_S(y) \geq I_S(x) + g^T(y - x) \quad \text{for all } y$$

- For $y \notin S$, $I_S(y) = \infty$

Connection to convex geometry

Convex set $S \subseteq \mathbb{R}^n$, consider indicator function $I_S : \mathbb{R}^n \to \mathbb{R}$,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$



For $x \in S$, $\partial I_S(x) = \mathcal{N}_S(x)$, the **normal cone** of S at x is, recall

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in S\}$$

Why? By definition of subgradient g,

$$I_S(y) \geq I_S(x) + g^T(y - x) \quad \text{for all } y$$

- For $y \notin S$, $I_S(y) = \infty$
- For $y \in S$, this means $0 \geq g^T(y - x)$

Optimality Condition

For any f (convex or not),

$$f(x^\star) = \min_x f(x) \quad \Longleftrightarrow \quad 0 \in \partial f(x^\star)$$

That is, $x^\star$ is a minimizer if and only if 0 is a subgradient of f at $x^\star$. This is called the **subgradient optimality condition**.

Why? Easy: $g = 0$ being a subgradient means that for all y

$$f(y) \geq f(x^\star) + 0^T(y - x^\star) = f(x^\star)$$

Note the implication for a convex and differentiable function f, with

$$\partial f(x) = \{\nabla f(x)\}$$

Derivation of first-order optimality

Example of the power of subgradients: we can use what we have learned so far to derive the **first-order optimality condition**. Recall

$$\min_x f(x) \text{ subject to } x \in S$$

is solved at x, for f convex and differentiable, if and only if

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{for all } y \in S$$

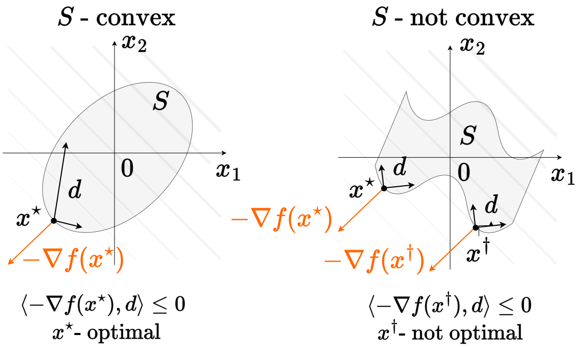
Intuitively: this says that the gradient increases as we move away from x. How to prove it? First, recast the problem as

$$\min_x f(x) + I_S(x)$$

Now apply subgradient optimality:

$$0 \in \partial(f(x) + I_S(x))$$

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



Derivation of first-order optimality

Observe

$$0 \in \partial(f(x) + I_S(x))$$

$$\Leftrightarrow 0 \in \{\nabla f(x)\} + \mathcal{N}_S(x)$$

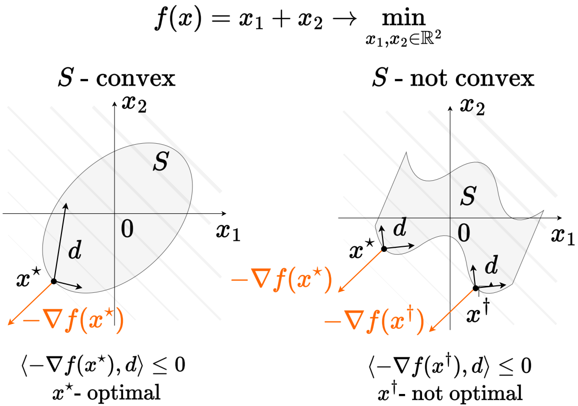
$$\Leftrightarrow -\nabla f(x) \in \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x)^T x \geq -\nabla f(x)^T y \text{ for all } y \in S$$

$$\Leftrightarrow \nabla f(x)^T (y - x) \geq 0 \text{ for all } y \in S$$

as desired.

Note: the condition $0 \in \partial f(x) + \mathcal{N}_S(x)$ is a **fully general condition** for optimality in convex problems. But it's not always easy to work with (KKT conditions, later, are easier).



