Convex Optimization Overview

Conner DiPaolo

Contents

1 Introduction

Convex optimization in a large way influences the way people think about and phrase machine learning problems. Almost all problems we will see in our studies are developed or can be viewed as optimization problems. Some problems, like the Support Vector Machine you will all see in the coming weeks, are almost entirely based in the heart of convex optimization.

We don't plan on bringing you all up to speed completely on the art of Convex Optimization, but hopefully in two sections you will know enough to be able to think about problems in new, interesting ways that will aid your studies in and out of machine learning.

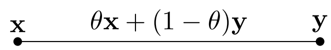

Figure 1: **Convex Combination**. The line segment between $\mathbf{x}$ and $\mathbf{y}$ above represents every possible combination $\{\theta\mathbf{x} + (1-\theta)\mathbf{y} : 0 \leq \theta \leq 1\}$.

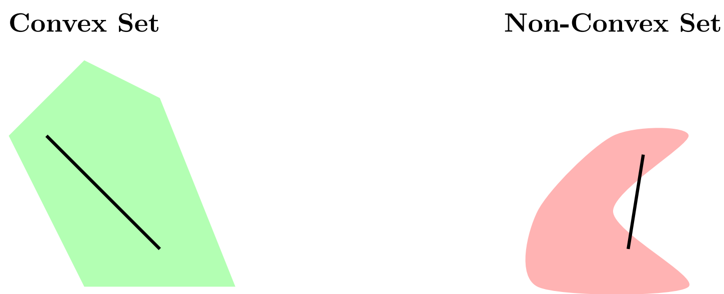

Figure 2: **Set convexity.** Intuitively, a set is convex if a line drawn between any two points within the set lies completely in the set. The convex set to the left is a *convex hull* of it's vertices, meaning the set is constructed as all possible convex combinations of its vertices. These sets are subsets of $\mathbb{R}^2$.

The only real prerequisite for this material is a strong confidence in linear algebra and familiarity with matrix calculus.

2 Convex Sets

The first step into examining convexity is defining what a **convex set** is when given, for example, a subset of the real numbers, or the set of matrices.

Definition 2.1 (Convex Combination)**.** *In the $n = 2$ case, the convex combination of points $\mathbf{x}$ and $\mathbf{y}$ in an affine space is*

$$\theta\mathbf{x} + (1-\theta)\mathbf{y}$$

where $0 \leq \theta \leq 1$. Intuitively, this is the line segment between $\mathbf{x}$ and $\mathbf{y}$ (consider $\theta = 0$ and $\theta = 1$), as seen in Figure 1. More generally, a convex combination of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in an affine space (vector spaces included) is the combination

$$\boldsymbol{\theta}_1\mathbf{x}_1 + \boldsymbol{\theta}_2\mathbf{x}_2 + \cdots + \boldsymbol{\theta}_n\mathbf{x}_n = \boldsymbol{\theta}^\top\mathbf{x}$$

where $\theta_i \geq 0$ and $\mathbf{1}^\top\theta = 1$. That is, $\boldsymbol{\theta}$ lies on the standard probability simplex.

With this definition of a convex combination we can define a convex set:

Definition 2.2 (Convex Set)**.** *A set C is convex if, given $\mathbf{x}, \mathbf{y} \in C$, every convex combination of $\mathbf{x}$ and $\mathbf{y}$ is still in C. Mathematically,*

$$\theta\mathbf{x} + (1-\theta)\mathbf{y} \in C. \qquad (0 \leq \theta \leq 1)$$

See Figure 2 for an illustration. Intuitively, this means that every line segment between any two points in C is contained entirely within C.

2.1 Examples of Convex and Non-Convex Sets

Example 2.1 (All of $\mathbb{R}^n$)**.** $\mathbb{R}^n$ *is convex.*

Proof. As a vector space, for any $\theta_1, \theta_2 \in \mathbb{R}$ and $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

$$\theta_1\mathbf{x} + \theta_2\mathbf{y} \in \mathbb{R}^n.$$

Thus, restricting $\theta_2 = 1 - \theta_1$ and $0 \le \theta_1 \le 1$ does not change this fact, and any convex combination is also in $\mathbb{R}^n$, making the set convex by definition. ■

Example 2.2 (The Non-Negative Orthant $\mathbb{R}^n_+$)**.** *The set of all vectors*

$$\{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \text{ and } \mathbf{x}_i \ge 0\}$$

is convex.

