Optimization (2MMD10/2DME20), lecture 2

Gerhard Woeginger

Technische Universiteit Eindhoven

Fall 2015, Q1

Program for this week

Convexity, convexity, convexity, ...

- Positive semi-definite matrices and functions
- Convex sets
- Convex functions
- Applications of convexity
- Useful inequalities

Positive semi-definite matrices (1)

Definition

A real matrix A is symmetric if $A^T = A$.
The set of symmetric $n \times n$ matrices is denoted by S^n.

Recall:

Theorem

For any matrix $A \in S^n$,
there exists an $n \times n$ matrix F and a diagonal matrix Λ
so that $F^T F = I$ and $F^T A F = \Lambda$.

Let $\lambda_1, \ldots, \lambda_n$ be the diagonal entries of Λ, and let $f_1, \ldots, f_n$ be the columns of F. Then

- $f_1, \ldots, f_n$ is an orthonormal basis of $\mathbb{R}^n$.
- $Af_i = \lambda_i f_i$ for all i.
- $A = \lambda_1 f_1 f_1^T + \cdots + \lambda_n f_n f_n^T$.

Positive semi-definite matrices (2)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is positive semi-definite
if $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

If $A \in S^n$, then the function $f(x) = x^T Ax = \sum_i \sum_j A_{ij} x_i x_j$
is a homogeneous quadratic function ($f : \mathbb{R}^n \to \mathbb{R}$).

Definition

Let $A \in S^n$. Then A is positive semi-definite (PSD) if

$$x^T Ax \geq 0 \text{ for all } x \in \mathbb{R}^n.$$

The set of positive semi-definite matrices is denoted by S^n_+.

- An $A \in S^n$ is positive definite (PD) if A is PSD and non-singular.
- The set of positive definite matrices is denoted by S^n_{++}.
- We write $A \succeq 0$ to denote that A is PSD, and $A \succ 0$ if A is PD.

Positive semi-definite matrices (3)

Most results on PSD matrices in this course are derived from this one:

Theorem

Let $A \in S^n$. The following three statements are equivalent:

1. *A is positive semi-definite.*
2. *each eigenvalue of A is ≥ 0.*
3. *there is some real matrix Z such that $A = Z^T Z$.*

In particular,

- A is PSD $\Longrightarrow \det(A) \geq 0$
- A is PSD $\Longrightarrow$ the diagonal entries of A are ≥ 0
- if A is diagonal, then: A is PSD $\Longleftrightarrow$ diagonal entries of A are ≥ 0
- if $A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$, then: A is PSD $\Longleftrightarrow$ both B and C are PSD.

Positive semi-definite matrices (4)

Matrices $A, B \in S^n$ are congruent,
if $B = U^T A U$ for some non-singular U.

Lemma

Let $A, B \in S^n$ be congruent. Then $A \succeq 0 \Longleftrightarrow B \succeq 0$.

Applying one or more of the following symmetric matrix operations to A yields a congruent matrix:

- scaling the i-th row and the i-the column by a $\lambda \neq 0$
- interchanging the i-th row with the j-th row and the i-th column with the j-th column
- adding $\lambda \times$ the i-th row to the j-th row and adding $\lambda \times$ the i-th column to the j-th column

By these operations, a matrix A may be transformed to a congruent diagonal matrix D. Then, $A \succeq 0 \Leftrightarrow D \succeq 0 \Leftrightarrow D \geq 0$.

Linear, affine, convex sets (1)

Definition

Let $x, y \in \mathbb{R}^n$, $\alpha, \beta \in \mathbb{R}$.
Then $z := \alpha x + \beta y$ is a linear combination of x and y.

- z lies on the plane through $0, x, y$
- if $\alpha + \beta = 1$, then z lies on the line through x, y
- if in addition $\alpha, \beta \geq 0$, then z lies between x and y

Linear, affine, convex sets (2)

Definition

A set $L \subseteq \mathbb{R}^n$ is

- linear, if $\alpha x + \beta y \in L$ for all $x, y \in L$ and all $\alpha, \beta \in \mathbb{R}$
- affine, if $\alpha x + \beta y \in L$ for all $x, y \in L$ and $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1$

Theorem

Let $L \subseteq \mathbb{R}^n$. The following are equivalent:

- *L is affine*
- *$L = \{x \mid Ax = b\}$ for some A, b*
- *$L = \{Cx + d \mid x\}$ for some C, d*

Linear, affine, convex sets (3)

Definition

A set $C \subseteq \mathbb{R}^n$ is
convex, if $\alpha x + \beta y \in C$ for all $x, y \in C$ and $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$

Example

- affine sets are convex.
- a hyperplane $H_{a,b} := \{x \in \mathbb{R}^n \mid a^T x = b\}$ is convex
- a halfspace $H^{\leq}_{a,b} := \{x \in \mathbb{R}^n \mid a^T x \leq b\}$ is convex
- the unit ball $B^n := \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ is convex

Linear, affine, convex sets (4)

Definition

A set $C \subseteq \mathbb{R}^n$ is a cone, if $\alpha x + \beta y \in C$ for all $x, y \in C$ and all $\alpha, \beta \geq 0$.

