Introduction

• Our goal is to study a class of mathematical optimization problems of the following type:

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, 2, \ldots, m \\
\text{and} & \quad h_i(x) = 0, \quad i = 1, 2, \ldots, p.
\end{align*}
\]

Here, \(x\) represents a vector of decision variables, \(f_0(x)\) is cost function to be minimized and \(f_i(x), h_i(x)\) represent the constraints which the decision variables must observe.

• The goal of optimization is to find an optimal vector \(\hat{x}\) which satisfies \(f_i(\hat{x}) \leq 0, h_i(\hat{x}) = 0\) and minimizes \(f_0\). The class of optimization problems which we are interested in are called convex optimization problems.

Outline

• Introduction to convex optimization
• Convex sets
• Convex functions
• Conditions for convexity
• Quasiconvex functions

Convex optimization problems

These problems are of special interest with OR/ applied mathematics for several reasons:

• The minimum solution in guaranteed to be unique, i.e. there is only one vector \(\hat{x}\) which solves the problem.
• A large number of problems in operations research, signal processing, process control etc can be formulated as convex optimization problems.
• Efficient numerical algorithms exist to solve several special types of convex optimization problems which are of practical importance.
• One can use convex relaxation to find good approximate solutions to many non-convex optimization problems relatively quickly.
Convex optimization: our road-map

We will now look at

- sets over which these problems are defined (convex sets), and
- the classes of functions for which these problems are defined (convex functions).

In subsequent lectures, we will move on to

- Different types of convex optimization problems
- Generic methods for solving some classes of these problems.

Affine sets and convex sets

- A set $C \subseteq \mathbb{R}^n$ is affine if, for any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$, we have $\theta x_1 + (1 - \theta) x_2 \in C$. In other words, the line joining any two points in an affine set $C$ lies entirely in $C$.
- Every affine set may be expressed as the solution set of a system of linear equations, $C = \{x | Ax = b\}$.
- A set $C$ is convex if the line segment between two points $x_1, x_2$ lies entirely in $C$, i.e., if for any $x_1, x_2 \in C$ and for any $\theta \in [0, 1]$, we have $\theta x_1 + (1 - \theta) x_2 \in C$.
- One can move from any point in a convex set $C$ to any other point via an unobstructed path within the set.
- Every affine set is convex, but the converse is not true.

Examples of convex sets I: hyperplanes

- A hyperplane is a set of the form
  $$\{x | a^\top x = b\},$$
  where $a \subseteq \mathbb{R}^n$, $a \neq 0$ and $b \in \mathbb{R}$. Alternatively, the hyperplane may be expressed as
  $$\{x | a^\top (x - x_0) = 0\},$$
  where $x_0$ is any vector such that $a^\top x_0 = b$.
- A hyperplane divides $\mathbb{R}^n$ into two convex half spaces
  $$\{x | \hat{a}^\top x \leq b\},$$
  with $\hat{a} = a$ for one half space and $\hat{a} = -a$ for another half space.
Hyperplanes (continued)

There are two important results in convexity theory related to hyperplanes:

- Suppose that \( C \) and \( D \) are two convex sets which do not intersect, i.e. \( C \cap D = \{\emptyset\} \). Then the **Separating hyperplane theorem** states that there exist \( a \neq 0 \) and \( b \) such that \( a^\top x \leq b \) for all \( x \in C \) and \( a^\top x \geq b \) for all \( x \in D \). In other words, the hyperplane \( \{x | a^\top x = b\} \) separates the two convex sets \( C \) and \( D \).

- Suppose that \( C \in \mathbb{R}^n \) and \( x_0 \) is on the boundary \( \text{bd} C \). If \( a \neq 0 \) satisfies \( a^\top x \leq a^\top x_0 \) for all \( x \in C \), the hyperplane \( \{x | a^\top x = a^\top x_0\} \) is called a supporting hyperplane to \( C \) at \( x_0 \). The **supporting hyperplane theorem** states that for any nonempty convex set \( C \) and any \( x_0 \in \text{bd} C \), there exists a supporting hyperplane for \( C \) at \( x_0 \).

Examples of convex sets II

- An **ellipsoid** is defined by
  \[
  \mathcal{E}(x_c, P) = \{x \mid (x - x_c)^\top P^{-1} (x - x_c) \leq 1\},
  \]
  where \( P \) is a symmetric positive definite matrix, i.e. it is symmetric and has all positive eigenvalues. We will represent this fact by \( P > 0 \). In \( n \)-dimensional space, ellipsoid has semi-axes with length equal to \( \sqrt{\lambda_i} \), where \( \lambda_i \) are the eigenvalues of \( P \).

- A **Euclidean ball** is an ellipsoid with \( P = r^2 I \), where \( I \) is the identity matrix. It represents a sphere in \( n \)-dimensional space with radius \( r \) and center at a point with coordinate vector \( x_c \).

