EE 381V: Large Scale Optimization **Fall 2012**

Lecture 1 — August 30

Lecturer: Caramanis & Sanghavi *Scribe: Sarabjot Singh & Ahmed Alkhateeb*

In this lecture we give the basic definition of a convex set, a convex function, and convex optimization. As an application of these ideas, we derive the characterization of an optimal solution to an unconstrained convex optimization problem.

1.1 Formulation of Convex Optimization Problems

A typical formulation of a convex optimization problem is as follows:

$$\begin{aligned} \min: \quad & f_o(x) \\ \text{s.t.}: \quad & f_i(x) \le 0, \qquad i = 1, \dots, m. \end{aligned}$$

We also often write the problem with an abstract (convex) set constraint:

$$\begin{aligned} \min: \quad & f_o(x) \\ \text{s.t.}: \quad & x \in \mathcal{X}, \end{aligned}$$

where f_o is a convex function and $\mathcal{X}$ is a convex set.

1.1.1 Convex Sets

Definition 1. *A set $\mathcal{X}$ is called a convex set if and only if the convex combination of any two points in the set belongs to the set. In symbols:*

$$\mathcal{X} \subseteq \mathbb{R}^n \text{ is convex } \Leftrightarrow \text{ for any } x_1, x_2 \in \mathcal{X},\ \lambda x_1 + (1-\lambda)x_2 \in \mathcal{X}\ \forall\, \lambda \in [0,1]. \tag{1.1}$$

Definition 2. *A hyperplane, $\mathcal{H}$ is a set defined as*

$$\mathcal{H} = \{x \in \mathbb{R}^n : s^T x = b\}(s \neq 0), \tag{1.2}$$

where b is the offset and s is the normal vector. If $b = 0$, $\mathcal{H}$ is a $(n-1)$ dimensional subspace.

Definition 3. *A halfspace, $\mathcal{H}_+$ is a set defined as*

$$\mathcal{H}_+ = \{x \in \mathbb{R}^n : s^T x \le b\}(s \neq 0), \tag{1.3}$$

where b is the offset and s is the normal vector.

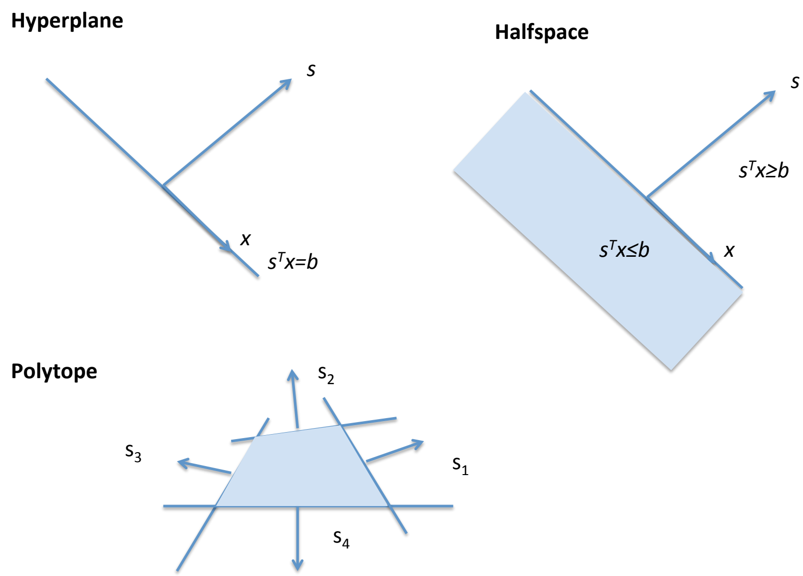

Figure 1.1. Example of convex sets

Definition 4. *A polyhedron is an intersection of finite number of halfspaces and hyperplanes. A polytope is a polyhedron that is also bounded. Thus, the positive orthant is a polyhedron but not a polytope.*

A very interesting set that turns out to be useful for many applications, and in general comes up frequently in convex optimization, is the set $\mathcal{S}^n_+$, of symmetric $n \times n$ matrices with non-negative eigenvalues. As a simple exercise, we can show that this set is in fact convex.