Proof. Left as an exercise to the reader. ■

Example 2.3 (Closed Intervals in $\mathbb{R}$)**.** *Let $C = [a, b]$ be a subset of the real numbers where $a \le b$. Then C is convex.*

Proof. Suppose, without loss of generality, that $x_1 \le x_2$ where $x_1, x_2 \in [a, b]$. Now let $0 \le \theta \le 1$. Then

$$\theta x_1 + (1 - \theta)x_2 \le \theta x_2 + (1 - \theta)x_2 = x_2$$

because $\theta x_1 \le \theta x_2$. Similarly,

$$\theta x_1 + (1 - \theta)x_2 \ge \theta x_1 + (1 - \theta)x_1 = x_1$$

because $(1 - \theta)x_2 \ge (1 - \theta)x_1$. Thus

$$a \le x_1 \le \theta x_1 + (1 - \theta)x_2 \le x_2 \le b,$$

and hence

$$\theta x_1 + (1 - \theta)x_2 \in [a, b].$$

Therefore, by the definition of convexity, $[a, b] \subseteq \mathbb{R}$ is convex for any $a \le b$. ■

Example 2.4 (The Set of All Complex Hermetian Matrices)**.** *Let $C = \{A : A \in \mathbb{C}^{n\times n}$ and $A^* = A\}$. C is convex.*

Proof. Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian matrices and $0 \le \theta \le 1$. Then

$$\begin{aligned}
(\theta A + (1 - \theta)B)^* &= (\theta A)^* + [(1 - \theta)B]^* && (1)\\
&= \theta A^* + (1 - \theta)B^* && (2)\\
&= \theta A + (1 - \theta)B && \text{(because } A^* = A \text{ and } B^* = B) \\
& && (3)
\end{aligned}$$

Thus every convex combination of Hermitian matrices is Hermitian, and by the definition of convexity the set is convex. ■

Example 2.5 (The Set of All Real Symmetric Matrices)**.** *Let $C = \{A : A \in \mathbb{R}^{n\times n}$ and $A^\top = A\}$. C is convex.*

Proof. Left as an exercise to the reader. ■

Example 2.6 (The Set of All Linear Matrix Inequalities)**.** *Let A_i and B be symmetric $n \times n$ matrices and $\mathbf{x} \in \mathbb{R}^n$. Let $C = \{\mathbf{x} : A(\mathbf{x}) \preceq B\}$ where $A(\mathbf{x}) = \mathbf{x}_1 A_1 + \cdots + \mathbf{x}_k A_k$. C is convex.*

Proof. Let $0 \le \theta \le 1$ and $\mathbf{x}, \mathbf{y} \in C$. Then

$$\begin{aligned}
A(\theta\mathbf{x} + (1 - \theta)\mathbf{y}) &= \sum_i [\theta\mathbf{x}_i + (1 - \theta)\mathbf{y}_i]\, A_i \\
&= \theta \sum_i \mathbf{x}_i A_i + (1 - \theta) \sum_i \mathbf{x}_i A_i \\
&= \theta A(\mathbf{x}) + (1 - \theta)A(\mathbf{y}) \\
&\le \theta B + (1 - \theta)B \\
&= B.
\end{aligned}$$

Thus any convex combination of elements of C is contained in C and by definition C is convex. ■

Example 2.7 (The Space of Probability Distributions)**.** *Let $\mathcal{P}$ be the space of continuous probability distributions over $\mathbb{R}^n$. That is, every element of $\mathcal{P}$ defines a unique probability density function $\mathbb{P}(x) \geq 0$ such that*

$$\int_{\mathbb{R}^n} \mathbb{P}(x)dx = 1.$$

$\mathcal{P}$ is convex.

Proof. Let f and h be valid probability distributions from $\mathcal{P}$. That is,

$$f, h \in \left\{ \mathbb{P}(x) : \int_{\mathbb{R}^n} \mathbb{P}(x)dx = 1 \text{ and } \mathbb{P}(x) \geq 0 \right\}.$$

Now let $0 \leq \theta \leq 1$. Then

$$\theta f(x) + (1-\theta)h(x) \geq 0$$

as a positive combination of positive functions. Similarly,

$$\begin{aligned} \int_{\mathbb{R}^n} [\theta f(x) + (1-\theta)h(x)]\, dx &= \theta \int_{\mathbb{R}^n} f(x)dx + (1-\theta) \int_{\mathbb{R}^n} h(x)dx \\ &= \theta + (1-\theta) = 1. \end{aligned}$$

Thus every convex combination of probability distributions over $\mathbb{R}^n$ is a valid distribution (often called a mixture), and therefore the set of all valid probability distributions over $\mathbb{R}^n$ is convex itself. ■

