MATH3027: Optimization 2022

Week 6: Convex Sets

Prof. Richard Wilkinson
Please send any comments or mistakes to r.d.wilkinson@nottingham.ac.uk

So far we have only considered unconstrained optimization problems. This week we begin our study of **constrained optimization** by discussing one of the most important concepts in continuous optimization: convexity. Solving general constrained optimization problems is hard[1], whereas in contrast, convex problems can be solved reliably and relatively quickly. Convex problems include linear programmes[2] and constrained least-squares problems, as well as many other types of problem that naturally occur throughout science, engineering, economics etc.

This week, we focus on the properties of convex sets and convex functions, establishing results that will be useful in the forthcoming weeks.

Convex Sets

We begin discussing convexity by defining what a convex set is.

Definition (Convex Set). *A set $C \subseteq \mathbb{R}^n$ is called convex if for any $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0,1]$ the point $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ belongs to C.*

[1] NP hard for those of you who know what that means.

[2] Linear programmes are of the form

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{subject to } & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

In other words, the above definition is equivalent to saying that for any $\mathbf{x}, \mathbf{y} \in C$, the line segment $[\mathbf{x}, \mathbf{y}]$ is also in C. See Figure 1.

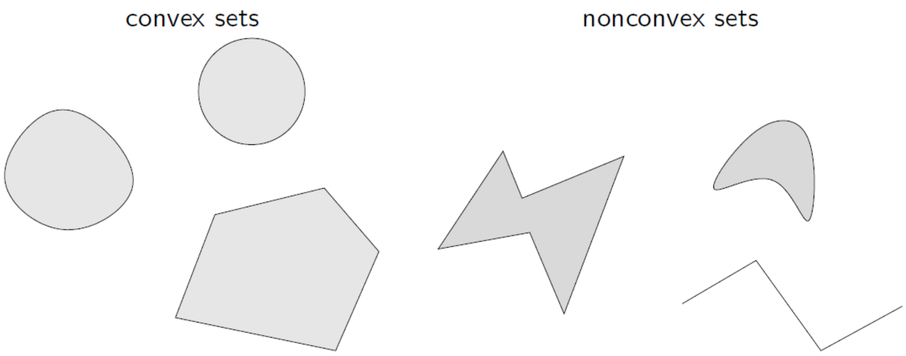

Figure 1: Examples of convex and nonconvex sets.

Important Convex Sets

- $\mathbb{R}^n$
- Lines in $\mathbb{R}^n$, which are of the form

$$L = \{\mathbf{z} + t\mathbf{d} : t \in \mathbb{R}\}$$

  where $\mathbf{z}, \mathbf{d} \in \mathbb{R}^n$ and $\mathbf{d} \neq \mathbf{0}$.

- The open and closed line segments $[\mathbf{x}, \mathbf{y}], (\mathbf{x}, \mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n (\mathbf{x} \neq \mathbf{y})$, and more generally, the open balls

$$B(\mathbf{c}, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\| < r\} \ ,$$

  and closed balls

$$B[\mathbf{c}, r] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\| \leq r\} \ ,$$

  are convex. Note that this is true for an arbitrary norm defined over $\mathbb{R}^n$, and that the shape of the set depends upon the particular norm. See Figure 2.

- Hyperplanes, which are sets of the form

$$H = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b\right\} \quad (\mathbf{a} \in \mathbb{R}^n \backslash \{\mathbf{0}\}, b \in \mathbb{R}) \ .$$

- The associated half-space is the set

$$H^- = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq b\right\} \ .$$

- Ellipsoids, which can be written as

$$E = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^\top \mathbf{Q} \mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c \leq 0\right\} \ ,$$

  where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is positive semidefinite, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