Example 1

Example

Find $\partial f(x)$, if $f(x) = |x-1| + |x+1|$

Example 1

Example

Find $\partial f(x)$, if $f(x) = |x-1| + |x+1|$

$$\partial f_1(x) = \begin{cases} -1, & x < 1 \\ [-1;1], & x = 1 \\ 1, & x > 1 \end{cases} \qquad \partial f_2(x) = \begin{cases} -1, & x < -1 \\ [-1;1], & x = -1 \\ 1, & x > -1 \end{cases}$$

So

$$\partial f(x) = \begin{cases} -2, & x < -1 \\ [-2;0], & x = -1 \\ 0, & -1 < x < 1 \\ [0;2], & x = 1 \\ 2, & x > 1 \end{cases}$$

Example 2

Find $\partial f(x)$ if $f(x) = [\max(0, f_0(x))]^q$. Here, $f_0(x)$ is a convex function on an open convex set S, and $q \geq 1$.

Example 2

Find $\partial f(x)$ if $f(x) = [\max(0, f_0(x))]^q$. Here, $f_0(x)$ is a convex function on an open convex set S, and $q \geq 1$. According to the composition theorem (the function $\varphi(x) = x^q$ is differentiable) and $g(x) = \max(0, f_0(x))$, we have:

$$
\partial f(x) = q(g(x))^{q-1} \partial g(x)
$$

By the theorem on the pointwise maximum:

$$
\partial g(x) = \begin{cases} \partial f_0(x), & f_0(x) > 0, \\ \{0\}, & f_0(x) < 0, \\ \{a \mid a = \lambda a',\ 0 \leq \lambda \leq 1,\ a' \in \partial f_0(x)\}, & f_0(x) = 0 \end{cases}
$$

Example 3. Subdifferential of the Norm

Let V be a finite-dimensional Euclidean space, and $x_0 \in V$. Let $\|\cdot\|$ be an arbitrary norm in V (not necessarily induced by the scalar product), and let $\|\cdot\|_*$ be the corresponding conjugate norm. Then,

$$
\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0,1), & \text{if } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{otherwise.} \end{cases}
$$

Where $B_{\|\cdot\|_*}(0,1)$ is the closed unit ball centered at zero with respect to the conjugate norm. In other words, a vector $s \in V$ with $\|s\|_* = 1$ is a subgradient of the norm $\|\cdot\|$ at point $x_0 \neq 0$ if and only if the Hölder's inequality $\langle s, x_0 \rangle \leq \|x_0\|$ becomes an equality.

Example 3. Subdifferential of the Norm

Let V be a finite-dimensional Euclidean space, and $x_0 \in V$. Let $\|\cdot\|$ be an arbitrary norm in V (not necessarily induced by the scalar product), and let $\|\cdot\|_*$ be the corresponding conjugate norm. Then,

$$
\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0,1), & \text{if } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{otherwise.} \end{cases}
$$

Where $B_{\|\cdot\|_*}(0,1)$ is the closed unit ball centered at zero with respect to the conjugate norm. In other words, a vector $s \in V$ with $\|s\|_* = 1$ is a subgradient of the norm $\|\cdot\|$ at point $x_0 \neq 0$ if and only if the Hölder's inequality $\langle s, x_0 \rangle \leq \|x_0\|$ becomes an equality.

Let $s \in V$. By definition, $s \in \partial\|\cdot\|(x_0)$ if and only if

$$
\langle s, x \rangle - \|x\| \leq \langle s, x_0 \rangle - \|x_0\|, \text{ for all } x \in V,
$$

or equivalently,

$$
\sup_{x \in V}\{\langle s, x \rangle - \|x\|\} \leq \langle s, x_0 \rangle - \|x_0\|.
$$

By the definition of the supremum, the latter is equivalent to

$$
\sup_{x \in V}\{\langle s, x \rangle - \|x\|\} = \langle s, x_0 \rangle - \|x_0\|.
$$

Example 3. Subdifferential of the Norm

Let V be a finite-dimensional Euclidean space, and $x_0 \in V$. Let $\|\cdot\|$ be an arbitrary norm in V (not necessarily induced by the scalar product), and let $\|\cdot\|_*$ be the corresponding conjugate norm. Then,

$$
\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0,1), & \text{if } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{otherwise.} \end{cases}
$$

Where $B_{\|\cdot\|_*}(0,1)$ is the closed unit ball centered at zero with respect to the conjugate norm. In other words, a vector $s \in V$ with $\|s\|_* = 1$ is a subgradient of the norm $\|\cdot\|$ at point $x_0 \neq 0$ if and only if the Hölder's inequality $\langle s, x_0 \rangle \leq \|x_0\|$ becomes an equality.

