Lecture 2: Problems and solutions. Optimality conditions for unconstrained optimization (continued)

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Summary of first-order conditions. A look ahead

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

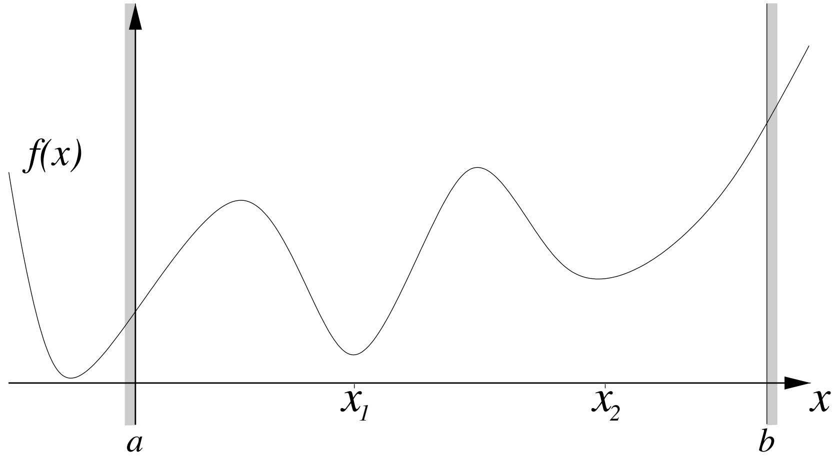
First-order necessary optimality conditions: $f \in \mathcal{C}^1(\mathbb{R}^n)$;

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.

Second-order optimality conditions (nonconvex fcts.)

Second-order necessary conditions: $f \in \mathcal{C}^2(\mathbb{R}^n)$;
x^* local minimizer of $f \implies \nabla^2 f(x^*)$ positive semidefinite, i.e.,
$s^T \nabla^2 f(x^*) s \geq 0$, for all $s \in \mathbb{R}^n$. [local convexity]

Example: $f(x) := x^3$, $x^* = 0$ not a local minimizer but
$f'(0) = f''(0) = 0$.

Second-order optimality conditions (nonconvex fcts.)

Second-order necessary conditions: $f \in \mathcal{C}^2(\mathbb{R}^n)$;

x^* local minimizer of $f \implies \nabla^2 f(x^*)$ positive semidefinite, i.e.,

$s^T \nabla^2 f(x^*) s \geq 0$, for all $s \in \mathbb{R}^n$. [local convexity]

Example: $f(x) := x^3$, $x^* = 0$ not a local minimizer but

$f'(0) = f''(0) = 0$.

The second order necessary conditions are not sufficient.

Second-order sufficient conditions: $f \in \mathcal{C}^2(\mathbb{R}^n)$;

$\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ positive definite, namely,

$$s^T \nabla^2 f(x^*) s > 0, \text{ for all } s \neq 0.$$

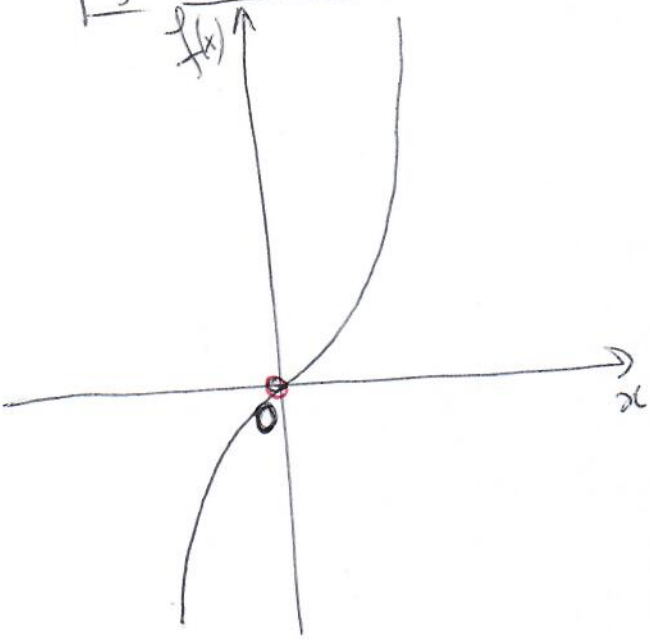
$\implies x^*$ (strict) local minimizer of f.

Example: $f(x) := x^4$, $x^* = 0$ is a (strict) local minimizer but

$f''(0) = 0$.

The second order sufficient conditions are not necessary.

