C477: Introduction to Optimality Conditions

Ruth Misener
r.misener@imperial.ac.uk

Panos Parpas
p.parpas@imperial.ac.uk

Computational Optimisation Group
Department of Computing

Imperial College London

computational optimisation group

19 October 2018

Review

Definition: Convex Optimisation Problem

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, m \\
& h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, p
\end{aligned}
$$

If f and $g_1, \ldots, g_m$ are convex on $\mathbb{R}^n$, and $h_1, \ldots, h_p$ are affine, then this is said to be a **convex optimisation problem**

Recall, from the lecture and tutorial on convexity that we can be more general than this. But for the purposes of this class, we will stick to the above definition.

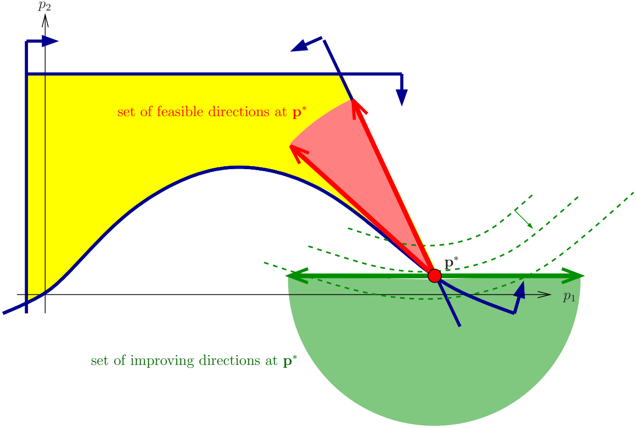
Outline

- **Topics**
  - Necessary Conditions for (Local) Optimality
    - First Order Condition
    - Second Order Condition
  - More on eigenvalues and positive semidefinite matrices;
  - Sufficient Condition for (Local) Optimality
  - This is a first pass on the subject; we will cover the Karush-Kuhn-Tucker condition in the second half of the class.
- **Example**
  - Designing a Wireless System
- **Reading**
  - Chapter 6.2 (Conditions for Local Minimizers) in *An Introduction to Optimization*, Chong & Żak, Third Edition.
- **Acknowledgements**

  - Parts of these slides were originally developed by Benoit Chachuat and Panos Parpas. LaTeX design and proof reading by Miten Mistry. Mistakes by Ruth Misener.

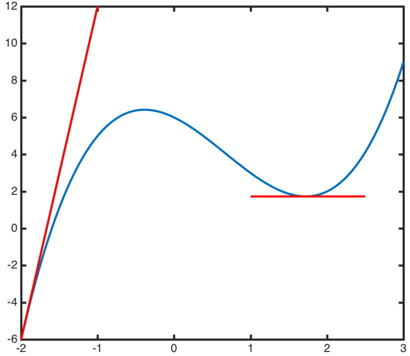
Necessary Conditions for Optimality

No optimisation model solution at which an improving feasible direction is available can be a local optimum



First Order **Necessary** Condition for Local Minimisers

- **First Order Condition:** Only use first order derivatives;
- Assume: f is $\mathcal{C}^1$, i.e., once continuously differentiable.



Reminder: The gradient of f is denoted by:

$$\nabla f(\boldsymbol{x}) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^\top.$$

The Jacobian of f is denoted by Df and $\nabla f = Df^\top$

Example

$f(x_1, x_2) = 5x_1 + 8x_2 + x_1 x_2 - x_1^2 - 2x_2^2$

Find the Jacobian $Df(\boldsymbol{x})$ and the Hessian, $\nabla^2 f(\boldsymbol{x})$:

$$Df(\boldsymbol{x}) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right] = [5 + x_2 - 2x_1, 8 + x_1 - 4x_2]$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}$$