Let $s \in V$. By definition, $s \in \partial\|\cdot\|(x_0)$ if and only if

$$
\langle s, x \rangle - \|x\| \leq \langle s, x_0 \rangle - \|x_0\|, \text{ for all } x \in V,
$$

or equivalently,

$$
\sup_{x \in V}\{\langle s, x \rangle - \|x\|\} \leq \langle s, x_0 \rangle - \|x_0\|.
$$

By the definition of the supremum, the latter is equivalent to

$$
\sup_{x \in V}\{\langle s, x \rangle - \|x\|\} = \langle s, x_0 \rangle - \|x_0\|.
$$

It is important to note that the expression on the left side is the supremum from the definition of the Fenchel conjugate function for the norm, which is known to be

$$
\sup_{x \in V}\{\langle s, x \rangle - \|x\|\} = \begin{cases} 0, & \text{if } \|s\|_* \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}
$$

Thus, equation is equivalent to $\|s\|_* \leq 1$ and $\langle s, x_0 \rangle = \|x_0\|$.

Example 3. Subdifferential of the Norm

Consequently, it remains to note that for $x_0 \neq 0$, the inequality $\|s\|_* \leq 1$ must become an equality since, when $\|s\|_* < 1$, Hölder's inequality implies $\langle s, x_0 \rangle \leq \|s\|_* \|x_0\| < \|x_0\|$.

The conjugate norm in Example above does not appear by chance. It turns out that, in a completely similar manner for an arbitrary function f (not just for the norm), its subdifferential can be described in terms of the dual object — the Fenchel conjugate function.

Optimality conditions

Background

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

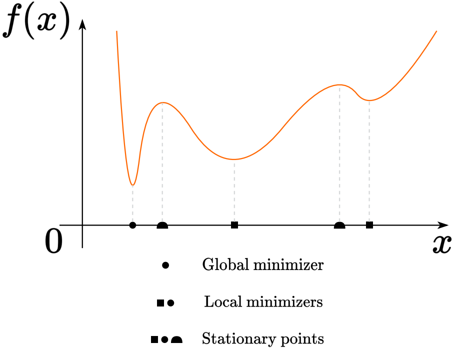


Figure 5: Illustration of different stationary (critical) points

Background

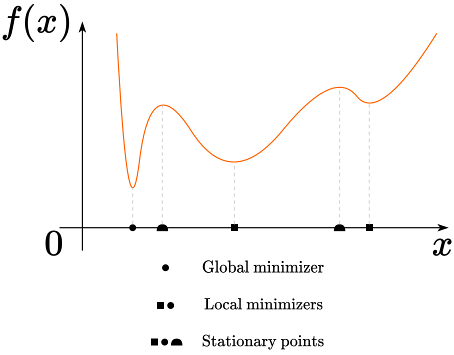


Figure 5: Illustration of different stationary (critical) points

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

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

Background

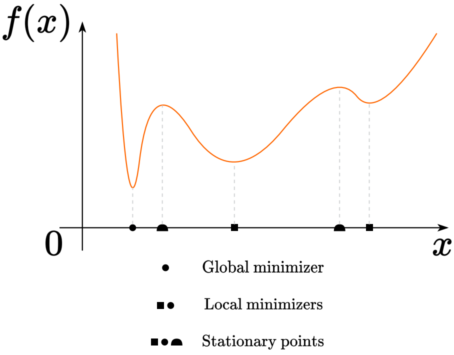


Figure 5: Illustration of different stationary (critical) points

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

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.

Background

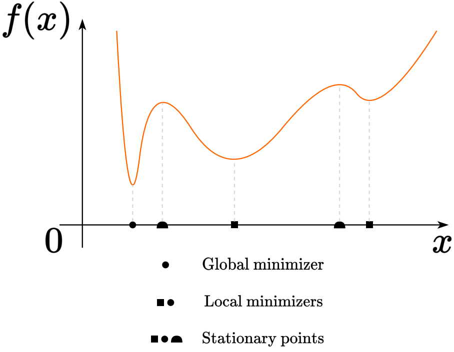


Figure 5: Illustration of different stationary (critical) points

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

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.

