Foundations of convexity

Dr Paresh Date

Brunel University, UK

June 2016

Outline

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

Introduction

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

$$\begin{aligned} &\text{minimize } f_0(\mathbf{x}) \\ &\text{subject to } f_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m \\ &\text{and } h_i(\mathbf{x}) = 0,\ i = 1, 2, \ldots, p. \end{aligned}$$

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

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

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 $\hat{\mathbf{x}}$ 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 $\mathcal{C} \in \mathbb{R}^n$ is *affine* if, for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$ and $\theta \in \mathbb{R}$, we have $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{C}$. In other words, the line joining any two points in an affine set $\mathcal{C}$ lies entirely in $\mathcal{C}$.
- Every affine set may be expressed as the solution set of a system of linear equations, $\mathcal{C} = \{\mathbf{x}|A\mathbf{x} = \mathbf{b}\}$.
- A set $\mathcal{C}$ is *convex* if the line segment between two points $\mathbf{x}_1, \mathbf{x}_2$ lies entirely in $\mathcal{C}$, i.e., if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$ and for any $\theta \in [0,1]$, we have $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{C}$.
- One can move from any point in a convex set $\mathcal{C}$ to any other point via an unobstructed path within the set.
- Every affine set is convex, but the converse is not true.

Convex hull

- A point of the form $\sum_{i=1}^k \theta_i\mathbf{x}_i$, $\sum_{i=1}^k \theta_i = 1$, $\theta_i \geq 0$ is called a *convex combination* of points $\mathbf{x}_i$, $i = 1, 2, \ldots, k$. For a (not necessarily convex) set $\mathcal{C}$, the set of convex combinations of all its points is called the *convex hull* of $\mathcal{C}$, denoted by $\mathrm{conv}\,\mathcal{C}$:

$$\mathrm{conv}\,\mathcal{C} := \left\{ \sum_{i=1}^k \theta_i\mathbf{x}_i \,|\, \mathbf{x}_i \in \mathcal{C}, \theta_i \geq 0, \sum_{i=1}^k \theta_i = 1 \right\}.$$

Examples of convex sets I: hyperplanes

- A *hyperplane* is a set of the form

$$\left\{\mathbf{x} \,|\, \mathbf{a}^\top\mathbf{x} = b\right\},$$

where $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq 0$ and $b \in \mathbb{R}$. Alternatively, the hyperplane may be expressed as

$$\left\{\mathbf{x}|\mathbf{a}^\top(\mathbf{x} - \mathbf{x}_0) = 0\right\},$$

where $\mathbf{x}_0$ is any vector such that $\mathbf{a}^\top\mathbf{x}_0 = b$.

- A hyperplane divides $\mathbb{R}^n$ into two convex *half spaces*:

$$\left\{\mathbf{x} \,|\, \hat{\mathbf{a}}^\top\mathbf{x} \leq b\right\},$$

with $\hat{\mathbf{a}} = \mathbf{a}$ for one half space and $\hat{\mathbf{a}} = -\mathbf{a}$ for another half space.

Hyperplanes (continued)

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

- Suppose that $\mathcal{C}$ and $\mathcal{D}$ are two convex sets which do not intersect, i.e. $\mathcal{C} \bigcap \mathcal{D} = \{\emptyset\}$. Then *Separating hyperplane theorem* states that there exist $\mathbf{a} \neq 0$ and $\mathbf{b}$ such that $\mathbf{a}^\top \mathbf{x} \leq \mathbf{b}$ for all $\mathbf{x} \in \mathcal{C}$ and $\mathbf{a}^\top \mathbf{x} \geq \mathbf{b}$ for all $\mathbf{x} \in \mathcal{D}$. In other words, the hyperplane $\{\mathbf{x} | \mathbf{a}^\top \mathbf{x} = \mathbf{b}\}$ separates the two convex sets $\mathcal{C}$ and $\mathcal{D}$.
- Suppose that $\mathcal{C} \in \mathbb{R}^n$ and $\mathbf{x}_0$ is on the boundary $\mathrm{bd}\,\mathcal{C}$. If $\mathbf{a} \neq 0$ satisfies $\mathbf{a}^\top \mathbf{x} \leq \mathbf{a}^\top \mathbf{x}_0$ for all $\mathbf{x} \in \mathcal{C}$, the hyperplane $\left\{\mathbf{x} \,|\, \mathbf{a}^\top \mathbf{x} = \mathbf{a}^\top \mathbf{x}_0\right\}$ is called a supporting hyperplane to $\mathcal{C}$ at $\mathbf{x}_0$. The *Supporting hyperplane theorem* states that for any nonempty convex set $\mathcal{C}$ and any $\mathbf{x}_0 \in \mathrm{bd}\,\mathcal{C}$, there exists a supporting hyperplane for $\mathcal{C}$ at $\mathbf{x}_0$.

