EE 227A: Convex Optimization and Applications	**January 24, 2012**

Lecture 3: **Convex Sets and Functions**

Lecturer: Laurent El Ghaoui

Reading assignment: Chapters 2 (except §2.6) and sections 3.1, 3.2, 3.3 of BV. You can also look at section 3.1 of the web textbook.

3.1 Convex Sets

Definition. A subset $\mathcal{C}$ of $\mathbf{R}^n$ is convex if and only if it contains the line segment between any two points in it:

$$\forall\, x_1, x_2 \in \mathcal{C},\ \ \forall\, \theta_1 \geq 0,\ \theta_2 \geq 0,\ \theta_1 + \theta_2 = 1\ :\ \theta_1 x_1 + \theta_2 x_2 \in \mathcal{C}.$$

Some important special cases of convex sets are the following.

- The set is said to be an *affine subspace* if it contains the entire line passing through any two points. This corresponds to the condition above, with θ_1, θ_2 arbitrary. Subspaces and affine subspaces are convex.
- The set is said to be a *convex cone* if the condition above holds, but with the restriction $\theta_1 + \theta_2 = 1$ removed.

Examples.

- The *convex hull* of a set of points $\{x_1, \ldots, x_m\}$ is defined as

$$\mathbf{Co}(x_1, \ldots, x_m) := \left\{ \sum_{i=1}^m \lambda_i x_i\ :\ \lambda \in \mathbf{R}^m_+,\ \sum_{i=1}^m \lambda_i = 1 \right\},$$

  and is convex. The *conic hull:*

$$\left\{ \sum_{i=1}^m \lambda_i x_i\ :\ \lambda \in \mathbf{R}^m_+ \right\}$$

  is a convex cone.

- For $a \in \mathbf{R}^n$, and $b \in \mathbf{R}$, the hyperplane $\mathcal{H} = \{x\ :\ a^T x = b\}$ is affine. The *half-space* $\{x\ :\ a^T x \leq b\}$ is convex.

- For a square, non-singular matrix $R \in \mathbf{R}^{n\times n}$, and $x_c \in \mathbf{R}^n$, the *ellipsoid* $\{x_c + Ru \ : \ \|u\|_2 \le 1\}$ is convex. (The center of the epllipsoid is x_c, and you can think of R as the "radius".) With $P = RR^T$, and assuming R is full rank, we can describe the ellipsoid as
$$\left\{x \ : \ (x - x_c)^T P^{-1}(x - x_c) \le 1\right\}.$$

- A *polyhedron* is a set described by a finite number of affine inequalities and equalities:
$$\mathcal{P} = \{x \ : \ Ax \le b, \ \ Cx = d\},$$
where A, C are matrices, b, d are vectors, and inequalities are understood component-wise. Sometimes bounded polyhedra are referred to as polytopes.

- The *probability simplex*
$$\left\{p \in \mathbf{R}^n_+ \ : \ \sum_{i=1}^n p_i = 1\right\}$$
is a special case of a polyhedron, and is useful to describe discrete probabilities.

- The *second-order cone*
$$\left\{(x,t) \in \mathbf{R}^{n+1} \ : \ t \ge \|x\|_2\right\} \tag{3.1}$$
is a convex cone. It is sometimes called "ice-cream cone", for obvious reasons. (We will prove the convexity of this set later.)

- The positive semi-definite cone
$$\mathcal{S}^n_+ := \left\{X = X^T \in \mathbf{R}^{n\times n} \ : \ X \succeq 0\right\}$$
is a convex cone. (Again, we will prove the convexity of this set later.)

Support and indicator functions. For a given set S, the function
$$\phi_S(x) := \max_{u\in S} \ x^T u$$
is called the *support function* of S. If S is the unit ball for some norm: $S = \{u \ : \ \|u\| \le 1\}$, then the support function of S is the dual norm. Another important function associated with S is the indicator function
$$I_S(x) = \begin{cases} 0 & x \in S, \\ +\infty & x \notin S. \end{cases}$$

Operations that preserve convexity. Two important operations that preserve convexity are:

- *Intersection:* the intersection of a (possibly infinite) family of convex sets is convex. We can use this property to prove that the semi-definite cone $\mathcal{S}_+^n$ is convex, since

$$\mathcal{S}_+^n = \left\{X = X^T \in \mathbf{R}^{n\times n} \ : \ \forall\, z \in \mathbf{R}^n, \ \ z^T X z \geq 0\right\},$$

  from which we see that the set is the intersection of the subspace of symmetric matrices with a set described by an infinite number of linear inequalities of the form $z^T X z \geq 0$, indexed by $z \in \mathbf{R}^n$. Likewise, the second-order cone defined in (3.1) is convex, since the condition $t \geq \|x\|_2$ is equivalent to the infinite number of affine inequalities $t \geq u^T x$, $\|u\|_2 \leq 1$.

