Convexity: an introduction

Geir Dahl
CMA, Dept. of Mathematics and Dept. of Informatics
University of Oslo

1. Introduction

- what is convexity
- where does it arise
- main concepts and results

Literature:

- Rockafellar: *Convex analysis*, 1970.
- Webster: *Convexity*, 1994.
- Grünbaum: *Convex polytopes*, 1967.
- Ziegler: *Lectures on polytopes*, 1994.
- Hiriart-Urruty and Lemaréchal: *Convex analysis and minimization algorithms*, 1993.
- Boyd and Vandenberghe: *Convex optimization*, 2004.

- roughly: a convex set in $\mathbb{R}^2$ (or $\mathbb{R}^n$) is a set "with no holes".
- more accurately, a convex set C has the following property: whenever we choose two points in the set, say $x, y \in C$, then all points in the line segment between x and y also lie in C.
- a sphere (ball), an ellipsoid, a point, a line, a line segment, a rectangle, a triangle, halfplane, the plane itself
- the union of two disjoint (closed) triangles is nonconvex.

Why are convex sets important?

Optimization:

- mathematical foundation for optimization
- feasible set, optimal set, ...
- objective function, constraints, value function
- closely related to the numerical solvability of an optimization problem

Statistics:

- statistics: both in theory and applications
- estimation: "estimate" the value of one or more unknown parameters in a stochastic model. To measure quality of a solution one uses a "loss function" and, quite often, this loss function is convex.
- statistical decision theory: the concept of risk sets is central; they are convex sets, so-called polytopes.

- The expectation operator: Assume that X is a discrete variable taking values in some finite set of real numbers, say $\{x_1, \ldots, x_r\}$ with probabilities p_i of the event $X = x_i$. Probabilities are all nonnegative and sum to one, so $p_j \geq 0$ and $\sum_{j=1}^{r} p_j = 1$.The expectation (or mean) of X is the number

$$EX = \sum_{j=1}^{r} p_j x_j.$$

  This as a weighted average of the possible values that X can attain, and the weights are the probabilities. We say that EX is a convex combination of the numbers $x_1, \ldots, x_r$.

- An extension is when the discrete random variable is a vector, so it attains values in a finite set $S = \{x_1, \ldots, x_r\}$ of points in $\mathbb{R}^n$. The expectation is defined by $EX = \sum_{j=1}^{r} p_j x_j$ which, again, is a convex combination of the points in S.

Approximation

- approximation: given some set $S \subset \mathbb{R}^n$ and a vector $z \notin S$, find a vector $x \in S$ which is as close to z as possible among all vectors in S.
- distance: Euclidean norm (given by ($\|x\| = (\sum_{j=1}^n x_j^2)^{1/2}$) or some other norm.
- convexity?
- norm functions, i.e., functions $x \to \|x\|$, are convex functions.
- a basic question is if a nearest point (to z in S) exists: yes, provided that S is a closed set.
- and: if S is a convex set (and the norm is the Euclidean norm), then the nearest point is unique.
- this may not be so for nonconvex sets.

Nonnegative vectors

- convexity deals with inequalities
- $x \in \mathbb{R}^n$ is nonnegative if each component x_i is nonnegative.
- we let $\mathbb{R}^n_+$ denote the set of all nonnegative vectors. The zero vector is written O.
- inequalities for vectors, so if $x, y \in \mathbb{R}^n$ we write

$$x \leq y \quad (\text{or } y \geq x)$$

and this means that $x_i \leq y_i$ for $i = 1, \ldots, n$.

2. Convex sets

- definition of convex set
- polyhedron
- connection to LP

Convex sets and polyhedra

- definition: A set $C \subseteq \mathbb{R}^n$ is called convex if $(1-\lambda)x_1 + \lambda x_2 \in C$ whenever $x_1, x_2 \in C$ and $0 \leq \lambda \leq 1$.
- geometrically, this means that C contains the line segment between each pair of points in C.
- examples: circle, ellipse, rectangle, certain polygons, pyramids
- how can we prove that a set is convex?
- later we learn some other useful techniques.
- how can we verify that a set S is not convex? Well, it suffices to find two points x_1 and x_2 and $0 \leq \lambda \leq 1$ with the property that $(1-\lambda)x_1 + \lambda x_2 \notin S$ (you have then found a kind of "hole" in S).

- the unit ball:

$$B = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

- to prove it is convex: let $x, y \in B$ and $\lambda \in [0, 1]$. Then

$$\begin{aligned} \|(1-\lambda)x + \lambda y\| &\leq \|(1-\lambda)x\| + \|\lambda y\| \\ &= (1-\lambda)\|x\| + \lambda\|y\| \\ &\leq (1-\lambda) + \lambda = 1 \end{aligned}$$

  Therefore B is convex. □

- we here used the triangle inequality which is a convexity property (we return to this): recall that the triangle ineq. may be shown from the Cauchy-Schwarz inequality:

$$|x \cdot y| \leq \|x\| \, \|y\| \quad \text{for } x, y \in \mathbb{R}^n.$$

- More generally: $B(a, r) := \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$ is convex (where $a \in \mathbb{R}^n$ and $r \geq 0$).