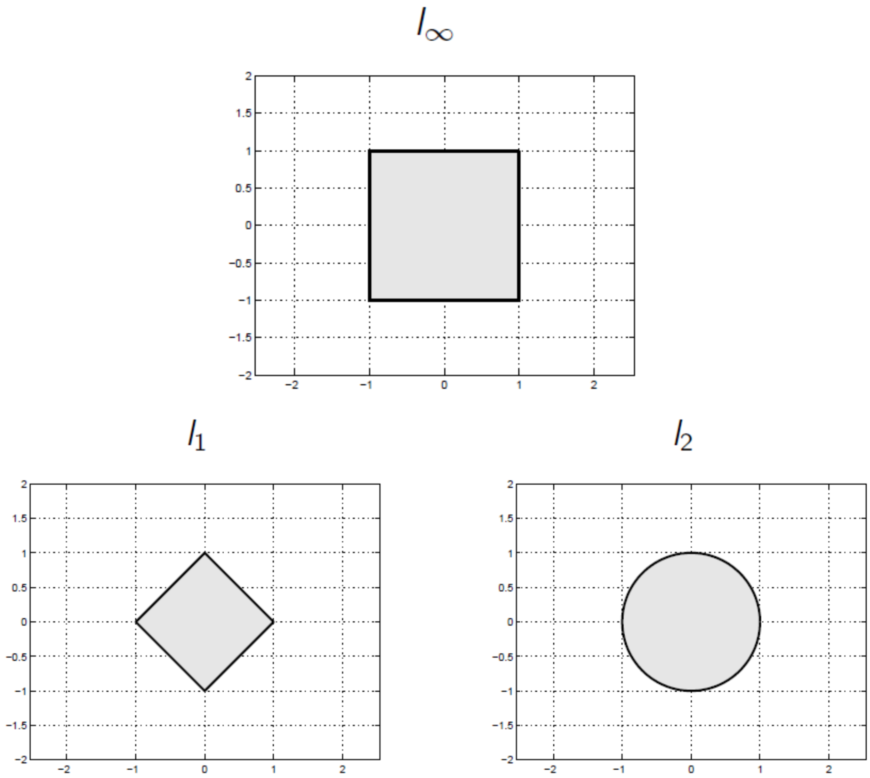

Figure 2: Different unit balls $\|\mathbf{x}\|_p \leq 1$ in $\mathbb{R}^2$.

We'll prove the result for ellipsoids (proofs for the other sets are a good exercise, but are in the video).

Proof. Write E as $E = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq 0\}$ where $f(\mathbf{x}) \equiv \mathbf{x}^\top \mathbf{Q}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c$. Then, take $\mathbf{x}, \mathbf{y} \in E$ and $\lambda \in [0, 1]$, and $f(\mathbf{x}) \leq 0, f(\mathbf{y}) \leq 0$. The vector $z = \lambda \mathbf{x} + (1 - \lambda)\mathbf{y}$ satisfies

$$\mathbf{z}^\top \mathbf{Q}\mathbf{z} = \lambda^2 \mathbf{x}^\top \mathbf{Q}\mathbf{x} + (1 - \lambda)^2 \mathbf{y}^\top \mathbf{Q}\mathbf{y} + 2\lambda(1 - \lambda)\mathbf{x}^\top \mathbf{Q}\mathbf{y}\,,$$

and using Cauchy-Schwartz it follows that

$$\begin{aligned} \mathbf{x}^\top \mathbf{Q}\mathbf{y} \leq \left\|\mathbf{Q}^{1/2}\mathbf{x}\right\| \cdot \left\|\mathbf{Q}^{1/2}\mathbf{y}\right\| &= \sqrt{\mathbf{x}^\top \mathbf{Q}\mathbf{x}}\sqrt{\mathbf{y}^\top \mathbf{Q}\mathbf{y}} \\ &\leq \frac{1}{2}\left(\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{y}^\top \mathbf{Q}\mathbf{y}\right)\,. \end{aligned}$$

Can you see why the second inequality holds? This implies that

$$\mathbf{z}^\top \mathbf{Q}\mathbf{z} \leq \lambda \mathbf{x}^\top \mathbf{Q}\mathbf{x} + (1 - \lambda)\mathbf{y}^\top \mathbf{Q}\mathbf{y}\,.$$

Finally,

$$\begin{aligned} f(\mathbf{z}) &= \mathbf{z}^\top \mathbf{Q}\mathbf{z} + 2\mathbf{b}^\top \mathbf{z} + c \\ &\leq \lambda \mathbf{x}^\top \mathbf{Q}\mathbf{x} + (1 - \lambda)\mathbf{y}^\top \mathbf{Q}\mathbf{y} + 2\lambda \mathbf{b}^\top \mathbf{x} + 2(1 - \lambda)\mathbf{b}^\top \mathbf{y} + c \\ &= \lambda\left(\mathbf{x}^\top \mathbf{Q}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c\right) + (1 - \lambda)\left(\mathbf{y}^\top \mathbf{Q}\mathbf{y} + 2\mathbf{b}^\top \mathbf{y} + c\right) \\ &= \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) \leq 0 \end{aligned}$$

establishing the desired result that $\mathbf{z} \in E$. □

Some important algebraic properties of convex sets are stated in the exercises. In particular, that convexity is preserved under intersections, linear maps, addition and Cartesian products.