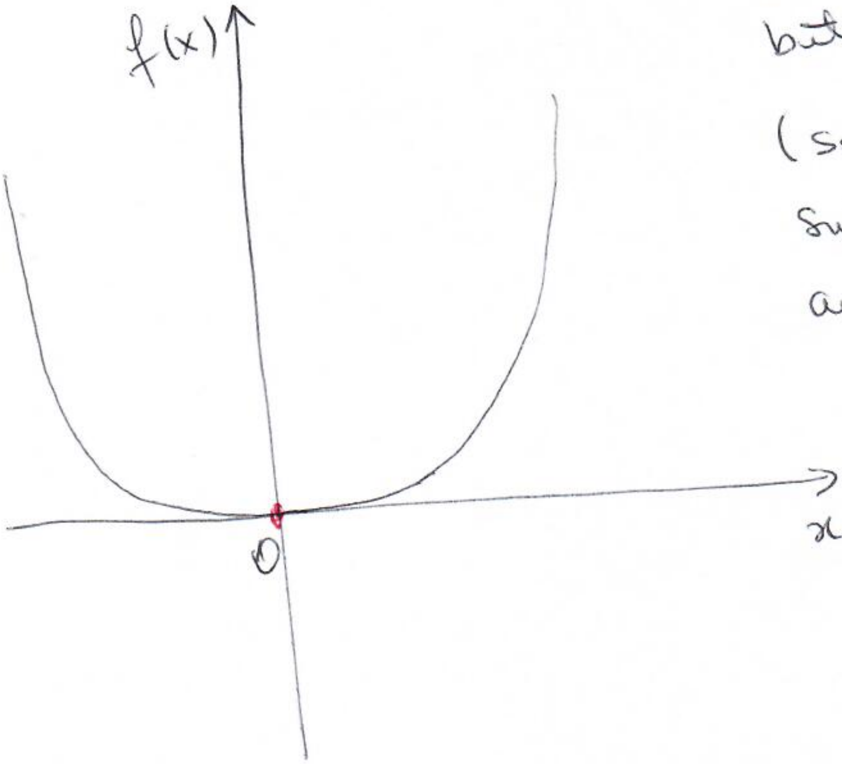
$\boxed{f(x) = x^3}$



$x^* = 0$ not a local min. but $f'(x^*) = f''(x^*) = 0$.

(so second-order necessary opt. conditions are satisfied).

$\boxed{f(x) = x^4}$



$x_* = 0$ local minimizer but $f'(x_*) = f''(x_*) = 0$.

(so second-order sufficient opt. conditions are not satisfied).

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

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

where $\mathcal{O}(\cdot)$ 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 $\nabla f=\left(\frac{\partial f}{\partial x_1}\cdots\frac{\partial f}{\partial x_n}\right)^T$.

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

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

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

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

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

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

$$\boxed{\phi(\alpha)=\phi(0)+\alpha\phi'(0)+\frac{\alpha^2}{2}\phi''(\tilde{\alpha})}, \text{ where } \tilde{\alpha}\in(0,\alpha). \quad (4)$$

(mean value theorem)

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

$$\boxed{f(x+\alpha s)=f(x)+\alpha\nabla f(x)^T s+\frac{1}{2}\alpha^2 s^T\nabla^2 f(x+\tilde{\alpha}s)s}, \quad (5)$$

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

Proof of second-order conditions

Recall second-order Taylor expansions (see (4) and (5) earlier, Lecture 1): let x and s in $\mathbb{R}^n$ be fixed; then for any $\alpha > 0$, we have

$$f(x + \alpha s) = f(x) + \alpha s^T \nabla f(x) + \frac{\alpha^2}{2} s^T \nabla^2 f(x + \tilde{\alpha} s) s \tag{5}$$

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

Proof of second-order conditions

Recall second-order Taylor expansions (see (4) and (5) earlier, Lecture 1): let x and s in $\mathbb{R}^n$ be fixed; then for any $\alpha > 0$, we have

$$f(x+\alpha s) = f(x) + \alpha s^T \nabla f(x) + \frac{\alpha^2}{2} s^T \nabla^2 f(x+\tilde{\alpha} s) s \tag{5}$$

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

Proof of second order necessary conditions. Assume there exists $s \in \mathbb{R}^n$ with $s^T \nabla^2 f(x^*) s < 0$.

