Convexity

Kaarel Hänni

August 2023

1 Introduction

The goals of this piece are the following:

1. to give a fairly solid introduction of the basic concepts required to make sense of the kind of business convex optimization is — namely, *convex sets* and *convex functions*;
2. to try to briefly explain why convex optimization is 'easy';
3. to briefly explain the pipeline for using convex optimization.

2 Convex sets

We say a subset C of $\mathbb{R}^n$ is *convex* if it contains all line segments between points in it:

Definition 2.1. A set $C \subseteq \mathbb{R}^n$ is *convex* if for any points x and y in C, and $t \in [0,1]$, the point $tx + (1-t)y$ is also in C.

Here are some more important examples of convex sets:

- A **subspace** $V \subseteq \mathbb{R}^n$.
- An **affine subspace** — for a point $a \in \mathbb{R}^n$ and a subspace V, the set $a + V \subseteq \mathbb{R}^n$.[1]
- A **convex cone** — a set $C \subseteq \mathbb{R}^n$ such that whenever $x, y \in C$, for any positive coefficients $a, b > 0$, also $ax + by \in C$.
- A **half-space** — for a vector $v \in \mathbb{R}^n$ and constant $b \in \mathbb{R}$, the set of all points $x \in \mathbb{R}^n$ with $x \cdot v \leq b$.
- A **polyhedron** — for a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$, the set of all $x \in \mathbb{R}^n$ such that $Ax \leq b$.[2]
- A **polytope** — a bounded polyhedron.[3]
- A **ball** — for some $R > 0$, the set of all $x \in \mathbb{R}^n$ with $\|x\| \leq R$

Here's an important basic property of convex sets:

Proposition 2.0.1. *The intersection of a collection $\mathcal{C}$ of convex sets is convex.*

Proof. If x, y are in the intersection $\bigcap_{C \in \mathcal{C}} C$, then x, y are also in each set $C \in \mathcal{C}$. Therefore, for any $t \in [0,1]$, the point $tx+(1-t)y$ is in each $C \in \mathcal{C}$ (since each C is convex). Thus, $tx+(1-t)y \in \bigcap_{C \in \mathcal{C}} C$, which is what we wanted to show. □

For any set, there is a canonical way to construct a corresponding convex set:

Definition 2.2. The convex hull $\mathrm{conv}(S)$ of a set $S \subseteq \mathbb{R}^n$ is the intersection of all convex subsets of $\mathbb{R}^n$ that contain S.[4]

[1] Equivalently, an affine subspace is the set of all solutions $x \in \mathbb{R}^n$ to a particular equation $Ax = b$ for some matrix $A \in \mathbb{R}^{m \times n}$ and vector $b \in \mathbb{R}^m$.

[2] In other words: a polyhedron is the intersection of a bunch of half-spaces.

[3] This turns out to be equivalent to being the convex hull of a finite set of points.

[4] Note that there is indeed at least one such convex subset, because $\mathbb{R}^n \subseteq \mathbb{R}^n$ itself is convex.

A few remarks on $\text{conv}(S)$:

1. Proposition 2.0.1 implies that $\text{conv}\, S$ is indeed convex.
2. What's more, since any convex set in the intersection defining $\text{conv}(S)$ contains S, we also have that $S \subseteq \text{conv}(S)$.
3. $\text{conv}(S)$ is contained in every convex set which contains S (because it is the intersection of all such sets).

In summary, $\text{conv}(S)$ is the smallest convex set containing S. Here's an analogue of linear combinations useful for thinking about convex sets:

Definition 2.3. We say $y \in \mathbb{R}^n$ is a *convex combination* of $x_1, \ldots, x_k \in \mathbb{R}^n$ if there are $0 \leq t_1, \ldots, t_k \leq 1$ such that $t_1 + t_2 + \cdots + t_k = 1$ and $y = t_1x_1 + t_2x_2 + \cdots + t_kx_k$.