Example 2.8 (Disjoint Intervals in $\mathbb{R}$)**.** *Let $N = [a, b] \cup [c, d]$ where $a \leq b < c \leq d$. N is **not** convex.*

Proof. Let $b, c \in N$ be as described above. Then for $0 < \theta < 1$ (not we aren't including inequality),

$$\theta b + (1-\theta)c \notin N.$$

Thus not *every* convex combination of elements in N is in N, and N is not convex. ■

Example 2.9 (The Set of All Stochastic Matrices)**.** *The set of all matrices A such that for all elements $0 \leq A_{ij} \leq 1$ and each row sums to 1,*

$$M = \{A : A \in \mathbb{R}^{m \times n} \text{ and } 0 \leq A_{ij} \leq 1 \text{ and } A\mathbf{1} = \mathbf{1}\},$$

is convex.

Note that this set is the set of all matrix representations of every possible Markov Chain.

Proof. Let $A, B \in M$ and $0 \leq \theta \leq 1$. Then

$$c = [\theta A + (1-\theta)B]_{ij} = \theta A_{ij} + (1-\theta)B_{ij}$$

satisfies $0 \leq c \leq 1$ as $A_{ij}, B_{ij} \in [0, 1]$ and closed intervals on $\mathbb{R}$ are convex as seen in Example 2.3.

Further, because $A\mathbf{1} = \mathbf{1}$ and $B\mathbf{1} = \mathbf{1}$,

$$[\theta A + (1-\theta)B]\, \mathbf{1} = \theta A\mathbf{1} + (1-\theta)B\mathbf{1} = \theta\mathbf{1} + (1-\theta)\mathbf{1} = \mathbf{1},$$

as desired. Thus every convex combination of elements withn M remains in M, and by definition the set M is convex. ■

Example 2.10 (Norm Balls)**.** *For any valid norm $||\cdot||$ and scalar $r \geq 0$, the set*

$$C = \{\mathbf{x} : ||\mathbf{x}|| \leq r\}$$

is convex.

Proof. Let $\mathbf{x}$ and $\mathbf{y}$ be elements from C and $0 \leq \theta \leq 1$. Then

$$||\theta\mathbf{x} + (1-\theta)\mathbf{y}|| \leq ||\theta\mathbf{x}|| + ||(1-\theta)\mathbf{y}|| = \theta||\mathbf{x}|| + (1-\theta)||\mathbf{y}|| \leq \theta r + (1-\theta)r = r,$$

as desired, where we used the Triangle Inequality for the first step and the second used the homogeneity of norms. Thus every convex combination of elements in C is within C, and therefore C is convex. ■

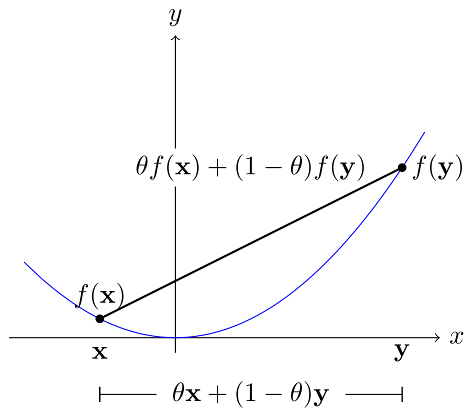

Figure 3: **Convex Functions**. This function $f : [a, b] \mapsto \mathbb{R}$ is convex because it's domain is convex and every line between two points on the function lies above the function.

3 Convex Functions

You likely have already seen convex functions from your time in Calculus I in high school or something similar. Nevertheless, our treatment will be much more rigorous (though not *too* rigorous) and be very applicable to our later studies.

The most important fact to know about convex functions is that they

(a) Have a unique global minimizer $\mathbf{x}^\star$.

(b) Generally have efficient algorithms for finding that minimizer.

3.1 Convex and Concave Functions

Our idea of a convex set will be useful in considering convex functions.

Definition 3.1 (Convex)**.** *Given a convex set C, a function $f : C \mapsto \mathbb{R}$ is convex if, for $0 \le \theta \le 1$, given any $\mathbf{x}, \mathbf{y} \in C$,*

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

Intuitively, this means that a convex function always lies under a line between any two points where it is evaluated. This is seen in Figure 3.