- A **polyhedron** is a solution set (or a feasible set) for a finite number of linear inequalities and equalities:
  \[
  \mathcal{P} = \{x \mid a_j^\top x \leq b_j, j = 1, 2, \ldots, m, c_j^\top x = d_j, j = 1, 2, \ldots, p\}.
  \]

Examples of convex sets III

- A convex set \( C \) is called a **cone** if for any \( x_1, x_2 \in C \) and \( \theta_1, \theta_2 \geq 0 \), we have \( \theta_1 x_1 + \theta_2 x_2 \in C \).

  In \( n \)-dimensional space, a cone has a shape of a pie-slice, with apex at the origin \( (\theta_1 = \theta_2 = 0) \) and passing through points \( x_1 (\theta_2 = 0) \), \( x_2 (\theta_1 = 0) \).

- A **positive semidefinite cone**, which is the set of symmetric positive semidefinite \( n \times n \) matrices:
  \[
  \mathcal{S}_{n^+} = \{X \in \mathbb{R}^{n \times n} | X \geq 0\}.
  \]

  Recall: a symmetric matrix \( A \) is said to be positive semidefinite if \( x^\top Ax \geq 0 \) for all \( x \in \mathbb{R}^n \), which in turn is equivalent to the fact that all the eigenvalues of \( A \) are real and nonnegative.
Examples of convex sets IV

• A hyperbolic set defined by
  \[ \{ x \in \mathbb{R}^2 \mid x_1 x_2 \geq 1 \} \]
  is convex. If \( x, y \) are such that \( \min(x_1 x_2, y_1 y_2) \geq 1 \), one can show that
  \( z_1 z_2 \geq 1 \), where \( z = \theta x + (1 - \theta) y, \theta \in (0,1) \). Proving this if
  \( (x_1 - y_1)(x_2 - y_2) < 0 \) depends on re-arranging \( z_1 z_2 \) as
  \[ (\theta x_1 + (1 - \theta) y_1)(\theta x_2 + (1 - \theta) y_2) = \frac{\theta x_1 x_2 + (1 - \theta) y_1 y_2}{\theta (1 - \theta)(x_1 - y_1)(x_2 - y_2)} - \theta (1 - \theta)(x_1 - y_1)(x_2 - y_2). \]

Operations on convex sets

• An intersection of a finite number of convex sets is always convex (as in
  the case of definition of polyhedron).
• A sum of a finite number of convex sets is convex. Sum of two sets is
  defined by
  \[ S_1 + S_2 = \{ x + y \mid x \in S_1, y \in S_2 \}. \]

Examples of convex functions I

• \( e^{ax} \) is convex on \( \mathbb{R} \), for any \( a \in \mathbb{R} \).
• \( x^a \) is convex on \( \mathbb{R}_+ \) (positive real line), if \( a \in (-\infty, 0] \cup [1, \infty) \).
• \( |x|^p, p \geq 1 \) is convex on \( \mathbb{R} \).
• \( -\log x \) is convex on \( \mathbb{R}_+ \).
• Every norm on \( \mathbb{R}^n \) is convex (by virtue of triangle inequality and
  homogeneity).
• A function defined by \( f(x) = \max \{ x_1, x_2, \ldots, x_n \} \) is convex on \( \mathbb{R}^n \).

Convex functions

• A function \( f : \mathbb{R}^n \to \mathbb{R} \) is convex if its domain is a convex set and if for
  all \( x, y \in \text{dom } f \) and \( \theta \in [0,1] \), we have
  \[ f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y). \] (1)
• We say that a function is strictly convex if a strict inequality holds in (1).
  A function \( f \) is concave (respectively, strictly concave) if \( -f \) is convex
  (respectively, strictly convex).
• An affine function (i.e. a function of the form \( f(x) = Ax + b \)) is both
  convex and concave, since the inequality in (1) is replaced by an equality.
Examples of convex functions II

- A function defined by
  \[ f(x) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n}) \]
  is convex over \( \mathbb{R}^n \). This function is an analytic approximation to the maximum eigenvalue of a symmetric matrix, e.g. \( \max \{ x_1, x_2, \ldots, x_n \} \in dom(f) \implies f(x) \leq \log(n) \) holds.

- A quadratic function given by
  \[ f(x) = \frac{1}{2} x^\top A x + b^\top x + c, \]
  with \( A \) being a symmetric matrix, \( b \in \mathbb{R}^n \) and \( c \in \mathbb{R} \) is convex if and only if \( A \geq 0 \).

Operations which preserve convexity II

- Nonnegative weighted sum of convex functions is convex, i.e. if \( f_i \), \( i = 1, 2, \ldots, n \) are convex, then \( \sum_i w_i f_i \) is also convex if \( w_i \geq 0 \), \( i = 1, 2, \ldots, n \).