Proof of second-order conditions

Recall second-order Taylor expansions (see (4) and (5) earlier, Lecture 1): let x and s in $\mathbb{R}^n$ be fixed; then for any $\alpha > 0$, we have

$$f(x+\alpha s) = f(x) + \alpha s^T \nabla f(x) + \frac{\alpha^2}{2} s^T \nabla^2 f(x+\tilde{\alpha}s)s \qquad (5)$$

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

Proof of second order necessary conditions. Assume there exists $s \in \mathbb{R}^n$ with $s^T \nabla^2 f(x^*)s < 0$. Then $s \neq 0$ and $f \in \mathcal{C}^2$ imply there exists $\hat{\alpha} > 0$ such that

$$s^T \nabla^2 f(x^* + \alpha s)s < 0 \text{ for all } \alpha \in [0, \hat{\alpha}]. \qquad (6)$$

Proof of second-order conditions

Recall second-order Taylor expansions (see (4) and (5) earlier, Lecture 1): let x and s in $\mathbb{R}^n$ be fixed; then for any $\alpha > 0$, we have

$$f(x+\alpha s) = f(x) + \alpha s^T \nabla f(x) + \frac{\alpha^2}{2} s^T \nabla^2 f(x+\tilde{\alpha} s) s \qquad (5)$$

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

Proof of second order necessary conditions. Assume there exists $s \in \mathbb{R}^n$ with $s^T \nabla^2 f(x^*) s < 0$. Then $s \neq 0$ and $f \in \mathcal{C}^2$ imply there exists $\hat{\alpha} > 0$ such that

$$s^T \nabla^2 f(x^* + \alpha s) s < 0 \text{ for all } \alpha \in [0, \hat{\alpha}]. \qquad (6)$$

Let $\alpha \in (0, \hat{\alpha})$. Then (5) with $x = x^*$ and $\nabla f(x^*) = 0$ imply

$$f(x^* + \alpha s) = f(x^*) + \frac{\alpha^2}{2} s^T \nabla^2 f(x^* + \tilde{\alpha} s) s. \qquad (7)$$

for some $\tilde{\alpha} \in (0, \alpha)$. Since $0 < \tilde{\alpha} < \alpha \leq \hat{\alpha}$, (6) implies that $s^T \nabla^2 f(x^* + \tilde{\alpha} s) s < 0$. Thus (7) implies $f(x^* + \alpha s) < f(x^*)$, and this holds for all $\alpha \in (0, \hat{\alpha}]$. Contradiction, as x^* is a local minimizer.$\square$

Proof of second-order conditions ...

Proof of second order sufficient conditions. $f \in \mathcal{C}^2$ and $\nabla^2 f(x^*) \succ 0$ imply that

Proof of second-order conditions ...

Proof of second order sufficient conditions. $f \in \mathcal{C}^2$ and $\nabla^2 f(x^*) \succ 0$ imply that there exists a neighbourhood $\mathcal{N}(x^*, \delta)$ of x^* such that

$$\nabla^2 f(x^* + s) \succ 0 \text{ for all } x^* + s \in \mathcal{N}(x^*, \delta). \quad (8)$$

Proof of second-order conditions ...

Proof of second order sufficient conditions. $f \in \mathcal{C}^2$ and $\nabla^2 f(x^*) \succ 0$ imply that there exists a neighbourhood $\mathcal{N}(x^*, \delta)$ of x^* such that

$$\nabla^2 f(x^* + s) \succ 0 \text{ for all } x^* + s \in \mathcal{N}(x^*, \delta). \qquad (8)$$

Use (5) with $x = x^*$, $\alpha = 1$ and for any s with $x^* + s \in \mathcal{N}(x^*, \delta)$:

$$f(x^* + s) = f(x^*) + s^T \nabla f(x^*) + \frac{1}{2} s^T \nabla^2 f(x^* + \tilde{\alpha} s) s \qquad (9)$$

for some $\tilde{\alpha} \in (0, 1)$.

Proof of second-order conditions ...

Proof of second order sufficient conditions. $f \in \mathcal{C}^2$ and $\nabla^2 f(x^*) \succ 0$ imply that there exists a neighbourhood $\mathcal{N}(x^*, \delta)$ of x^* such that

$$\nabla^2 f(x^* + s) \succ 0 \text{ for all } x^* + s \in \mathcal{N}(x^*, \delta). \qquad (8)$$

Use (5) with $x = x^*$, $\alpha = 1$ and for any s with $x^* + s \in \mathcal{N}(x^*, \delta)$:

$$f(x^* + s) = f(x^*) + s^T \nabla f(x^*) + \frac{1}{2} s^T \nabla^2 f(x^* + \tilde{\alpha} s) s \qquad (9)$$

for some $\tilde{\alpha} \in (0, 1)$.