Definition 3.2 (Concave Functions)**.** *Given a convex set C, function $f : C \mapsto \mathbb{R}$ is concave if $-f$ is convex. That is, for $0 \le \theta \le 1$, given any $\mathbf{x}, \mathbf{y} \in C$,*

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

Theorem 3.1 (Restriction To a Line)**.** *A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex if and only if f is convex when restricted to any line that intersects its domain. Mathematically, f is convex if and only if for all $x \in C$ and any $\mathbf{v} \in \mathbb{R}^n$ and $t \in \mathbb{R}$,*

$$g(t) = f(\mathbf{x} + t\mathbf{v})$$

is convex where $g : \{t : \mathbf{x} + t\mathbf{v} \in C\} \mapsto \mathbb{R}$

Proof. Omitted. See *Convex Optimization* by Boyd and Vandenberghe. ■

3.2 First Order Conditions: $f(x) \geq f(y) + \nabla f(y)^\top (x - y)$

Generally speaking, determining convexity from the definition is hard. In this section we will develop more tractable conditions that will help us determine if an arbitrary function is convex. The following necessary and sufficient condition for convexity is called the First Order Condition because it relies only on the first derivative.

Theorem 3.2 (First Order Conditions)**.** *Let f be a function mapping from some convex set C to $\mathbb{R}$. Then f is convex if and only if, given $\mathbf{x}$ and $\mathbf{y}$ from C,*

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

Proof. Omitted. See *Convex Optimization* by Boyd and Vandenberghe. ■

3.2.1 Examples On Determining Convexity

Theorem 3.3 (Linear Functions are Both Concave and Convex)**.** *Given a function $f : C \mapsto \mathbb{R}$ where C is a convex set and for any $a, b \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}$*

$$f(a\mathbf{x} + b\mathbf{y}) = af(\mathbf{x}) + bf(\mathbf{y}),$$

f is both convex and concave. Note that f is the definition of a linear function.

Proof. Let $\mathbf{x}, \mathbf{y}$ be any elements of C and $0 \leq \theta \leq 1$. Then by linearity

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

Thus by definition f is both convex and concave. ■

Example 3.1. *$f(x) = x^2$ is convex.*

Proof. We will use the much more convenient second order condition for this problem later. Consider $x, y \in \mathbb{R}$ and $0 \leq \theta \leq 1$. Then

$$\begin{aligned} f(\theta x + (1-\theta)y) &= (\theta x + (1-\theta)y)^2 \\ &= (\theta x + (1-\theta)y)(\theta x + (1-\theta)y) \\ &= \theta^2 x^2 + 2\theta(1-\theta)xy + (1-\theta)^2 y^2. \end{aligned}$$

f will be convex if and only if

$$\theta f(x) + (1-\theta) f(y) - f(\theta x + (1-\theta)y) \geq 0$$

for all x, y. We have

$$\begin{aligned} g = \theta f(x) + (1-\theta) f(y) - f(\theta x + (1-\theta)y) &= \theta x^2 + (1-\theta)y^2 - \theta^2 x^2 - 2\theta(1-\theta)xy - (1-\theta)^2 y^2 \\ &= \theta(1-\theta)x^2 - 2\theta(1-\theta)xy + \theta(1-\theta)y^2 \\ &= \theta(1-\theta)(x-y)^2 \end{aligned}$$

We know $(x-y)^2 \geq 0$ for all x and y. Similarly, when $\theta \in [0,1]$, both θ and $1-\theta$ are positive, so this expression is positive. Thus f must be convex, as desired. ■

3.3 Second Order Conditions: $\nabla^2 f \succeq 0$

Here we discuss the most useful condition for determining convexity.

Theorem 3.4 (Second Order Conditions)**.** *A twice-differentiable function $f : C \mapsto \mathbb{R}$ is convex if and only if and only if the Hessian*

$$\nabla^2 f \succeq 0$$

Proof. Omitted. ■

3.3.1 More Examples On Determining Convexity

Example 3.2 (Quadratic Functions)**.** *The function $f : \mathbb{R}^n \mapsto \mathbb{R}$ defined by*

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

is convex if $A \succeq 0$.

Proof. We know

$$\nabla^2 f = 2A.$$

Similarly we know that $2A \succeq 0$ if and only if $A \succeq 0$. Thus by the second order conditions f is convex if and only if $A \succeq 0$.

