EECS 127
Convex Optimization
Notes

Instructor: El Ghaoui

Kelvin Lee

UC Berkeley

Contents

1

Vectors and Functions

1.1 Basics

Definition 1 (Vector)**.** A **vector** is a collection of numbers, arranged in a column or a row, representing the **coordinates** of a point in n-dimensional space. We write vectors in column format:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

where each element x_i is the ***i*-th component** of vector x and n is the **dimension** of x. If x is a *real* vector, then we write $x \in \mathbb{R}^n$. If x is a *complex* vector, then we write $x \in \mathbb{C}^n$.

Definition 2 (Transpose)**.** The **transpose** of a vector x is defined as

$$x^\top = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array} \right]$$

and the transpose of the transpose of x is itself, i.e., $x^{\top\top} = x$.

1.2 Vector Spaces

Definition 3 (Vector Space)**.** A **vector space** $\mathcal{V}$ is a set of vectors on which two operations: **vector addition** and **scalar multiplication**, are defined.

1.2.1 Subspaces and Span

Definition 4 (Subspace)**.** A nonempty subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is a **subspace** of $\mathcal{V}$ if, for $x, y \in \mathcal{S}$ and any scalars $\alpha, \beta \in \mathbb{R}$,

$$\alpha x + \beta y \in \mathcal{S}.$$

In other words, $\mathcal{S}$ is *closed* under addition and scalar multiplication.

Definition 5 (Linear Combination)**.** A **linear combination** of a set of vectors $S = \{x^{(1)}, \ldots, x^{(m)}\}$ in a vector space $\mathcal{X}$ is a vector

$$x = \sum_{i=1}^{m} \alpha x^{(i)},$$

where each α_i is a given scalar.

Definition 6 (Span)**.** The **span** of a set of vectors $S = \{x^{(1)}, \dots, x^{(m)}\}$ in a vector space $\mathcal{X}$ is the set of all vectors that is a linear combination of that set of vectors

$$\text{span}(S) = \left\{ x \mid \exists \alpha_1, \dots, \alpha_m \text{ s.t. } x = \sum_{i=1}^{m} \alpha_i x^{(i)} \right\}.$$

Definition 7 (Direct Sum)**.** Given two subspaces $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$, the **direct sum** of $\mathcal{X}, \mathcal{Y}$, denoted by $\mathcal{X} \oplus \mathcal{Y}$, is the set of vectors of the form $x + y$, where $x \in \mathcal{X}, y \in \mathcal{Y}$. The direct sum is itself a subspace.

1.2.2 Bases and Dimensions

Definition 8 (Linearly Independent)**.** A set of vectors $x^{(1)}, \dots, x^{(m)}$ in a vector space $\mathcal{X}$ is **linearly independent** if

$$\sum_{i=1}^{m} \alpha_i x^{(i)} = 0 \implies \alpha_1 = \ldots = \alpha_n = 0.$$

Definition 9 (Basis)**.** Given a subspace of $\mathcal{S}$ of a vector space $\mathcal{X}$, a **basis** of $\mathcal{S}$ is a set $\mathcal{B}$ of vectors of minimal cardinality, such that $\text{span}(\mathcal{B}) = \mathcal{S}$.

Definition 10 (Dimension)**.** The **dimension** of a subspace is the cardinality of a basis of that subspace.

If we have a basis $\{x^{(1), \dots, x^{(d)}}\}$ for a subspace $\mathcal{S}$, then any element in the subspace can be expressed as a linear combination of the elements in the basis. That is, any $x \in \mathcal{S}$ can be written as

$$x = \sum_{i=1}^{d} \alpha_i x^{(i)},$$

for some scalars α_i.

1.2.3 Affine Sets

Definition 11 (Affine Set)**.** An **affine set** is a set of the form

$$\mathcal{A} = \left\{ x \in \mathcal{X} \mid x = v + x^{(0)}, v \in \mathcal{V} \right\}$$

where $x^{(0)}$ is a given point and $\mathcal{V}$ is a given subspace of $\mathcal{X}$. Subspaces are just affine spaces containing the origin.

Geometric interpretation: An affine set is a flat plane passing through $x^{(0)}$.

The dimension of an affine set $\mathcal{A}$ is defined as the dimension of its generating subspace $\mathcal{V}$.