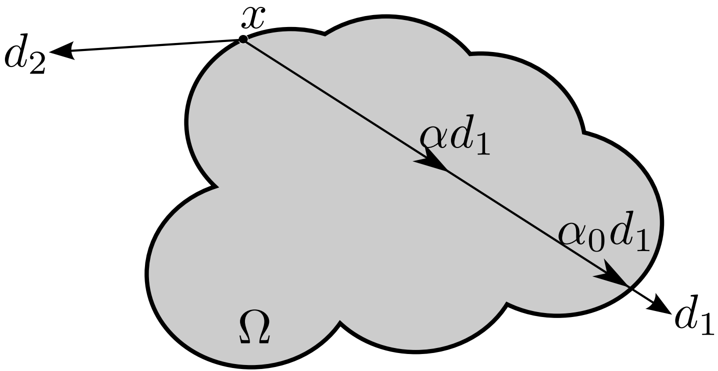
Sanity Check

Is $\min_{\boldsymbol{x}} f(\boldsymbol{x})$ a convex optimisation problem? Are local optimisers necessarily global optimisers?

Feasible Directions

Definition (Feasible Direction)

A vector $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d} \neq \boldsymbol{0}$ is a feasible direction at $\boldsymbol{x} \in \Omega$ if there exists an $\alpha_0 > 0$ such that $\boldsymbol{x} + \alpha\, \boldsymbol{d} \in \Omega$ for all $\alpha \in [0,\, \alpha_0]$.



Examples of Feasible Directions

Are the following directions feasible at the origin $(0, 0)$?

(a) $\Omega = \{(x_1,\, x_2) \mid x_1^2 + x_2 \leq 2\}$, $\boldsymbol{d} = (1,\, 2)$.

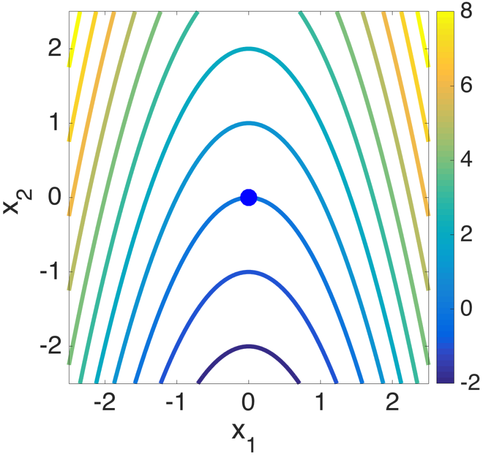
Yes, since for $\alpha = \frac{1}{2}$, $\alpha^2 + 2\alpha = \frac{5}{4} < 2$.

(b) $\Omega = \{(x_1,\, x_2) \mid x_1^2 + x_2 \leq 3\}$, $\boldsymbol{d} = (5, 10)$.

Yes, since for $\alpha = \frac{1}{5}$, $25\alpha^2 + 10\alpha = 3$.

(c) $\Omega = \{(x_1,\, x_2) \mid x_1^2 + x_2 \leq 0\}$, $\boldsymbol{d} = (-2, 1)$.

No, since no $\alpha > 0$ exists such that $4\alpha^2 + \alpha \leq 0$.



Directional Derivative

Definition (Directional Derivative)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real valued function, and let $\boldsymbol{d} \in \mathbb{R}^n \backslash \mathbf{0}$. The directional derivative of f in the direction $\boldsymbol{d}$ is defined as,

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{d}) - f(\boldsymbol{x})}{\alpha}$$

Calculating Directional Derivatives

Suppose that $\boldsymbol{x}$ and $\boldsymbol{d}$ are given, then

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \left. \frac{\partial f}{\partial \alpha}(\boldsymbol{x} + \alpha \boldsymbol{d}) \right|_{\alpha=0}$$

Using the chain rule,

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \left. \frac{\partial f}{\partial \alpha}(\boldsymbol{x} + \alpha \boldsymbol{d}) \right|_{\alpha=0} = \nabla f(\boldsymbol{x})^\top \boldsymbol{d} = \langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \rangle = \boldsymbol{d}^\top \nabla f(\boldsymbol{x})$$

Directional Derivative Example

Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\boldsymbol{x}) = x_1 x_2 x_3$ and let,

$$\boldsymbol{d} = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right]^\top$$