- If \( g(x) \) is convex on \( \mathbb{R}^n \), so is \( \exp(g(x)) \). If \( g(x) \) is convex and nonnegative, \( (g(x))^p \) is convex for \( p \geq 1 \).

- If \( f : \mathbb{R}^n \to \mathbb{R} \) is convex, so is \( g : \mathbb{R}^k \to \mathbb{R} \) defined by \( g(x) = f(Ax + b) \), where \( A \in \mathbb{R}^{n \times m} \), \( b \in \mathbb{R}^n \) and
  \[ \text{dom}(g) = \{ x | Ax + b \in \text{dom}(f) \} . \]

Operations which preserve convexity III

- If \( f_1, f_2, \ldots, f_n \) are convex, then so is \( f(x) = \max \{ f_1(x), f_2(x), \ldots, f_n(x) \} \), where \( \text{dom}(f) = \text{dom}(f_1) \cap \cdots \cap \text{dom}(f_n) \).

- Pointwise supremum of a family of convex functions is always convex, e.g. the maximum eigenvalue of a symmetric matrix,
  \[ f : \mathbb{R}^{n \times n} \to \mathbb{R}; f(X) = \sup \{ y^\top X y | \|y\|_2 = 1 \} \]
  is convex in \( X \). Conversely, a pointwise infimum of concave functions is concave; a fact which will prove useful when we study duality.

- If \( g : \mathbb{R} \to \mathbb{R} \) is convex and non-negative, so is \( h = (\sum_{i=1}^n g(x_i))^p \) for any \( p \geq 1 \).

Examples of convex functions II

- A function defined by
  \[ f(x) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n}) \]
  is convex over \( \mathbb{R}^n \). This function is an analytic approximation to the max function, since
  \[ \max \{ x_1, x_2, \ldots, x_n \} \leq f(x) \leq \max \{ x_1, x_2, \ldots, x_n \} + \log n \] holds.

- A quadratic function given by
  \[ f(x) = \frac{1}{2} x^\top A x + b^\top x + c, \]
  with \( A \) being a symmetric matrix, \( b \in \mathbb{R}^n \) and \( c \in \mathbb{R} \) is convex if and only if \( A \geq 0 \).

Operations which preserve convexity II

- Nonnegative weighted sum of convex functions is convex, i.e. if \( f_i \), \( i = 1, 2, \ldots, n \) are convex, then \( \sum_i w_i f_i \) is also convex if \( w_i \geq 0 \), \( i = 1, 2, \ldots, n \).

- If \( g(x) \) is convex on \( \mathbb{R}^n \), so is \( \exp(g(x)) \). If \( g(x) \) is convex and nonnegative, \( (g(x))^p \) is convex for \( p \geq 1 \).

- If \( f : \mathbb{R}^n \to \mathbb{R} \) is convex, so is \( g : \mathbb{R}^m \to \mathbb{R} \) defined by \( g(x) = f(Ax + b) \), where \( A \in \mathbb{R}^{n \times m} \), \( b \in \mathbb{R}^n \) and
  \[ \text{dom}(g) = \{ x | Ax + b \in \text{dom}(f) \} . \]
Conditions for convexity I

- Suppose that $f$ is differentiable over its (open) domain, $\text{dom}(f)$. Then $f(x)$ is convex if and only if $\text{dom}(f)$ is convex and
  
  $$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

  holds for all $x, y \in \text{dom}(f)$. Note that the right hand side of the inequality is the first order Taylor approximation of $f$ in the neighbourhood of $x$.

- For a convex function, the above inequality states that a first order Taylor approximation always underestimates $f(y)$ irrespective of how near or far $y$ is from $x$ (in terms of appropriate metric).

Conditions for convexity II

- As a special case,
  
  $$\text{if } \nabla f(x^*) = 0 \text{ for some } x^* \in \text{dom}(f) \iff f(x^*) = \min_{x \in \text{dom}(f)} f(x).$$

  Minimising a convex differentiable function on its domain is equivalent to finding a point where its gradient is zero.

- Suppose that $f$ is twice differentiable, i.e. the Hessian matrix $\nabla^2 f(x) = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j} \right]$ exists at each point in $\text{dom}(f)$. Then $f$ is convex if and only if $\text{dom}(f)$ is convex and its Hessian is positive semidefinite for all $x \in \text{dom}(f)$, i.e. $\nabla^2 f(x) \succeq 0$.

- For twice differentiable $f : \mathbb{R} \rightarrow \mathbb{R}$, this means that the slope of tangent to $f$ is always increasing as $x$ increases.

Conditions for convexity III

- How do we know if a function is convex?
  
  - Use definition, or prove from first principles.
  