- Note: cones are convex sets.

Example

- linear sets are cones
- the Lorentz cone $L^{n+1} := \{(x, t) \mid x \in \mathbb{R}^n, t \in \mathbb{R}, \|x\| \leq t\}$ is a cone
- the positive semi-definite (PSD) matrices

$$S_+^n := \{A \in S^n \mid A \succeq 0\}$$

  form a cone

Linear, affine, convex sets (5)

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is a norm if

- $f(x) \geq 0$ for all $x \in \mathbb{R}^n$
- $f(x) = 0 \iff x = 0$
- $f(\lambda x) = \lambda f(x)$ for all $\lambda \in \mathbb{R}^+$, $x \in \mathbb{R}^n$
- $f(x + y) \leq f(x) + f(y)$ for all $x, y \in \mathbb{R}^n$

Definition

If f is a norm,

- then the norm ball is $\{x \in \mathbb{R}^n \mid f(x) \leq 1\}$
- and the norm cone is $\{(x, t) \in \mathbb{R}^{n+1} \mid f(x) \leq t\}$.

For any norm, the norm ball is a convex set and the norm cone is a cone.

Making convex sets (1)

Intersection of convex sets:

Lemma

Let $C_\alpha \subseteq \mathbb{R}^n$ be convex for all $\alpha \in A$.
Then $\bigcap_{\alpha \in A} C_\alpha$ is convex.

Example

The set of copositive polynomials of degree n:

$$P_+^n := \{(p_0, \ldots, p_n) \mid 0 \leq p_0 + p_1 x + \cdots + p_n x^n \text{ for all } x \in [0, \infty)\}$$

can be written as $P_+^n = \bigcap_{x \in [0,\infty)} P_x^n$, where

$$P_x^n := \{(p_0, \ldots, p_n) \mid 0 \leq p_0 + p_1 x + \cdots + p_n x^n\}.$$

Each P_x^n is a halfspace in $\mathbb{R}^{n+1}$, hence P_+^n is convex.

Making convex sets (2)

Polyhedra:

Definition

A polyhedron is
a set $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ for some linear inequalities $Ax \leq b$.

- Example: the n-simplex $\{x \in \mathbb{R}^n \mid x \geq 0, \sum x_i = 1\}$
- Polyhedra are convex sets

Making convex sets (3a)

Balls and Ellipsoids:

Example

The unit ball $B^n = \{x \in \mathbb{R}^n | \ ||x|| \leq 1\}$ is convex.

Definition

Let Z be a non-singular $n \times n$ matrix; let $c \in \mathbb{R}^n$.
Then $E(Z, c) := \{c + Zx \mid \|x\| \leq 1\}$ is an ellipsoid.

So ellipsoids are scaled, rotated and shifted balls.

Lemma

A set $E \subseteq \mathbb{R}^n$ is an ellipsoid if and only if

$$E = \{y \in \mathbb{R}^n \mid (y - c)^T A^{-1}(y - c) \leq 1\}$$

for some $c \in \mathbb{R}^n$ and some positive definite A.

Making convex sets (3b)

Balls and Ellipsoids:

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be affine (i.e. $f : x \mapsto Ax + b$); let $C \subseteq \mathbb{R}^n$ be convex.
Then $f[C] := \{f(x) \mid x \in C\}$ *is convex.*

Example

Consider an ellipsoid $E(Z, c) = \{Zx + c \mid \|x\| \leq 1\}$.
For $f : x \mapsto Zx + c$, we have $E(Z, c) = f[B^n]$.
Hence ellipsoids are convex sets.

Making convex sets (4)

Convex hulls:

Definition

Let $a_1, \dots, a_m \in \mathbb{R}^n$. Let $\lambda_1, \dots, \lambda_m \geq 0$ and $\sum_i \lambda_i = 1$.
Then $\lambda_1 a_1 + \cdots + \lambda_m a_m$ is a convex combination of $a_1, \dots, a_m$.