Compute the directional derivative of f in the direction $\boldsymbol{d}$

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^\top \boldsymbol{d} = [x_2 x_3 \; x_1 x_3 \; x_1 x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix}$$

$$= \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}$$

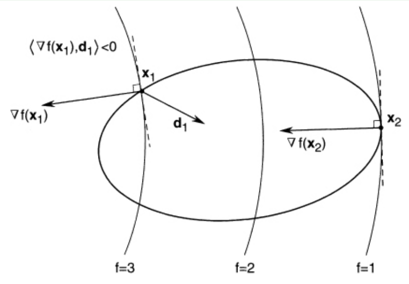
First Order Condition (FONC)

Theorem (First Order Necessary Condition)

Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real valued function on Ω. If $\boldsymbol{x}^*$ is a local minimiser of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$,

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

Example: $\boldsymbol{x}_1$ does not satisfy the FONC, $\boldsymbol{x}_2$ does



First Order Condition (FONC)

Theorem (First Order Necessary Condition)

Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real valued function on Ω. If $\boldsymbol{x}^*$ is a local minimiser of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$,

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

Proof.

Since $\boldsymbol{x}^*$ is a local minimiser, for any feasible direction $\boldsymbol{d}$ there exists an $\bar{\alpha}$ such that for all $\alpha \in (0, \bar{\alpha})$, $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^* + \alpha\, \boldsymbol{d})$. Hence for all $\alpha \in (0, \bar{\alpha})$.

$$\frac{f(\boldsymbol{x}^* + \alpha\, \boldsymbol{d}) - f(\boldsymbol{x}^*)}{\alpha} \geq 0.$$

Taking the limit $\alpha \to 0$ and the fact that $\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \boldsymbol{d}^\top \nabla f(\boldsymbol{x})$ we obtain $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0$. $\square$

First Order Condition (FONC): Interior case

Corollary (First Order Necessary Condition (Interior case))

Let Ω *be a subset of* $\mathbb{R}^n$ *and* $f \in \mathcal{C}^1$ *a real valued function on* Ω*. If* $\boldsymbol{x}^*$ *is a local minimiser of* f *over* Ω*, and* $\boldsymbol{x}^*$ *is an interior point of* Ω *then*

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0}.$$

Proof.

Since $\boldsymbol{x}^*$ is a local minimiser, and an interior point of Ω, the set of feasible directions at $\boldsymbol{x}$ is the whole of $\mathbb{R}^n$. Therefore for any $\boldsymbol{d} \in \mathbb{R}^n$ we must have,

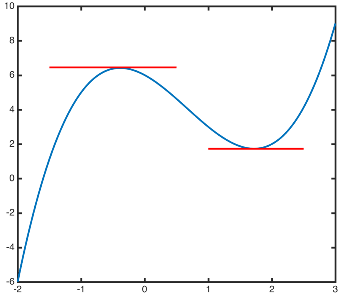
$$\begin{aligned} \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) &\geq 0 \\ -\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) &\geq 0 \end{aligned}$$

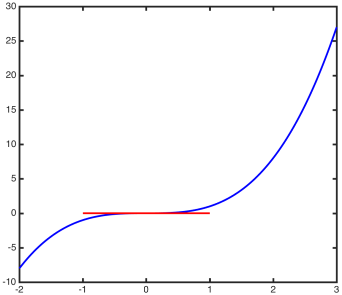
Hence $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$ for all $\boldsymbol{d} \in \mathbb{R}^n$, therefore $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$. $\square$

Why is the **FONC** not **Sufficient**?

Sanity Check

Cases where points satisfying the FONC are not local minimisers?





Most algorithms will test the FONC as a termination criteria. But state-of-the-art codes often have other, additional, tests.

