Weekly Overview Slides of Statistical Machine Learning CSE 575, Spring 2023

Moses A. Boudourides[1]

SPA and SCAI
Arizona State University

[1] Moses.Boudourides@asu.edu

Week 11

Review of Convex Optimization

March 23, 2023

Affine Sets

Definitions and Basic Results

- Given two points $x_1, x_2 \in \mathbb{R}^n$ such that $x_1 \neq x_2$, points of the form $y = \theta x_1 + (1 - \theta)x_2$, where $\theta \in \mathbb{R}$, form the **line** passing from x_1 and x_2.
- A set $C \subseteq \mathbb{R}^n$ is **affine** if (all the points of) the line through any two distinct points in C lies in C.
- Given $k \geq 2$ distinct points $x_1, \ldots, x_k \in \mathbb{R}^n$, a point of the form $\theta_1 x_1 + \ldots + \theta_k x_k$, where $\theta_1, \ldots, \theta_k \in \mathbb{R}$ and $\theta_1 + \ldots + \theta_k = 1$, is an **affine combination** of the points $x_1, \ldots, x_k$.
- A set is affine if and only if it contains every affine combination of its points.
- The set of all affine combinations of points in some set $C \subseteq \mathbb{R}^n$ is called the **affine hull** of C and denoted as $\mathbf{aff}\, C$. The affine hull of C is the smallest affine set that contains C.
- The solution set of a system of linear equations, $C = \{x \in \mathbb{R}^n : Ax = b\}$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, is an affine set in $\mathbb{R}^n$.

Convex Sets

Definitions and Basic Results

- Given $x_1, x_2 \in \mathbb{R}^n$ such that $x_1 \neq x_2$, the **line segment** between x_1 and x_2 is the subset of the line $y = \theta x_1 + (1-\theta)x_2$ from x_1 and x_2, when $0 \leq \theta \leq 1$.
- A set $C \subseteq \mathbb{R}^n$ is **convex** if (all the points of) the line segment between any two points in C lies in C.
- Given $k \geq 2$ distinct points $x_1, \ldots, x_k \in \mathbb{R}^n$, a point of the form $\theta_1 x_1 + \ldots + \theta_k x_k$, where $\theta_1 + \ldots + \theta_k = 1$ and $\theta_1, \ldots, \theta_k > 0$, is a **convex combination** of the points $x_1, \ldots, x_k$.
- A set is convex if and only if it contains every convex combination of its points.
- The set of all convex combinations of points in some set $C \subseteq \mathbb{R}^n$ is called the **convex hull** of C and denoted as $\mathbf{conv}\, C$. The convex hull of C is the smallest convex set that contains C.
- The intersection of two convex sets is a convex set.
- A set $C \subseteq \mathbb{R}^n$ is convex if and only if its intersection with an arbitrary line is convex.

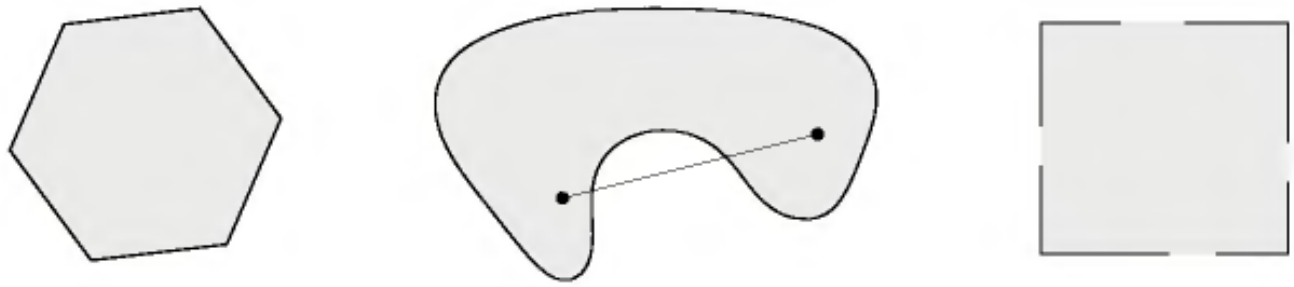

Figure 2.2 Some simple convex and nonconvex sets. *Left.* The hexagon, which includes its boundary (shown darker), is convex. *Middle.* The kidney shaped set is not convex, since the line segment between the two points in the set shown as dots is not contained in the set. *Right.* The square contains some boundary points but not others, and is not convex.