Linear systems and polyhedra

- By a linear system we mean a finite set of linear equations and/or linear inequalities involving variables $x_1, \ldots, x_n$.
- Example: the linear system $x_1 + x_2 = 3$, $x_1 \geq 0$, $x_2 \geq 0$ in the variables x_1, x_2.
- equivalent form is $x_1 + x_2 \leq 3$, $-x_1 - x_2 \leq -3$, $-x_1 \leq 0$, $-x_2 \leq 0$. Here we only have $\leq$–inequalities
- definition: we define a polyhedron in $\mathbb{R}^n$ as a set of the form $\{x \in \mathbb{R}^n : Ax \leq b\}$ where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Here m is arbitrary, but finite. So: the solution set of a linear system.

Proposition

Every polyhedron is a convex set.

Proposition

The intersection of convex sets is a convex set. The sum of convex sets if also convex.

Note:

- $\{x \in \mathbb{R}^n : Ax = b\}$: affine set; if $b = O$: linear subspace
- the dimension of an affine set $z + L$ is defined as the dimension of the (uniquely) associated subspace L
- each affine set is a polyhedron
- of special interest: affine set of dimension $n-1$, i.e.

$$H = \{x \in \mathbb{R}^n : a^T x = \alpha\}$$

where $a \in \mathbb{R}^n$, $a \neq O$ and $\alpha \in \mathbb{R}$, i.e., solution set of one linear equation. Called a hyperplane.

LP and convexity

Consider a linear programming (LP) problem

$$\max\{c^T x : Ax \leq b,\ x \geq O\}$$

- Then the feasible set $\{x \in \mathbb{R}^n : Ax \leq b,\ x \geq O\}$ is a polyhedron, and therefore convex.
- Assume that there is a finite optimal value v^*. Then the set of optimal solutions $\{x \in \mathbb{R}^n : Ax \leq b,\ x \geq O,\ c^T x = v^*\}$ is a polyhedron.
- This is (part of) the convexity in LP.

Convex hulls

- convex hull
- Carathéodory's theorem
- polytopes
- linear optimization over polytopes

Convex hulls

Goal:

- convex combinations are natural linear combinations to work with in convexity: represent "mixtures".
- convex hull gives a smallest convex set containing a given set S. Makes it possible to approximate S by a nice set.

- consider vectors $x_1, \ldots, x_t \in \mathbb{R}^n$ and nonnegative numbers (coefficients) $\lambda_j \geq 0$ for $j = 1, \ldots, t$ such that $\sum_{j=1}^t \lambda_j = 1$. Then the vector $x = \sum_{j=1}^t \lambda_j x_j$ is called a convex combination of $x_1, \ldots, x_t$. Thus, a convex combination is a special linear combination.
- convex comb. of two points (vectors), three, ...

Proposition

A set $C \subseteq \mathbb{R}^n$ is convex if and only if it contains all convex combinations of its points.

Proof: Induction on number of points. □

Definition. Let $S \subseteq \mathbb{R}^n$ be any set. Define the convex hull of S, denoted by $\mathrm{conv}\,(S)$ as the set of all convex combinations of points in S.

- the convex hull of two points x_1 and x_2 is the line segment between the two points, $[x_1, x_2]$.
- an important fact is that $\mathrm{conv}\,(S)$ is a convex set, whatever the set S might be.

Proposition

Let $S \subseteq \mathbb{R}^n$. Then $\mathrm{conv}\,(S)$ is equal to the intersection of all convex sets containing S. Thus, $\mathrm{conv}\,(S)$ is is the smallest convex set containing S.

A "special kind" of convex hull

- what happens if we take the convex hull of a finite set of points?

Definition. A set $P \subset \mathbb{R}^n$ is called a polytope if it is the convex hull of a finite set of points in $\mathbb{R}^n$.

- polytopes have been studied a lot during the history of mathematics
- Platonian solids
- important in many branches of mathematics, pure and applied.
- in optimization: highly relevant in, especially, linear programming and discrete optimization.

Linear optimization over polytopes

Consider

$$\max\{c^T x : x \in \mathrm{conv}\,(\{x_1, \ldots, x_t\}\}$$

where $c \in \mathbb{R}^n$.

