Convex Optimization: Fundamental Concepts

Oliver Zhao

1 Convex Sets

1.1 Definitions

Definition 1.1 (Convex Set)**.** A *convex set* C is a subset of an Euclidean space such that for any two points, it contains a whole line segment within the set C that joins them. In other words, $C \subseteq \mathbb{R}^n$ such that

$$x, y \in C \quad \Rightarrow \quad tx + (1-t)y \in C \text{ for all } 0 \leq t \leq 1. \tag{1}$$

Definition 1.2 (Convex Combination)**.** A *convex combination* is a linear combination of points (including vectors) where all coefficients are nonnegative and sum to unity. More specifically, a convex combination of the points $x_1, \ldots, x_k \in \mathbb{R}^n$ is any linear combination

$$\theta_1 x_1 + \cdots + \theta_k x_k, \text{ with } \theta_i \geq 0, \text{ and } \sum_{i=1}^{k} \theta_i = 1. \tag{2}$$

Notice that by definition of a convex combination, a convex set contains all convex combinations of its points.

Definition 1.3 (Convex Hull)**.** A *convex hull* of a convex set C is defined as all the convex combinations of elements in C. The convex hull is always convex.

Definition 1.4 (Convex Cone)**.** A *cone* is a set $C \subseteq \mathbb{R}^n$ where for any $x \in C$, we have $tx \in C$ for all $t \geq 0$. A *convex cone* is a cone that is also convex, or

$$x_1, x_2 \in C \Rightarrow t_1 x_1 + t_2 x_2 \in C \text{ for all } t_1, t_2 \geq 0. \tag{3}$$

Definition 1.5 (Conic Combination)**.** A *conic combination* of $x_1, \ldots, x_k \in \mathbb{R}$ is any linear combination

$$\theta_1 x_1 + \cdots + \theta_k x_k, \text{ with } \theta_i \geq 0. \tag{4}$$

Definition 1.6 (Conic Hull)**.** The *conic hull* of set C consists of all conic combinations of elements in C.

1.2 Properties of Convex Sets

1. **Separating Hyperplane Theorem:** two disjoint convex sets have a hyperplane separating them. If C and D are nonempty convex sets with $C \cap D = \emptyset$, then there exists a and b such that

$$C \subseteq \{x : a^T x \leq b\} \text{ and } D \subseteq \{x : a^T x \geq b\}. \tag{5}$$

2. **Supporting Hyperplane Theorem:** a boundary point of a convex set has a supporting hyperplane passing through it. Formally, if C is a nonempty convex set, and $x_0 \in \mathrm{bd}(C)$, then there exists a such at

$$C \subseteq \{x : a^T x \leq a^T x_0\}. \tag{6}$$

1.3 Operations Conserving Convexity

1. **Intersection:** the intersection of convex sets is convex.

2. **Scaling and Translation:** if C is convex, then $aC + b = \{ax + b : x \in C\}$ is convex for any a, b.

3. **Affine Images and Preimages:** if $f(x) = Ax + b$ and C is convex, then $f(C) = \{f(x) : x \in C\}$ is convex, and if D is convex, then $f^{-1}(D) = \{x : f(x) \in D\}$ is convex.

4.

1.4 Examples of Convex Sets

- **Norm ball:** $\{x : ||x|| \leq r\}$, for given norm $||\cdot||$ and radius r.
- **Hyperplane:** $\{x : a^T x = b\}$ for given a, b.
- **Halfspace:** $\{x : a^T x \leq b\}$.
- **Affine space:** $\{x : Ax = b\}$ for given A, b.
- **Polyhyderon:** $\{x : Ax \leq b\}$, where $\leq$ is interpreted componentwise.