Examples of convex sets II

- An *ellipsoid* is defined by

$$\mathcal{E}(\mathbf{x}_c, P) = \left\{\mathbf{x} \,|\, (\mathbf{x} - \mathbf{x}_c)^\top P^{-1} (\mathbf{x} - \mathbf{x}_c) \leq 1\right\},$$

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 λ_i are the eigenvalues of P.

- A *Euclidian 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 $\mathbf{x}_c$.
- A *polyhedron* is a solution set (or a feasible set) for a finite number of linear inequalities and equalities:

$$\mathcal{P} = \left\{\mathbf{x} \,|\, \mathbf{a}_j^\top \mathbf{x} \leq \mathbf{b}_j,\, j = 1, 2, \ldots, m,\, \mathbf{c}_j^\top \mathbf{x} = \mathbf{d}_j,\, j = 1, 2, \ldots, p\right\}.$$

Examples of convex sets II (continued)

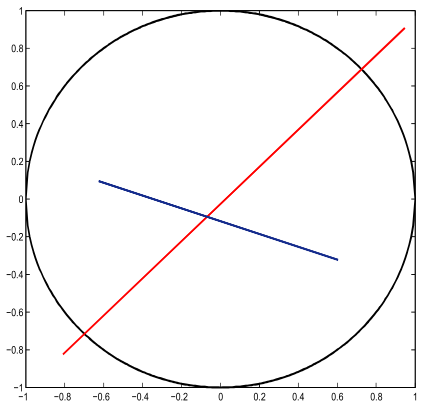

Figure : The *inside* of a circle in $\mathbb{R}^2$ is convex, *outside* isn't

Examples of convex sets III

- A convex set $\mathcal{C}$ is called a *cone* if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$ and $\theta_1, \theta_2 \geq 0$, we have

$$\theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 \in \mathcal{C}.$$

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 $\mathbf{x}_1$ ($\theta_2 = 0$), $\mathbf{x}_2$ ($\theta_1 = 0$).

- A *positive semidefinite cone*, which is the set of symmetric positive semidefinite $n \times n$ matrices:

$$\mathcal{S}_{n+} = \left\{X \in \mathbb{R}^{n \times n} \,|\, X \geq 0\right\}.$$

Recall: a symmetric matrix A is said to be positive semidefinite if $\mathbf{x}^\top A \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, 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

$$\{\mathbf{x} \in \mathbb{R}^2_+ \mid x_1 x_2 \geq 1\}$$

is convex. If $\mathbf{x}, \mathbf{y}$ 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 $\mathbf{z} = \theta\mathbf{x} + (1-\theta)\mathbf{y}$, $\theta \in (0,1)$.

- Proving this if $(x_1 - y_1)(x_2 - y_2) < 0$ depends on re-arranging $z_1 z_2$ as

$$(\theta x_1 + (1-\theta)y_1)(\theta x_2 + (1-\theta)y_2) =$$
$$\underbrace{\{\theta x_1 x_2 + (1-\theta)y_1 y_2\}}_{\geq 1} - \underbrace{\theta(1-\theta)(x_1 - y_1)(x_2 - y_2)}_{\leq 0}.$$

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

$$\mathcal{S}_1 + \mathcal{S}_2 = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{S}_1,\ \mathbf{y} \in \mathcal{S}_2\}.$$

Convex functions

- A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is *convex* if its domain is a convex set and if for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}\, f$ and $\theta \in [0,1]$, we have

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}). \tag{1}$$

- We say that a function is *strictly convex* is 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 $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$) is both convex and concave, since the inequality in (1) is replaced by an equality.