The Convex Hull

Definition (Convex Combinations). *Given m points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$, a convex combination of these m points is a vector of the form $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \cdots + \ldots + \lambda_m \mathbf{x}_m$, where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are nonnegative numbers satisfying $\lambda_1 + \lambda_2 + \ldots + \lambda_m = 1$ (i.e., $\boldsymbol{\lambda} \in \Delta_m$ the unit simplex).*

We defined a convex set to be a set for which any convex combination of two points from the set is also in the set. We will now show that a convex combination of any number of points from a convex set is in the set.

Theorem. *Let $C \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in C$. Then for any $\boldsymbol{\lambda} \in \Delta_m$, the relation $\sum_{i=1}^m \lambda_i \mathbf{x}_i \in C$ holds.*

Proof. Proof by induction on m. For $m = 1$ the result is obvious. The induction hypothesis is that for any m vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in C$ and any $\boldsymbol{\lambda} \in \Delta_m$, the vector $\sum_{i=1}^m \lambda_i \mathbf{x}_i$ belongs to C. We will now prove the theorem for $m + 1$ vectors. Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1} \in C$ and that $\boldsymbol{\lambda} \in \Delta_{m+1}$. We will show that $\mathbf{z} \equiv \sum_{i=1}^{m+1} \lambda_i \mathbf{x}_i \in C$. For this, if $\lambda_{m+1} = 1$, then $\mathbf{z} = \mathbf{x}_{m+1} \in C$ and the result obviously follows. Otherwise, if $\lambda_{m+1} < 1$, then

$$
\begin{aligned}
\mathbf{z} &= \sum_{i=1}^{m} \lambda_i \mathbf{x}_i + \lambda_{m+1} \mathbf{x}_{m+1} \\
&= (1 - \lambda_{m+1}) \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} \mathbf{x}_i + \lambda_{m+1} \mathbf{x}_{m+1} \,.
\end{aligned}
$$

Since $\sum_{i=1}^m \frac{\lambda_i}{1-\lambda_{m+1}} = \frac{1-\lambda_{m+1}}{1-\lambda_{m+1}} = 1$, it follows that $\mathbf{v} = \sum_{i=1}^m \frac{\lambda_i}{1-\lambda_{m+1}} \mathbf{x}_i$ is a convex combination of m points from C, and hence by the induction hypotheses we have that $\mathbf{v} \in C$. Thus, by the definition of a convex set, $\mathbf{z} = (1 - \lambda_{m+1})\,\mathbf{v} + \lambda_{m+1} \mathbf{x}_{m+1} \in C$. ◻

Definition (The Convex Hull). *Let $S \subseteq \mathbb{R}^n$. The convex hull of S, denoted by $\operatorname{conv}(S)$, is the set comprising all the convex combinations of vectors from S:*

$$
\operatorname{conv}(S) \equiv \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i : \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in S, \boldsymbol{\lambda} \in \Delta_k \right\}
$$

The convex hull $\operatorname{conv}(S)$ is "smallest" convex set containing S, in the sense that if another convex set T contains S, then $\operatorname{conv}(S) \subset T$.

The following well-known result, called the Carethéodory theorem, states that any element in the convex hull of a subset of a given set $S \subset \mathbb{R}^n$ can be expressed as a **convex combination** of no more than $n + 1$ vectors from S.

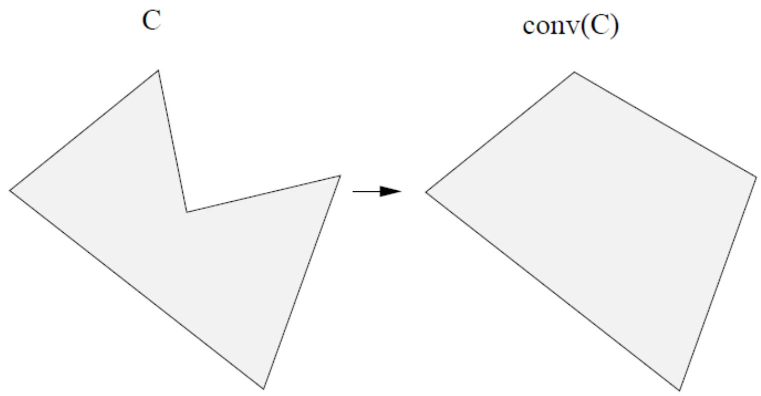

Figure 3: The convex hull of a non-convex set.

