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

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization
Let \( \phi : \mathbb{R} \rightarrow \mathbb{R} \), \( \phi \in C^1(\mathbb{R}) \) i.e. continuously differentiable.

Then for any \( \alpha \in \mathbb{R} \), we have
\[
\phi(\alpha) = \phi(0) + \alpha \phi'(0) + O(\alpha^2) \tag{1}
\]
where \( O(\cdot) \) implies an upper bound that is a multiple of \( \alpha^2 \).

Also,
\[
\phi(\alpha) = \phi(0) + \alpha \phi'(0) + \frac{\alpha^2}{2} \phi''(\xi) \tag{2}
\]
[mean-value theorem]

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

Let \( X = (x_1, \ldots, x_n)^T \) and \( S = (s_1, \ldots, s_n)^T \) \( E \mathbb{R}^n \), fixed.

Then \( f(X+S) = f(X) + \nabla f(X) \cdot S + \frac{1}{2} S^T \nabla^2 f(X) S \).

By chain rule,
\[
\phi'(\alpha) = \sum_{i=1}^n \frac{d}{dx_i} f(x_1 + \alpha s_1, \ldots, x_n + \alpha s_n) \quad \text{(by chain rule)}
\]

Thus the first-order Taylor expansion of \( \phi \) gives from (2),
\[
f(x + \alpha S) = f(x) + \alpha \nabla f(x + \alpha S) \cdot S, \quad \text{for some } \alpha \in (0, 1).
\tag{3}
\]

SECOND ORDER TAYLOR EXPANSION

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

Then for any \( \alpha \in \mathbb{R} \), we have
\[
\phi(\alpha) = \phi(0) + \alpha \phi'(0) + \frac{\alpha^2}{2} \phi''(\xi), \tag{4}
\]
where \( \xi \in (0, \alpha) \), (mean value theorem)

Thus the second-order Taylor expansion of \( \phi(\alpha) \) gives
\[
f(x + \alpha S) = f(x) + \alpha \nabla f(x) \cdot S + \frac{1}{2} \alpha^2 S^T \nabla^2 f(x + \xi S) S, \quad \text{for some } \xi \in (0, \alpha). \tag{5}
\]
Unconstrained optimization problems and solutions

minimize \( f(x) \) subject to \( x \in \mathbb{R}^n \). \hspace{1cm} (UP)

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

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

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

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

The points \( x_1 \) and \( x_2 \) are (unconstrained) local minimizers of \( f \) (for example).
Optimality conditions for unconstrained problems

== algebraic characterizations of solutions → 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 → 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):

\[ f \in C^1(\mathbb{R}^n); \]
\[ x^* \text{ a local minimizer of } f \implies \nabla f(x^*) = 0. \]
\[ \nabla f(x) = 0 \iff x \text{ stationary point of } f. \]
Lemma 1. Let \( f \in C^1 \), \( x \in \mathbb{R}^n \) and \( s \in \mathbb{R}^n \) with \( s \neq 0 \). Then
\[
\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \quad \forall \alpha > 0 \text{ sufficiently small.}
\]
Lemma 1. Let $f \in C^1$, $x \in \mathbb{R}^n$ 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. $f \in C^1$ and the gradient is continuous \[ \implies \exists \bar{\alpha} > 0 \text{ such that } \nabla f(x + \alpha s)^T s < 0, \quad \forall \alpha \in [0, \bar{\alpha}] \hspace{1cm} (\Diamond) \]
**Lemma 1.** Let $f \in C^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then
\[
\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \quad \forall \alpha > 0 \text{ sufficiently small.}
\]

**Proof.**
\[
f \in C^1 \implies \exists \alpha > 0 \text{ such that } \nabla f(x + \alpha s)^T s < 0, \quad \forall \alpha \in [0, \alpha]. \quad (\diamond)
\]

Taylor’s/Mean value theorem: by First order Taylor expansion revision slide, see equation (3)
\[
f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s, \text{ for some } \tilde{\alpha} \in (0, \alpha).
\]
Lemma 1. Let $f \in C^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then

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

Proof. $f \in C^1 \implies \exists \alpha > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \; \forall \alpha \in [0, \alpha]. \quad (\diamond)$$

Taylor’s/Mean value theorem:

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

$(\diamond) \implies f(x + \alpha s) < f(x), \; \forall \alpha \in (0, \alpha]. \; \square$
Lemma 1. Let $f \in C^1$, $x \in \mathbb{R}^n$ 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. $f \in C^1 \implies \exists \overline{\alpha} > 0$ such that
$$\nabla f(x + \alpha s)^T s < 0, \forall \alpha \in [0, \overline{\alpha}].$$

Taylor’s/Mean value theorem:
$$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s, \text{ for some } \tilde{\alpha} \in (0, \alpha).$$

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

• $s$ descent direction for $f$ at $x$ if $\nabla f(x)^T s < 0$. 
Optimality conditions for unconstrained problems...