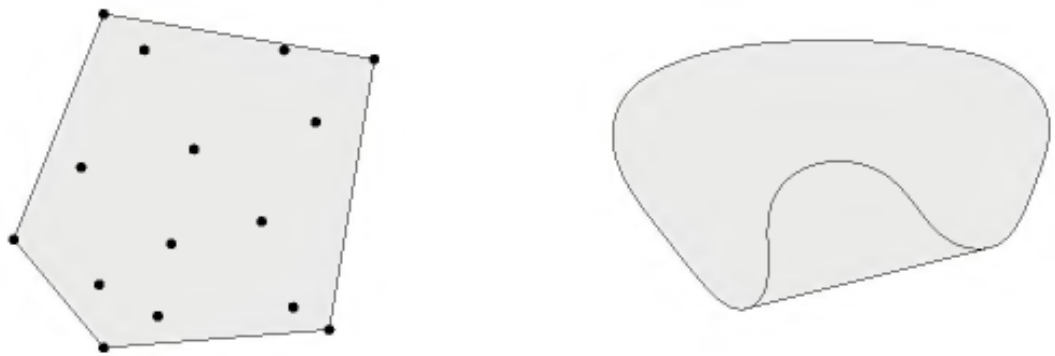

Figure 2.3 The convex hulls of two sets in $\mathbf{R}^2$. *Left.* The convex hull of a set of fifteen points (shown as dots) is the pentagon (shown shaded). *Right.* The convex hull of the kidney shaped set in figure 2.2 is the shaded set.

Examples of Convex Sets

Examples

- Trivially, an affine set is convex (but not the opposite).
- The empty set $\varnothing$, any single point $\{x_0\}$ and the whole $\mathbb{R}^n$ are convex (subsets of $\mathbb{R}^n$; affine too).
- A **hyperplane** is a set of the form $\{x \in \mathbb{R}^n : a^T x = b\}$, where $a \in \mathbb{R}^n, a \neq 0$ and $b \in \mathbb{R}$. A hyperplane divides $\mathbb{R}^n$ into two **halfspaces**. A (closed) halfspace is a set of the form $\{x \in \mathbb{R}^n : a^T x \leq b\}$, where $a \in \mathbb{R}^n, a \neq 0$ and $b \in \mathbb{R}$. Halfspaces are convex, but not affine.
- A **polyhedron** $\mathcal{P}$ is defined as the solution set of a finite number of linear equalities and inequalities, $\mathcal{P} = \{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \ldots, m, c_j^T x = d_j, j = 1, \ldots, p\}$. Thus, a polyhedron is the intersection of a finite number of halfspaces and hyperplanes.
- The **nonnegative orthant** $\mathbb{R}^n_+$ is the set of points with nonnegative components is convex (subset of $\mathbb{R}^n$).
- **Norm balls** $C = \{x \in \mathbb{R}^n : ||x - x_0|| \leq r\}$, for some $x_0 \in \mathbb{R}^n$ (*center*) and $r > 0$ (*radius*). (**Ellipses** too.)
- Denoting by $\mathbb{S}^n$ the set of symmetric $n \times n$ matrices, the set of **positive semidefinite matrices** $\mathbb{S}^n_+ = \{A \in \mathbb{S}^n : x^T A x \geq 0, x \in \mathbb{R}^n\}$ is convex (subset of $\mathbb{S}^n$).

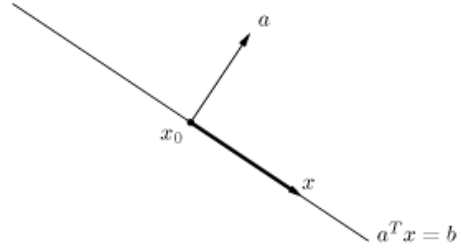

Figure 2.6 Hyperplane in $\mathbf{R}^2$, with normal vector a and a point x_0 in the hyperplane. For any point x in the hyperplane, $x - x_0$ (shown as the darker arrow) is orthogonal to a.

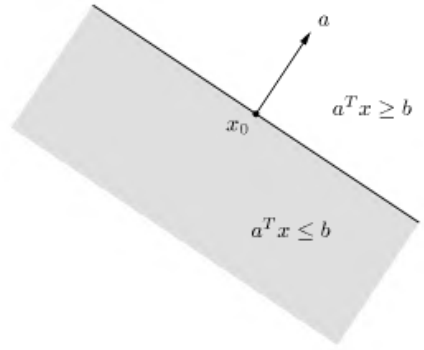

Figure 2.7 A hyperplane defined by $a^T x = b$ in $\mathbf{R}^2$ determines two halfspaces. The halfspace determined by $a^T x \geq b$ (not shaded) is the halfspace extending in the direction a. The halfspace determined by $a^T x \leq b$ (which is shown shaded) extends in the direction $-a$. The vector a is the outward normal of this halfspace.

