Lecture 1: Problems and solutions. Optimality conditions for unconstrained optimization (continued)

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Taylor expansions (REVISION)

FIRST ORDER TAYLOR EXPANSION

Let $\phi : \mathbb{R} \to \mathbb{R}$, $\phi \in C^1(\mathbb{R})$ i.e. continuously differentiable.

Then for any $\alpha \in \mathbb{R}$, we have

$$\phi(\alpha) = \phi(0) + \alpha \phi'(0) + O(\alpha^2) \quad (1)$$

where $O(\cdot)$ implies an upper bound that is a multiple of α^2.

Also, [mean-value theorem]

$$\boxed{\phi(\alpha) = \phi(0) + \alpha \phi'(\tilde{\alpha})}, \text{ for some } \tilde{\alpha} \in (0, \alpha). \quad (2)$$

Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$ with gradient $\nabla f = \left(\frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_n}\right)^T$.

Let $x = (x_1 \dots x_n)^T$ and $s = (s_1 \dots s_n)^T \in \mathbb{R}^n$, fixed.

Let $\phi(\alpha) := f(x + \alpha s)$, $\alpha \in \mathbb{R}$ (so $\phi : \mathbb{R} \to \mathbb{R}$).

Then $\phi'(\alpha) = \frac{d}{d\alpha} f(\overbrace{x_1 + \alpha s_1, \dots, x_n + \alpha s_n}^{= x + \alpha s})$ [by chain rule]

$$= \frac{\partial f}{\partial x_1}(x+\alpha s) \cdot \underbrace{\frac{d}{d\alpha}(x_1 + \alpha s_1)}_{= s_1} + \frac{\partial f}{\partial x_2}(x+\alpha s) \cdot \underbrace{\frac{d}{d\alpha}(x_2 + \alpha s_2)}_{= s_2} + \dots + \frac{\partial f}{\partial x_n}(x+\alpha s) \cdot \underbrace{\frac{d}{d\alpha}(x_n + \alpha s_n)}_{= s_n}$$

$$= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x+\alpha s) \cdot s_i = \nabla f(x+\alpha s)^T s.$$

Thus first-order Taylor expansion of ϕ gives from (2),

$$\boxed{f(x+\alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s}, \text{ for some } \tilde{\alpha} \in (0, \alpha). \quad (3)$$

SECOND ORDER TAYLOR EXPANSION

Let $\phi : \mathbb{R} \to \mathbb{R}$, $\phi \in C^2(\mathbb{R})$ i.e. twice continuously differentiable.

Then for any $\alpha \in \mathbb{R}$, we have

$$\boxed{\phi(\alpha) = \phi(0) + \alpha \phi'(0) + \frac{\alpha^2}{2} \phi''(\tilde{\alpha})},$$

where $\tilde{\alpha} \in (0, \alpha)$. (4)

(mean value theorem)

Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(\mathbb{R}^n)$ with Hessian $\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$ $n \times n$ symmetric matrix

$\xrightarrow{\phi(\alpha) := f(x+\alpha s)} \phi'(0) = \nabla f(x)^T s$, $\phi''(\alpha) = s^T \nabla^2 f(x + \alpha s) s$.

Thus the second order Taylor expansion of ϕ (4) gives

$$\boxed{f(x+\alpha s) = f(x) + \alpha \nabla f(x)^T s + \frac{1}{2}\alpha^2 s^T \nabla^2 f(x + \tilde{\alpha} s) s},$$

for some $\tilde{\alpha} \in (0, \alpha)$. (5)

Unconstrained optimization problems and solutions

$$\text{minimize } f(x) \quad \text{subject to } x \in \mathbb{R}^n. \qquad \text{(UP)}$$