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

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

sufficiently small.

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

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

(♦)

Taylor’s/Mean value theorem:

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

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

(♦) $\implies$ $f(x + \alpha s) < f(x), \quad \forall \alpha \in [0, \bar{\alpha}].$ □

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

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

Lecture 1: Problems and solutions. Optimality conditions for unconstrained optimization – p. 9/17
Lemma 1. Let $f \in C^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then
\[ \nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \quad \forall \alpha > 0 \text{ sufficiently small.} \]

Proof. $f \in C^1 \implies \exists \overline{\alpha} > 0$ such that
\[ \nabla f(x + \alpha s)^T s < 0, \quad \forall \alpha \in [0, \overline{\alpha}]. \quad (\diamond) \]

Taylor’s/Mean value theorem:
\[ f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s, \text{ for some } \tilde{\alpha} \in (0, \alpha). \]
(\diamond) $\implies f(x + \alpha s) < f(x), \quad \forall \alpha \in [0, \overline{\alpha}]. \quad \square$

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

Proof of 1st order necessary conditions. assume $\nabla f(x^*) \neq 0$.
\[ s := -\nabla f(x^*) \text{ is a descent direction for } f \text{ at } x = x^*. \]
Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in C^1$, $x \in \mathbb{R}^n$ 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. $f \in C^1 \implies \exists \bar{\alpha} > 0$ such that
\[ \nabla f(x + \alpha s)^T s < 0, \quad \forall \alpha \in [0, \bar{\alpha}]. \quad (\diamond) \]

Taylor’s/Mean value theorem:
\[ f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s, \quad \text{for some } \tilde{\alpha} \in (0, \alpha). \]

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

• $s$ 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 $a = 0$. 

Lecture 1: Problems and solutions. Optimality conditions for unconstrained optimization – p. 9/17
Lemma 1. Let $f \in C^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then
\[ \nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \ \forall \alpha > 0 \text{ sufficiently small.} \]

Proof. $f \in C^1 \implies \exists \overline{\alpha} > 0$ such that
\[ \nabla f(x + \alpha s)^T s < 0, \ \forall \alpha \in [0, \overline{\alpha}]. \] (♦)

Taylor’s/Mean value theorem:
\[ f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s, \ \text{for some } \tilde{\alpha} \in (0, \alpha). \] (♦) $\implies f(x + \alpha s) < f(x), \ \forall \alpha \in (0, \overline{\alpha}]$. □

• $s$ 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 $f$. □
• $-\nabla f(x)$ 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:

$$\langle -\nabla f(x), s \rangle \in [0, \pi/2).$$
Summary of first-order conditions. A look ahead

minimize $f(x)$ subject to $x \in \mathbb{R}^n$. \hspace{0.5cm} (UP)

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

\[ \ldots \text{except for convex functions.} \]
Illustration of the definition of convexity of $f$:

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

Let $x_1, y \in \mathbb{R}^n$ ($n=1$),

$z = x + \alpha_0 (y-x)$, $\alpha_0 \in (0,1)$

If $f \in C^2(\mathbb{R}^n)$, then $f$ convex ($\Rightarrow$) $D^2f$ is positive semi-definite matrix.

(see pg. 8 sheet 1).

For convex $f$, any stationary point is a global minimizer of $f$. 
Optimality conditions for convex problems

- \( f \) convex \( \iff \quad f(x + \alpha(y - x)) \leq f(x) + \alpha(f(y) - f(x)), \) for all \( x, y \in \mathbb{R}^n, \alpha \in [0, 1]. \)

- \( \iff \quad \nabla^2 f(x) \) 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

- \( x^* \) local minimizer \( \implies \) \( x^* \) global minimizer.
- \( x^* \) stationary point \( \implies \) \( x^* \) global minimizer.
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

- \( x^* \) local minimizer \implies \( x^* \) global minimizer.
- \( x^* \) stationary point \implies \( x^* \) 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 \( q. \)