Background

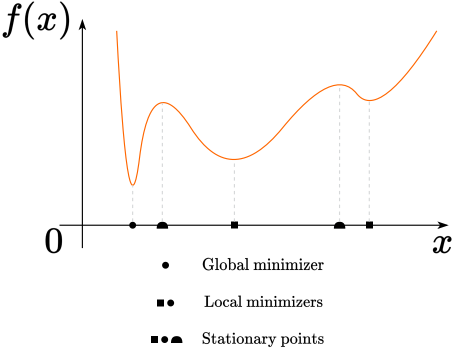


Figure 5: Illustration of different stationary (critical) points

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

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 $x \in N$.

Background

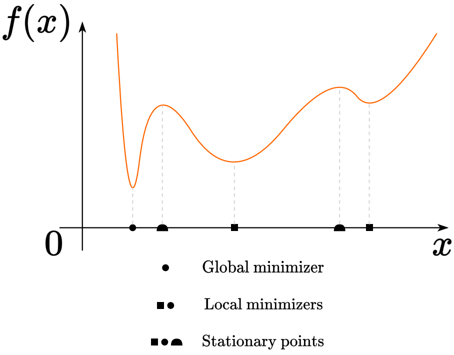


Figure 5: Illustration of different stationary (critical) points

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

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 $x \neq x^*$.

Background

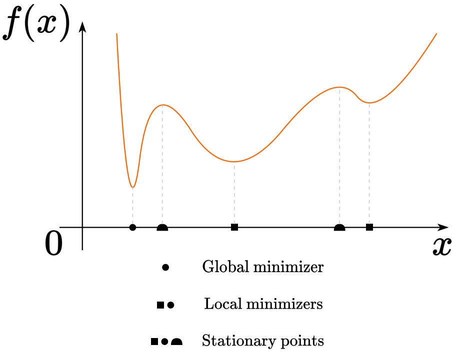


Figure 5: Illustration of different stationary (critical) points

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

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.

Extreme value (Weierstrass) theorem

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.

Extreme value (Weierstrass) theorem

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.

Figure 6: A lot of practical problems are theoretically solvable

Extreme value (Weierstrass) theorem

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.

Figure 6: 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:

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

Extreme value (Weierstrass) theorem

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.

Figure 6: 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 $t \in (0,1)$.

Unconstrained optimization

Necessary Conditions

First-Order Necessary Conditions

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

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

Necessary Conditions

> 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

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

Necessary Conditions

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

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

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

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

Necessary Conditions

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

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

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

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

Necessary Conditions

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

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

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

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

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.

Sufficient Conditions

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.

Sufficient Conditions

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

Sufficient Conditions

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

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

Sufficient Conditions

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

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

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

Sufficient Conditions

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

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

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

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.

Peano counterexample

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.

Peano counterexample

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.

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

Peano counterexample

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.

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

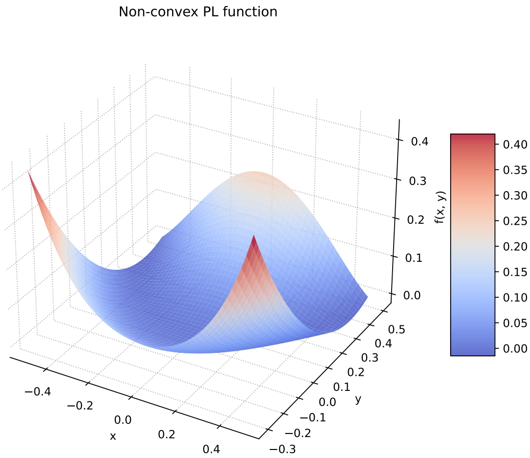
Although the surface does not have a local minimizer at the origin, its intersection with any vertical plane through the origin (a plane with equation $y = mx$ or $x = 0$) is a curve that has a local minimum at the origin. In other words, if a point starts at the origin $(0, 0)$ of the plane, and moves 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.

Peano counterexample

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 $y = mx$ 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.



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 of S.

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 of 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^*.

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 of 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 $\nabla f(x^*)^\top d \geq 0$.