Each $x \in P$ may be written as $x = \sum_{j=1}^t \lambda_j x_j$ for some $\lambda_j \geq 0$, $j = 1, \ldots, t$ where $\sum_j \lambda_j = 1$. Define $v^* = \max_j c^T x_j$. Then

$$c^T x = c^T \sum_j \lambda_j x_j = \sum_{j=1}^t \lambda_j c^T x_j \leq \sum_{j=1}^t \lambda_j v^* = v^* \sum_{j=1}^t \lambda_j = v^*.$$

- The set of optimal solutions is

  $$\mathrm{conv}\,(\{x_j : j \in J\})$$

  where J is the set of indices j satisfying $c^T x_j = v^*$.
- This is a subpolytope of the given polytope (actually a so-called *face*). Computationally OK if "few" points.

Carathéodory's theorem

The following result says that a convex combination of "many" points may be reduced by using "fewer" points.

Theorem

Let $S \subseteq \mathbb{R}^n$. Then each $x \in \mathrm{conv}\,(S)$ may be written as a convex combination of (say) m affinely independent points in S. In particular, $m \leq n + 1$.

Try to construct a proof!

Two consequences

- $k + 1$ vectors $x_0, x_1, \dots, x_k \in \mathbb{R}^n$ are called affinely independent if the k vectors $x_1 - x_0, \dots, x_k - x_0$ are linearly independent.
- A simplex is the convex hull of a affinely independent points.

Proposition

Every polytope in $\mathbb{R}^n$ can be written as the union of a finite number of simplices.

Proposition

Every polytope in $\mathbb{R}^n$ is compact, i.e., closed and bounded.

4. Projection and separation

- nearest points
- separating and supporting hyperplanes
- Farkas' lemma

Projection

Approximation problem: Given a set S and a point x outside that set, find a nearest point to x in S !

- Question 1: does a nearest point exist?
- Question 2: if it does, is it unique?
- Question 3: how can we compute a nearest point?
- convexity is central here!

Let S be a closed subset of $\mathbb{R}^n$. Recall: S is closed if and only if S contains the limit point of each convergent sequence of points in S. Thus, if $\{x^{(k)}\}_{k=1}^{\infty}$ is a convergent sequence of points where $x^{(k)} \in S$, then the limit point $x = \lim_{k\to\infty} x^{(k)}$ also lies in S.

For $S \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$ we define the distance function

$$d_S(x) = \inf\{\|x - s\| : s \in S\}$$

where $\|\cdot\|$ is the Euclidean norm.

Nearest point

Proposition

Let $S \subseteq \mathbb{R}^n$ be a nonempty closed set and let $x \in \mathbb{R}^n$. Then there is a nearest point $s \in S$ to x, i.e., $\|x - s\| = d_S(x)$.

Proof. There is a sequence $\{s^{(k)}\}_{k=1}^{\infty}$ of points in S such that $\lim_{k\to\infty} \|x - s^{(k)}\| = d_S(x)$. This sequence is bounded and has a convergent subsequence, and the limit point must lie in S. Then, by continuity, $d_S(x) = \lim_{j\to\infty} \|x - s^{(i_j)}\| = \|x - s\|$. □

Thus, closedness of S assures that a nearest point exists. But such a point may not be unique.

Good news for convex sets

Theorem

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, for every $x \in \mathbb{R}^n$, the nearest point x_0 to x in C is unique. Moreover, x_0 is the unique solution of the inequalities

$$(x - x_0)^T (y - x_0) \leq 0 \quad \text{for all } y \in C. \tag{1}$$

Proof: Let x_0 be a nearest point to x in C. Let $y \in C$ and let $0 < \lambda < 1$. Since C is convex, $(1-\lambda)x_0 + \lambda y \in C$ and since x_0 is a nearest point we have that $\|(1-\lambda)x_0 + \lambda y - x\| \geq \|x_0 - x\|$, i.e., $\|(x_0 - x) + \lambda(y - x_0)\| \geq \|x_0 - x\|$. This implies $\|x_0 - x\|^2 + 2\lambda(x_0 - x)^T(y - x_0) + \lambda^2\|y - x_0\|^2 \geq \|x_0 - x\|^2$. We now subtract $\|x_0 - x\|^2$ on both sides, divide by λ, let $\lambda \to 0^+$ and finally multiply by -1. This proves that the inequality (1) holds for every $y \in C$. Let now x_1 be another nearest point to x in C; we want to show that $x_1 = x_0$. By letting $y = x_1$ in (1) we get

$$(*_1) \;\; (x - x_0)^T(x_1 - x_0) \leq 0.$$

Proof, cont.: By symmetry we also get that

$$(*_2)\ \ (x - x_1)^T(x_0 - x_1) \le 0.$$

By adding the inequalities $(*_1)$ and $(*_2)$ we obtain $\|x_1 - x_0\|^2 = (x_1 - x_0)^T(x_1 - x_0) \le 0$ which implies that $x_1 = x_0$. Thus, the nearest point is unique. $\square$

The variational inequality (1) has a nice geometrical interpretation: the angle between the vectors $x - x_0$ and $y - x_0$ (both starting in the point x_0) is obtuse, i.e., larger that 90°.