  - If it is differentiable, check if $f(y) \geq f(x) + \nabla f(x)^T(y - x)$ holds for all $x, y \in \text{dom}(f)$ ($1^{\text{st}}$ order characterization).
  
  - If it is twice differentiable, check if $\nabla^2 f(x) \succeq 0$ holds ($2^{\text{nd}}$ order characterization).
  
  - Check if you can construct it from more elementary convex functions (e.g., pointwise maximum, affine translation, non-negative weighted sum etc).
  
  - ··· or, try $0^{\text{th}}$ order characterization.
$0^{\text{th}}$ order characterization for convexity

- A function $f(x)$ is convex if and only if $g(t) = f(x + tv)$ is convex in $t$, where $\text{dom}(g) = \{t | x + tv \in \text{dom}(f), x \in \text{dom}(f), v \in \mathbb{R}^n\}$.
- This allows checking convexity for $f(x)$ by checking convexity of a scalar function $g(t)$.
- Example: $f(X) = -\log \det(X)$, $\text{dom}(f) = \{X \in \mathbb{R}^{n \times n}, X > 0\}$. Then $g(t) = -\log \det(X + tV) = -\log \det(X) - \log \det(I + tX^{-0.5}VX^{-0.5}) = -\log \det(X) - \sum t \lambda_i$, where $\lambda_i$ are eigenvalues of $X^{-0.5}VX^{-0.5}$. $g(t)$ is convex as $g''(t) > 0$; hence so is $f(X)$.

Quasiconvex functions II: some examples

- Given a cash-flow $x_0 < 0$, $x_0 + x_1 + \cdots + x_n > 0$, internal rate of return $\text{IRR}(x)$ is defined by

$$\text{IRR}(x) = \inf \left\{ r \left| \sum_{i=0}^{n} \frac{x_i}{(1+r)^i} = 0 \right. \right\}.$$  

$\text{IRR}(x)$ is quasiconcave; the superlevel sets $\text{IRR}(x) \geq \alpha$ are convex for each $\alpha$ $\text{IRR} \geq \alpha$ means $\sum_{i=0}^{n} x_i (1+r)^{-i} \geq 0$ for $r \in [0, \alpha]$.
- A function $f(x) = p(x)/q(x)$ is quasiconvex over $\{x | x \in \text{dom}(q) \cap \text{dom}(p), q(x) > 0\}$ whenever $p$ is convex, $q$ is affine. Note that $f(x) \leq \alpha \Leftrightarrow p(x) - \alpha q(x) \leq 0$, so that all sublevel sets of $f(x)$ are convex.

Quasiconvex functions I: definitions

- A function $f$ is quasiconvex if its domain and its sublevel sets,

$$S_\alpha := \{x \in \text{dom}(f) \mid f(x) \leq \alpha\}$$

are convex for $\alpha \in \mathbb{R}$.
- All convex functions are quasiconvex, but converse is not true. On $\mathbb{R}$, all monotonic functions (increasing or decreasing) are also quasiconvex; this includes many concave (e.g. $\log x$ over $\mathbb{R}_+$) functions.
- If $f$ is quasiconvex, $-f$ is quasiconcave. Superlevel sets of quasiconcave functions are convex.
- A function is quasiconvex if and only if

$$f(\theta x + (1-\theta)y) \leq \max(\theta f(x), (1-\theta)f(y)). \quad (2)$$

Quasiconvex functions III: one more example

- Given $a, b \in \mathbb{R}^n$, distance ratio function

$$f(x) = \frac{||x - a||_2}{||x - b||_2}$$

is quasiconvex over domain $\{x | ||x - a||_2 \leq ||x - b||_2\}$. 
Next steps

• Now that we know what convex sets and convex/quasiconvex functions are,
• we are now ready to look at different types of convex and quasiconvex optimization problems.
• Main reference (for this lecture and the next two lectures): *Convex Optimization*, by Stephen Boyd and Lieven Vandenberghe, Cambridge University Press, 2009 (available as a free download online).

A couple of exercises

• State whether the following function is convex or concave (or neither):

\[ f(x) = \frac{1}{x_1 x_2}, \quad x \in \mathbb{R}_+^2 \]

• Suppose that \( f : \mathbb{R} \mapsto \mathbb{R} \) is differentiable and convex with \( \mathbb{R}_+ \subseteq \text{dom } f \).
Show that the *running average* function defined by

\[ F(x) = \frac{1}{x} \int_0^x f(t)dt, \quad \text{dom } F = \mathbb{R}_+ \]

is convex. Hint: use the first order condition for convexity of \( f(x) \), to verify the second order condition of convexity for \( F(x) \); note that

\[ F'(x) = -\frac{1}{x^2} \int_0^x f(t)dt + \frac{f(x)}{x} \]