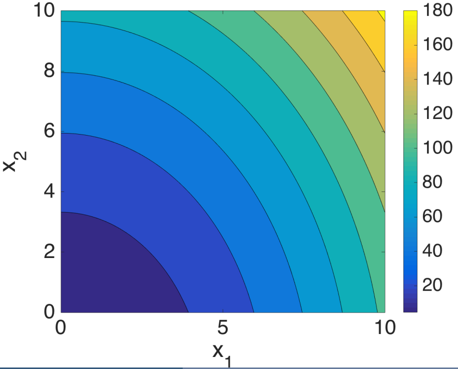
First Order Necessary Condition (FONC): Example

Example

$$\begin{aligned} \min_{x_1, x_2} \ & x_1^2 + \frac{1}{2}x_2^2 + 3x_2 + 4.5 \\ \text{s.t. } & x_1 \geq 0, \\ & x_2 \geq 0. \end{aligned}$$

FONC for a local minimiser satisfied at these points?

- (a) $\boldsymbol{x} = [1, 3]^\top$
- (b) $\boldsymbol{x} = [0, 3]^\top$
- (c) $\boldsymbol{x} = [1, 0]^\top$
- (d) $\boldsymbol{x} = [0, 0]^\top$



FONC Example: Parts (a) & (b)

$\boldsymbol{x} = [1,\, 3]^\top$ (Interior point)

For an interior point, the FONC requires that $\nabla f(\boldsymbol{x}) = \mathbf{0}$. We have:

$$\nabla f(\boldsymbol{x}) = [2x_1,\, x_2 + 3]^\top.$$

Substituting, $\nabla f\left([1,\, 3]^\top\right) = [2,\, 6]^\top \neq \mathbf{0}$, and the point does not satisfy the FONC.

$\boldsymbol{x} = [0,\, 3]^\top$ (Boundary point)

At this point we have $\nabla f(\boldsymbol{x}) = [0,\, 6]^\top$, hence $\nabla f(\boldsymbol{x})^\top \boldsymbol{d} = 6d_2$, where $\boldsymbol{d} = [d_1,\, d_2]$. For $\boldsymbol{d}$ to be feasible we need $d_1 \geq 0$ and d_2 can be arbitrary. But if d_2 is negative then $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) < 0$. For example, if we take $\boldsymbol{d} = [1,\, -1]^\top$ then $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = -6 < 0$.

FONC Example: Parts (c) & (d)

$\boldsymbol{x} = [1,\, 0]^\top$ (Boundary point)

At this point we have $\nabla f(\boldsymbol{x}) = [2,\, 3]^\top$ and hence $\nabla f(\boldsymbol{x})^\top \boldsymbol{d} = 2d_1 + 3d_2$. For $\boldsymbol{d}$ to be feasible we need $d_2 \geq 0$ and d_1 can be arbitrary. If we take $\boldsymbol{d} = [-5,\, 1]^\top$, then $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = -7 < 0$. Hence this point does not satisfy the FONC either.

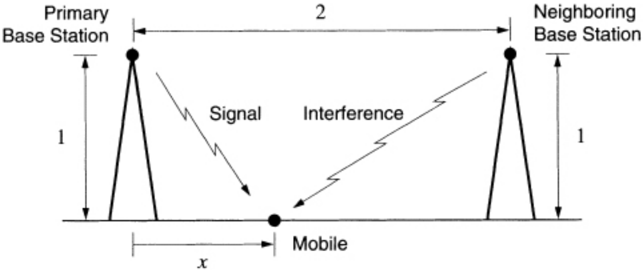
$\boldsymbol{x} = [0,\, 0]^\top$ (Boundary point)

At this point we have $\nabla f(\boldsymbol{x}) = [0,\, 3]^\top$ and hence $\nabla f(\boldsymbol{x})^\top \boldsymbol{d} = 3\, d_2$. For $\boldsymbol{d}$ to be feasible we need $d_1 \geq 0$ and $d_2 \geq 0$. Hence this point does satisfy the FONC.

Sanity Check

Are the FONC sufficient on the boundary of the feasible set?