- $p_C(x)$ denotes the (unique) nearest point to x in C.

What's next?

We shall now discuss supporting hyperplanes and separation of convex sets.

Why is this important?

- leads to another representation of closed convex sets
- may be used to approximate convex functions by simpler functions
- may be used to prove Farkas' lemma, and the linear programming duality theorem
- used in statistics (e.g. decision theory), mathematical finance, economics, game theory.

Hyperplanes: definitions

- Hyperplane: has the $H = \{x \in \mathbb{R}^n : a^T x = \alpha\}$ for some nonzero vector a and a real number α.
- a is called the normal vector of the hyperplane.
- Every hyperplane is an affine set of dimension $n - 1$.
- Each hyperplane divides the space into two sets $H^+ = \{x \in \mathbb{R}^n : a^T x \geq \alpha\}$ and $H^- = \{x \in \mathbb{R}^n : a^T x \leq \alpha\}$.
- These sets H^+ and H^- are called halfspaces.

Definition: Let $S \subset \mathbb{R}^n$ and let H be a hyperplane in $\mathbb{R}^n$.

- If S is contained in one of the halfspaces H^+ or H^- and $H \cap S$ is nonempty, we say that H is a supporting hyperplane of S.
- We also say that H supports S at x, for each $x \in H \cap S$.

Supporting hyperplanes

Note:

- We now restrict the attention to closed convex sets.
- Recall that $p_C(x)$ is the (unique) nearest point to $x$in C.
- Then each point outside our set C gives rise to a supporting hyperplane as the following lemma tells us.

Proposition

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and let $x \in \mathbb{R}^n \setminus C$. Consider the hyperplane H containing $p_C(x)$ and having normal vector $a = x - p_C(x)$. Then H supports C at $p_C(x)$ and C is contained in the halfspace $H^- = \{y : a^T y \leq \alpha\}$ where $\alpha = a^T p_C(x)$.

The proof

Note that a is nonzero as $x \notin C$ while $p_C(x) \in C$. Then H is the hyperplane with normal vector a and given by $a^T y = \alpha = a^T p_C(x)$. We shall show that C is contained in the halfspace H^-. So, let $y \in C$. Then, by (1) we have $(x - p_C(x))^T (y - p_C(x)) \leq 0$, i.e., $a^T y \leq a^T p_C(x) = \alpha$ as desired. □

Separation

Define:

$$H_{a,\alpha} := \{x \in \mathbb{R}^n : a^T x = \alpha\};$$
$$H^-_{a,\alpha} := \{x \in \mathbb{R}^n : a^T x \le \alpha\};$$
$$H^+_{a,\alpha} := \{x \in \mathbb{R}^n : a^T x \ge \alpha\}.$$

We say that the hyperplane $H_{a,\alpha}$ separates two sets S and T if $S \subseteq H^-_{a,\alpha}$ and $T \subseteq H^+_{a,\alpha}$ or vice versa.

Note that both S and T may intersect the hyperplane $H_{a,\alpha}$ in this definition.

We say that the hyperplane $H_{a,\alpha}$ strongly separates S and T if there is an $\epsilon > 0$ such that $S \subseteq H^-_{a,\alpha-\epsilon}$and $T \subseteq H^+_{a,\alpha+\epsilon}$ or vice versa. This means that

$$a^T x \le \alpha - \epsilon \quad \text{for all } x \in S;$$
$$a^T x \ge \alpha + \epsilon \quad \text{for all } x \in T.$$

Strong separation

Theorem

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and assume that $x \in \mathbb{R}^n \setminus C$. Then C and x can be strongly separated.

Proof. Let H be the hyperplane containing $p_C(x)$ and having normal vector $x - p_C(x)$. From the previous proposition we know that H supports C at $p_C(x)$. Moreover $x \neq p_C(x)$ (as $x \notin C$). Consider the hyperplane H^* which is parallel to H (i.e., having the same normal vector) and contains the point $(1/2)(x + p_C(x))$. Then H^* strongly separates x and C. □

An important consequence

Exterior description of closed convex sets:

Corollary

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then C is the intersection of all its supporting halfspaces.

Another application: Farkas' lemma

Theorem

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then there exists an $x \geq O$ satisfying $Ax = b$ if and only if for each $y \in \mathbb{R}^m$ with $y^T A \geq O$ it also holds that $y^T b \geq 0$.