We can also show that f is concave if $A \preceq 0$. ■

Example 3.3. *$f(x) = a\ e^x$ is convex if $a \geq 0$.*

Proof. The Hessian

$$\nabla^2 f = \frac{d^2 f}{dx^2} = a\ e^x \geq 0$$

as long as $a \geq 0$. Thus by the second order conditions f is convex if $a \geq 0$. ■

Example 3.4. *The function $f : \mathbb{R}^n \mapsto \mathbb{R}$ defined by*

$$f(\mathbf{x}) = e^{\mathbf{x}^\top \mathbf{x}} = e^{\mathbf{x}_1^2 + \cdots + \mathbf{x}_n^2}$$

is convex.

Proof. Each element of the Hessian

$$\nabla^2 f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} = 4\mathbf{x}_i \mathbf{x}_j e^{\mathbf{x}^\top \mathbf{x}}.$$

Therefore

$$\nabla^2 f = 4\mathbf{x}\mathbf{x}^\top e^{\mathbf{x}^\top \mathbf{x}}.$$

Is this positive semidefinite? Consider $\mathbf{z} \in \mathbb{R}^n$. Then

$$\mathbf{z}^\top \nabla^2 f(x) \mathbf{z} = \mathbf{z}^\top 4\mathbf{x}\mathbf{x}^\top e^{\mathbf{x}^\top \mathbf{x}} \mathbf{z} = 4(\mathbf{z}^\top \mathbf{x})^2 e^{\mathbf{x}^\top \mathbf{x}} \geq 0$$

because $\exp(\cdot)$ and $(\mathbf{z}^\top \mathbf{x})^2$ are both non-negative. Thus $\nabla^2 f \succeq 0$ and by the second order conditions f is convex. ■

3.4 Operations Preserving Convexity

In this section we examine some handy tools for constructing convex functions from other convex functions. There are *many* more properties than those shown here. See *Convex Optimization* by Boyd and Vandenberghe for many others.

Theorem 3.5 (Non-Negative Weighted Sum)**.** *The function $F : C \mapsto \mathbb{R}$ where C is a convex set, $f_i : C \mapsto \mathbb{R}$ is a convex function, $w_i \geq 0$ and*

$$F(\mathbf{x}) = \sum_i w_i f_i(\mathbf{x})$$

is convex.

Proof. Let $\mathbf{x}, \mathbf{y} \in C$ and $0 \leq \theta \leq 1$. Then

$$\begin{aligned} F(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= \sum_i w_i f_i(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \\ &\leq \sum_i w_i \left[\theta f_i(\mathbf{x}) + (1-\theta) f_i(\mathbf{y})\right] \\ &= \theta F(\mathbf{x}) + (1-\theta) F(\mathbf{y}), \end{aligned}$$

as desired. ■

Theorem 3.6 (Pointwise Maximum)**.** *Given convex functions $f_i : C \mapsto \mathbb{R}$, and $F : C \mapsto \mathbb{R}$ defined as*

$$F(\mathbf{x}) = \max_i f_i(\mathbf{x}),$$

F is convex.

Proof. By the definition of convexity, given $\mathbf{x}, \mathbf{y} \in C$ and $0 \leq \theta \leq 1$ we have

$$\begin{aligned} F(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= \max_i f_i(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \\ &\leq \max_i \{\theta f_i(\mathbf{x}) + (1-\theta) f_i(\mathbf{y})\} && \text{(because } f_i \text{ is convex)} \\ &\leq \theta \max_i f_i(\mathbf{x}) + (1-\theta) \max_i f_i(\mathbf{y}) \\ &= \theta F(\mathbf{x}) + (1-\theta) F(\mathbf{y}), \end{aligned}$$

as desired. ■

4 Optimization Problems

An optimization problem is a problem of the form

$$\begin{aligned} \text{minimize:} \quad & f_0(x) \\ \text{subj. to:} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned}$$

where $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$ is called the **objective function**, $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ is an **inequality constraint**, and $h_i : \mathbb{R}^n \mapsto \mathbb{R}$ is an **equality constraint**.

4.1 Convex Optimization Problems

4.1.1 Linear Programs

4.1.2 Quadratic Programs

4.1.3 Global Optimality of a Convex Optimum

4.2 Optimal Flow As a Convex Optimization Problem

5 Algorithms for Solving Convex Problems

5.1 Gradient Descent

5.2 Newton's Method

6 Duality

6.1 Duality Gap

6.2 The Lagrange Dual Function

6.3 Finding the Dual

6.3.1 The Dual of The Linear Program

6.4 Solving Problems Using Duality

6.4.1 Minimize a Quadratic Under Quadratic Constraints

6.5 Complementary Slackness of Duality

6.5.1 Solving The Dual Linear Program via a Primal Solution

6.6 The KKT (Karush-Kuhn-Tucker) Conditions