1.2.4 Euclidean Length

Definition 12 (Euclidean Length)**.** The **Euclidean length** of a vector $x \in \mathbb{R}^n$ is defined as

$$\|x\|_2 \doteq \sqrt{\sum_{i=1}^{n} x_i^2}.$$

1.2.5 Norms

Definition 13 (Norm)**.** A **norm** on a vector space $\mathcal{X}$ is a real-valued function with special properties that maps any element $x \in \mathcal{X}$ into a real number $\|x\|$.

Definition 14. A function from $\mathcal{X}$ to $\mathbb{R}$ is a **norm**, if

- $\forall x \in \mathcal{X}, \|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$;
- $\forall x, y \in \mathcal{X}, \|x + y\| \leq \|x\| + \|y\|$ (**triangle inequality**);
- $\forall x \in \mathcal{X}, \|\alpha x\| = |\alpha| \|x\|$ for any scalar α.

Definition 15 (ℓ_p norms)**.** $\boldsymbol{\ell_p}$ **norms** are defined as

$$\|x\|_p \doteq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \qquad 1 \leq p < \infty.$$

For $p = 2$, we have the **Euclidean length**

$$\|x\|_2 \doteq \sqrt{\sum_{i=1}^{n} x_i^2},$$

or $p = 1$ we get the **sum-of-absolute-values length**

$$\|x_1\| \doteq \sum_{i=1}^{n} |x_i|.$$

The limit case $p = \infty$ defines the ℓ_∞ **norm** (**max absolute value norm**, or **Chebyshev norm**)

$$\|x\|_\infty \doteq \max_{i=1,\ldots,n} |x_i|.$$

The cardinality of a vector x is called the $\boldsymbol{\ell_0}$ (**pseudo**) **norm** and denoted by $\|x\|_0$.

1.3 Inner Product

Definition 16 (Inner Product)**.** An **inner product** on a real vector space $\mathcal{X}$ is a real-valued function which maps any pair of elements $x, y \in \mathcal{X}$ into a scalar denoted as $\langle x, y \rangle$. It satisfies the following axioms: for any $x, y, z \in \mathcal{X}$ and scalar α

(i) $\langle x, x \rangle \geq 0$;

(ii) $\langle x, x \rangle = 0$ if and only if $x = 0$;

(iii) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$;

(iv) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$;

(v) $\langle x, y \rangle = \langle y, z \rangle$.

Definition 17 (Standard Inner Product)**.** The **standard inner product**, also called the **dot product** is defined as

$$\langle x, y \rangle = x^\top y = \sum_{i=1}^{n} x_i y_i.$$

An inner product naturally induces an associated **norm**: $\|x\| = \sqrt{\langle x, x \rangle}$.

1.3.1 Angle between vectors

The angle between x and y is defined via the relation

$$\cos\theta = \frac{x^\top y}{\|x\|_2\|y\|_2}.$$

There is a right angle between x and y when $x^\top y = 0$, i.e., x and y are **orthogonal**.
When $\theta = 0^\circ$, or $\pm 180^\circ$, then $y = \alpha x$ for some scalar α, i.e. x and y are **parallel**. Then $|x^\top y|$ achieves its **maximum value** $|\alpha|\|x\|_2^2$.

1.3.2 Cauchy-Schwartz and Hölder Inequality

Theorem 18 (Cauchy-Schwartz's Inequality)**.** For any vectors $x, y \in \mathbb{R}^n$, it holds that

$$|\langle x, y\rangle| = |x^\top y| \leq \|x\|_2\|y\|_2,$$

Proof. Note that $|\cos\theta| \leq 1$, then using the angle equation, we have

$$|\cos\theta| = \frac{|x^\top y|}{\|x\|_2\|y\|_2} \leq 1 \implies |x^\top y| \leq \|x\|_2\|y\|_2.$$

□

Theorem 19 (Hölder's Inequality)**.** For any vectors $x, y \in \mathbb{R}^n$ and for any $p, q \geq 1$ such that $1/p + 1/q = 1$, it holds that

$$|\langle x, y\rangle| = |x^\top y| \leq \sum_{i=1}^n |x_i y_i| \leq \|x\|_p\|y\|_q.$$

1.4 Orthogonality and Orthonormality

1.4.1 Orthogonal Vectors

Definition 20 (Orthogonal)**.** Two vectors x, y in an inner product space $\mathcal{X}$ are **orthogonal** if $\langle x, y\rangle = 0$, i.e., $x \perp y$.