2 Convex Functions

2.1 Definitions

Definition 2.1 (Convex Function)**.** A *convex function* exists if the line segment between any two points on the graph of the function lives above the graph between the two points. In other words, a convex function is a function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\text{dom}(f) \subseteq \mathbb{R}^n$ convex and

$$f(tx+(1-t)y) \leq tf(x)+(1-t)f(y) \text{ for all } 0 \leq t \leq 1 \text{ and all } x, y, \epsilon \text{ dom}(f). \quad (7)$$

Definition 2.2 (Concave Function)**.** Opposite inequality of the convex function, or that

$$f \text{ concave } \Leftrightarrow -f \text{ convex.} \quad (8)$$

Definition 2.3 (Strictly Convex)**.** If f is convex and has greater curvature than a linear function, or

$$f(tx + 1 - t)y) < tf(x) + (1 - t)f(y), \text{ for all } x \neq y \text{ and } 0 < t < 1. \quad (9)$$

Definition 2.4 (Strongly Convex)**.** If f is at least as convex as a quadratic function, or with parameter $m > 0$, $f(-\frac{m}{2}||x||_2^2)$ is convex. Or in other words, if

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq m||x - y||_2^2. \quad (10)$$

2.2 Properties of Convex Functions

1. A function is convex if and only if its restriction to any line is convex.
2. **Epigraph Characterization:** a function f is convex if and only if its epigraph is a convex set, or

$$\text{epi}(f) = (x, t) \; \epsilon \; \text{dom}(f) \times \mathbb{R} : f(x) \leq t. \quad (11)$$

3. **Convex Sublevel Sets:** if f is convex, then its sublevel sets

$$x \; \epsilon \; \text{dom}(f) : f(x) \leq t \quad (12)$$

   Are convex for all $t \; \epsilon \; \mathbb{R}$.

4. **First-Order Characterization:** if f is differentiable, then f is convex if and only if dom(f) is convex and

$$f(y) \geq f(x) + \nabla f(x)^T(y - x), \text{ for all } x, y \epsilon \text{ dom}(f). \tag{13}$$

   Therefore if a differentiable convex function $\nabla f(x) = 0 \Leftrightarrow x$ minimizes f.

5. **Second-Order Characterization:** if f is twice differentiable, then f is convex if and only if dom(f) is convex and $\nabla^2 f(x) \geq 0$ for all $x \epsilon$ dom(f).

6. **Jensen's Inequality:** if f is convex, and X is a random variable supported on dom(f), then $f(E[X]) \leq E[f(X)]$.

7. **Long-Sum-Exp Function:** Also often called the soft max, since it smoothly approximates $\max_{i=1,\ldots,k}(a_i^T x + b_i)$, states

$$g(x) = \log\left(\sum_{i=1}^{k} e^{q_i^T x + b_i}\right) \text{ for fixed } a_i, b_i \tag{14}$$

2.3 Operations Preserving Convexity

1. **Nonnegative Linear Combination:** $f_1, \ldots, f_m$ convex implies $a_1 f_1 + \cdots + a_m f_m$ is also convex for any $a_1, \ldots, a_m \geq 0$.

2. **Pointwise Maximization:** if f_s is convex for any $s \epsilon S$, then $f(x) = \max_{s \epsilon S}$ is also convex. Note the set S is the number of functions f_x, which can be infinite.

3. **Partial Minimization:** if $g(x, y)$ is convex in x, y and C is convex, then $f(x) = \min_{y \epsilon C} g(x, y)$ is convex.

4. **Affine Composition:** if f is convex, then $g(x) = f(Ax + b)$ is convex.

5. **General Composition:** suppose $f = hg$, where $g : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$, and $f : \mathbb{R}^n \to \mathbb{R}$. Then

   (a) f is convex if h is convex and nondecreasing, g is convex

   (b) f is convex if h is convex and nondecreasing, g is concave

   (c) f is concave if h is concave and nondecreasing, g is concave

   (d) f is concave if h is convex and nondecreasing, g is convex

6. **Vector Composition:** suppose that

$$f(x) = h(g(x)) = h(g_1(x), \ldots, g_k(x)) \tag{15}$$

   Where $g : \mathbb{R}^n \to \mathbb{R}^k$, $h : \mathbb{R}^k \to \mathbb{R}$, and $f : \mathbb{R}^n \to \mathbb{R}$. Then

   (a) f is convex if h is convex and nondecreasing in each argument, g is convex

(b) f is convex if h is convex and nonincreasing in each argument, g is concave

(c) f is concave if h is concave and nondecreasing in each argument, g is concave