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 $a \in (-\infty, 0] \bigcup [1, \infty)$.
- $|x|^p$, $p \geq 1$ 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 $f(\mathbf{x}) = \max\{x_1, x_2, \ldots, x_n\}$ is convex on $\mathbb{R}^n$.

Examples of convex functions II

- A function defined by

$$f(\mathbf{x}) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$$

is convex over $\mathbb{R}^n$. This function is an analytic approximation to the max function, since

$$\max\{x_1, x_2, \ldots, x_n\} \leq f(\mathbf{x}) \leq \max\{x_1, x_2, \ldots, x_n\} + \log n \text{ holds.}$$

- A quadratic function given by

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{x} + c,$$

with A being a symmetric matrix, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ is convex if and only if $A \geq 0$.

Operations which preserve convexity I

- 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 $\mathbb{R}$, so is $exp(g(x))$. If $g(x)$ is convex and nonnegative, $(g(x))^p$ is convex for $p \geq 1$.
- If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex, so is $g : \mathbb{R}^m \mapsto \mathbb{R}$ defined by $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, where $A \in \mathbb{R}^{n\times m}$, $\mathbf{b} \in \mathbb{R}^n$ and

$$\text{dom}(g) = \{\mathbf{x} | A\mathbf{x} + \mathbf{b} \in \text{dom}(f)\}.$$

Operations which preserve convexity II

- If $f_1, f_2, \ldots, f_n$ are convex, then so is $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})\}$, where $\text{dom}(f) = \text{dom}(f_1) \bigcap \text{dom}(f_2) \cdots \bigcap \text{dom}(f_n)$.
- Pointwise supremum of a family of convex functions is always convex, *e.g.* the maximum eigenvalue of a symmetric matrix,

$$f : \mathbb{R}^{n\times n} \mapsto \mathbb{R},\ f(X) = \sup\{\mathbf{y}^\top X\mathbf{y} \mid \|\mathbf{y}\|_2 = 1\}$$

is convex in X. Conversely, a pointwise infimum of concave functions is concave; a fact which will prove useful when we study duality.

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

Operations which preserve convexity III

- $f(\mathbf{x}) = g_0(g_1(\mathbf{x}), g_2(\mathbf{x}), g_k(\mathbf{x}))$, with $g_i : \mathbb{R}^n \mapsto \mathbb{R}$, $g_0 : \mathbb{R}^k \mapsto \mathbb{R}$, is convex if
  - g_0 convex and nondecreasing in each argument; $g_i, i = 1, 2, \cdots k$ convex OR
  - g_0 convex and nonincreasing in each argument; $g_i, i = 1, 2, \cdots k$ concave.
- Examples:
  - $f(\mathbf{x}) = \max_i\{g_i(\mathbf{x})\}$ is convex if each g_i is convex;
  - $f(\mathbf{x}) = 1/(g(\mathbf{x}))$ is convex if $g(\mathbf{x})$ is positive and concave;
  - $f(\mathbf{x}) = (g(\mathbf{x}))^p$ is convex for $p \geq 1$ if $g(\mathbf{x})$ is non-negative and convex.

Operations which preserve convexity IV

- If $h : \mathbb{R} \mapsto \mathbb{R}$ is concave and nondecreasing, $g : \mathbb{R}^n \mapsto \mathbb{R}$ is concave, $f(\mathbf{x}) = h(g(\mathbf{x}))$ is concave.
- Similarly, if $h : \mathbb{R} \mapsto \mathbb{R}$ is convex and nondecreasing, $g : \mathbb{R}^n \mapsto \mathbb{R}$ is convex, $f(\mathbf{x}) = h(g(\mathbf{x}))$ is convex.
- In both the cases, if $h'(x) \neq 0$, extremum of f and the corresponding extremum of g are attained by the same $\mathbf{x}$.