Example: Designing a Wireless System



- Two base station antennas, one primary and one neighbouring;
- Both stations have equal power;
- Power of the received signal measured by the receiver (mobile) is the reciprocal of the squared distance from the associated antenna;
- Find the receiver position maximising the signal-to-interference ratio, i.e., the ratio of the signal power received from the primary station to the signal power received from the neighbouring base station.

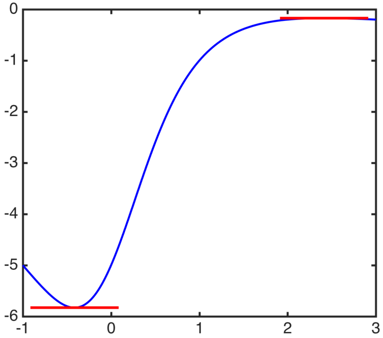
Example: Designing a Wireless System

The squared distance to the primary antenna is $1 + x^2$ and the squared distance to the neighbouring antenna is $1 + (2 - x)^2$. Signal-to-interference ratio:

$$f(x) = \frac{1 + (2 - x)^2}{1 + x^2}$$

Optimisation problem

$$-\min_x \left[-1 \cdot \frac{1 + (2 - x)^2}{1 + x^2} \right]$$

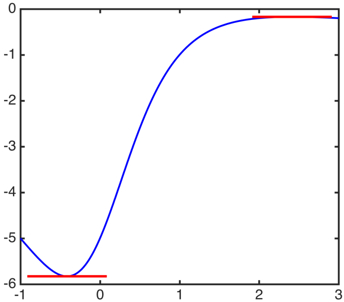


Example: Designing a Wireless System

The FONC for this problem is $\frac{df}{dx} = 0$,

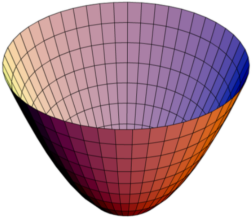
$$\frac{df}{dx} = \frac{-2(2-x)(1+x^2) - 2x\left(1+(2-x)^2\right)}{(1+x^2)^2} = \frac{4(x^2-2x-1)}{(1+x^2)^2}.$$

Therefore, $x^* = 1 \pm \sqrt{2}$, by evaluating the solution at the two points we find that $x^* = 1 - \sqrt{2}$.



Second Order **Necessary** Condition for Local Minimisers

- **Second Order Condition:** Also use second order derivatives;
- Assume: f is $\mathcal{C}^2$, i.e., twice continuously differentiable.



Given a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $\boldsymbol{x} \in \mathbb{R}^n$, recall the Hessian, $\boldsymbol{H}(\boldsymbol{x})$, the matrix of second partial derivatives

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{H}(\boldsymbol{x}) \triangleq \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{x}) \end{pmatrix}$$

Review of Eigenvalues & Eigenvectors

Recall Eigenvalues & Eigenvectors

- An **eigenvector** of square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a vector $\boldsymbol{v} \in \mathbb{R}^n \backslash \mathbf{0}$ such that the product $\boldsymbol{Av}$ is equal to a scalar multiple ($\lambda \in \mathbb{R}$) of $\boldsymbol{v}$:

$$\boldsymbol{Av} = \lambda \boldsymbol{v}$$

- The scalars λ are called **eigenvalues**; a matrix is positive definite if all eigenvalues are positive.
- For λ to be an eigenvalue it is necessary and sufficient for the determinant of matrix $\boldsymbol{A} - \lambda \boldsymbol{I}$ to be 0, that is:

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = \left| \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} - \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix} \right|$$

Review of Positive (Semi)Definite Matrices

- A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is **positive semidefinite** if for all $\boldsymbol{d} \in \mathbb{R}^n$,

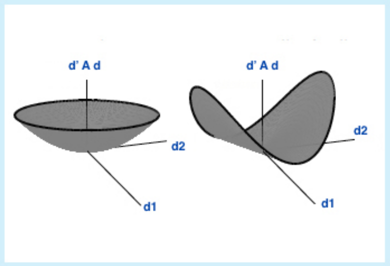
$$\boldsymbol{d}^\top \boldsymbol{A} \boldsymbol{d} \geq 0.$$

  We say $\boldsymbol{A} \succeq \mathbf{0}$; all eigenvalues of $\left(\boldsymbol{A} + \boldsymbol{A}^\top\right)/2$ are non-negative.