Proof: We use the fact that a symmetric matrix M is in $\mathcal{S}^n_+$ iff $x^T M x \geq 0 \; \forall x \in \mathbb{R}^n$. Using this, checking convexity becomes straightforward. For a convex combination of any two matrices $M_1, M_2 \in \mathcal{S}^n_+$ we have

$$\begin{aligned} x^T(\lambda M_1 + (1-\lambda)M_2)x &= \lambda x^T M_1 x + (1-\lambda)x^T M_2 x \\ &\geq 0 \end{aligned}$$

Hence, $\lambda M_1 + (1-\lambda)M_2 \in \mathcal{S}^n_+ \; \forall \, \lambda \in [0,1]$. □

Definition 5. *A subspace in* $\mathbb{R}^n$ *is a set which is closed under linear combination of the constituent vectors, or*

$$V = \{w : \sum_{i=1}^{K} \lambda_i v_i = w, \; \forall \lambda_i \in \mathbb{R}, \; v_i \in \mathcal{V}\} \tag{1.4}$$

Remark 1. *The nullspace of a $m \times n$ matrix A given by*

$$\text{null}(A) = \{v : Av = 0\}$$

is a subspace. This can be shown by using the distributivity of matrix multiplication over addition.

Remark 2. *The range of a $m \times n$ matrix A given by*

$$\text{range}(A) = \{Av \in \mathcal{W}\}$$

is a subspace as

$$\sum_{i=1}^{K} \lambda_i T v_i = T\left(\sum_{i=1}^{K} \lambda_i v_i\right) = T v_0 \in \mathcal{W},$$

where v_0 is defined as the $\sum_{i=1}^{K} \lambda_i v_i$.

Definition 6. *An affine subspace in $\mathbb{R}^n$ is a set which is closed under affine combination of the constituent vectors, or*

$$\mathcal{V} = \{w : \sum_{i=1}^{K} \lambda_i v_i = w,\ \forall \lambda_i,\ s.t \sum_{i=1}^{K} \lambda_i = 1,\ v_i \in \mathcal{V}\} \tag{1.5}$$

Several operations preserve convexity. Here are a few:

- If $\mathcal{C}_1$ and $\mathcal{C}_2$ are two convex sets, then $\mathcal{C}_1 \cap \mathcal{C}_1$ remains convex. In fact, the same holds true for arbitrary intersections of convex sets.

  Proof: Let $x_1, x_2 \in \mathcal{C}_1 \cap \mathcal{C}_1$, and let x be a point on the line segment between x_1 and x_2. Then, x lies in both $\mathcal{C}_1$ & $\mathcal{C}_2$ as they are convex. Consequently, $x \in \mathcal{C}_1 \cap \mathcal{C}_1$. □

- If $\mathcal{C}_1$ and $\mathcal{C}_2$ are two convex sets, then their cartesian product is convex, i.e.

  $$\mathcal{C}_1 \times \mathcal{C}_2 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x \in \mathcal{C}_1, y \in \mathcal{C}_2 \right\}$$

  is convex.

  Proof: Let $u, v \in \mathcal{C}_1 \times \mathcal{C}_2$ where $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, and let $w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda u + (1-\lambda)v$. Now, as $w_1 \in \mathcal{C}_1$ and $w_2 \in \mathcal{C}_2$, then $w \in \mathcal{C}_1 \times \mathcal{C}_2$.

  □

Definition 7. *A convex hull of a set $\mathcal{C}$ is the set of all convex combinations of points in $\mathcal{C}$. See Fig. 1.2 for an illustration.*

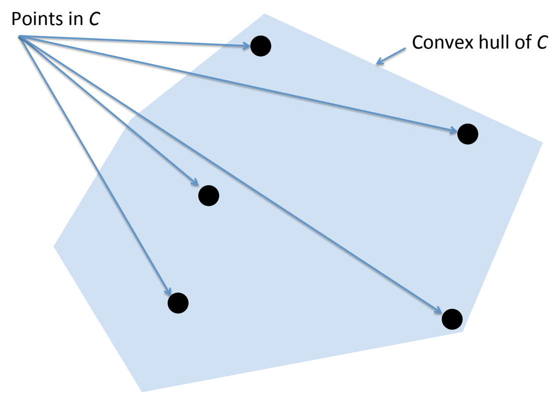

Figure 1.2. Convex hull

1.1.2 Convex functions

We give three definitions of convex functions: A basic definition that requires no differentiability of the function; a first-order definition of convexity that uses the gradient; and a second-order condition of convexity. We show that (for smooth functions) these three are equivalent. Later in the course, it will be important to deal with non-differentiable functions.