Definition 21 (Mutually Orthogonal)**.** Nonzero vectors $x^{(1)}, \ldots, x^{(d)}$ are said to be **mutually orthogonal** if $\langle x^{(i)}, x^{(j)}\rangle = 0$ whenever $i \neq j$. In other words, each vector is orthogonal to all other vectors in the collection.

Proposition 22. Mutually orthogonal vectors are linearly independent.

Definition 23 (Orthonormal)**.** A collection of vectors $S = \{x^{(1)}, \ldots, x^{(d)}\}$ is **orthonormal** if, for $i, j = 1, \ldots, d$

$$\left\langle x^{(i)}, x^{(j)}\right\rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

In words, S is orthonormal if every element has unit norm, and all elements are orthogonal to each other. A collection of orthonormal vectors S forms an **orthonormal basis** for the span of S.

1.4.2 Orthogoanl Complement

Definition 24 (Orthogonal Complement)**.** The set of vectors in $\mathcal{X}$ that are orthogonal to $\mathcal{S}$ is called the **orthogonal complement** of $\mathcal{S}$, denoted by $S^{\perp}$.

Theorem 25 (Orthogonal Decomposition)**.** If $\mathcal{S}$ is a subspace of an inner product space $\mathcal{X}$, then any vector $x \in \mathcal{X}$ can be written in an **unique** way as the sum of an element in $\mathcal{S}$ and one in the orthogonal complement $\mathcal{S}^{\perp}$:

$$\mathcal{X} = \mathcal{S} \oplus \mathcal{S}^{\perp}$$

for any subspace $\mathcal{S} \subseteq \mathcal{X}$.

1.4.3 Projections

Definition 26.

Theorem 27 (Projection Theorem)**.** Let $\mathcal{X}$ be an inner product space, let x be a given element in $\mathcal{X}$, and let $\mathcal{S}$ be a subspace of $\mathcal{X}$. Then, there exists a unique vector $x^* \in \mathcal{S}$ which is solution to the problem

$$\min_{y \in \mathcal{S}} \|y - x\|$$

Moreover, a necessary and sufficient condition for x^* being the optimal solution for this problem is that

$$x^* \in \mathcal{S}, \quad (x - x^*) \perp \mathcal{S}$$

2

Convex Sets

2.1 Affine Sets

Definition 28 (Affine Set)**.** A set $C \subseteq \mathbb{R}^n$ is **affine** if the line through any two distinct points in C lies in C, i.e. if for any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$, we have $\theta x_1 + (1-\theta)x_2 \in C$.

Definition 29 (Affine Combination)**.** A point of the form

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

where $\sum_{i=1}^k \theta_i = 1$ is an **affine combination** of the points $x_1, \ldots, x_k$.

Definition 30 (Affine Hull)**.** The set of all affine combinations of points in some set $C \subseteq \mathbb{R}^n$ is called the **affine hull** of C, denoted **aff** C:

$$\textbf{aff } C = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_1, \ldots, x_k \in C, \sum_{i=1}^k \theta_i = 1 \right\}.$$

2.2 Convex Sets

Definition 31 (Convex Set)**.** A set C is **convex** if for any $x_1, x_2 \in C$ and any $\theta \in [0, 1]$:

$$\theta x_1 + (1-\theta)x_2 \in C,$$

i.e., the **line segment** between any two points in C lies in C.

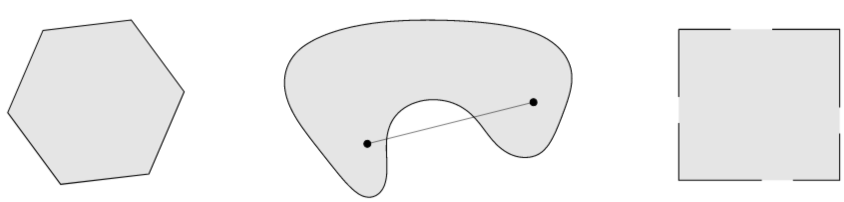

Figure 2.1: Convex and nonconvex sets. *Left.* Convex. *Middle.* Not convex as the line segment between the two points in the set is not contained. *Right.* Not convex as it contains some boundary points but not other.

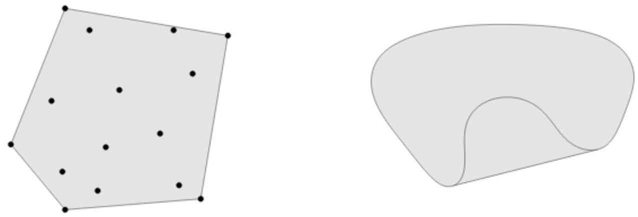

Figure 2.2: *Left.* The convex hull of a set of 15 points is the pentagon. *Right.* The convex hull of the kidney shped set is the shaded set.