- *Affine transformation:* If a function is affine (that is, it is the sum of a linear function and a constant), and $\mathcal{C}$ is convex, then the set

$$f(\mathcal{C}) := \{f(x) \ : \ x \in \mathcal{C}\}$$

  is convex. A particular example is projection on a subspace, which preserves convexity.

Separation theorems. There are many versions of separation theorems in convex analysis. One of them is the separating hyperplane theorem:

Theorem 1 (Separating hyperplane). *If $\mathcal{C}$, $\mathcal{D}$ are two convex subsets of $\mathbf{R}^n$ that do not intersect, then there is an hyperplane that separates them, that is, there exit $a \in \mathbf{R}^n$, $a \neq 0$, and $b \in \mathbf{R}$, such that $a^T x \leq b$ for every $x \in \mathcal{C}$, and $a^T x \geq b$ for every $x \in \mathcal{D}$.*

Another result involves the separation of a set from a point on its boundary:

Theorem 2 (Supporting hyperplane). *If $\mathcal{C} \subseteq \mathbf{R}^n$ is convex and non-empty, then for any x_0 at the boundary of $\mathcal{C}$, there exist a supporting hyperplane to $\mathcal{C}$ at x_0, meaning that there exist $a \in \mathbf{R}^n$, $a \neq 0$, such that $a^T(x - x_0) \leq 0$ for every $x \in \mathcal{C}$.*

3.2 Convex Functions

Domain of a function. The *domain* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the set $\mathbf{dom} f \subseteq \mathbf{R}^n$ over which f is well-defined, in other words:

$$\mathbf{dom}\, f := \{x \in \mathbf{R}^n : -\infty < f(x) < +\infty\}.$$

Here are some examples:

- The function with values $f(x) = \log(x)$ has domain $\mathbf{dom}\, f = \mathbf{R}_{++}$.
- The function with values $f(X) = \log\det(X)$ has domain $\mathbf{dom}\, f = \mathcal{S}_{++}^n$ (the set of positive-definite matrices).

Definition of convexity. A function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if

i) **dom** f is convex;

ii) $\forall\, x, y \in \mathbf{dom}\, f$ and $\forall\, \lambda \in [0,1]$, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$.

Note that the convexity of the domain is required. For example, the function $f : \mathbf{R} \to \mathbf{R}$ defined as

$$f(x) = \begin{cases} x & \text{if } x \notin [-1,1] \\ +\infty & \text{otherwise} \end{cases}$$

is *not* convex, although is it linear (hence, convex) on its domain $]-\infty, -1) \cup (1, +\infty[$.

We say that a function is *concave* if $-f$ is convex.

Here are some examples:

- The support function of any set is convex.
- The indicator function of a set is convex if and only if the set is convex.
- The quadratic function $f(x) = x^T P x + 2q^T x + r$, with $P \in \mathcal{S}^n_{++}$, is convex. (For a proof, see later.)
- The function $f : \mathbf{R} \to \mathbf{R}$ defined as $f(x) = 1/x$ for $x > 0$ and $f(x) = +\infty$ is convex.

Alternate characterizations of convexity. Let $f : \mathbf{R}^n \to \mathbf{R}$. The following are equivalent conditions for f to be convex.

- *Epigraph condition:* f is convex if and only if its *epigraph*

  $$\mathbf{epi} f := \left\{ (x,t) \in \mathbf{R}^{n+1} \ : \ t \geq f(x) \right\}$$

  is convex. We can us this result to prove for example, that the largest eigenvalue function $\lambda_{\max} : \mathcal{S}^n \to \mathbf{R}$, which to a given $n \times n$ symmetric matrix X associates its largest eigenvalue, is convex, since the condition $\lambda_{\max}(X) \leq t$ is equivalent to the condition that $tI - X \in \mathcal{S}^n_+$.

- *Restriction to a line:* The function f is convex if and only if its restriction to *any* line is convex, meaning that for every $x_0 \in \mathbf{R}^n$, and $v \in \mathbf{R}^n$, the function $g(t) := f(x_0 + tv)$ is convex.

  For example, the function $f(X) = \log\det X$ is convex. (Prove this as an exercise.) You can also use this to prove that the quadratic function $f(x) = x^T P x + 2q^T x + r$ is convex if and only if $P \succeq 0$.

- *First-order condition:* If f is differentiable (that is, **dom**f is open and the gradient exists everywhere on the domain), then f is convex if and only if

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

  The geometric interpretation is that the graph of f is bounded below everywhere by anyone of its tangents.