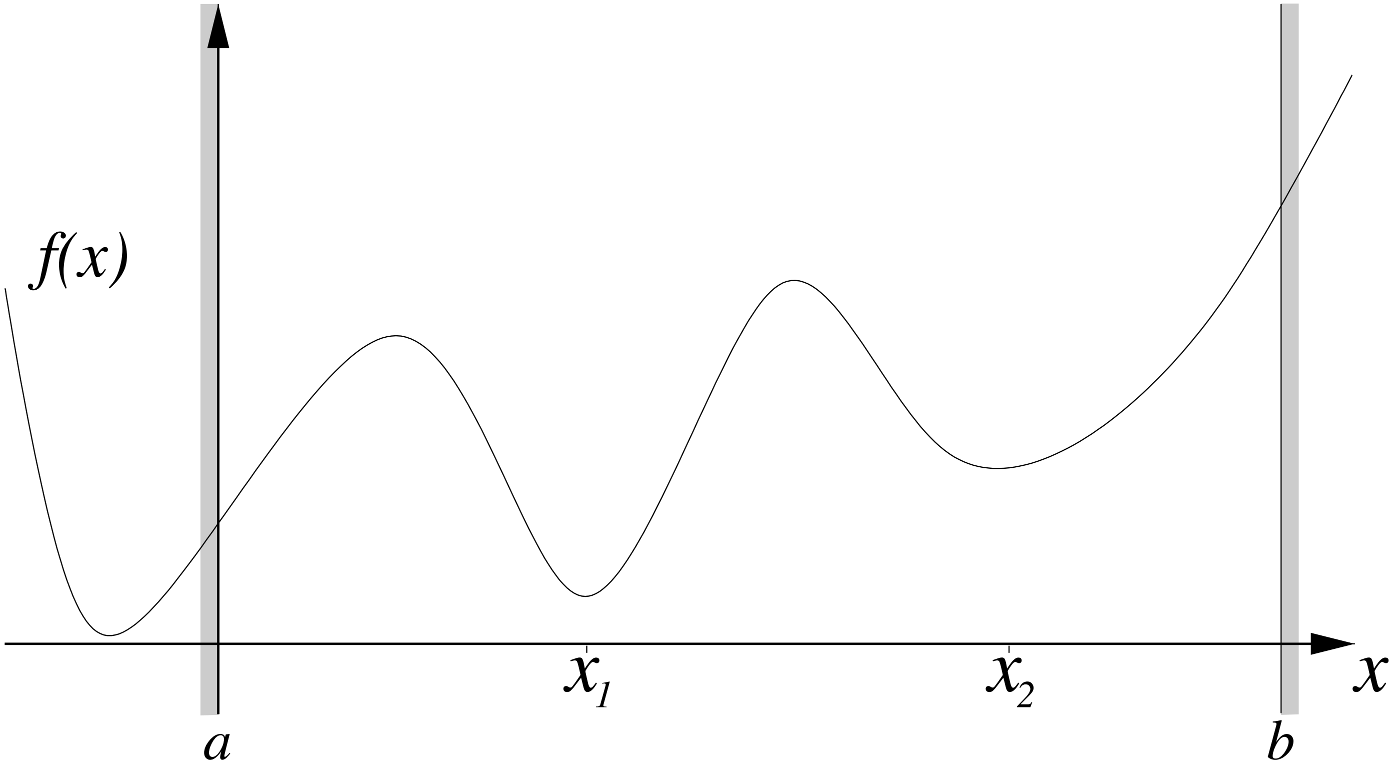
- $f : \mathbb{R}^n \to \mathbb{R}$ is (sufficiently) smooth ($f \in \mathcal{C}^i(\mathbb{R}^n)$, $i \in \{1, 2\}$).
- f objective; x variables.

x^* global minimizer of f (over $\mathbb{R}^n$) $\iff$ $f(x) \geq f(x^*)$, $\forall x \in \mathbb{R}^n$.

x^* local minimizer of f (over $\mathbb{R}^n$) $\iff$ there exists $\mathcal{N}(x^*, \delta)$ such that $f(x) \geq f(x^*)$, for all $x \in \mathcal{N}(x^*, \delta)$, where $\mathcal{N}(x^*, \delta) := \{x \in \mathbb{R}^n : \|x - x^*\| \leq \delta\}$ and $\|\cdot\|$ is the Euclidean norm.

Example problem in one dimension

$$\textbf{Example}: \quad \min f(x) \quad \text{subject to} \quad x \in \mathbb{R}.$$



- The points x_1 and x_2 are (unconstrained) local minimizers of f (for example).

Optimality conditions for unconstrained problems

== algebraic characterizations of solutions $\longrightarrow$ suitable for computations.

- provide a way to guarantee that a candidate point is optimal (sufficient conditions)
- indicate when a point is not optimal (necessary conditions)

Optimality conditions for unconstrained problems

== algebraic characterizations of solutions $\longrightarrow$ suitable for computations.

- provide a way to guarantee that a candidate point is optimal (sufficient conditions)
- indicate when a point is not optimal (necessary conditions)

First-order necessary conditions for (UP): $f \in \mathcal{C}^1(\mathbb{R}^n)$;

x^* a local minimizer of $f \implies \nabla f(x^*) = 0$.

$\nabla f(x) = 0 \iff x$ stationary point of f.

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then
$\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \;\; \forall \alpha > 0$ sufficiently small.