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 of 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 $\nabla f(x^*)^\top d \geq 0$.
2. If, additionally, S is convex then

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

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 of 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 $\nabla f(x^*)^\top d \geq 0$.
2. If, additionally, S is convex then

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

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 of 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 $\nabla f(x^*)^\top d \geq 0$.
2. If, additionally, S is convex then

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

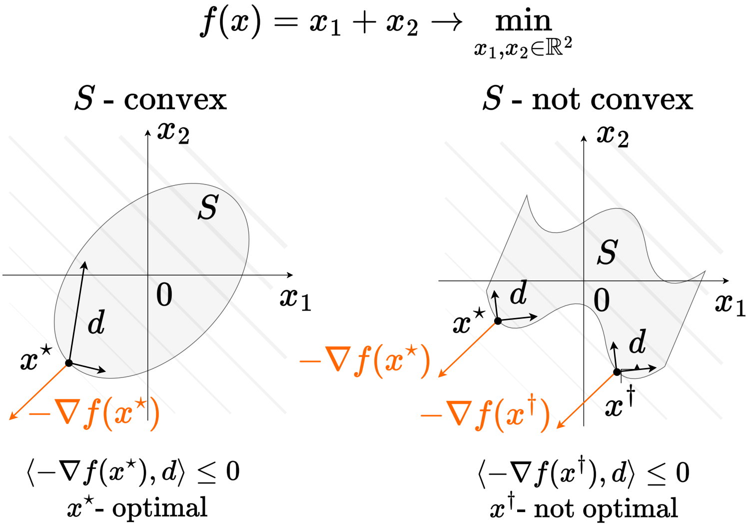


Figure 7: General first order local optimality condition

Convex case

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

Convex case

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:

Convex case

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 minima is the global one.

Convex case

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

Convex case

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 minima 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 $S^* = \{x^*\}$.

Optimization with equality constraints

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

Optimization with equality constraints

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

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

Optimization with equality constraints

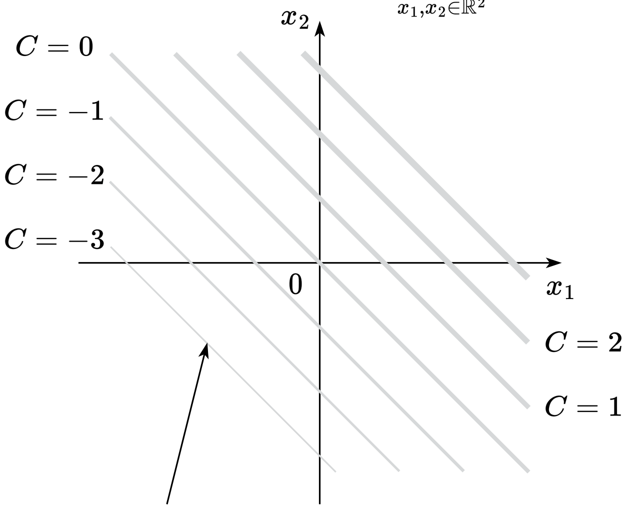
Things are pretty simple and intuitive in unconstrained problems. In this section, we will add one equality constraint, i.e.