Proof: Consider the closed convex cone (define!!)
$C = \mathrm{cone}\,(\{a^1, \ldots, a^n\}) \subseteq \mathbb{R}^m$. Observe: $Ax = b$ has a nonnegative solution simply means simply (geometrically) that $b \in C$.

Assume now that $Ax = b$ and $x \geq O$. If $y^T a \geq O$, then
$y^T b = y^T(ax) = (y^T a)x \geq 0$.

Proof, cont.: Conversely, if $Ax = b$ has no nonnegative solution, then $b \notin C$. But then, by Strong Separation Theorem, C and b can be strongly separated, so there is a nonzero vector $y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$with $y^T x \geq \alpha$ for each $x \in C$ and $y^T b < \alpha$. As $O \in C$, we have $\alpha \leq 0$. Moreover $y^T a^j \geq 0$ so $y^T a \geq O$. Since $y^T b < 0$ we have proved the other direction of Farkas' lemma.

5. Representation of convex sets

- study (very briefly) the structure of convex sets
- involves the notions: faces, extreme points and extreme halflines
- an important subfield: the theory (and application) of polyhedra and polytopes

Faces

Definition. Let C be a convex set in $\mathbb{R}^n$. A convex subset F of C is a face of C whenever the following condition holds:

- if $x_1, x_2 \in C$ is such that $(1-\lambda)x_1 + \lambda x_2 \in F$ for some $0 < \lambda < 1$, then $x_1, x_2 \in F$.

So: if a relative interior point of the line segment between two points of C lies in F, then the whole line segment between these two points lies in F.

Note: the empty set and C itself are (trivial) faces of C.

Example:

- faces of the unit square and unit circle

Exposed faces

Definition. Let $C \subseteq \mathbb{R}^n$ be a convex set and H a supporting hyperplane of C. Then the intersection $C \cap H$ is called an exposed face of C.

Relation between faces and exposed faces:

- Let C be a nonempty convex set in $\mathbb{R}^n$. Then each exposed face of C is also a face of C.
- For polyhedra: exposed faces and faces are the same!

Extreme points and extreme halflines

Definition. If $\{x\}$ is a face of a convex set C, then x is called an extreme point of C. (So: face of dimension 0)

- Equivalently: $x \in C$ is an extreme point of C if and only if whenever $x_1, x_2 \in C$ satisfies $x = (1/2)x_1 + (1/2)x_2$, then $x_1 = x_2 = x$.
- Example: what are the extreme points if a polytope $P = \mathrm{conv}\,(\{x_1, x_2, \ldots, x_t\})$?

Definition. Consider an unbounded face F of C that has dimension 1. Since F is convex, F must be either a line segment, a line or a halfline (i.e., a set $\{x_0 + \lambda z : \lambda \geq 0\}$). If F is a halfline, we call F an extreme halfline of C.

Inner description of closed convex sets

Theorem

Let $C \subseteq \mathbb{R}^n$ be a nonempty and line-free closed convex set. Then C is the convex hull of its extreme points and extreme halflines.

The bounded case is called Minkowski's theorem.

Corollary

If $C \subseteq \mathbb{R}^n$ is a compact convex set, then C is the convex hull of its extreme points.

Representation of polyhedra

Consider a polyhedron

$$P = \{x \in \mathbb{R}^n : Ax \leq b\}$$

A point $x_0 \in P$ is called a vertex of P if x_0 is the (unique) solution of n linearly independent equations from the system $Ax = b$.

The following says: Extreme point = vertex

Proposition

Let $x_0 \in P$. Then x_0 is a vertex of P if and only if x_0 is an extreme point of P.

Main theorem for polyhedra

Theorem

Each polyhedron $P \subseteq \mathbb{R}^n$ may be written as

$$P = \mathrm{conv}\,(V) + \mathrm{cone}\,(Z)$$

for finite sets $V, Z \subset \mathbb{R}^n$. In particular, if P is pointed, we may here let V be the set of vertices and let Z consist of a direction vector of each extreme halfline of P.
Conversely, if V and Z are finite sets in $\mathbb{R}^n$, then the set $P = \mathrm{conv}\,(V) + \mathrm{cone}\,(Z)$ is a polyhedron. i.e., there is a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ for some m such that

$$\mathrm{conv}\,(V) + \mathrm{cone}\,(Z) = \{x \in \mathbb{R}^n : Ax \leq b\}.$$

6. Convex functions

- convex functions of a single variable
- ... of several variables
- characterizations
- properties, and optimization

Convex function - one variable

Definition. Let $f : \mathbb{R} \to \mathbb{R}$. We say that f is **convex** if

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y)$$

holds for every $x, y \in \mathbb{R}$ and every $0 \le \lambda \le 1$. Extension: $f : [a, b] \to \mathbb{R}$

Geometric interpretation: "graph below secant".