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

$\nabla f(x)^T s < 0 \quad \Longrightarrow \quad f(x + \alpha s) < f(x), \quad \forall \alpha > 0$ sufficiently small.

Proof. $f \in \mathcal{C}^1 \overset{\text{gradient is}}{\underset{\text{continuous}}{\Longrightarrow}} \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \ \forall \alpha \in [0, \overline{\alpha}]. \qquad (\Diamond)$$

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

$\nabla f(x)^T s < 0 \quad \Longrightarrow \quad f(x + \alpha s) < f(x), \quad \forall \alpha > 0$ sufficiently small.

Proof. $\quad f \in \mathcal{C}^1 \quad \Longrightarrow \quad \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \;\; \forall \alpha \in [0, \overline{\alpha}]. \qquad (\Diamond)$$

Taylor's/Mean value theorem: by First order Taylor expansion revision slide, see equation (3)

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

$\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \quad \forall \alpha > 0$ sufficiently small.

Proof. $\quad f \in \mathcal{C}^1 \implies \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \; \forall \alpha \in [0, \overline{\alpha}]. \qquad (\lozenge)$$

Taylor's/Mean value theorem:

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

$(\lozenge) \implies f(x + \alpha s) < f(x), \forall \alpha \in (0, \overline{\alpha}]$. $\square$

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

$\nabla f(x)^T s < 0 \quad \Longrightarrow \quad f(x + \alpha s) < f(x), \quad \forall \alpha > 0$ sufficiently small.

Proof. $\quad f \in \mathcal{C}^1 \quad \Longrightarrow \quad \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \;\; \forall \alpha \in [0, \overline{\alpha}]. \qquad (\Diamond)$$

Taylor's/Mean value theorem:

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

$(\Diamond) \;\; \Longrightarrow \;\; f(x + \alpha s) < f(x), \forall \alpha \in (0, \overline{\alpha}]$. $\Box$

- s descent direction for f at x if $\nabla f(x)^T s < 0$.

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

$\nabla f(x)^T s < 0 \quad \Longrightarrow \quad f(x + \alpha s) < f(x), \quad \forall \alpha > 0$ sufficiently small.

Proof. $\quad f \in \mathcal{C}^1 \quad \Longrightarrow \quad \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \; \forall \alpha \in [0, \overline{\alpha}]. \qquad (\lozenge)$$

Taylor's/Mean value theorem:

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

$(\lozenge) \;\Longrightarrow\; f(x + \alpha s) < f(x)$, $\forall \alpha \in (0, \overline{\alpha}]$. $\square$

- s descent direction for f at x if $\nabla f(x)^T s < 0$.

Proof of 1st order necessary conditions. assume $\nabla f(x^*) \neq 0$.

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

$\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \;\; \forall \alpha > 0$ sufficiently small.

Proof. $\quad f \in \mathcal{C}^1 \implies \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \; \forall \alpha \in [0, \overline{\alpha}]. \qquad (\Diamond)$$

Taylor's/Mean value theorem:

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

$(\Diamond) \implies f(x + \alpha s) < f(x)$, $\forall \alpha \in (0, \overline{\alpha}]$. $\square$

- s descent direction for f at x if $\nabla f(x)^T s < 0$.

Proof of 1st order necessary conditions. assume $\nabla f(x^*) \neq 0$.

$s := -\nabla f(x^*)$ is a descent direction for f at $x = x^*$:

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then $\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \quad \forall \alpha > 0$ sufficiently small.

Proof. $\quad f \in \mathcal{C}^1 \implies \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \ \forall \alpha \in [0, \overline{\alpha}]. \qquad (\Diamond)$$

Taylor's/Mean value theorem:

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

$(\Diamond) \implies f(x + \alpha s) < f(x), \forall \alpha \in (0, \overline{\alpha}]. \ \Box$

- s descent direction for f at x if $\nabla f(x)^T s < 0$.

Proof of 1st order necessary conditions. assume $\nabla f(x^*) \neq 0$.
$s := -\nabla f(x^*)$ is a descent direction for f at $x = x^*$:
$\nabla f(x^*)^T(-\nabla f(x^*)) = -\nabla f(x^*)^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$
since $\nabla f(x^*) \neq 0$ and $\|a\| \geq 0$ with equality iff $a = 0$.

Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

$\nabla f(x)^T s < 0 \quad \Longrightarrow \quad f(x + \alpha s) < f(x), \quad \forall \alpha > 0$ sufficiently small.