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

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

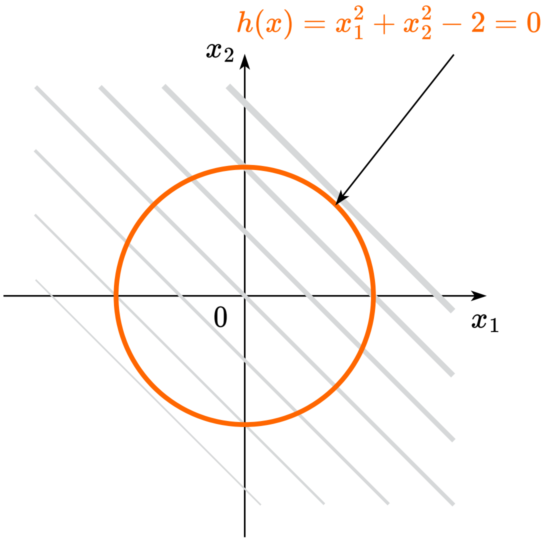
Optimization with equality constraints

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

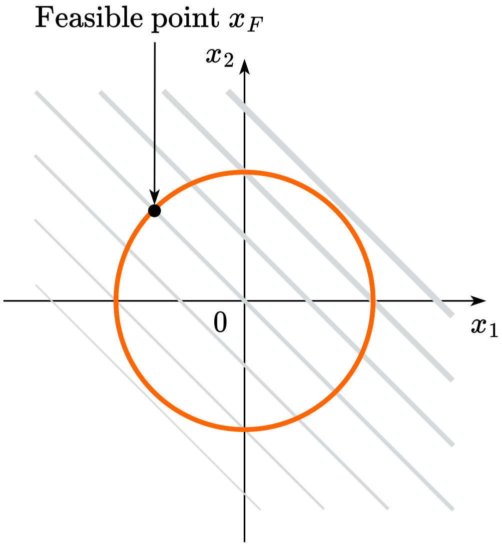


Contour lines of $f(x) = x_1 + x_2 = C$

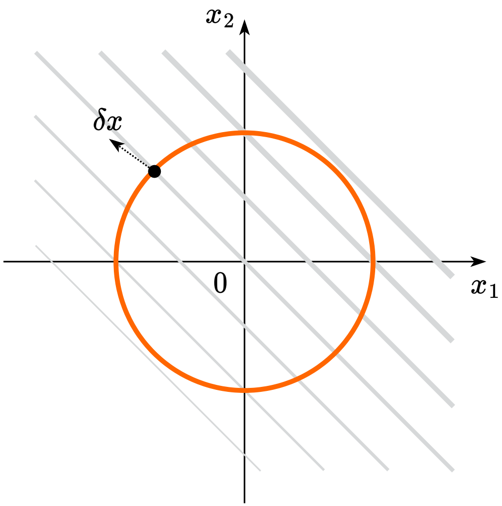
Optimization with equality constraints



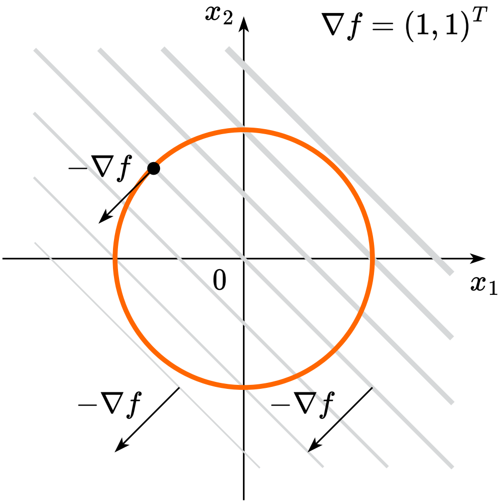
Optimization with equality constraints



Optimization with equality constraints

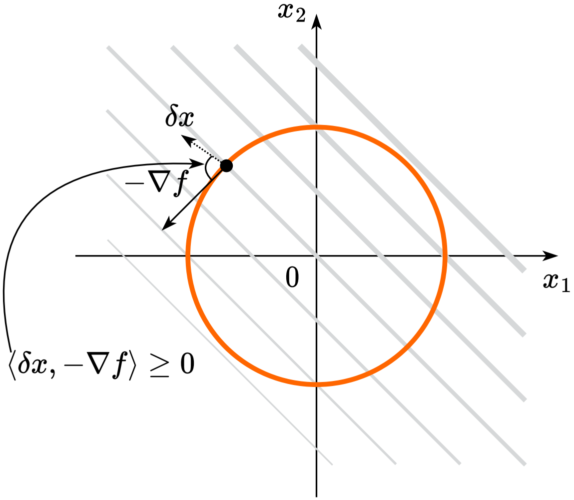


Optimization with equality constraints



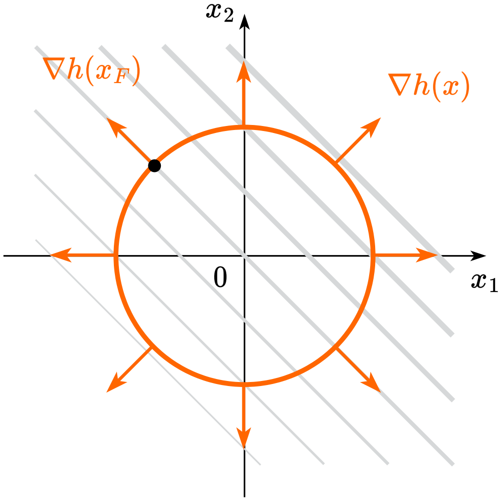
Optimization with equality constraints

We want: $f(x_F + \delta x) \leq f(x_F)$

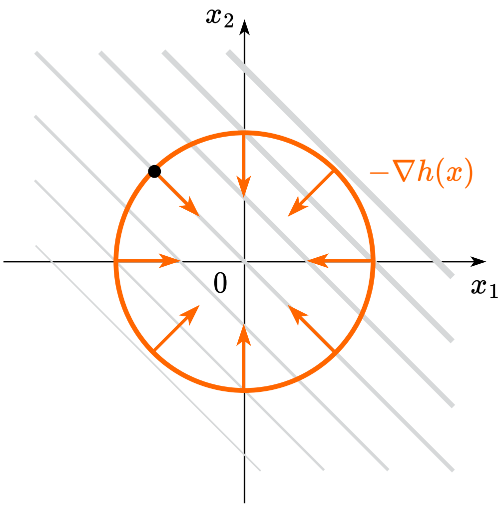


Optimization with equality constraints

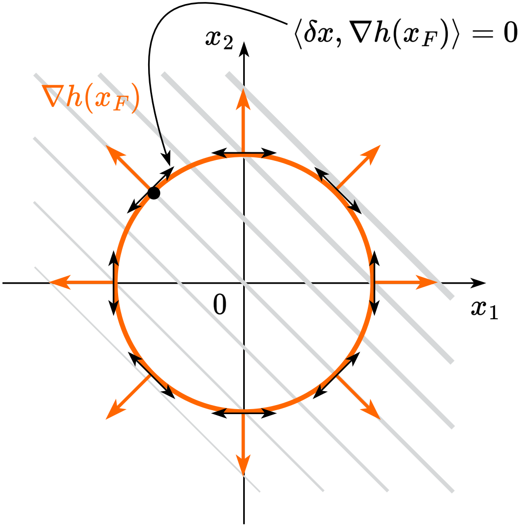
$$\nabla h = (2x_1, 2x_2)^T$$



Optimization with equality constraints



Optimization with equality constraints



Optimization with equality constraints

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

Optimization with equality constraints

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

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

Optimization with equality constraints

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

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

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

Optimization with equality constraints

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

Optimization with equality constraints

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

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

Optimization with equality constraints

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

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

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

Optimization with equality constraints

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

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

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