Definition

The convex hull of $a_1, \dots, a_m$ is

$$\text{conv}\{a_1, \dots, a_m\} := \{\sum_i \lambda_i a_i \mid \sum_i \lambda_i = 1, \lambda_i \geq 0\}.$$

For the affine function $f : \lambda \mapsto \sum_i \lambda_i a_i$ and
for the convex set $C := \{\lambda \mid \sum_i \lambda_i = 1, \lambda_i \geq 0\}$,
we have $\text{conv}\{a_1, \dots, a_m\} = f[C]$.
Hence $\text{conv}\{a_1, \dots, a_m\}$ is convex.

Making convex sets (5)

Inverse image of an affine function:

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be affine (i.e. $f : x \mapsto Ax + b$); let $C \subseteq \mathbb{R}^m$ be convex. Then $f^{-1}[C] := \{x \in \mathbb{R}^n \mid f(x) \in C\}$ is convex.

Example

Let $A_0, \ldots, A_n \in S^n$. Then the set

$$X := \{x \in \mathbb{R}^n \mid A_0 + x_1 A_1 + \cdots + x_m A_m \succeq 0\}$$

is convex, as $X = f^{-1}[S^n_+]$, where $f : \mathbb{R}^m \to S^n$ is the affine map

$$f : x \mapsto A_0 + x_1 A_1 + \cdots + x_m A_m.$$

More about convex sets (1)

Definition

Let $C \subseteq \mathbb{R}^n$ be convex and non-empty.
A point $x \in C$ is an extreme point of C,
if $x = \lambda x_1 + (1 - \lambda)x_2$ with $x_1, x_2 \in C$ and $0 < \lambda < 1$
implies $x = x_1 = x_2$.

Example

What are the extreme points of
(a) a closed disk in $\mathbb{R}^2$ (b) a convex polygon in $\mathbb{R}^2$?

Lemma

Let $P \subseteq \mathbb{R}^n$ be a polyhedron. Then P has finitely many extreme points.

Krein-Milman theorem

A compact convex set C is the convex hull of its extreme points.

More about convex sets (2)

Theorem

Let $C \subseteq \mathbb{R}^n$ be a closed, convex set. Let $x_0 \notin C$.
Then there exists a nonzero $y \in \mathbb{R}^n$ and a $z \in \mathbb{R}$ such that

$$y^T x > z \text{ for all } x \in C \qquad \text{and} \qquad y^T x_0 < z.$$

Theorem

Let $C, D \subseteq \mathbb{R}^n$ be convex sets with $C \cap D = \emptyset$.
Then there exist a nonzero vector $y \in \mathbb{R}^n$ and a $z \in \mathbb{R}$ such that

$$y^T x \leq z \text{ for all } x \in C \qquad \text{and} \qquad y^T x \geq z \text{ for all } x \in D.$$

The hyperplane $H = \{x \mid y^T x = z\}$ is said to separate C from D.

Theorem

Let $C \subseteq \mathbb{R}^n$ be a convex set, and let x_0 lie on the boundary of C.
Then there exist a nonzero vector $y \in \mathbb{R}^n$ and a $z \in \mathbb{R}$ such that

$$y^T x \leq z \text{ for all } x \in C \qquad \text{and} \qquad y^T x_0 = z.$$

The hyperplane $H = \{x \mid y^T x = z\}$ is said to support C at x_0.

Convex functions (1)

Definition

A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex if

$$f(\alpha x + \beta y) \le \alpha f(x) + \beta f(y)$$

for all $x, y \in \mathbb{R}^n$ and $\alpha, \beta \ge 0$ with $\alpha + \beta = 1$.

- Function f strictly convex: strict inequality if $\alpha, \beta > 0$
- Function f concave: if $-f$ is convex.

Example

- Norm functions are convex
- If $C \subseteq \mathbb{R}^n$ is a convex set, then the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise} \end{cases}$$

is convex.

Convex functions (2)

Definition

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a function. Then the epigraph of f is

$$\text{Epi}(f) := \{(x,t) \mid x \in \mathbb{R}^n, t \in \mathbb{R}, f(x) \le t\}.$$

Theorem

f is a convex function $\iff$ Epi(f) is a convex set.

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be convex and let $\gamma \in \mathbb{R}$.
Then the sublevel set $\{x \in \mathbb{R}^n \mid f(x) \le \gamma\}$ is a convex set.

First-order condition (1)

The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the vector

$$\nabla f = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})^T$$

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function.
Then f is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

for all $x, y \in \mathbb{R}^n$.