Proof. $\quad f \in \mathcal{C}^1 \quad \Longrightarrow \quad \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \; \forall \alpha \in [0, \overline{\alpha}]. \qquad (\lozenge)$$

Taylor's/Mean value theorem:

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

$(\lozenge) \;\Longrightarrow\; f(x + \alpha s) < f(x), \forall \alpha \in (0, \overline{\alpha}]$. $\square$

- s descent direction for f at x if $\nabla f(x)^T s < 0$.

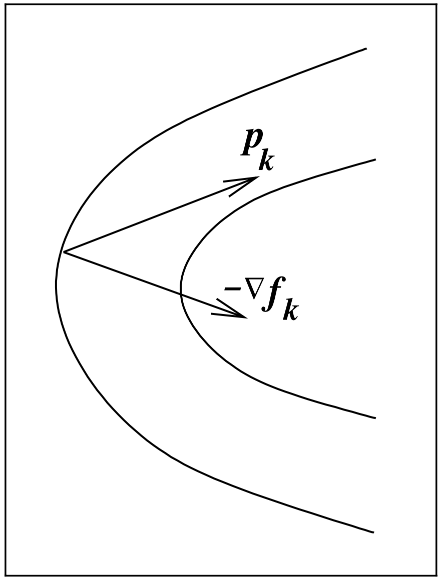
Proof of 1st order necessary conditions. assume $\nabla f(x^*) \neq 0$.
$s := -\nabla f(x^*)$ is a descent direction for f at $x = x^*$:
$\nabla f(x^*)^T(-\nabla f(x^*)) = -\nabla f(x^*)^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$
since $\nabla f(x^*) \neq 0$ and $\|a\| \geq 0$ with equality iff $a = 0$.
Thus, by Lemma 1, x^* is not a local minimizer of f. $\square$

Optimality conditions for unconstrained problems...

- $-\nabla f(x)$ is a descent direction for f at x whenever $\nabla f(x) \neq 0$.
- s descent direction for f at x if $\nabla f(x)^T s < 0$, which is equivalent to

$$\cos\langle -\nabla f(x), s\rangle = \frac{(-\nabla f(x))^T s}{\|\nabla f(x)\| \cdot \|s\|} = \frac{|\nabla f(x)^T s|}{\|\nabla f(x)\| \cdot \|s\|} > 0,$$

and so:
$\langle -\nabla f(x), s\rangle \in [0, \pi/2)$.



A descent direction p_k.

Summary of first-order conditions. A look ahead

minimize $f(x)$ subject to $x \in \mathbb{R}^n$. (UP)

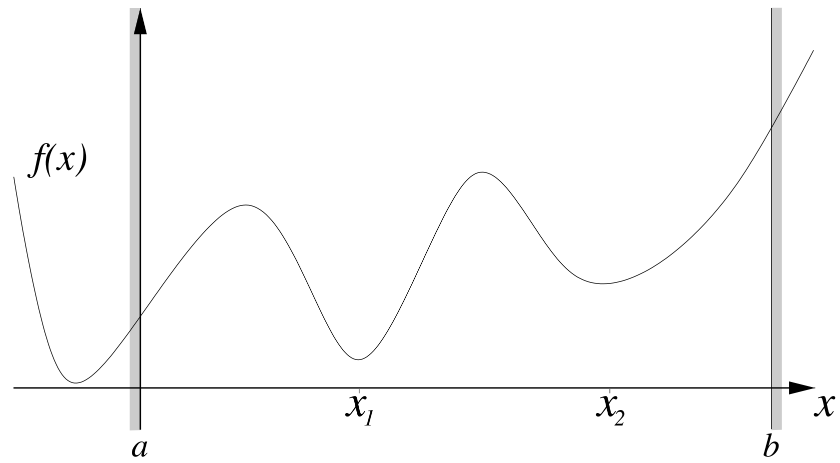
First-order necessary optimality conditions: $f \in \mathcal{C}^1(\mathbb{R}^n)$;

x^* a local minimizer of $f \implies \nabla f(x^*) = 0$.