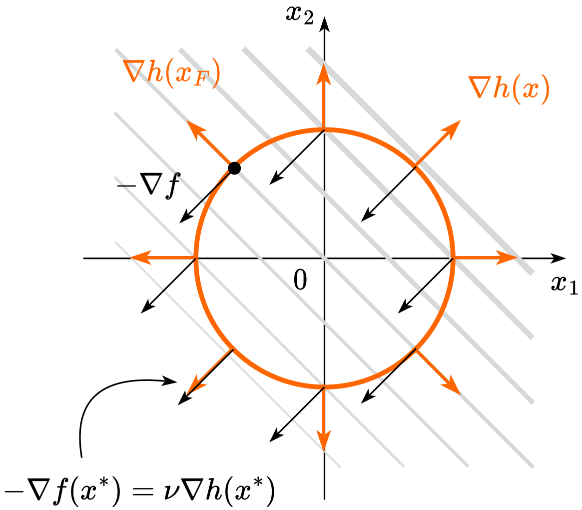
Let's assume, that in the process of such a movement, we have come to the point where

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

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

Then we came to 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 :)

Optimization with equality constraints



Lagrangian

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

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

Lagrangian

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^*:

Lagrangian

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

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

Lagrangian

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

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

Lagrangian

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

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

Lagrangian

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

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

Lagrangian

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

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

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

Lagrangian

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 $L(x^*, \nu^*) = f(x^*)$.

Equality constrained problem

$$
\begin{split}
& f(x) \to \min\limits_{x \in \mathbb{R}^n} \\
\text{s.t. } & h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
\tag{ECP}
$$

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

Let $f(x)$ and $h_i(x)$ be twice differentiable at the point x^* and continuously differentiable in some neighborhood 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

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

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

Linear Least Squares

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

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):

- $m < n$
- $m = n$

Linear Least Squares

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):

- $m < n$
- $m = n$
- $m > n$