Convex Functions

Definitions

- Let a function $f : \mathbb{R}^n \to \mathbb{R}$ and denote by $\mathbf{dom}\, f$ the domain (of definition) of f.
- f is a **convex function** if $\mathbf{dom}\, f$ is convex (subset of $\mathbb{R}^n$) and, for all $x, y \in \mathbf{dom}\, f$ and θ with $0 \leq \theta \leq 1$, we have

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

  which means geometrically that the line segment between $(x, f(x))$ and $(y, f(y))$ (i.e., the *chord* from x to y) lies above the graph of f.
- f is **strictly convex** if strict inequality holds above, whenever $x \neq y$ and $0 < \theta < 1$.
- f is **concave** if $-f$ is convex and **strictly concave** if $-f$ is strictly convex.

Convexity Conditions

First–order convexity condition

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable (i.e., the **gradient** $\nabla f = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)$ exists at each point $x \in \mathbf{dom}\, f$, which is assumed to be open). Then f is convex if and only if $\mathbf{dom}\, f$ is convex and $f(y) \geq f(x) + \nabla f(x)^T (y - x)$, for all $x, y \in \mathbf{dom}\, f$.

Corollary

If f is convex differentiable and $\nabla f(x) = 0$ at some point $x \in \mathbf{dom}\, f$, then $f(x)$ is the minimum of f.

Second–order convexity condition

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable (i.e., the **Hessian** $(n \times n)$ matrix $\nabla^2 f = \left\{ \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right\}_{i,j=1,\ldots,n}$ exists at each point $x \in \mathbf{dom}\, f$, which is assumed to be open). Then f is convex if and only if $\mathbf{dom}\, f$ is convex and its Hessian is positive semidefinite, for all $x \in \mathbf{dom}\, f$.

Examples of Convex–Concave Functions

Examples

- $f: \mathbb{R} \to \mathbb{R}$ is an **affine function** if $f(x) = ax + b$, where $a, b \in \mathbb{R}$. Affine functions are *both* convex and concave.
- The exponential function (on $\mathbb{R}$) $f(x) = e^{ax}$, where $a \in \mathbb{R}$, is convex.
- The power function (on positive reals) $f(x) = x^a$ is convex, when $a \geq 1$ or $a \leq 0$, and it is concave, for $0 \leq a \leq 1$. However, $f(x) = x^{-1}$ is not convex on $\mathbb{R} \setminus \{0\}$, because $\mathbf{dom}\, f$ is not convex.
- The power of absolute value function (on $\mathbb{R}$) $f(x) = |x|^p$, where $p \geq 1$, is convex.
- The logarithmic function (on positive reals) $f(x) = \log x$ is concave.
- The quadratic form $f(x) = \frac{1}{2}x^T Ax + b^T x + c$, where $A \in \mathbb{S}^n$, is convex if and only if A is positive semidefinite.
- **Composition of scalar functions**: Let $g: \mathbb{R}^n \to \mathbb{R}, h: \mathbb{R} \to \mathbb{R}$ and $f(x) = h(g(x))$. Then f is convex if g is convex and h is convex and nondecreasing. In addition, f is convex if g is concave and h is convex and nonincreasing.
- Recall that the **graph** of a function $f: \mathbb{R}^n \to \mathbb{R}$ is defined as $\{(x, f(x)) : x \in \mathbf{dom}\, f\}$ and the **epigraph** is defined as $\{(x, y) : x \in \mathbf{dom}\, f, y \geq f(x)\}$. Then f is convex if and only if the epigraph of f is convex.
- If, for each i in an index set $\mathcal{I}$, f_i is convex, the function $f(x) = \sup_{i \in \mathcal{I}} f_i(x)$ is convex too.

Convex Optimization Problems

Standard form of convex optimization

The standard form of a convex oprimization problem, denoted using the functions (f_0, f_i, h_i), is when the objective function f_0 is convex, the inequality constraints f_i are convex and the equality constraints h_i are affine. Typically, this is represented as follows:

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

where $f_i : \mathbb{R}^n \to R, f_0, f_1, \ldots, f_m$ are convex and $h_i : \mathbb{R}^n \to R$ are affine. Furthermore, we assume that $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \bigcap_{i=1}^{p} \mathbf{dom}\, h_i \neq \varnothing$.

Why is it nice to be convex?

Proposition 1

A convex function on a closed area attains its maximum at the boundary of the area.

Proposition 2

If a point is a local minimizer of a convex optimization problem, then it is a global minimizer.