I claim without proof that a set C is convex if and only if it contains all convex combinations of finite collections of its points.[5] The notion of convex combinations lets us provide a more explicit definition of the convex hull:

Proposition 2.0.2. *The convex hull of $S \subseteq \mathbb{R}^n$ is the set of all convex combinations of all finite collections of points in S.*[6]

Proof. Let Q be the set of all convex combinations of finite collections of points in S. We will first show that $Q \subseteq \text{conv}(S)$. Note that if C is convex with $S \subseteq C$, then C must contain all convex combinations of its points by the unproven claim above, and must therefore, in particular, contain all convex combinations of points of S. In other words, $Q \subseteq C$ for any convex C containing S. Therefore, Q is in the intersection of all $C \supseteq S$, and so $Q \subseteq \text{conv}(S)$.

Let us now show that $\text{conv}(S) \subseteq Q$. Note that Q is convex. This is because a convex combination of convex combinations of points of S is a convex combination of points of S, so Q contains any convex combination of its points. Note also that $S \subseteq Q$. So Q is a convex set containing S. Therefore, $\text{conv}(S) \subseteq Q$.

We've established that $Q \subseteq \text{conv}(S)$ and that $\text{conv}(S) \subseteq Q$. It follows that $Q = \text{conv}(S)$. □

3 Convex functions

A convex function is one with a graph such that all chords drawn on the graph are above the graph:

Definition 3.1. Let $C \subseteq \mathbb{R}^n$ be a convex set. We say a function $f\colon C \to \mathbb{R}$ is convex if for any $x, y \in C$ and $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

We say a function g is *concave* if $-g$ is convex, that is, if

$$f(tx + (1-t)y) \geq tf(x) + (1-t)f(y).$$

Here are a number of alternative characterizations of convexity:

Theorem 3.1.

1. *(Jensen's inequality — discrete version) A function $f\colon C \to \mathbb{R}$ is convex iff for any $0 \leq t_1, \ldots, t_k \leq 1$ with $t_1 + \cdots + t_k = 1$ and $x_1, \ldots, x_k \in \mathbb{R}^n$, we have*

$$f\left(t_1x_1 + \cdots + t_kx_k\right) \leq t_1f(x_1) + \cdots + t_kf(x_k).$$

2. *(Jensen's inequality)*[7] *A function $f\colon C \to \mathbb{R}$ is convex iff for any random variable X taking values in C,*

$$f\left(\mathbb{E}[X]\right) \leq \mathbb{E}[f(X)]$$

3. *(First derivative condition) A continuously differentiable function $f\colon C \to \mathbb{R}$ is convex iff for all x, y, we have*

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

[5] The problem set asks you to prove this.

[6] In fact, convex combinations of only $n + 1$ points suffice. See `https://en.wikipedia.org/wiki/Carath%C3%A9odory%27s_theorem_(convex_hull)`.

[7] Note that the case above is the special case of this where X takes values in a finite subset of C.

4. *(Second derivative condition) A twice continuously differentiable function $f\colon C \to \mathbb{R}$ is convex iff at all x, the Hessian $H_f(x)$ is positive-semidefinite.*

Two remarks:

- Analogous statements, i.e. the above with each inequality having its sign flipped, hold for concave functions.
- The equality cases of these inequalities have simple characterizations. For instance, for a convex function, $f(\mathbb{E}[X]) = \mathbb{E}[f(X)]$ iff f is equal to a linear function on the support of X.