- If the above inequality is satisfied strictly, i.e. if

$$\boldsymbol{d}^\top \boldsymbol{A} \boldsymbol{d} > 0, \ \forall\, \boldsymbol{d} \in \mathbb{R}^n \backslash \mathbf{0},$$

  then $\boldsymbol{A}$ is called **positive definite**. We say $\boldsymbol{A} \succ \mathbf{0}$. All eigenvalues of $\left(\boldsymbol{A} + \boldsymbol{A}^\top\right)/2$ are positive.



Notation Alert!

Chong & Żak write $\boldsymbol{A} \geq 0$; we write $\boldsymbol{A} \succeq \mathbf{0}$. The Chong & Żak notation is uncommon.

How do I know that eigenvalues exist and are real numbers?

Symmetric Matrix

$$\boldsymbol{A} = \boldsymbol{A}^\top$$

FACTS (Please post to Piazza if you want to prove these)

1. A symmetric matrix has **real eigenvalues**;
2. There are up to n distinct eigenvalues in a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$.

Assume (for the purposes of C477) that matrices are symmetric

For testing if a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is positive (semi)definite, assume without loss of generality that $\boldsymbol{A}$ is symmetric. If $\boldsymbol{A}$ was not symmetric, we could take its *symmetric part*:

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = \boldsymbol{x}^\top \left(\frac{\boldsymbol{A}^\top + \boldsymbol{A}}{2} \right) \boldsymbol{x}$$

Symmetric Positive (Semi)Definite Matrices

- A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is **positive semidefinite** if $\forall\, \boldsymbol{d} \in \mathbb{R}^n$,

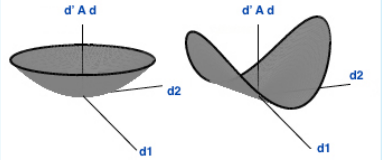
$$\boldsymbol{d}^\top \boldsymbol{A} \boldsymbol{d} \geq 0.$$

  We say $\boldsymbol{A} \succeq \mathbf{0}$; all eigenvalues of $\boldsymbol{A}$ are non-negative.

- If the above inequality is satisfied strictly, i.e. if

$$\boldsymbol{d}^\top \boldsymbol{A} \boldsymbol{d} > 0, \ \forall\, \boldsymbol{d} \in \mathbb{R}^n \backslash \mathbf{0},$$

  then $\boldsymbol{A}$ is called **positive definite**. We say $\boldsymbol{A} \succ \mathbf{0}$. All eigenvalues of $\boldsymbol{A}$ are positive.



Notation Alert!

Chong & Żak write $\boldsymbol{A} \geq 0$; we write $\boldsymbol{A} \succeq \mathbf{0}$. The Chong & Żak notation is uncommon.

Easy tests to know if a matrix is Positive Definite?

Symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is positive definite if & only if ...

- All n eigenvalues are positive;
- Sylvester's criterion: all n upper left determinants positive;
- $\boldsymbol{d}^\top \boldsymbol{A}\boldsymbol{d} > 0,\ \forall\, \boldsymbol{d} \in \mathbb{R}^n \backslash \mathbf{0}$.

These three tests are equivalent; use whatever one is easiest!

Sylvester's criterion (Symmetric)

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{12} & A_{22} & A_{23} & \cdots & A_{2n} \\ A_{13} & A_{23} & A_{33} & \cdots & A_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & A_{3n} & \cdots & A_{nn} \end{pmatrix} \succ \mathbf{0}$$

$$A_{11} > 0$$

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{vmatrix} > 0$$

$$\ldots$$

All upper left must be positive!

Easy tests to know if a matrix is Positive Semidefinite?

Symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is positive semidefinite if & only if ...