Examples:

- $f(x) = x^2$ (or $f(x) = (x-a)^2$)
- $f(x) = x^n$ for $x \ge 0$
- $f(x) = |x|$
- $f(x) = e^x$
- $f(x) = -\log x$
- $f(x) = -x \log x$

Increasing slopes

Here is a characterization of convex functions. And it also works even when f is not differentiable!

Proposition

A function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if for each $x_0 \in \mathbb{R}$ the slope function

$$x \to \frac{f(x) - f(x_0)}{x - x_0}.$$

is increasing on $\mathbb{R} \setminus \{x_0\}$.

Differentiability

The left-sided derivative of f at x_0 is defined by

$$f'_{-}(x_0) := \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}.$$

provided this limit exists. Similar: right-sided derivative $f'_{+}(x_0)$.

Theorem

Let $f : I \to \mathbb{R}$ be a convex function defined on an interval I. Then f has both left-and right-sided derivatives at every interior point of I. Moreover, if $x, y \in I$ and $x < y$, then

$$f'_{-}(x) \leq f'_{+}(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_{-}(y) \leq f'_{+}(y).$$

In particular, both f'_{-} and f'_{+} are increasing functions.

Criterion: derivatives

Theorem

Let $f : I \to \mathbb{R}$ be a continuous function defined on an open interval I.

(i) If f has an increasing left-derivative (or an increasing right-derivative) on, then f is convex.

(ii) If f is differentiable, then f is convex if and only if f' is increasing. If f is two times differentiable, then f is convex if and only if $f'' \geq 0$ in I.

Convex functions are "essentially continuous"!

Corollary

Let $f : [a, b] \to \mathbb{R}$ *be convex and define* $M = \max\{-f'_+(a), f'_-(b)\}$. *Then*

$$|f(y) - f(x)| \le M\,|y - x| \quad \text{for all } x, y \in [a, b].$$

In particular, f *is continuous at every interior point of* I.

Generalized derivative: the subdifferential

- Differentiability: one can show that each convex function is differentiable almost everywhere; the exceptional set is countable.
- We now look further at derivatives of convex functions.

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. For each $x \in \mathbb{R}$ we associate the closed interval

$$\partial f(x) := [f'_-(x), f'_+(x)].$$

which is called the subdifferential of f at x. Each point $s \in \partial f(x)$ is called a subderivative of f at x.

- By a previous result: $\partial f(x)$ is a nonempty and finite (closed) interval for each $x \in \mathbb{R}$.
- Moreover, f is differentiable at x if and only if $\partial f(x)$ contains a single point, namely the derivative $f'(x)$.

Corollary

Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function and let $x_0 \in \mathbb{R}$. Then, for every $s \in \partial f(x_0)$, the inequality

$$f(x) \geq f(x_0) + s \cdot (x - x_0)$$

holds for every $x \in \mathbb{R}$.

Proof: Let $s \in \partial f(x_0)$. Due to Theorem 9 the following inequality holds for every $x < x_0$:

$$(f(x) - f(x_0))/(x - x_0) \leq f'_-(x_0) \leq s.$$

Thus, $f(x) - f(x_0) \geq s \cdot (x - x_0)$. Similarly, if $x > x_0$ then

$$s \leq f'_+(x_0) \leq (f(x) - f(x_0))/(x - x_0)$$

so again $f(x) - f(x_0) \geq s \cdot (x - x_0)$ and we are done.

Support

Consider again the inequality:

$$f(x) \geq f(x_0) + s \cdot (x - x_0) = L(x)$$

- L can be seen as a linear approximation to f at x_0. We say that L supports f at x_0; this means that $L(x_0) = f(x_0)$ and $L(x) \leq f(x)$ for every x.
- So L underestmates f everywhere!

Global minimum

We call x_0 a global minimum if

$$f(x_0) \leq f(x) \quad \text{for all } x \in \mathbb{R}.$$

Weaker notion: local minimum: smallest function value in some neighborhood of x_0.

- In general it is hard to find a global minimum of a function.
- But when f is convex this is much easier!

The following result may be derived from

$$f(x) \geq f(x_0) + s \cdot (x - x_0) = f(x_0).$$

Corollary

Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then the following three statements are equivalent.

(i) x_0 *is a* *local minimum* *for* f.

(ii) x_0 *is a* *global minimum* *for* f.

(iii) $0 \in \partial f(x_0)$.

Jensen's inequality

Theorem

Let $f : I \to \mathbb{R}$ be a convex function defined on an interval I. If $x_1, \ldots, x_r \in I$ and $\lambda_1, \ldots, \lambda_r \geq 0$ satisfy $\sum_{j=1}^r \lambda_j = 1$, then

$$f\Big(\sum_{j=1}^{r} \lambda_j x_j\Big) \leq \sum_{j=1}^{r} \lambda_j f(x_j).$$

The arithmetic geometric mean inequality follows from this by using $f(x) = -\log x$:

$$\Big(\prod_{j=1}^{r} x_j\Big)^{1/r} \leq (1/r) \sum_{j=1}^{r} x_j$$

Convex functions of several variables

- many results from the univariate case extends to the general case of n variables.