First, we define the set of points where a function is finite.

Definition 8. *The domain of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *is denoted* $\mathrm{dom}(f)$*, and is defined as the set of points where a function* f *is finite:*

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n \ : \ f(x) < \infty\}.$$

Definition 9. *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is convex if for any* $x_1, x_2 \in \mathrm{dom}(f) \subseteq \mathbb{R}^n$*,* $\lambda \in [0, 1]$*, we have:*

$$\lambda f(x_1) + (1 - \lambda) f(x_2) \ \geq f (\lambda x_1 + (1 - \lambda) x_2) . \tag{1.6}$$

This inequality is illustrated in Figure 1.3.

As a simple application of this definition, we can show that indeed, the feasible set of the convex optimization problem written above is indeed convex. That is, the set $\{x \ : \ f(x) \leq a\}$, is convex. This is called the *level set* of a function. It is convex for any value of a (note that the empty set is convex by convention), and for any convex function f. We show this now.

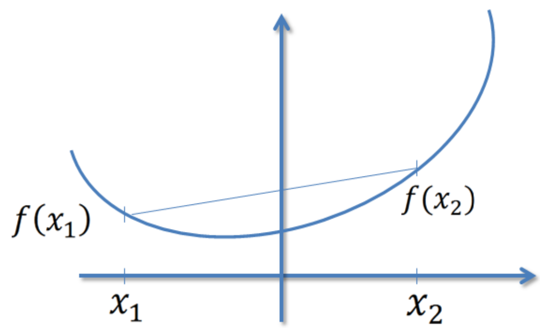

Figure 1.3. Convex functions

Proof: Basically, it is to be shown that if $x_1, x_2 \in \mathcal{C}$ then any convex combination $\lambda x_1 + (1-\lambda)x_2 \in \mathcal{C}$. If $x_1, x_2 \in \mathcal{C}$ then

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &\overset{(a)}{\leq} \lambda f(x_1) + (1-\lambda)f(x_2) \ \forall \lambda \in [0,1] \\ &\overset{(b)}{\leq} \lambda a + (1-\lambda)a \ \forall \lambda \in [0,1] \\ &= a, \end{aligned}$$

where (a) follows by using the convexity of f and (b) follows from using the definition of the level set. □

We now give the first-order condition for convexity.

Definition 10. *Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then it is convex if and only if*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x). \tag{1.7}$$

Proposition 1. *For differentiable functions, the two definitions above are equivalent.*

Proof: Consider first a univariate function. Suppose a function f satisfies the derivative-free definition of convexity. Then we know that for any $x_1, x_2 \in \text{dom}(f)$, and any $\lambda \in [0,1]$,

$$f(x_1 + \lambda(x_2 - x_1)) \leq (1-\lambda)f(x_1) + \lambda f(x_2).$$

Rearranging terms and dividing by λ, we obtain:

$$f(x_2) \geq f(x_1) + \frac{f(x_1 + \lambda(x_2 - x_1)) - f(x_1)}{\lambda}.$$

Taking λ to zero, we obtain the first-order condition for convexity.

Conversely, suppose now that function f satisfies the first-order condition. Pick points $x_1 < x_2$, and let $\bar{x} = \lambda x_1 + (1-\lambda)x_2$ be some point in the convex hull ($\lambda \in (0,1)$). Drawing the picture of the first-order condition from $\bar{x}$, and how both x_1 and x_2 are underestimated, should immediately convey the intuition of the equivalence. To see it algebraically: the two underestimates are:

$$\begin{aligned} f(x_1) &\geq f(\bar{x}) + f'(\bar{x}) \cdot (\bar{x} - x_1), \\ f(x_2) &\geq f(\bar{x}) + f'(\bar{x}) \cdot (\bar{x} - x_2). \end{aligned}$$

Multiplying the first inequality by λ, the second by $(1-\lambda)$ and adding, we recover the derivative-free condition.