- All n eigenvalues are nonnegative;
- Sylvester's criterion: all principal minors are nonnegative;
- $\boldsymbol{d}^\top \boldsymbol{A}\boldsymbol{d} \geq 0, \ \forall\, \boldsymbol{d} \in \mathbb{R}^n$.

These three tests are equivalent; use whatever one is easiest!

Sylvester's criterion (Symmetric)

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{12} & A_{22} & A_{23} & \cdots & A_{2n} \\ A_{13} & A_{23} & A_{33} & \cdots & A_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & A_{3n} & \cdots & A_{nn} \end{pmatrix} \succeq \mathbf{0}$$

$$A_{11} \geq 0$$

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{vmatrix} \geq 0$$

$$\ldots$$

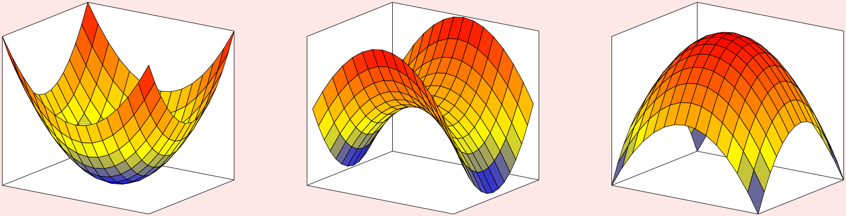
All principal minors must be nonnegative!

Example Matrices

A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is ...

Positive semidefinite	$\boldsymbol{d}^\top \boldsymbol{A}\boldsymbol{d} \geq 0, \; \forall\, \boldsymbol{d} \in \mathbb{R}^n$	$\boldsymbol{A} \succeq \mathbf{0}$
Positive definite	$\boldsymbol{d}^\top \boldsymbol{A}\boldsymbol{d} > 0, \; \forall\, \boldsymbol{d} \in \mathbb{R}^n \backslash \mathbf{0}$	$\boldsymbol{A} \succ \mathbf{0}$
Negative semidefinite	$-\boldsymbol{A}$ is PSD	$\boldsymbol{A} \preceq \mathbf{0}$
Negative definite	$-\boldsymbol{A}$ is PD	$\boldsymbol{A} \prec \mathbf{0}$
Indefinite	Neither PSD nor NSD	

Sanity Check: Definiteness of matrix?



Second Order **Necessary** Condition

Theorem

Let $\Omega \subset \mathbb{R}^n$, and $f \in \mathcal{C}^2$, $\boldsymbol{x}^*$ be a local minimiser of f over Ω and $\boldsymbol{d}$ be a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then:

$$\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} \geq 0,$$

where $\nabla^2 f$ is the Hessian matrix of f.

Given a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $\boldsymbol{x} \in \mathbb{R}^n$, recall the Hessian, $\boldsymbol{H}(\boldsymbol{x})$, the matrix of second partial derivatives

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{H}(\boldsymbol{x}) \triangleq \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{x}) \end{pmatrix}$$

Second Order **Necessary** Condition

Proof.

Suppose, to get a contradiction, that there is a feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$ such that $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, but $\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*)\,\boldsymbol{d} < 0$. Let $\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha\,\boldsymbol{d}$ and define the composite function $\phi(\alpha) = f(\boldsymbol{x}^* + \alpha\,\boldsymbol{d}) = f(\boldsymbol{x}(\alpha))$. Then by Taylor's theorem,

$$\phi(\alpha) = \phi(0) + \phi''(0)\frac{\alpha^2}{2} + r(\alpha).$$

Note that we have used the assumption that $\phi'(\alpha) = \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$ and $\phi''(0) = \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*)\,\boldsymbol{d}$. Since $\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*)\,\boldsymbol{d} < 0$, it follows that if α is sufficiently small,

$$\phi(\alpha) - \phi(0) = \phi''(0)\frac{\alpha^2}{2} + r(\alpha) < 0,$$

implying $f(\boldsymbol{x}^* + \alpha\,\boldsymbol{d}) < f(\boldsymbol{x}^*)$, which contradicts that $\boldsymbol{x}^*$ is a local minimiser. $\square$

Second Order **Necessary** Condition (Interior Case)

Corollary

Let $\boldsymbol{x}^*$ be an interior point of Ω. If $\boldsymbol{x}^*$ is a local minimiser of $f : \Omega \to \mathbb{R}$ and $f \in \mathcal{C}^2$, then

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0},$$

and the Hessian of f is positive semidefinite at the point $\boldsymbol{x}^*$,

$$\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} \geq 0 \quad \forall\, \boldsymbol{d}$$

Proof.