Let $f : C \rightarrow \mathbb{R}$ where $C \subseteq \mathbb{R}^n$ is a convex set. We say that f is convex if

$$f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y)$$

holds for every $x, y \in \mathbb{R}^n$and every $0 \leq \lambda \leq 1$.

- note: need C to be a convex set here
- every linear, or affine, function from $\mathbb{R}^n$ to $\mathbb{R}$ is convex.
- Assume that $f : \mathbb{R}^n \rightarrow \mathbb{R}$is convex and $h : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is affine. Then the composition $f \circ h$ is convex (where $(f \circ h)(x) := f(h(x))$)

Jensen's inequality, more generally

Theorem

Let $f : C \to \mathbb{R}$ be a convex function defined on a convex set $C \subseteq \mathbb{R}^n$. If $x_1, \ldots, x_r \in C$ and $\lambda_1, \ldots, \lambda_r \geq 0$ satisfy $\sum_{j=1}^r \lambda_j = 1$, then

$$f(\sum_{j=1}^{r} \lambda_j x_j) \leq \sum_{j=1}^{r} \lambda_j f(x_j).$$

Note: in (discrete) probability this means

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

The epigraph

Let $f : C \to \mathbb{R}$ where $C \subseteq \mathbb{R}^n$ is a convex set. Define the following set in $\mathbb{R}^{n+1}$ associated with f:

$$\mathrm{epi}\ (f) = \{(x, y) \in \mathbb{R}^{n+1} : y \geq f(x)\}.$$

It is called the epigraph of f.

The following result makes it possible to use results for convex sets to obtain results for convex function (and vice versa). relation.

Theorem

Let $f : C \to \mathbb{R}$ where $C \subseteq \mathbb{R}^n$ is a convex set. Then f is a convex function if and only if $\mathrm{epi}\ (f)$ is a convex set.

Supremum of convex functions

Corollary

Let f_i $(i \in I)$ be a nonempty family of convex functions defined on a convex set $C \subseteq \mathbb{R}^n$. Then the function f given by

$$f(x) = \sup_{i \in I} f_i(x) \quad \text{for } x \in C$$

(the pointwise supremum) is convex.

Example:

- Pointwise supremum of affine functions, e.g. (finite case)

$$f(x) = \max_{i \leq n} (a_i^T x + b_i)$$

- Note: such a function if not differentiable in certain points!

The support function

Let P be a polytope in $\mathbb{R}^n$, say $P = \mathrm{conv}\,(\{v_1, \ldots, v_t\})$. Define

$$\psi_P(c) := \max\{c^T x : x \in P\}.$$

which is the optimal value of this LP problem. This function ψ_P is called the support function of P.

- ψ_P is a convex function! Because it is the pointwise supremum of the linear functions $c \to c^T v_j$ ($j \leq t$). This maximum is attained in a vertex (since the objective function is linear).
- More generally: the support function ψ_C of a compact convex set C is convex. Similar proof, but we take the supremum of an infinite family of linear functions; one for each extreme point of C.
- Here we used Minkowski's theorem saying that a compact convex set is the convex hull of its extreme points.

Directional derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and let $x_0 \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$, $z \neq O$. The directional derivative of f at x_0 is

$$f'(x_0; z) = \lim_{t \to 0} \frac{f(x_0 + tz) - f(x_0)}{t}$$

provided the limit exists. Special case: $f'(x_0; e_j) = \frac{\partial f(x)}{\partial x_j}$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and consider a line $L = \{x_0 + \lambda z : \lambda \in \mathbb{R}\}$ where x_0 is a point on the line and z is the direction vector of L. Define the function $g : \mathbb{R} \to \mathbb{R}$ by

$$g(t) = f(x_0 + tz) \quad \text{for } t \in \mathbb{R}.$$

One can prove that g is a convex function (of a single variable).

- Thus, the restriction g of a convex function f to any line is another convex function.
- A consequence of this result is that a convex function $f : \mathbb{R}^n \to \mathbb{R}$ has one-sided directional derivatives:

$$\begin{aligned} g'_+(0) &= \lim_{t \to 0^+} (g(t) - g(0))/t \\ &= \lim_{t \to 0^+} (f(x_0 + tz) - f(x_0))/t \\ &= f'_+(x_0; z) \end{aligned}$$

Continuity

Theorem

Let $f : C \to \mathbb{R}$ be a convex function defined on an open convex set $C \subseteq \mathbb{R}^n$. Then f is continuous on C.