Theorem (Carathéodory). *Let* $S \subseteq \mathbb{R}^n$ *and let* $\mathbf{x} \in \operatorname{conv}(S)$. *Then, there exist* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1} \in S$ *such that* $\mathbf{x} \in \operatorname{conv}(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1}\})$, *that is, there exist* $\lambda \in \Delta_{n+1}$ *such that*

$$\mathbf{x} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i$$

We present this proof as it provides a construction mechanism.

Proof. Let $\mathbf{x} \in \operatorname{conv}(S)$. Then, there exist vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in S$ and $\boldsymbol{\lambda} \in \Delta_k$ such that

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i .$$

We can assume that $\lambda_i > 0$ for all $i = 1, 2, \ldots, k$. If $k \leq n+1$, the result is proven. Otherwise, if $k \geq n+2$, then the vectors $\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \ldots, \mathbf{x}_k - \mathbf{x}_1$, are necessarily linearly dependent (as they are in $\mathbb{R}^n$ and there are at least $n+1$ of them) implying that there exist $\mu_2, \mu_3, \ldots, \mu_k$ not all zeros such that

$$\sum_{i=2}^{k} \mu_i (\mathbf{x}_i - \mathbf{x}_1) = \mathbf{0} .$$

Defining $\mu_1 = -\sum_{i=2}^{k} \mu_i$, we obtain that

$$\sum_{i=1}^{k} \mu_i \mathbf{x}_i = \mathbf{0} .$$

Note that not all of the coefficients $\mu_1, \mu_2, \ldots, \mu_k$ are zeros and $\sum_{i=1}^{k} \mu_i = 0$. Thus, there exists an index i for which $\mu_i < 0$. Let $\alpha \in \mathbb{R}_+$. Then,

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i + \alpha \sum_{i=1}^{k} \mu_i \mathbf{x}_i = \sum_{i=1}^{k} (\lambda_i + \alpha \mu_i) \mathbf{x}_i .$$

We have $\sum_{i=1}^{k}(\lambda_i + \alpha\mu_i) = 1$, so the equation above is a convex combination if and only if

$$\lambda_i + \alpha\mu_i \geq 0 \text{ for all } i = 1, \dots, k.$$

But since $\lambda_i > 0$ for all i, it follows that these inequalities are satisfied for $\alpha = \min_{i:\mu_i<0}\left\{-\frac{\lambda_i}{\mu_i}\right\}$. In addition, we must have $\lambda_j + \varepsilon\mu_j = 0$ for $j \in \operatorname*{argmin}_{i:\mu_i<0}\left\{-\frac{\mu_i}{\lambda_i}\right\}$. This means that we found a representation of $\mathbf{x}$ as a convex combination of $k-1$ (or less) vectors. This process can be carried on until a representation of $\mathbf{x}$ as a convex combination of no more than $n+1$ vectors is derived. $\square$

Convex Polytopes and Basic Feasible Solutions

This representation theorem has an important application to convex polytopes[3] of the form

$$P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}, \quad \text{where } \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m.$$

As we will see in the coming weeks, this is a standard formulation of the constraints of a linear programming problem. We will assume without loss of generality that the rows of A are linearly independent[4].

An important property of nonempty convex polytopes of the form P is that they contain at least one vector with at most m nonzero elements (m being the number of constraints). An important concept related to this is the idea of a **basic feasible solution**.

Definition (Basic Feasible Solution). $\overline{\mathbf{x}}$ *is a basic feasible solution (bfs) of* P *if the columns of* $\mathbf{A}$ *corresponding to the indices of the positive values of* $\overline{\mathbf{x}}$ *are linearly independent.*

This definition seems unintuitive at first, but as we will see, if a linear programme has an optimal solution then there is an optimal bfs. So to find the optimal solution, we only need to find the basic feasible solutions.