This first part follows is just the first order condition for the interior case, and the second part follows from the fact that if $\boldsymbol{x}^*$ is interior then all directions are feasible. □

Second Order Necessary Condition Example

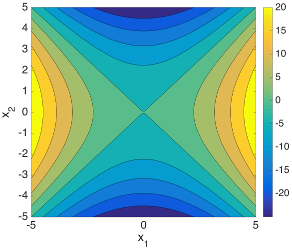
$f(\boldsymbol{x}) = x_1^2 - x_2^2$

Are the FONC & SONC satisfied at $\boldsymbol{x} = [0,\ 0]^\top$?
The point $\boldsymbol{x} = [0,\ 0]^\top$ satisfies the FONC:

$$\nabla f(\boldsymbol{x}) = \left[\frac{df}{dx_1},\ \frac{df}{dx_2}\right]^\top = [2x_1,\ -2x_2]^\top = \mathbf{0}.$$

But the Hessian (matrix of second derivatives) is:

$$H(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

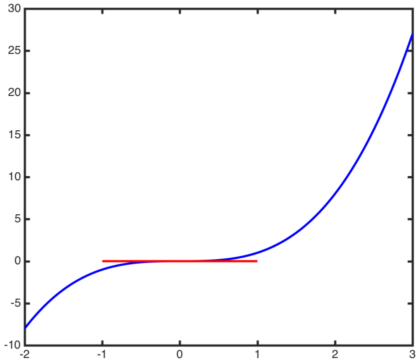


For $\boldsymbol{d} = [1,\ 0]^\top$, $\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{d} > 0$. But for $\boldsymbol{d} = [0,\ 1]^\top$, $\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{d} < 0$.
So the SONC is not satisfied, and hence $\boldsymbol{x} = [0,\ 0]^\top$ is not a minimiser.

Second Order **Necessary** Condition

Sanity

Is the second order necessary condition sufficient for optimality?



Consider the one dimensional function $f(x) = x^3$, then:

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

thus the FONC and SONC are satisfied at $x = 0$ (why?). But $x = 0$ is not a minimiser.

Second Order **Sufficient** Condition (Interior Case)

Theorem

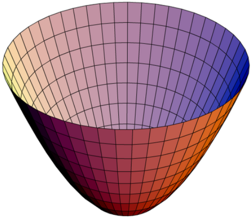
Suppose that $f \in \mathcal{C}^2$ in a region where $\boldsymbol{x}^*$ is an interior point. Suppose that,

1. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.
2. $\nabla^2 f(\boldsymbol{x}^*) \succ \mathbf{0}$, *i.e., the Hessian is positive definite at the point $\boldsymbol{x}^*$.*

Then $\boldsymbol{x}^*$ is a strict local minimiser of f.

Sanity Check

Cases where the second order sufficient condition misses a local (or even a global!!) minimum?



Example of using FONC & SOSC

$f(\boldsymbol{x}) = x_1^2 + x_2^2$ at the point $\boldsymbol{x} = \mathbf{0}$?

We have $\nabla f(\boldsymbol{x}) = [2x_1,\, 2x_2]^\top = \mathbf{0}$ if $\boldsymbol{x} = \mathbf{0}$. For all $\boldsymbol{x}$ we have (why?),

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \succ \mathbf{0}$$

Therefore, the point $[0, 0]$ satisfies the first order necessary and sufficient conditions for a local minimum (in fact it is a strict global minimum).