(d) f is concave if h is concave and nonincreasing in each argument, g is convex

2.4 Examples of Convex and Concave Functions

- **Exponential function:** e^{ax} is convex for any a over $\mathbb{R}$.
- **Power function:** x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ and is concave for $0 \leq a \leq 1$ over $\mathbb{R}_+$.
- **Logarithmic function:** $\log(x)$ is concave over $\mathbb{R}_{++}$.
- **Affine function:** $a^T x + b$ is both convex and concave.
- **Quadratic function:** $\frac{1}{2}x^T Qx + b^T x + c$ is convex provided that Q is positive semidefinite or $Q \geq 0$.
- **Least squares loss:** $||y - Ax||_2^2$ is always convex since $A^T A$ is always positive semidefinite.
- $||x||$ is convex for any normal, such as l_p norms where

$$||\mathbf{x}||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \quad \text{for } p \geq 1 \tag{16}$$

  and operator (spectral) and trace (nuclear) norms

$$||X||_{op} = \sigma_1(X) \qquad ||X|_{tr} = \sum_{i=1}^{n} \sigma_r(X) \tag{17}$$

  where $\sigma_1(X) \geq \cdots \geq \sigma_r(X) \geq 0$ are the singular values of the matrix X.

- **Indicator function:** if C is convex, then its indicator function is convex, where

$$I_C(x) = \begin{cases} 0 \text{ if } x \epsilon C, \\ \infty \text{ if } x \not\epsilon C. \end{cases} \tag{18}$$

- **Support function:** for any st C (convex or not), its support function is convex, where

$$I_C^*(x) = \max_{y \epsilon C} x^T y \tag{19}$$

- **Max function:** $f(x) = \max\{x_1, \ldots, x_n\}$ is convex.

3 Convex Optimization Problem

A convex optimization problem is of the form

$$\begin{aligned} \underset{x \epsilon D}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{20}$$

Where f and g_i are all convex, and the optimization domain $D = \text{dom}(f) \cap \bigcap_{i=1}^{m} \text{dom}(g_i)$.

3.1 Optimization Terminology

1. f is called the *criterion* or *objective* function.
2. g_i is called the *inequality constraint* function.
3. If $x \epsilon D, g_i(x) \leq 0, i = 1, \ldots, m$ and $Ax = b$, then x is called a *feasible* point.
4. The minimum of $f(x)$ over all feasible points $x \epsilon D$ is called the *optimal value*, written as f^*.
5. If x is feasible and $f(x) = f^*$, then x is called *optimal*, *solution*, or *minimizer*.
6. If x is feasible and $f(x) \leq f^* + \epsilon$, then x is called $\epsilon - \textit{suboptimal}$.
7. If x is feasible and $g_i(x) = 0$, then we say g_i is *active* at x.
8. Convex minimization can be reposed as concave maximization, where Equation (20) is equivalent to

$$\begin{aligned} \underset{x \epsilon D}{\text{maximize}} \quad & -f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{21}$$

3.2 Solution Set

Let X_{opt} be the set of all solutions of a given convex problem, written as

$$
\begin{aligned}
X_{\text{opt}} = \quad \underset{x \epsilon D}{\operatorname{argmin}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\
& Ax = b
\end{aligned} \tag{22}
$$

Lemma 3.1. X_{opt} is a convex set.

Proof. Using definitions, if $x, y \; \epsilon \; X_{\text{opt}}$, then for $0 \leq t \leq 1$,

- $g_i(tx + (1-t)y) \leq t g_i(x) + (1-t) g_i(y) \leq 0$
- $A(tx + (1-t)y) = tAx + (1-t)Ay = b$
- $f(tx + (1-t)y) \leq t f(x) + (1-t) f(y) = f^*$

It follows that $tx + (1-t)y$ is also a solution.

Lemma 3.2. If f is strictly convex, then the solution is unique and X_{opt} contains only one element.

3.3 Rewriting Constraints

Previously we wrote the optimization problem as

$$
\begin{aligned}
\underset{x \epsilon D}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\
& Ax = b
\end{aligned} \tag{23}
$$

But this is equivalent to writing

$$
\min_x f(x) \text{ subject to } x \; \epsilon \; C \tag{24}
$$

Where $C = \{x : g_i(x) \leq 0, i = 1, \ldots, m, Ax = b\}$ is the feasible set. Another way of writing the same problem is

$$
\min_x f(x) + I_C(x) \tag{25}
$$

3.4 Local Minima are Global Minima

Any local solution is globally optimal in convex optimization problems. Whenever f is a convex function, if there exists an $R > 0$ such that $f(x) \leq f(y)$ whenever $||x - y||_2 \leq R$ then $f(x) \leq f(y)$ for all y.