A pattern that is often useful is to show that the Hessian is positive-definite, and to conclude from this that Jensen's inequality holds; here's an example:

Theorem 3.2 (Weighted AM≥GM)**.** *Let $a_1, \dots, a_k \geq 0$, and $0 \leq w_1, \dots, w_k \leq 1$ with $w_1 + \cdots + w_k = 1$. Then*

$$\sum_{i=1}^{k} a_i w_i \geq \prod_{i=1}^{k} a_i^{w_i}.$$

Proof. Since $(\log x)'' = \left(\frac{1}{x}\right)' = -\frac{1}{x^2} < 0$, we have that $\log x$ is concave. Taking logarithms of both sides of the desired inequality, it remains to show that $\log\left(\sum_{i=1}^{k} a_i w_i\right) \geq \sum w_i \log a_i$. Note that this is the discrete version of Jensen's inequality for $\log x$. □

4 Convex optimization

4.1 Briefly on why convex optimization is easy

A *convex optimization problem* is the problem of minimizing a convex function (on a convex domain). Here's a central reason why convex optimization is easy:

Theorem 4.1. *Let $C \subseteq \mathbb{R}^n$ be a convex set, and let $f\colon C \to \mathbb{R}$ be a convex function. If $x_0 \in C$ is a local minimum of f, then $x_0 \in C$ is also a global minimum of f.*

Proof. Suppose x_0 is not a global minimum of f; then there is $x_1 \neq x_0$ with $f(x_1) < f(x_0)$. Then for any $t \in [0, 1]$, we have $f((1-t)x_0 + tx_1) \leq (1-t)f(x_0) + tf(x_1) < (1-t)f(x_0) + tf(x_0) = f(x_0)$. Since if we pick t to be arbitrarily small, the point $(1-t)x_0 + tx_1$ is in an arbitrarily small ball around x_0, there is no ball around x_0 in which x_0 is a minimum of f. Hence, x_0 is not a local minimum of f, a contradiction. □

(Analogously, maximizing a concave function on a convex domain is also easy.)

4.2 Briefly on what convex optimization can usually be made to look like

The following is called the standard form of a convex optimization problem:

$$\min_x f(x)$$

$$\text{subject to } g_i(x) \leq 0$$

$$\text{and } h_j(x) = 0$$

where the functions $g_1, \dots, g_m\colon \mathbb{R}^n \to \mathbb{R}$ are convex and the functions $h_1, \dots, h_p\colon \mathbb{R}^n \to \mathbb{R}$ are affine.

4.3 Briefly on solving convex optimization problems

There is a zoo of different types of convex optimization problems of various levels of generality, with various solution methods. But here is what I think is a fairly general strategy. In the unconstrained case, one can pretty much just do gradient descent, get to a local minimum, and conclude one has found a global minimum. [8] One can get rid of equality constraints by reparametrizing. And one can handle inequality constraints by adding a convex barrier function to the loss — that is, a function that is mostly negligible inside the feasible region, but that blows up near the boundary. Modulo finding at least one feasible point to start from, this turns everything else to the unconstrained case.

[8] I believe there is a large literature on properties of this, e.g. convergence speed, that I will not cover.

4.4 Briefly on turning problems into convex optimization problems

I read somewhere that while there are remaining research problems in convex optimization, solving convex optimization problems is essentially a technology by now. I'm also told that there is some remaining art in reformulating a problem as a linear programming problem. (Of course, there is also a bag of standard tricks for this step.) For instance, for a data set $x_1, \ldots, x_m \in \mathbb{R}^n$ with corresponding labels $y_1, \ldots, y_m \in \mathbb{R}$, consider the following problem (called the *Chebyshev approximation problem*):

$$\min_{a \in \mathbb{R}^n} \max_{i=1,\ldots,k} |a^T x_i - y_i|.$$

This is just like the least-squares problem, but instead of minimizing the L^2 norm of the vector of differences between predictions and labels, we are minimizing the L^∞ norm of this vector. We can reformulate this as a linear programming problem by introducing an additional variable λ that we think of as the maximum:

$$\min_{a \in \mathbb{R}^n, \lambda \in \mathbb{R}} \lambda$$

$$\text{subject to } a^T x_i - b_i \leq \lambda \text{ and } -(a^T x_i - b_i) \leq \lambda \text{ for all } 1 \leq i \leq m.$$