- *Second-order condition:* If f is twice differentiable, then it is convex if and only if its Hessian $\nabla^2 f$ is positive semi-definite everywhere. This is perhaps the most commonly known characterization of convexity.

  For example, the function $f(x,t) = x^T x/t$ with domain $\{(x,t) \ : \ t > 0\}$, is convex. (Check this!) Other examples include the log-sum-exp function, $f(x) = \log \sum_{i=1}^n \exp x_i$, and the quadratic function alluded to above.

Operations that preserve convexity.

- The nonnegative weighted sum of convex functions is convex.

- The composition with an affine function preserves convexity: if $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$ and $f : \mathbf{R}^m \to \mathbf{R}$ is convex, then the function $g : \mathbf{R}^n \to \mathbf{R}$ with values $g(x) = f(Ax + b)$ is convex.

- The *pointwise maximum* of a family of convex functions is convex: if $(f_\alpha)_{\alpha\in\mathcal{A}}$ is a family of convex functions index by α, then the function

$$f(x) := \max_{\alpha\in\mathcal{A}} \ f_\alpha(x)$$

  is convex. For example, the dual norm

$$x \to \max_{y\,:\,\|y\|\le 1} \ y^T x$$

  is convex, as the maximum of convex (in fact, linear) functions (indexed by the vector y). Another example is the largest singular value of a matrix A: $f(A) = \sigma_{\max}(A) = \max_{x\,:\,\|x\|_2=1} \|Ax\|_2$. Here, each function (indexed by $x \in \mathbf{R}^n$) $A \to \|Ax\|_2$ is convex, since it is the composition of the Euclidean norm (a convex function) with an affine function $A \to Ax$.

  Also, this can be used to prove convexity of the function we introduced in lecture 2,

$$\|x\|_{1,k} := \sum_{i=1}^k |x|_{[i]} = \max_u \ u^T|x| \ : \ \sum_{i=1}^n u_i = k, \ \ u \in \{0,1\}^n,$$

  where we use the fact that for any u feasible for the maximization problem, the function $x \to u^T|x|$ is convex (since $u \ge 0$).

- If f is a convex function in $x = (y,z)$, then the function $g(y) := \min_z f(y,z)$ is convex. (Note that joint convexity in (y,z) is essential.)

- If f is convex, its *perspective* $g(x,t) := tf(x/t)$ with domain $\mathbf{dom} g = \{(x,t) \ : \ x \in \mathbf{dom} f, \ t > 0\}$, is convex. You can use this to prove convexity of the function $f(x,t) = x^T x/t$, with domain $\{(x,t) \ : \ t > 0\}$.

- The composition with another function does not always preserve convexity. However, if the functions $g_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, k$ are convex and $h : \mathbf{R}^k \to \mathbf{R}$ is convex and non-decreasing in each argument, with $\mathbf{dom} g_i = \mathbf{dom} h = \mathbf{R}$, then $x \to (h \circ g)(x) := h(g_1(x), \ldots, g_k(x))$ is convex.

  For example, if g_i's are convex, then $\log \sum_i \exp g_i$ also is.

Conjugate function. The conjugate function of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function defined as

$$f^*(y) = \max_x \; x^T y - f(x) \;:\; x \in \mathbf{dom} f.$$

The function f^* is convex (even if f is not). The conjugate function plays a very important role in convex optimization, similar to the Fourier transform in signal processing.

For example, the conjugate of the convex quadratic function $f(x) = (1/2)x^T Q x$, with $Q \succ 0$, is $f^*(y) = (1/2)y^T Q^{-1} y$. Another important example is the conjugate of a norm, which is the indicator function of the unit ball for the dual norm:

$$f^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \leq 1, \\ +\infty & \text{otherwise.} \end{cases}$$

The conjugate of a conjugate is not always the original function. However, if f is convex, and closed (meaning that its epigraph is), then $f^{**} = f$.

Exercises

1. *Supporting hyperplane.* Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex, differentiable function, with domain $\mathbf{R}^n$. Assume that the set $\mathcal{F} := \{x \in \mathbf{R}^n \;:\; f(x) \leq 0\}$ is not empty, and let x_0 be a point on its boundary (so that $f(x_0) = 0$). Show that the set

   $$\mathcal{H} := \left\{x \in \mathbf{R}^n \;:\; \nabla f(x_0)^T (x - x_0) = 0\right\}$$

   defines a supporting hyperplane of $\mathcal{F}$ at x_0. (*Hint:* express the first-order convexity condition for f at x_0.)

2. *Maximum of a convex function over a polyhedron.* Show that the maximum of a convex function f over the polyhedron $\mathcal{P} = \mathbf{conv}\{v_1, \ldots, v_k\}$ is achieved at one of its vertices, i.e.,

   $$\max_{x \in \mathcal{P}} \; f(x) = \max_{i=1,\ldots,k} \; f(v_i)$$

   Hint: Assume the statement is false, and use Jensen's inequality.