3.5 First Order Optimality Condition

For a convex problem

$$\min_{x} f(x) \text{ subject to } x \epsilon C \tag{26}$$

And a differentiable f, a feasible point is optimal if and only if

$$\nabla f(x)^T (y - x) \geq 0. \tag{27}$$

In short, all feasible directions from x are aligned with the gradient. When the optimization is unconstrained, this reduces to $\nabla f(x) = 0$.

3.6 Partial Optimization

Previously we saw that if a function $f(x, y)$ is convex in both arguments and if C is a convex set, then the function $g(x) = \min_{y \epsilon C} f(x, y)$ is also convex in x. This allows us to partially optimize a convex problem and still retain convexity guarantees. For example

$$\begin{aligned} \underset{x_1, x_2}{\text{minimize}} \quad & f(x_1, x_2) \\ \text{subject to} \quad & g_1(x_1) \leq 0 \\ & g_2(x_2) \leq 0 \end{aligned} \tag{28}$$

Is equivalent to

$$\begin{aligned} \underset{x_1}{\text{minimize}} \quad & \tilde{f}(x_1) \\ \text{subject to} \quad & g_1(x_1) \leq 0 \end{aligned} \tag{29}$$

Where $\tilde{f}(x_1) = \min\{f(x_1, x_2) : g(x_2) \leq 0\}$.

3.7 Eliminating Equality Constraints

An important special case of change of variables involves eliminating equality constraints. Given the problem

$$\begin{aligned} \underset{x \epsilon D}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{30}$$

We can always express any feasible point as $x = My + x_0$, where $Ax_0 = b$ and $\text{col}(M) = \text{null}(A)$. Hence, the above is equivalent to

$$\begin{aligned} \underset{x \epsilon D}{\text{minimize}} \quad & f(My + x_0) \\ \text{subject to} \quad & g_i(My + x_0) \leq 0, \quad i = 1, \ldots, m \end{aligned} \tag{31}$$

3.8 Introducing Slack Variables

Opposite to eliminating the equality constraints, we can introduce slack variables. Given the problem

$$\begin{aligned} \underset{x \epsilon D}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{32}$$

We can transform the inequality constraints via

$$\begin{aligned} \underset{x \epsilon D}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & s_i \geq 0, \quad i = 1, \ldots, m \\ & g_i(x) + s_i = 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{33}$$

3.9 Hierarchy of Convex Programs

There are sub-classes of convex programs, which can be related as follows: Linear Programs $\subset$ Quadratic Programs $\subset$ Semidefinite Programs $\subset$ Conic Programs $\subset$ Convex Programs.

3.9.1 Linear Program

A *linear program* (LP) is a special type of convex program that can be formulated as shown below. Linear programs are commonly solved with the simplex algorithm (non-iteratively) and interior point methods (iteratively).

$$\begin{array}{ll} & \textbf{Basic Form} \\ \min\limits_{x} & c^T x \\ \text{subject to} & Dx \leq d \\ & Ax = b \end{array} \qquad \begin{array}{ll} & \textbf{Standard Form} \\ \min\limits_{x} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \tag{34}$$

3.9.2 Quadratic Program

A *quadratic program* (QP) is an optimization problem of the form

$$\begin{array}{ll} & \textbf{Basic Form} \\ \min\limits_{x} & c^T x + \frac{1}{2} x^T Q x \\ \text{subject to} & Dx \leq d \\ & Ax = b \end{array} \qquad \begin{array}{ll} & \textbf{Standard Form} \\ \min\limits_{x} & c^T x + \frac{1}{2} x^T Q x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \tag{35}$$

Where $Q \succcurlyeq 0$ (positive semidefinite). If $Q \not\succcurlyeq 0$, then the optimization problem is not convex.