$$\tilde{x} = \arg\max_{x \in \mathbb{R}^n} f(x)$$

$$\Downarrow$$

$$\nabla f(\tilde{x}) = 0.$$

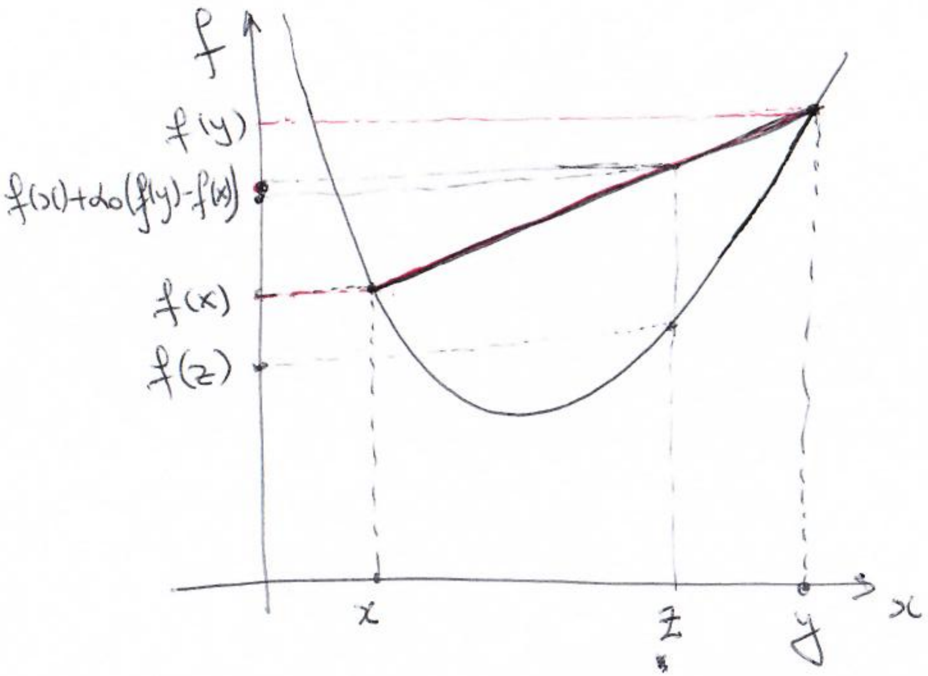


- Look at higher-order derivatives to distinguish between minimizers and maximizers.

. . . except for convex functions.

Illustration of the definition of convexity of f:

$f: \mathbb{R}^n \to \mathbb{R}$, f convex $\overset{\text{def.}}{\Leftrightarrow}$ $f(x+\alpha(y-x)) \leq f(x) + \alpha(f(y)-f(x))$, $\forall x, y \in \mathbb{R}^n$ and $\alpha \in [0,1]$.



Let $x, y \in \mathbb{R}$ $(n=1)$.
$z = x + \alpha_0(y-x)$, $\alpha_0 \in (0,1)$

If $f \in C^2(\mathbb{R}^n)$, then f convex $\Leftrightarrow$ $\nabla^2 f$ is positive semidefinite matrix.
(see Pb. sheet 1).

For convex f, any stationary point is a global minimizer of f.

Optimality conditions for convex problems

- f convex $\iff f(x + \alpha(y - x)) \leq f(x) + \alpha(f(y) - f(x))$,
  for all $x, y \in \mathbb{R}^n$, $\alpha \in [0, 1]$.
- $\iff \nabla^2 f(x)$ positive semidefinite, for all $x \in \mathbb{R}^n$, i.e.,
  - $s^T \nabla^2 f(x^*) s \geq 0$, $\forall\, s \in \mathbb{R}^n$; equivalently,
  - eigenvalues $\lambda_i(\nabla^2 f(x^*)) \geq 0$, $\forall i \in \{1, \ldots, n\}$.

If f convex, then [Pb Sheet 1]

$$x^* \text{ local minimizer} \implies x^* \text{ global minimizer.}$$

$$x^* \text{ stationary point} \implies x^* \text{ global minimizer.}$$

Optimality conditions for convex problems

- f convex $\iff$ $f(x+\alpha(y-x)) \le f(x)+\alpha(f(y)-f(x))$,
  for all $x,\ y \in \mathbb{R}^n$, $\alpha \in [0,1]$.
- $\iff$ $\nabla^2 f(x)$ positive semidefinite, for all $x \in \mathbb{R}^n$, i.e.,
  - $s^T\nabla^2 f(x^*)s \ge 0$, $\forall\, s \in \mathbb{R}^n$; equivalently,
  - eigenvalues $\lambda_i(\nabla^2 f(x^*)) \ge 0$, $\forall i \in \{1,\ldots,n\}$.

If f convex, then [Pb Sheet 1]

$$x^* \text{ local minimizer} \Longrightarrow x^* \text{ global minimizer.}$$

$$x^* \text{ stationary point} \Longrightarrow x^* \text{ global minimizer.}$$

Quadratic functions: $q(x) := g^T x + \frac{1}{2}x^T H x$.

$\nabla^2 q(x) = H$, for all x; if H is positive semidefinite, then q convex; any stationary point x^* is a global minimizer of q.