Theorem (Existence of Basic Feasible Solutions). *Let* $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, *where* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$. *If* $P \neq \emptyset$, *then it contains at least one* **bfs**.

Extreme points

We can characterize basic feasible solutions as extreme points, which are points in the set that cannot be represented as a nontrivial convex combination of two different points in S.

[3] A polytope is a geometric object with flat sides.
[4] If not, we can remove a constraint without changing the optimization problem

Definition (Extreme Point). *Let $S \subseteq \mathbb{R}^n$ be a convex set. A point $\mathbf{x} \in S$ is called an extreme point of S if there do not exist $\mathbf{x}_1, \mathbf{x}_2 \in S$ $(\mathbf{x}_1 \neq \mathbf{x}_2)$ and $\lambda \in (0,1)$, such that $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. The set of extreme point is denoted by $\mathrm{ext}(S)$.*

For example, the set of extreme points of a convex polytope consists of all its vertices - see Figure 4. There is an equivalence between basic feasible solutions and extreme points.

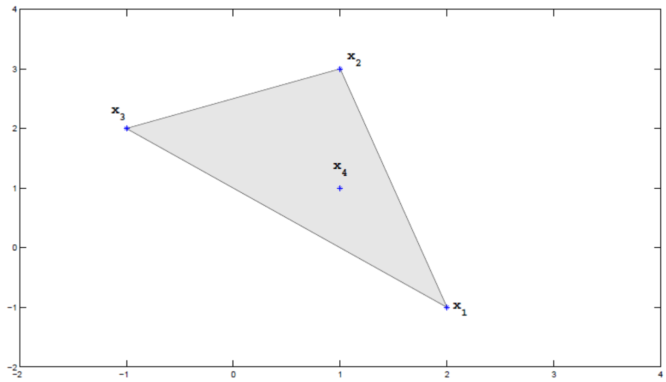

Figure 4: The extreme points of this triangle are given by its vertices $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$. The point $\mathbf{x}_4$ is not an extreme point.

Theorem (Equivalence between Basic Feasible Solutions and Extreme Points). *Let $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ has linearly independent rows and $\mathbf{b} \in \mathbb{R}^m$. The $\overline{\mathbf{x}}$ is a basic feasible solution of P if and only if it is an extreme point of P.*

Finally, we state an important result which says that a compact convex set is the convex hull of its extreme points.

Theorem (The Krein-Milman Theorem). *Let $S \subseteq \mathbb{R}^n$ be a compact convex set. Then*

$$S = \mathrm{conv}(\mathrm{ext}(S)) \,.$$

Next week, we will see that linear programmes, which are optimization problems of the form

$$(\mathbf{LP}) : \quad \begin{array}{ll} \min_{\mathbf{x}} & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array}$$

have at least one optimal solution which is an extreme point, or equivalently, a basic feasible solution. So if you can find basic feasible solutions, then you can solve (small) linear programmes by enumerating them all and testing to see which has the lowest value of $\mathbf{c}^\top \mathbf{x}$.

Convex Functions

We begin by giving a definition of a convex function.

Definition (Convex Function). *A function $f : C \longrightarrow \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is **convex** if*

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in C, \lambda \in [0,1].$$

*f is **strictly convex** if*

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) \text{ for any } \mathbf{x} \neq \mathbf{y} \in C, \lambda \in (0,1).$$

*We say f is **concave** if $-f$ is convex.*

Examples of Convex Functions

- Affine Functions. $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} + b$, where $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Take $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Then

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \mathbf{a}^\top(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + b \\ &= \lambda\left(\mathbf{a}^\top\mathbf{x}\right) + (1-\lambda)\left(\mathbf{a}^\top\mathbf{y}\right) + \lambda b + (1-\lambda)b \\ &= \lambda\left(\mathbf{a}^\top\mathbf{x} + b\right) + (1-\lambda)\left(\mathbf{a}^\top\mathbf{y} + b\right) \\ &= \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}). \end{aligned}$$

  Note that affine functions are also concave.