3.9.3 Second-Order Conic Program

A *second-order conic program* (SOCP) is an optimization problem of the form

$$\begin{array}{ll} \underset{x}{min} & c^T x \\ \text{subject to} & ||D_i x + d_i||_2 \leq e_i^T x + f_i, \quad i = 1, \ldots, p \\ & Ax = b \end{array} \tag{36}$$

Where

- $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$.
- $D_i : \mathbb{R}^i \to Y_i$ and $d_i \in Y_i$ for Euclidean space Y.

This is indeed a cone program because recall the second-order cone

$$Q = \{(x,t) : ||x||_2 \leq t\}. \tag{37}$$

So we have

$$||D_i x + d_i||_2 \leq e_i^T x + f_i \Leftrightarrow (D_i x + d_i, e_i^T x + f_i + f_i) \in Q_i \tag{38}$$

For second-order cone Q_i of appropriate dimensions. Now we take $K = Q_1 \times \cdots \times Q_p$.

3.9.4 Semidefinite Program

Consider the basic form of linear programming in Equation (34). We can generalize the problem by changing the $\leq$ to a different order.

A *semidefinite program* (SDP) is an optimization problem of the form

Basic Form	**Standard Form**
$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & x_1 F_1 + \cdots + x_n F_n \preccurlyeq F_0 \\ & Ax = b \end{aligned}$$	$$\begin{aligned} \min_x \quad & C \cdot X \\ \text{subject to} \quad & A_i \cdot X = b_i, \quad i = 1, \ldots, m \\ & X \succcurlyeq 0 \end{aligned}$$

(39)

Where $F_j \in \mathbb{S}^d$ ($d \times d$ symmetrical matrix), for $j = 0, 1, \ldots, n$ and $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$.

3.9.5 Conic Program

A *conic program* is an optimization problem of the form

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & D(x) + d \in K \end{aligned} \tag{40}$$

Where

- $c, x \, \epsilon \, \mathbb{R}^n$, $A \, \epsilon \, \mathbb{R}^{m \times n}$, and $b \, \epsilon \, \mathbb{R}^m$.
- $D : \mathbb{R}^n \to Y$ is a linear map and $d \, \epsilon \, Y$ for Euclidean space Y.
- $K \subseteq Y$ is a closed convex cone.

Both LPs and SDPs are special cases of conic programming, where for LPs, $K = \mathbb{R}^n_+$; and for SDPs, $K = \mathbb{S}^n_+$.

3.9.6 Geometric Program

A monomial is a function $f : \mathbb{R}^n_{++} \to \mathbb{R}$ of the form:

$$f(x) = \gamma x_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \tag{41}$$

For $\gamma > 0, a_1, a_2, \dots, a_n \, \epsilon \, \mathbb{R}$.

A posynomial is a sum of monomials,

$$f(x) = \sum_{k=1}^{p} \gamma_k x_1^{a_{k1}} x_2^{a_{k2}} \dots x_n^{a_{kn}} \tag{42}$$

A geometric program is of the form

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \quad i = 1, \dots, m \\ & h_j(x) = 1, \quad i = 1, \dots, r \end{aligned} \tag{43}$$

Where $f, g_i, i = 1, \dots, m$ are posynomials and $h_j, j = 1, \dots, r$ are monomials. It is important to note that a geometric program is non-convex. However, it can be made into a convex optimization problem.

Given $f(x) = \gamma x_1^{a_1} x_2^{a_2} \dots x_n^{a_n}$, let $y_i = \log x_i$ and rewrite this as

$$\gamma (e^{y_1})^{a_1} (e^{y_2})^{a_2} \dots (e^{y_n})^{a_n} = e^{a^T y + b} \tag{44}$$

For $b = \log \gamma$. Meanwhile, a posynomial can be written as

$$\sum_{k=1}^{p} e^{a_k^T y + b_k}. \tag{45}$$

With this variable substitution, and after taking logs, a geometric program is equivalent to

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \log\left(\sum_{k=1}^{p_0} e^{a_{0k}^T y + b_{0k}}\right) \\ \text{subject to} & \log\left(\sum_{k=1}^{p_i} e^{a_{ik}^T y + b_{ik}}\right) \leq 0, \quad i = 1, \ldots, m \\ & c_j^T y + d_j = 0, \quad j = 1, \ldots, r \end{array} \tag{46}$$

Which is not a convex problem (recall the convexity of softmax functions).