For convex f,
$f(x^*) = \min\{f(y) \mid y \in \mathbb{R}^n\} \iff \nabla f(x^*) = 0$

First-order condition (2)

Example

For a matrix $A \in S^n$, a vector $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$,
the quadratic function $f(x) = x^T Ax + bx + c$ is convex
if and only if A is positive semi-definite.

Proof:

- First-order condition $f(y) \geq f(x) + \nabla f(x)^T (y - x)$
- $\nabla f(x) = 2x^T A + b$
- $y^T Ay + by + c \geq x^T Ax + bx + c + (2x^T A + b)(y - x)$
  is equivalent to $(y - x)^T A(y - x) \geq 0$

Well-known special case

For $a, b, c \in \mathbb{R}$, the univariate quadratic function $f(x) = ax^2 + bx + c$ is convex if and only if $a \geq 0$.

Second-order condition (1)

The Hessian of a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is a symmetric matrix $\nabla^2 f \in S^n$ such that

$$(\nabla^2 f)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Example

Let $Q \in S^n$ and let c be a vector.
Then the Hessian of $f(x) = \frac{1}{2}x^T Q x + c^T x$ is $\nabla^2 f(x) = Q$.

Second-order condition (2)

Recall: a function is analytic, if it has a Taylor series for each point x in its domain that converges to the function in an open neighborhood of x.

Univariate case

For univariate analytic functions $f : \mathbb{R} \to \mathbb{R}$ we have:

$$f(x+h) \;=\; f(x) + f'(x)h + \frac{1}{2}f''(x+\alpha h)h^2$$

for some $\alpha \in [0,1]$.

Multivariate case

For multivariate analytic functions $f : \mathbb{R}^n \to \mathbb{R}$ we have:

$$f(x+h) \;=\; f(x) + \nabla f(x)^T h + \frac{1}{2}h^T \nabla^2 f(x+\alpha h)h$$

for some $\alpha \in [0,1]$.

Second-order condition (3)

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function.
Then f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^n$.

The proof uses:

Lemma

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex,
if and only if $g(\lambda) := f(x + \lambda d)$ is convex for all $x, d \in \mathbb{R}^n$.

Examples

- $x \mapsto \exp(ax)$ is convex on $\mathbb{R}$, for all $a \in \mathbb{R}$
- $x \mapsto x^a$ is convex on $\mathbb{R}^+$, for all $a \leq 0$ and $a \geq 1$ (concave otherwise)
- $x \mapsto \log(x)$ is concave on $\mathbb{R}^+$
- $x \mapsto x \log(x)$ is convex on $\mathbb{R}^+$
- $(x, y) \mapsto \frac{x^2}{y}$ is convex on $\{(x, y) \mid y > 0\}$
- $x \mapsto \log(\sum_i \exp(x_i))$ is convex on $\mathbb{R}^n$
- $x \mapsto (\prod_i x_i)^{1/n}$ is concave on $\mathbb{R}^n_+$

Operations that preserve convexity

Lemma

If $f, g : \mathbb{R}^n \to \mathbb{R}$ *are convex functions, then so is*

$$\alpha f + \beta g$$

for all $\alpha, \beta \geq 0$

Lemma

If $f, g : \mathbb{R}^n \to \mathbb{R}$ *are convex, then so is* $\max\{f, g\}$.

Lemma

If $f : \mathbb{R}^n \to \mathbb{R}$ *is convex and* $g : \mathbb{R}^m \to \mathbb{R}^n$ *is affine, then* $f \circ g$ *is convex.*

Example

Three lines in space have a unique waist

Given: three fixed lines in space that are made of iron wire;
these lines are pairwise disjoint and pairwise non-parallel.
We stretch an elastic band around the lines, which then by elasticity
will slip to a position where its total circumference is minimal.

Show: final position of band does not depend on initial position.

Mathematical formulation

Let ℓ_1, ℓ_2, ℓ_3 be three lines in $\mathbb{R}^3$.

Find $\min \; f(p_1, p_2, p_3) = ||p_1 - p_2|| + ||p_2 - p_3|| + ||p_3 - p_1||$
such that $p_i \in \ell_i$ for $i = 1, 2, 3$

- $f(p_1, p_2, p_3)$ is strictly convex

Inequalities from convexity

Almost all elementary (and many other) inequalities follow from convexity.

A well-known inequality

$e^z \geq 1 + z$ for all real numbers z.