- Norms. $g(\mathbf{x}) = \|\mathbf{x}\|$, take $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Then

$$\begin{aligned} g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\| \\ &\leq \|\lambda\mathbf{x}\| + \|(1-\lambda)\mathbf{y}\| \text{ by the triangle inequality} \\ &= \lambda\|\mathbf{x}\| + (1-\lambda)\|\mathbf{y}\| \\ &= \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}). \end{aligned}$$

As with convex sets, the definition of convexity can be extended from two points to many points:

Theorem (Jensen's Inequality[5]). *Let $f : C \longrightarrow \mathbb{R}$ be a convex function where $C \subseteq \mathbb{R}^n$ is a convex set. Then, for any $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in C$ and $\boldsymbol{\lambda} \in \Delta_k$, the following inequality holds:*

$$f\left(\sum_{i=1}^{k} \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i).$$

[5] You may have seen this result stated in probability courses as

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$$

where X is a random variable.

First and Second Order Characterizations of Convex Functions

Convex functions are not necessarily differentiable, but when they are, we can characterize them as functions for which the tangent hyperplane[6] underestimates the function. See Figure 5.

Theorem (The Gradient Inequality)**.** *Let $f : C \longrightarrow \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if*

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) \textit{ for any } \mathbf{x}, \mathbf{y} \in C\,. \tag{1}$$

An analogous result holds for strictly convex functions (with a strict inequality).

Proof. Suppose first that f is convex. Let $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in (0, 1]$. If $\mathbf{x} = \mathbf{y}$, then the inequality trivially holds. Assume that $\mathbf{x} \neq \mathbf{y}$. Then by the definition of convexity for f

$$\frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda} \leq f(\mathbf{y}) - f(\mathbf{x})\,.$$

Taking $\lambda \longrightarrow 0^+$, we obtain

$$f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$$

where $f'(\mathbf{x}; \mathbf{d})$ is the directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{d}$. Since f is continuously differentiable, $f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) = \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x})$, and the inequality follows.

To prove the converse, assume that that the gradient inequality holds. Let $\mathbf{z}, \mathbf{w} \in C$, and let $\lambda \in (0, 1)$. We will show that $f(\lambda\mathbf{z} + (1 - \lambda)\mathbf{w}) \leq \lambda f(\mathbf{z}) + (1 - \lambda)f(\mathbf{w})$, Let $\mathbf{u} = \lambda\mathbf{z} + (1 - \lambda)\mathbf{w} \in C$. Then

$$\mathbf{z} - \mathbf{u} = \frac{\mathbf{u} - (1 - \lambda)\mathbf{w}}{\lambda} - \mathbf{u} = -\frac{1 - \lambda}{\lambda}(\mathbf{w} - \mathbf{u})\,.$$

By invoking the gradient inequality on the pairs $\mathbf{u}, \mathbf{z}$ and $\mathbf{u}, \mathbf{w}$, and substituting for $\mathbf{w} - \mathbf{u}$ using the above, we have

$$\begin{aligned} f(\mathbf{u}) + \nabla f(\mathbf{u})^\top(\mathbf{z} - \mathbf{u}) &\leq f(\mathbf{z})\,, \\ f(\mathbf{u}) - \frac{\lambda}{1 - \lambda}\nabla f(\mathbf{u})^\top(\mathbf{z} - \mathbf{u}) &\leq f(\mathbf{w})\,. \end{aligned}$$

Thus,

$$f(\mathbf{u}) \leq \lambda f(\mathbf{z}) + (1 - \lambda)f(\mathbf{w})$$

□

[6] The tangent hyperplane of f at x is, by the linear approximation theorem,

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

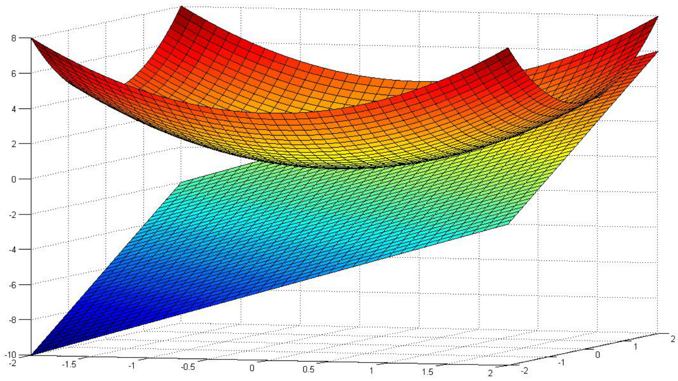

Figure 5: For a convex function f, the tangent plane at every point is always below f.