Remark. Subspaces, halfspaces, and affine sets, such as lines and hyperplanes are convex, as they contain the entire line between any two distinct points in it, and thus also the line segment in $\mathbb{R}^2$. Hence, **every affine set is convex. However, not every convex set is affine.**

Definition 32 (Convex Combination)**.** A point of the form

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

where $\sum_{i=1}^k \theta_i = 1$ and $\theta_i \geq 0$ is a **convex combination** of the points $x_1, \ldots, x_k$.

Definition 33 (Convex Hull)**.** The **convex hull** of a set C, denoted **conv** C, is the set of all convex combinations of points in C:

$$\textbf{conv } C = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in C, \theta_i \geq 0, i = 1, \ldots, k, \sum_{i=1}^k \theta_i = 1 \right\}.$$

Remark. conv C is always convex and it is the smallest convex set that contains C, i.e., if B is any convex set that contains C, then **conv** $C \subseteq B$.

2.3 Cones

Definition 34 (Cone)**.** A set C is a **cone** if $\theta x \in C$ for every $x \in C$ and $\theta \geq 0$.

Definition 35 (Convex Cone)**.** A set C is a **convex cone** if it is convex and a cone. Mathematically, it means that for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have

$$\theta_1 x_1 + \theta_2 x_2 \in C.$$

Definition 36 (Conic Combination)**.** A point of the form

$$\theta_1 x_2 + +\theta_k x_k$$

with $\theta_1, \ldots, \theta_k \geq 0$ is a **conic combination** of $x_1 \ldots, x_k$.

Definition 37 (Conic Hull)**.** The **conic hull** of a set C is the set of all conic combinations of points in C, i.e.,

$$\left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in C, \theta_i \geq 0, i = 1, \ldots, k \right\},$$

which is also the smallest convex cone that contains C.

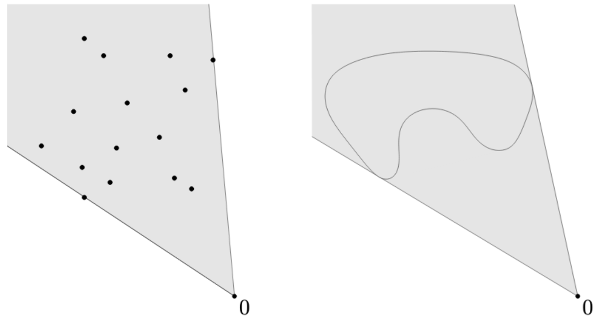

Figure 2.3: The conic hulls of the two sets of figure 1.2.

2.4 Hyperplanes and Halfspaces

Definition 38 (Hyperplane)**.** A **hyperplane** is a set of the form

$$\{x \mid a^\top x = b\},$$

where $a \in \mathbb{R}^n, a \neq 0$, and $b \in \mathbb{R}$, i.e., the solutions set of a nontrivial linear equation among the components of x (and hence an affine set).

Geometric interpretation: The hyperplane is a set of points with a constant inner product to a given vector a, which can also be viewed as a **normal vector**; the constant b determines the offset from the origin.

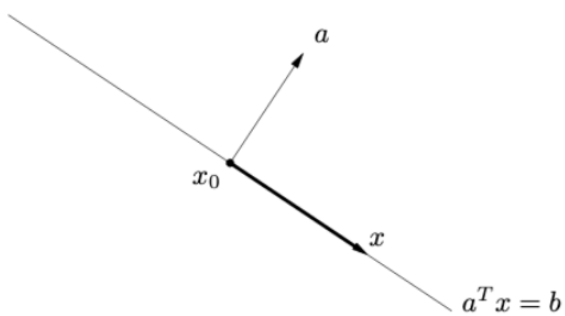

Figure 2.4: Hyperplane in $\mathbb{R}^2$, with normal vector a. $x - x_0$ (arrow) is orthogonal to a for any x in the hyperplane.

A hyperplane divides $\mathbb{R}^n$ into two **halfspaces**, defined as follows:

Definition 39 (Halfspace)**.** A **halfspace** is a set of the form

$$\{x \mid a^\top x \leq b\},$$

where $a \neq 0$, i.e., the solution set of a nontrivial linear inequality.