Proof:

- First-order condition $f(y) \geq f(x) + \nabla f(x)^T (y - x)$
- Choose $f(t) = e^t$, $x = 0$ and $y = z$

Inequalities: Jensen (1)

- Johan Jensen (1859–1925): Danish mathematician and engineer

Theorem (Jensen, 1906)

For a convex function $f : \mathbb{R} \to \mathbb{R}$ and real numbers $x_1, \ldots, x_n$, we have

$$\frac{1}{n} \cdot \sum_{i=1}^{n} f(x_i) \;\geq\; f\left(\frac{1}{n} \sum_{i=1}^{n} x_i\right).$$

- Question: When does equality hold for strictly convex f?
- If f is concave: then the inequality holds with $\leq$ instead of $\geq$

Theorem

For a convex function $f : \mathbb{R} \to \mathbb{R}$ and real numbers $x_1, \ldots, x_n$, and positive real numbers $a_1, \ldots, a_n$, we have

$$\frac{\sum_{i=1}^{n} a_i f(x_i)}{\sum_{i=1}^{n} a_i} \;\geq\; f\left(\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i}\right).$$

Inequalities: Jensen (2)

Theorem

For positive real numbers $a_1, \ldots, a_n$, *we have*

$$(a_1 a_2 \cdots a_n)^{1/n} \;\leq\; \frac{a_1 + a_2 + \cdots + a_n}{n}$$

Proof: let $x_i = \ln a_i$, and use Jensen with $f(x) = e^x$

Inequalities: Jensen (3)

Information theory considers the information content of a system that produces messages m_k with probability p_k,
where $p_1, p_2, \ldots, p_n \geq 0$ and $\sum_{k=1}^{n} p_k = 1$.

The entropy of a probability distribution is defined as

$$H(p) \;=\; -\sum_{k=1}^{n} p_k \log p_k.$$

The entropy satisfies the bound

$$H(p) \;\leq\; \log n.$$

Inequalities: Master (1)

Master inequality

Let $g : \mathbb{R} \to \mathbb{R}$ be a strictly concave function, and let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be the function that is defined by

$$f(x, y) \;=\; y \cdot g\left(\frac{x}{y}\right).$$

Then all real numbers $x_1, \ldots, x_n$ and all positive real numbers $y_1, \ldots, y_n$ satisfy the inequality

$$\sum_{i=1}^{n} f(x_i, y_i) \;\leq\; f\left(\sum_{i=1}^{n} x_i\,, \sum_{i=1}^{n} y_i\right)$$

Equality holds if and only if the two sequences x_i and y_i are proportional (that is, if there exists a real number t such that $x_i/y_i = t$ for all i).

Inequalities: Master (2)

How to prove the master inequality

- by induction on n
- case $n = 1$ holds with equality
- case $n = 2$ holds with $\alpha = y_1/(y_1 + y_2)$ and $\beta = y_2/(y_1 + y_2)$:

$$
\begin{aligned}
f(x_1, y_1) + f(x_2, y_2) &= \\
&= y_1 \cdot g\left(\frac{x_1}{y_1}\right) + y_2 \cdot g\left(\frac{x_2}{y_2}\right) \\
&= (y_1 + y_2)\left\{\frac{y_1}{y_1 + y_2} \cdot g\left(\frac{x_1}{y_1}\right) + \frac{y_2}{y_1 + y_2} \cdot g\left(\frac{x_2}{y_2}\right)\right\} \\
&\leq (y_1 + y_2) \cdot g\left(\frac{x_1 + x_2}{y_1 + y_2}\right) = f\left(x_1 + x_2,\ y_1 + y_2\right).
\end{aligned}
$$

- As g is strictly concave, equality holds if and only if $x_1/y_1 = x_2/y_2$.
- The inductive step for $n \geq 3$ also follows from the inequality.

Inequalities: Master (3)

Cauchy inequality

For real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, we have

$$(a_1^2 + a_2^2 + \cdots + a_n^2)\,(b_1^2 + b_2^2 + \cdots + b_n^2) \;\geq\; (a_1 b_1 + a_2 b_2 + \cdots + a_n b_n)^2.$$

Proof:

use the strictly concave function $g(x) = \sqrt{x}$
set $x_i = a_i^2$ and $y_i = b_i^2$

Homework 2

- Read chapters 2 and 3 in the book of Boyd & Vandenberghe
- Recommended exercises:
  25, 28, 30, 34, 36, 37, 39, 42

Collection of exercises can be downloaded from:
`http://www.win.tue.nl/~gwoegi/optimization/`

Attention!
Weeks 2-5 (Sep 8; Sep 15; Sep 22; Sep 29):

- Tuesday 1+2: instructions
- Tuesday 3+4: lecture