Convexity + Stationarity $\Rightarrow$ Global Optimality!

A direct result of the gradient inequality is that the first order optimality condition $\nabla f(\mathbf{x}^*) = 0$ is sufficient for global optimality.

Theorem (Stationarity Implies Global Optimality)**.** *Let f be a continuously differentiable function which is convex over a convex set $C \subseteq \mathbb{R}^n$. Suppose that $\nabla f(\mathbf{x}^*) = \mathbf{0}$ for some $\mathbf{x}^* \in C$. Then $\mathbf{x}^*$ is the global minimizer of f over C.*

Proof. This is a direct consequence of the gradient inequality. □

We can now extend our link between convexity and optimality conditions to second-order characterizations.

Theorem (Second-Order Characterization of Convexity)**.** *Let f be a twice continuously differentiable function over an open convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in C$.*

In addition, if $\nabla^2 f(\mathbf{x}) \succ 0 \; \forall \; \mathbf{x}$ then f is strictly convex. The converse is false[7].

We can now revisit optimality conditions for quadratic functions. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function given by $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then f is convex if and only if $\mathbf{A} \succeq \mathbf{0}$ (and strictly convex ifF $\mathbf{A} \succ \mathbf{0}$).

[7] Consider $f(x) = x^4$. This is a strictly convex function, but $f''(0) = 0$.

Example. Convexity of the log-sum-exp function:

$$f(\mathbf{x}) = \log\left(e^{x_1} + e^{x_2} + \ldots + e^{x_n}\right), \quad \mathbf{x} \in \mathbb{R}^n .$$

The gradient is given by:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad i = 1, 2, \ldots, n .$$

Therefore, the Hessian is

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} -\frac{e^{x_i} e^{x_j}}{\left(\sum_{j=1}^n e^{x_j}\right)^2}, & i \neq j \\ -\frac{e^{x_i} e^{x_j}}{\left(\sum_{j=1}^n e^{x_j}\right)^2} + \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, & i = j \end{cases} .$$

We can thus write the Hessian matrix as

$$\nabla^2 f(\mathbf{x}) = \operatorname{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^\top, \qquad \text{with} \qquad \mathbf{w} = \left(\frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}\right)_{i=1}^n \in \Delta_n .$$

For any $\mathbf{v} \in \mathbb{R}^n$:

$$\mathbf{v}^\top \nabla^2 f(\mathbf{x}) \mathbf{v} = \sum_{i=1}^n w_i v_i^2 - \left(\mathbf{v}^\top \mathbf{w}\right)^2 \geq 0 ,$$

since defining $s_i = \sqrt{w_i} v_i$, $t_i = \sqrt{w_i}$, we have

$$\left(\mathbf{v}^\top \mathbf{w}\right)^2 = \left(\mathbf{s}^\top \mathbf{t}\right)^2 \leq \|\mathbf{s}\|^2 \|\mathbf{t}\|^2 = \left(\sum_{i=1}^n w_i v_i^2\right)\left(\sum_{i=1}^n w_i\right) = \sum_{i=1}^n w_i v_i^2 .$$

Thus, $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ and hence f is convex over $\mathbb{R}^n$.

Operations Preserving Convexity

- Let f be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$ and let $\alpha \geq 0$. Then αf is a convex function over C.
- Let $f_1, f_2, \ldots, f_p$ be convex functions over a convex set $C \subseteq \mathbb{R}^n$. Then the sum function $f_1 + f_2 + \ldots + f_p$ is convex over C.
- Let f be a convex function defined on a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the function g defined by

  $$g(\mathbf{y}) = f(\mathbf{A}\mathbf{y} + \mathbf{b})$$

  is convex over the convex set $D = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{A}\mathbf{y} + \mathbf{b} \in C\}$.

- Let $f : C \longrightarrow \mathbb{R}$ be a convex function defined over the convex set $C \subseteq \mathbb{R}^n$. Let $g : I \longrightarrow \mathbb{R}$ be a one-dimensional nondecreasing convex function over the interval $I \subseteq \mathbb{R}$. Assume that the image of C under f is contained in $I : f(C) \subseteq I$. Then the composition of g with f defined by

  $$h(\mathbf{x}) \equiv g(f(\mathbf{x}))$$

  is convex over C.