Proposition 3

For a convex function f, $\nabla f(x) = 0$ if and only if x is a global minimizer of $f(x)$.

The Lagrangian

Definition of the Lagrangian

Given a convex optimization problem in the standard form (f_0, f_i, h_i), the **Lagrangian** (function) is defined as a mapping $\mathcal{L}\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ such that

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

where we refer to $\lambda_i > 0$ (for $i = 1, \ldots, m$) as the **Lagrange multiplier** associated with the ith inequality constraint $f_i(x) \leq 0$, and to ν_i (for $i = 1, \ldots, p$) as the **Lagrange multiplier** associated with the ith equality constraint $h_i(x) = 0$. The vectors $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^p$ are called the **dual variables** or **Lagrange multiplier vectors** associated with the optimization problem (f_0, f_i, h_i).

Proposition

The Lagrangian satisfies

$$\sup_{\lambda_i > 0, i=1,\ldots,m} \mathcal{L}(x, \lambda, \nu) = \begin{cases} f_0(x), & \text{for feasible } x \in \mathbb{R}^n, \\ \infty, & \text{otherwise.} \end{cases}$$

The Dual Problem

Definition of the dual function

The **Lagrange dual function** (or just **dual function**) is defined as the mapping $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ giving the minimum value of the Lagrangian over $x \in \mathbb{R}^n$, for $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$,

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \text{ feasible}} \mathcal{L}(x, \lambda, \nu) \\ &= \inf_{x \text{ feasible}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right). \end{aligned}$$

When the Lagrangian is unbounded below in x, the dual function becomes $-\infty$. Since the dual function is the positive infinum of a family of affine functions of (λ, ν), it is concave, even when the optimization problem is not convex. Moreover, the dual function yields lower bound on the optimal value p^* of the optimization problem (f_0, f_i, h_i), as one can show that for any $\lambda \succeq 0$ and ν we $g(\lambda, \nu)$ gives a lower bound to p^* (i.e., $g(\lambda, \nu) \leq p^*$).

The Dual Problem (cont.)

Thus, the **dual problem** is to find the best lower bound of the Lagrange dual function, which is written in the following form:

$$\text{minimize } g(\lambda, \nu)$$
$$\text{subject to } \lambda \succeq 0.$$

Notice that the dual is concave (essentially because λ, ν are affine and using the convexity propert of the pointwise supremum of convex functions). Therefore, the dual problem can be solved efficiently with convex optimization. In this way, we may obtain the following great property of the dual.

Lemma (Weak Duality)

If $\lambda \succeq 0$, then the following **max–min inequality** holds

$$d^* = \sup_{\lambda \succeq 0, \nu} \inf_{x \text{ feasible}} \mathcal{L}(x, \lambda, \nu) \leq \inf_{x \text{ feasible}} \sup_{\lambda \succeq 0, \nu} \mathcal{L}(x, \lambda, \nu) = p^*,$$

where d^* is the dual's best lower bound of the optimal solution p^* of the problem (f_0, f_i, h_i).

The Dual Problem (cont.)

Thus, the question is under what conditions $d^* = p^*$ (in which case the dual would solve our problem). A condition where $d^* = p^*$ is called **Slater's constraint qualification**.

Theorem (Strong Duality)

If there exists a feasible x such that the constraint

$$f_i(x) < 0, i = 1, \ldots, m,$$

holds, then Slater's constraint qualification is satisfied:

$$d^* = \sup_{\lambda \succeq 0, \nu} \inf_{x \text{ feasible}} \mathcal{L}(x, \lambda, \nu) = \inf_{x \text{ feasible}} \sup_{\lambda \succeq 0, \nu} \mathcal{L}(x, \lambda, \nu) = p^*.$$

The Dual Problem (cont.)

Strong duality is useful not only because it implies that $d^* = p^*$, but also because it may provide an excellent checklist to know whether we can solve the original problem with duality according to the next result:

Theorem (Karush–Kuhn–Tucker)

The following conditions, called **Karush–Kuhn–Tucker** (or **KKT**) **conditions**,

$$\begin{aligned} f_i(x^*) &\leq 0, \ i = 1, \ldots, m, \\ h_i(x^*) &= 0, \ i = 1, \ldots, p, \\ \lambda^* &\succeq 0, \\ \lambda^* f_i(x^*) &= 0, \ i = 1, \ldots, m, \end{aligned}$$

$$\nabla \left(f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \right) = 0,$$

hold if and only if x^* and (λ^*, ν^*) are optimal for the original convex optimization problem and the dual problem (resp.) and the validity of the strong duality $d^* = p^*$ too.