For the multivariate case, we use the fact that if a function is convex along all line segments, then it is convex, in which case this reduces to the above proof. □

We now give the final definition of convexity.

Definition 11. *Suppose that a function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then f is convex iff its Hessian is positive semidefinite.*

As an example, consider the function:

$$f(x) = \|Ax - b\|_2^2.$$

Expanding, we have $f(x) = x^\top A^\top Ax - b^\top Ax - x^\top A^\top b + \|b\|_2^2$. The Hessian of this is $A^\top A$, which is positive semidefinite, since for any vector x, $x^\top(A^\top A)x = \|Ax\|_2^2 \geq 0$.

Proposition 2. *The definition give above is equivalent to the previous two definitions. That is, a twice differentiable function f is convex if*

$$\nabla^2 f(x) \in \mathcal{S}_+^n. \tag{1.8}$$

Proof: One way to prove this is via Taylor's theorem, using the Lagrange form of the remainder. To prove that a function with positive semidefinite Hessian is convex, using a second order Taylor expansion we have:

$$f(y) = f(x) + \nabla f(x)^T (y-x) + (y-x)^T \nabla^2 f(x + \alpha(y-x))(y-x) \tag{1.9}$$

for some value of $\alpha \in [0,1]$. Now, since the Hessian is positive semidefinite,

$$(y-x)^T \nabla^2 f(x + \alpha(y-x))(y-x) \geq 0, \tag{1.10}$$

which leads to

$$f(y) \geq f(x) + \nabla f(x)^T (y-x), \tag{1.11}$$

which is the first-order condition of convexity. Hence, this proves that $f(x)$ is convex. For the reverse direction, showing that convexity implies positive semi definiteness of the Hessian, again we can use Taylor's theorem. However, there is a slightly delicate issue because of the Hessian. For this, we need to use some properties of symmetric matrices, to claim that if for some orthonormal basis $\{x_i\}$, a matrix A satisfies $x^\top Ax \geq 0$, then A is positive semidefinite. This completes the proof. □

Now let us revisit the basic convex optimization problem we saw at the beginning of the lecture:

$$\begin{aligned} \text{min} : \quad & f_o(x) \\ \text{s.t.} : \quad & f_i(x) \leq 0, \qquad i = 1, \ldots, m. \end{aligned}$$

If all functions are convex, then the feasible set is convex. The fundamental consequence of convexity is that local optimality implies global optimality. This notion will prove extremely useful. In terms of analysis, it implies that we can characterize the optimal solution of a convex optimization problem using only local conditions. In terms of algorithms, it means that doing things that are locally optimal will result in a globally optimal solution. See also the problem in Homework One.

1.1.3 Optimality Conditions: Unconstrained Optimization

We now give optimality conditions for unconstrained optimization. These are the familiar first-order conditions of optimality for convex optimization. While simple, the condition is extremely useful both algorithmically and for analysis.

Unconstrained optimization

The formulation of an unconstrained optimization problem is as follows

$$\text{min} : f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and convex. In these problems, the necessary and sufficient condition for the optimal solution x_o is

$$\nabla f(x) = 0 \text{ at } x = x_o. \tag{1.12}$$

The intuition here is simple: if there are no local directions of descent, then a point must be locally optimal, which in turn (thanks to convexity) implies that it is globally optimal. We can see more precisely that the condition $\nabla f(x_0) = 0$ implies global optimality by using the first order conditions for convexity.

Proof: Recall that if f is convex, we have

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

At $x = x_o$, $\nabla f(x_o) = 0$ and hence this reduces to

$$f(y) \geq f(x_o) \ \forall y,$$

which indeed is the statement that x_o is a global optimum. □

Note of course that this need not be unique. For uniqueness, we need a stronger version of convexity. This will be addressed in the next lecture.

Constrained optimization

Next we consider a general constrained convex optimization problem.

$$\begin{aligned} &\min f(x) \\ &\text{s.t } x \in \mathcal{X}, \end{aligned}$$

where $f : \mathcal{X} \to \mathbb{R}$ is convex and smooth. The same intuition used above still applies here: a point x_o is globally optimal if it is locally optimal. Locally optimal means that there are no directions of descent that are feasible, i.e., that stay inside the feasible set $\mathcal{X}$. We make this precise in the next lecture.