Note that $\|x^* + \tilde{\alpha} s - x^*\| = \tilde{\alpha}\|s\| \leq \delta$ since $\tilde{\alpha} \in (0, 1)$ and $x^* + s \in \mathcal{N}(x^*, \delta)$ (so that $\|s\| \leq \delta$); thus $x^* + \tilde{\alpha} s \in \mathcal{N}(x^*, \delta)$ which ensures that $\nabla^2 f(x^* + \tilde{\alpha} s) \succ 0$ due to (8).

Proof of second-order conditions ...

Proof of second order sufficient conditions. $f \in \mathcal{C}^2$ and $\nabla^2 f(x^*) \succ 0$ imply that there exists a neighbourhood $\mathcal{N}(x^*, \delta)$ of x^* such that

$$\nabla^2 f(x^* + s) \succ 0 \text{ for all } x^* + s \in \mathcal{N}(x^*, \delta). \qquad (8)$$

Use (5) with $x = x^*$, $\alpha = 1$ and for any s with $x^* + s \in \mathcal{N}(x^*, \delta)$:

$$f(x^* + s) = f(x^*) + s^T \nabla f(x^*) + \tfrac{1}{2} s^T \nabla^2 f(x^* + \tilde{\alpha} s) s \qquad (9)$$

for some $\tilde{\alpha} \in (0, 1)$.

Note that $\|x^* + \tilde{\alpha} s - x^*\| = \tilde{\alpha}\|s\| \leq \delta$ since $\tilde{\alpha} \in (0, 1)$ and $x^* + s \in \mathcal{N}(x^*, \delta)$ (so that $\|s\| \leq \delta$); thus $x^* + \tilde{\alpha} s \in \mathcal{N}(x^*, \delta)$ which ensures that $\nabla^2 f(x^* + \tilde{\alpha} s) \succ 0$ due to (8).

This and (9), as well as $\nabla f(x^*) = 0$, imply $f(x^* + s) > f(x^*)$ for all $s \neq 0$ with $x^* + s \in \mathcal{N}(x^*, \delta)$, q.e.d. □

Stationary points of quadratic functions

- $H \in \mathbb{R}^{n\times n}$ symmetric (wlog), $g \in \mathbb{R}^n$: q quadratic function

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

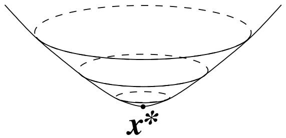
$\nabla q(x^*) = 0 \iff Hx^* + g = 0$: linear system.

$\nabla^2 q(x) = H$ for all $x \in \mathbb{R}^n$.

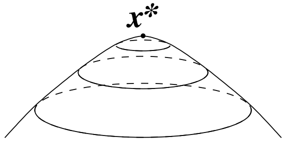
  - H nonsingular: $x^* = -H^{-1}g$ unique stationary point.
    - H positive definite $\Longrightarrow$ x^* minimizer (a), e)).
    - H negative definite $\Longrightarrow$ x^* maximizer (b), e)).
    - H indefinite $\Longrightarrow$ x^* saddle point (c), f)).
  - H singular and $g + Hx = 0$ consistent:
    - H positive semidefinite $\Longrightarrow$ infinitely many global minimizers (d), g)).
    - Similarly H negative semidefinite or indefinite.

General f: approximately locally quadratic around x^* stationary.

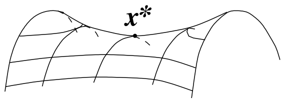
Stationary points of quadratic functions...



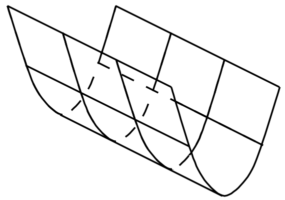
(a) Minimum



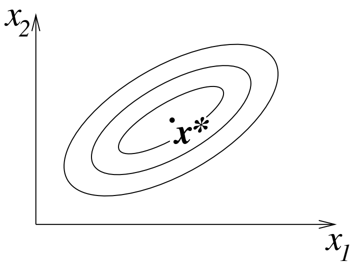
(b) Maximum



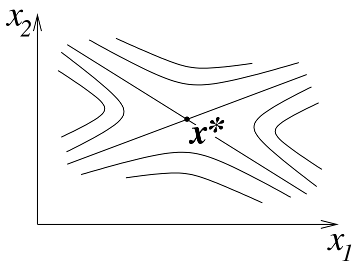
(c) Saddle



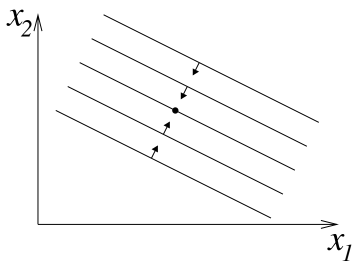
(d) Semidefinite



(e) Maximum or minimum



(f) Saddle



(g) Semidefinite