3. Prove that the function with matrix variable $f : \mathcal{S}^n_{++} \to \mathbf{R}$, with values on its domain (the positive-definite cone, $\mathcal{S}^n_{++}$) given by $f(X) = \log \det X$, is concave. *Hint:* use the property involving the restriction of f to a line.

4. Show that the conjugate of the convex quadratic function $f(x) = (1/2)x^T Qx$, with $Q \succ 0$, is $f^*(y) = (1/2)y^T Q^{-1} y$.

5. For a matrix $X \in \mathbf{R}^{n\times m}$, the nuclear norm $\|X\|_*$ is the sum of its singular values. Show that this function is convex, by proving that it is the conjugate of the largest singular value norm.

6. Find the conjugate functions of the univariate functions $f : \mathbf{R} \to \mathbf{R}$, with values given below.

   (a) $f(x) = \log(1 + \exp(x))$.

   (b) $f(x) = \max(0, 1 - x)$.

   (c) $f(x) = (1/2)(x - x_0)^2$, with $x_0 \in \mathbf{R}$ given.

7. *Composition rule:* Using an epigraph argument, prove the result mentioned in lecture: if the functions $g_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, k$ are convex and $h : \mathbf{R}^k \to \mathbf{R}$ is convex and non-decreasing in each argument, with $\mathbf{dom} g_i = \mathbf{dom} h = \mathbf{R}$, then $x \to (h \circ g)(x) := h(g_1(x), \ldots, g_k(x))$ is convex.

8. *A general vector composition rule:* suppose

$$f(x) = h(g_1(x), g_2(x), ..., g_k(x))$$

   where $h : \mathbf{R}^k \to \mathbf{R}$ is convex, and $g_i : \mathbf{R}^n \to \mathbf{R}$. Suppose that for each i, one of the following holds:

   - h is nondecreasing in the i-th argument, and g_i is convex;
   - h is nonincreasing in the i-th argument, and g_i is concave;
   - g_i is affine.

   Show that f is convex. (This composition rule subsumes all the one given above, and is the one used in software systems such as CVX.)

9. *Arbitrage and theorems of alternatives.* Consider an event (for example, a sports game, political elections, the evolution of the stock market over a certain period) with m possible outcomes. Suppose that n wagers on the outcome are possible. If we bet an amount x_j on wager j, and the outcome of the event is i ($i = 1, \ldots, m$), then our return will be equal to $r_{ij}x_j$ (this amount does not include the stake, *i.e.*, we pay x_j initially, and receive $(1 + r_{ij})x_j$ if the outcome of the event is i, so $r_{ij}x_j$ is the net gain). We allow the bets x_j to be positive, negative, or zero. The interpretation of a negative bet is as follows. If $x_j < 0$, then initially we *receive* an amount of money $|x_j|$, with an obligation to *pay* $(1+r_{ij})|x_j|$ if outcome i occurs. In that case, we lose $r_{ij}|x_j|$, *i.e.*, our net is gain $r_{ij}x_j$ (a negative number).

We call the matrix $R \in \mathbf{R}^{m \times n}$ with elements r_{ij} the *return matrix*. A *betting strategy* is a vector $x \in \mathbf{R}^n$, with as components x_j the amounts we bet on each wager. If we use a betting strategy x, our total return in the event of outcome i is equal to $\sum_{j=1}^n r_{ij} x_j$, *i.e.*, the ith component of the vector Rx.

Suppose you are given a return matrix R. Prove the following *arbitrage theorem*: there is a betting strategy $x \in \mathbf{R}^n$ for which

$$Rx \succ 0 \tag{3.2}$$

if and only if there exists no vector $p \in \mathbf{R}^m$ that satisfies

$$R^T p = 0, \quad p \succeq 0, \quad p \neq 0. \tag{3.3}$$

We can interpret this theorem as follows. If $Rx \succ 0$, then the betting strategy x guarantees a positive return for all possible outcomes, *i.e.*, it is a sure-win betting scheme. In economics, we say there is an *arbitrage opportunity*.

If we normalize the vector p in (3.3) so that $\mathbf{1}^T p = 1$, we can interpret it as a probability vector on the outcomes. The condition $R^T p = 0$ means that the expected return

$$\mathbf{E}\, Rx = p^T Rx = 0$$

for all betting strategies. We can therefore rephrase the arbitrage theorem as follows. There is no sure-win betting strategy (or arbitrage opportunity) if and only if there is a probability vector on the outcomes that makes all bets fair (*i.e.*, the expected gain is zero).