Conditions for convexity I

- Suppose that f is differentiable over its (open) domain, dom (f). Then $f(\mathbf{x})$ is convex if and only if dom(f) is convex and

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$$

  holds for all $\mathbf{x}, \mathbf{y} \in$ dom (f). Note that the right hand side of the inequality is the first order Taylor approximation of f in the neighbourhood of $\mathbf{x}$.
- For a convex function, the above inequality states that a first order Taylor approximation always *underestimates* $f(\mathbf{y})$ irrespective of how near or far $\mathbf{y}$ is from $\mathbf{x}$ (in terms of appropriate metric).

Conditions for convexity II

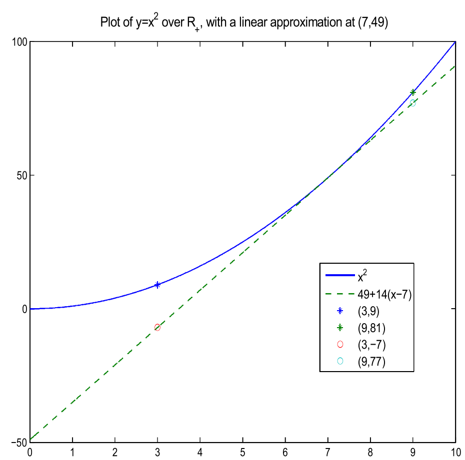

Figure : Illustration of convexity condition over $\mathbb{R}$

Conditions for convexity III

- As a special case,

$$\text{if } \nabla f(\mathbf{x}^\star) = 0 \text{ for some } \mathbf{x}^\star \in \text{dom}(f) \quad \Leftrightarrow \quad f(\mathbf{x}^\star) = \min_{\mathbf{x} \in \text{dom}\,(f)} f(\mathbf{x}).$$

  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 $\nabla^2 f(\mathbf{x}) = [\frac{\partial^2 f}{\partial x_i x_j}]$ exists at each point in dom (f). Then f is convex if and only if dom (f) is convex and its Hessian is positive semidefinite for all $\mathbf{x} \in$ dom (f), *i.e.* $\nabla^2 f(\mathbf{x}) \geq 0$.
- For twice differentiable $f : \mathbb{R} \mapsto \mathbb{R}$, this means that the slope of tangent to f is always increasing as x increases.

How do we know if a function is convex?

- Use definition, or prove from first principles.
- If it is differentiable, check if $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$ holds for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$ (1^{st} order characterization).
- If it is twice differentiable, check if $\nabla^2 f(\mathbf{x}) \geq 0$ holds (2^{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*).
- $\cdots$ or, try 0^{th} order characterization.

0^{th} order characterization for convexity

- A function $f(\mathbf{x})$ is convex if and only if $g(t) = f(\mathbf{x} + t\mathbf{v})$ is convex in t, where $\mathrm{dom}(g) = \{t | \mathbf{x} + t\mathbf{v} \in \mathrm{dom}(f)\}$, $\mathbf{x} \in \mathrm{dom}(f), \mathbf{v} \in \mathbb{R}^n$.
- This allows checking convexity for $f(\mathbf{x})$ by checking convexity of a scalar function $g(t)$.
- Example: $f(X) = -\log det(X)$, $\mathrm{dom}(f) = \{X \in \mathbb{R}^{n\times n}, X > 0\}$. Then

$$\begin{aligned} g(t) &= -\log det(X + tV) = -\log det(X) - \log det(I + tX^{-0.5}VX^{-0.5}) \\ &= -\log det(X) - \sum_i \log(1 + t\lambda_i), \end{aligned}$$

where λ_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 I: definitions

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

$$S_\alpha := \{\mathbf{x} \in \mathrm{dom}(f) \,|\, f(\mathbf{x}) \leq \alpha\}$$

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

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \max\left(\theta f(\mathbf{x}), (1-\theta) f(\mathbf{y})\right). \qquad (2)$$

Quasiconvex functions II: some examples

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

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

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

Quasiconvex functions III: one more example

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

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

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

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).

∘ ∘ ∘