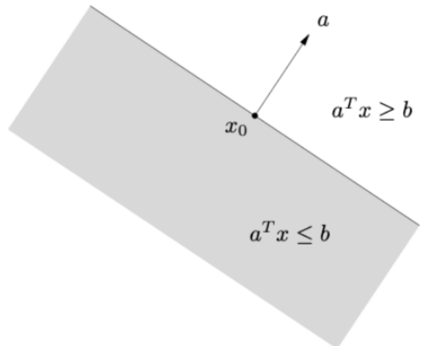

Figure 2.5: The halfspace determined by $a^\top x \leq b$ (shaded) extends in the direction $-a$.

Remark. Halfspaces are **convex**, but not affine.

Proof. Let x_1, x_2 be two points in a halfspace. Then for any $\theta \in [0,1]$, we have

$$\begin{aligned} a^\top(\theta x_1 + (1-\theta)x_2) &= \theta a^\top x_1 + (1-\theta)a^\top x_2 \\ &\leq \theta b + (1-\theta)b \\ &= b. \end{aligned}$$

Thus, halfspaces are convex. □

2.5 Operations preserving convexity

2.5.1 Intersection

Theorem 40. If $C_1, \ldots, C_m$ are convex sets, then their intersection

$$C = \bigcap_i^m C_i$$

is also a convex set.

Proof. Let $\{C_i\}_{i=1}^m$ be convex sets. For any $x_1, x_2 \in \bigcap_{i=1}^m C_i, \theta \in [0,1]$, $x_1 \in C_i$ and $x_2 \in C_i$ implies

$$\theta x_1 + (1-\theta)x_2 \in C_i$$

for $i = 1, 2, \ldots, m$ by convexity of C_i. Hence,

$$\theta x_1 + (1-\theta)x_2 \in \bigcap_{i=1}^m C_i.$$

Thus, $\bigcap_{i=1}^m C_i$ is convex. □

Remark. This also holds for possibly infinite families of convex sets.

3

Convex Functions

3.1 Basic properties and examples

Definition 41 (Domain)**.** The **domain** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set over which the function is well-defined:

$$\textbf{dom } f = x \in \mathbb{R}^n \mid -\infty < f(x) < \infty\}.$$

Definition 42 (Convex)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if **dom** f is a convex set and if **dom** f is a convex set, and for all $x, y \in$ **dom** f and all $\theta \in [0, 1]$ it holds that

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).$$

Figure 3.1: A convex function with the *chord* lying above the graph.

Definition 43 (Concave)**.** A function f is **concave** if $-f$ is convex.

3.1.1 Sublevel Sets

Definition 44 (α-sublevel set)**.** The **α-sublevel set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$C_\alpha = \{x \in \textbf{dom } f \mid f(x) \leq \alpha\}.$$

Theorem 45. For any value of α, sublevel sets of a convex function are convex.

Proof. Suppose $x, y \in C_\alpha$, then $f(x) \leq \alpha$ and $f(y) \leq \alpha$, and so by the definition of convexity for $\theta \in [0, 1]$:

$$\begin{aligned} f(\theta x + (1-\theta)y) &\leq \theta f(x) + (1-\theta)f(x) \\ &\leq \theta\alpha + (1-\theta)\alpha \\ &= \alpha, \end{aligned}$$

so we have $\theta x + (1-\theta)y \in C_\alpha$. $\square$

Remark. The converse is not true: a function may not be a convex function to have all its sublevel sets convex.

Example 46. One example would be $f(x) = \log x$, which is concave, but its sublevel sets are the intervals $(0, e^\alpha]$, which are convex.

3.1.2 Epigraph

Definition 47 (Epigraph)**.** The **epigraph** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\textbf{epi } f = \{(x, t) \mid x \in \textbf{dom } f, f(x) \leq t\},$$

('epi' means 'above' so epigraph means the set of points lying above the graph of the function).

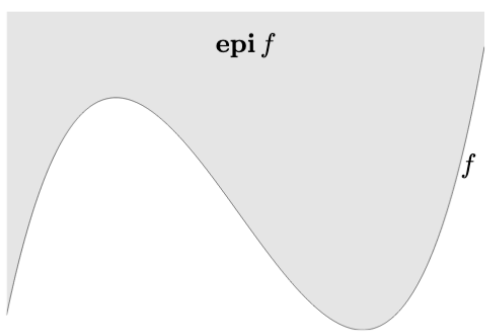

Figure 3.2: Epigraph (shaded region) of a function f.

Remark. A function is convex if and only if its epigraph is a convex set.