In the computer lab, we will look at software for solving convex optimization problems. To do this, we have to write functions in such a way that the software is able to validate that the function is indeed convex.

Checklist

The idea of this checklist is to help you to self-evaluate your progress and understanding of the subject, and to give you some guidance on where to focus. If you can tick all the boxes it means you're doing alright, otherwise you need to study a bit more, grab a book, watch the videos, or seek help from classmates, the lecturers, or the demonstrators. Try to fill as many gaps as quickly as possible.

Learning Outcome	Check
I can state the definition of a convex set and verify the examples in the notes.	
I understand the concept of a convex combination and its relation to the unit simplex presented in Week 2.	
I understand the notion of convex hull of a set S as the *smallest* convex set containing S.	
I have studied the proof of the representation theorem for convex hulls and I can use it.	
I understand what a convex polytope is.	
I understand the definition of basic feasible solution and I can verify it.	
I understand the relation between bfs and extreme points, and its relation to feasibility in optimization.	
I can state the definition of convex function and its different characterizations.	
I can show whether a function is convex or not.	

Exercises

1. **Intersection of convex sets is convex:** Prove that if $C_i \subseteq \mathbb{R}^n$ are convex sets for any $i = 1, \ldots, I$, then the set $\bigcap_{i \in I} C_i$ is also convex.

2. **Linear maps of convex sets are convex:** Let $M \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Prove that

$$\mathbf{A}(M) = \{\mathbf{Ax} : \mathbf{x} \in M\}$$

is convex.

3. Consider the four vectors

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

and let $\mathbf{x} \in \operatorname{conv}(\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\})$ be given by

$$\mathbf{x} = \frac{1}{8}\mathbf{x}_1 + \frac{1}{4}\mathbf{x}_2 + \frac{1}{2}\mathbf{x}_3 + \frac{1}{8}\mathbf{x}_4 = \begin{pmatrix} \frac{13}{8} \\ \frac{11}{8} \end{pmatrix}$$

Is it possible to represent $\mathbf{x}$ as a convex combination of no more than 3 vectors?

4. Find all the basic feasible solutions of the linear system:

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ x_2 + x_4 &= 3 \\ x_1, x_2, x_3, x_4 &\geq 0. \end{aligned}$$

5. Consider the linear system $\mathbf{Ax} = \mathbf{y}$ where $\mathbf{x} \geq \mathbf{0}$. If $\mathbf{A}$ is $m \times n$, then how many basic feasible solutions may there be to this system?

6. Find the extreme points of the system

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ x_2 + x_4 &= 3 \\ x_1, x_2, x_3, x_4 &\geq 0. \end{aligned}$$

7. Show that the quad-over-lin function

$$f(x_1, x_2) = \frac{x_1^2}{x_2}$$

defined over $\mathbb{R} \times \mathbb{R}_{++} = \{(x_1, x_2) : x_2 > 0\}$ is convex.

8. Prove that the following functions are convex:

- The generalized quad-over-lin function

$$g(\mathbf{x}) = \frac{\|\mathbf{A}\mathbf{x} + \mathbf{b}\|^2}{\mathbf{c}^\top \mathbf{x} + d} \quad \left(\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n, d \in \mathbb{R}\right)$$

  over the domain $D = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^\top \mathbf{x} + d > 0\}$.

- $h(\mathbf{x}) = e^{\|\mathbf{x}\|^2}$.

9. Let $f_1, f_2, \ldots, f_p$ be convex functions over a convex set $C \subseteq \mathbb{R}^n$. Prove that the sum function $f_1 + f_2 + \ldots + f_p$ is also convex over C.

10. **Point-Wise Maximum of Convex Functions** Let $f_1, f_2, \ldots, f_p : C \to \mathbb{R}$ be p convex functions over the convex set $C \subseteq \mathbb{R}^n$. Prove that the maximum function

$$f(\mathbf{x}) \equiv \max_{i=1,2,\ldots,p} \{f_i(\mathbf{x})\}$$

    is convex over C.

11. Proving the following functions are all convex

    - $f(\mathbf{x}) = \max\{x_1, x_2, \ldots, x_n\}$.
    - For a given vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$, let $x_{[i]}$ denote the i -th largest value in $\mathbf{x}$. For any $k \in \{1, 2, \ldots, n\}$ let

$$h_k(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[k]}.$$