Characterization of convexity

We now recall a concept from linear algebra: a symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if

$$x^T A x = \sum_{i,j} a_{ij} x_i x_j \geq 0 \quad \text{for each } x \in \mathbb{R}^n.$$

A useful fact is that A is positive semidefinite if and only if all the eigenvalues of A are (real and) nonnegative.

Theorem (Characterization via the Hessian)

Let f be a real-valued function defined on an open convex set $C \subseteq \mathbb{R}^n$ and assume that f has continuous second-order partial derivatives on C.

Then f is convex if and only if the Hessian matrix $H_f(x)$ is positive semidefinite for each $x \in C$.

Examples

- Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix which is positive semidefinite and consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = x^T A x = \sum_{i,j} a_{ij} x_i x_j.$$

  Then it is easy to check that $H_f(x) = A$ for each $x \in \mathbb{R}^n$. Therefore, f is a convex function.
- A symmetric $n \times n$ matrix A is called diagonally dominant if

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}| \quad (i \leq n)$$

  If all these inequalities are strict, A is strictly diagonally dominant. These matrices arise in many applications, e.g. splines and differential equations.
- It can be shown that every symmetric diagonally dominant matrix with positive diagonal is positive semidefinite.

Differentiability

A function f defined on an open set in $\mathbb{R}^n$ is said to be differentiable at a point x_0 in its domain if there is a vector d such that

$$\lim_{h \to O} (f(x_0 + h) - f(x_0) - d^T h)/\|h\| = 0.$$

Then d is unique; called the gradient of f at x_0.

Assume that f is differentiable at x_0 and the gradient at x_0 is d. Then, for each nonzero vector z,

$$f'(x_0; z) = d^T z.$$

Partial derivatives, gradients

Theorem

Let f be a real-valued convex function defined on an open convex set $C \subseteq \mathbb{R}^n$. Assume that all the partial derivatives exist at a point $x \in C$. Then f is differentiable at x.

Theorem

Let $f : C \rightarrow \mathbb{R}$ be a differentiable function defined on an open convex set $C \subseteq \mathbb{R}^n$. Then the following conditions are equivalent:

(i) f *is convex.*

(ii) $f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$ for all $x, x_0 \in C$.

(iii) $(\nabla f(x) - \nabla f(x_0))^T (x - x_0) \geq 0$ for all $x, x_0 \in C$.

Consider a convex function f and an affine function h, both defined on a convex set $C \subseteq \mathbb{R}^n$.
We say that $h : \mathbb{R}^n \to \mathbb{R}$ supports f at x_0 if $h(x) \leq f(x)$ for every x and $h(x_0) = f(x_0)$.

Theorem

Let $f : C \to \mathbb{R}$ be a convex function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f has a supporting (affine) function at every point. Moreover, f is the pointwise supremum of all its (affine) supporting functions.

Global minimum

Corollary

Let $f : C \to \mathbb{R}$ be a differentiable convex function defined on an open convex set $C \subseteq \mathbb{R}^n$. Let $x^* \in C$. Then the following three statements are equivalent.

(i) x^* *is a local minimum for* f.

(ii) x^* *is a global minimum for* f.

(iii) $\nabla f(x^*) = O$ *(i.e., all partial derivatives at* x^* *are zero).*

Subgradients

Definition. Let f be a convex function and $x_0 \in \mathbb{R}^n$. Then $s \in \mathbb{R}^n$ is called a subgradient of f at x_0 if

$$f(x) \geq f(x_0) + s^T(x - x_0) \quad \text{for all } x \in \mathbb{R}^n$$

- The set of all subgradients of f at x_0 is called the subdifferential of f at x_0, and it is denoted by $\partial f(x_0)$.

Here is the basic result on the subdifferential.

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and $x_0 \in \mathbb{R}^n$. Then $\partial f(x_0)$ is a nonempty, compact and convex set in $\mathbb{R}^n$.

Global minimum, again

Moreover, we have the following theorem on minimum of convex functions.

Corollary

Let $f : C \to \mathbb{R}$ be a convex function defined on an open convex set $C \subseteq \mathbb{R}^n$. Let $x^* \in C$. Then the following three statements are equivalent.

(i) x^* *is a* *local minimum* *for* f.

(ii) x^* *is a* *global minimum* *for* f.

(iii) $O \in \partial f(x^*)$ *(*O *is a subgradient).*

Final comments ...

- This means that convex problems are attractive, and sometimes other problems are reformulated/modfied into convex problems
- Algorithms exist for minimizing convex functions, with or wiothout constraints.
- So gradient-like methods for differentiable functions are extended into subgradient methods for general convex functions.
- More complicated, but efficient methods exist.