Lecture 5: September 10

Lecturer: Geoff Gordon/Ryan Tibshirani

Scribes: Jun Ma

Fall 2013

Note: LaTeX template courtesy of UC Berkeley EECS dept.

Disclaimer: These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.

This lecture's notes illustrate some uses of various LATEX macros. Take a look at this and imitate.

5.1 Convex Sets

5.1.1 Closed and open sets

Let $C \subseteq \mathbb{R}^n$.

Definition 5.1 The affine hull of C is the smallest affine set that contains C. $\mathbf{aff}(C) = \{\sum_{i=1}^{k} \theta_i x_i | x_i \in C, \sum_{i=1}^{k} \theta_i = 1, \theta_i \in \mathbb{R}\}.$

An example of an affine set is the solution set of a system of linear equations, $C = \{x | Ax = b\}, A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^n$. If $x_1, x_2 \in C, \theta \in \mathbb{R}$, then $A(\theta x_1 + (1 - \theta)x_2) = \theta Ax_1 + (1 - \theta)Ax_2 = \theta b + (1 - \theta)b = b$. Let x be some point in \mathbb{R}^n , and $B(x, \epsilon)$ be a ball of radius ϵ centered at x. Then

Definition 5.2 x is on the boundary of C, ∂C , if for all $\epsilon > 0$, $B(x, \epsilon) \cap C \neq \emptyset$ and $B(x, \epsilon) \cap C^c \neq \emptyset$.

Definition 5.3 x is in the interior of C, int C, if $\exists \epsilon > 0 : B(x, \epsilon) \subset C$.

Definition 5.4 x is in the relative interior of C, rel int C, if $\exists \epsilon > 0 : B(x, \epsilon) \cap \operatorname{aff}(C) \subseteq C$.

Definition 5.5 The closure of C, $\mathbf{cl} C = C \cup \partial C$.

Definition 5.6 The relative boundary of C, rel $\partial C = \operatorname{cl} C \setminus \operatorname{rel} \operatorname{int} C$.

Definition 5.7 *C* is closed if $\partial C \subset C$, open if $\partial C \cap C = \emptyset$, and compact iff it is closed and bounded in \mathbb{R}^n .

5.1.2 Convex sets and examples



Figure 5.1: Example of a convex set (left) and a non-convex set (right).

Definition 5.8 A set $C \subseteq \mathbb{R}^n$ is convex iff the line segment between any two points in C is completely contained in C.

 $C \text{ is convex } \iff \forall x_1, x_2 \in C, \forall \theta \in [0, 1], \theta x_1 + (1 - \theta) x_2 \in C.$

This definition of convexity holds for any number of points in C, even infinite countable sums: C is convex $\iff \forall x_i \in C, \theta_i \ge 0, i = 1, 2, ..., \infty, \sum_{i=1}^{\infty} \theta_i = 1$, the convex combination $\sum_{i=1}^{\infty} \theta_i x_i \in C$, if the series converges.

In general, C is convex iff for any random variable X over C, $P(X \in C) = 1$, its expectation is also in C: $\mathbb{E}(X) = \int_C x P(x) dx \in C$, if the integral exists.

Definition 5.9 *C* is strictly convex $\iff \forall x_1 \neq x_2 \in C, \forall \theta \in (0,1), \theta x_1 + (1-\theta)x_2 \in \text{int } C.$

Intuitively, C is a strictly convex set iff a line segment between two points on the boundary, x_1 and x_2 , intersects the boundary only at x_1 and x_2 .



Figure 5.1: L_p norm balls are convex sets for $p \ge 1$, and non-convex for p < 1.

Some examples of convex sets:

- The empty set \emptyset , the singleton set $\{x_0\}$, and the complete space \mathbb{R}^n .
- Lines $\{x \in \mathbb{R}^n | x = \theta x_1 + (1 \theta x_2), x_1, x_2 \in \mathbb{R}^n, \theta \in \mathbb{R}\}$, and line segments $\{x \in \mathbb{R}^n | x = \theta x_1 + (1 \theta x_2), x_1, x_2 \in \mathbb{R}^n, \theta \in [0, 1]\}$
- Hyperplanes $\{x \in \mathbb{R}^n | a^T x = b, a \in \mathbb{R}^n, b \in \mathbb{R}\}$, and halfspaces $\{x \in \mathbb{R}^n | a^T x \leq b, a \in \mathbb{R}^n, b \in \mathbb{R}\}$
- Euclidean balls $B(x_0, r) = \{x \in \mathbb{R}^n | \|x_0 x\|_2 \le r\}$
- L_p balls, $p \ge 1.\{x \in \mathbb{R}^n | \|x_0 x\|_p \le r\}$, where $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$. L_p balls for $p \in (0,1)$ are not convex.
- Polyhedron: the solution set of a finite number of linear equalities and inequalities. $P = \{x \in \mathbb{R}^n | a_j^T x \le b_j, j = 1, \dots, m, c_i^T x = d_i, i = 1, \dots, p\}$, or $P = \{x | Ax \le b, Cx = d\}$
- Polytope (bounded polyhedron), intersection of halfspaces and hyperplanes.

5.1.3 Convex and conic hulls

Definition 5.10 The convex hull of a set C, $\operatorname{conv}[C] = \{\sum_{i=1}^{k} \theta_i x_i | x_i \in C, \theta_i \ge 0 \ \forall i = 1, \dots, k, \sum_{i=1}^{k} \theta_i = 1, k \in \mathbb{Z}_+\}.$

Properties of convex hull:

- $\operatorname{conv}[C]$ is the smallest convex set that contains C.
- **conv**[C] is convex.
- $C \subseteq \mathbf{conv}[C]$
- $\forall C, C' \text{ convex sets}, C \subset C' \implies \mathbf{conv}[C] \subseteq C'$

Definition 5.11 *C* is a cone $\iff x \in C, \theta \ge 0 \implies \theta x \in C$. A convex cone is a cone which is also a convex set.

Definition 5.12 The conic hull of a set C, $\operatorname{cone}[C] = \{x | x = \theta_1 x_1 + \ldots + \theta_k x_k, \theta_i \ge 0, x_i \in C\}$

An example of a convex cone is the set of symmetric PSD matrices, $S^n_+ = \{A \in \mathbb{R}^{n \times n} | A \succeq 0\}$. If $A, B \in S^n_+, \theta \ge 0$, then $x^T(\theta A + (1 - \theta)B)x = \theta x^T A x + (1 - \theta)x^T B x \ge 0$.

5.1.4 Convex set representations



Figure 5.1: Representation of a convex set as the convex hull of a set of points (left), and as the intersection of a possibly infinite number of halfspaces (right).

5.1.4.1 Convex hull representation

Let $C \subseteq \mathbb{R}^n$ be a closed convex set. Then C can be written as $\mathbf{conv}(X)$, the convex hull of possibly infinitely many points $(|X| = \infty)$. Also, any closed convex set is the convex hull of itself.

5.1.4.2 Dual representation with halfspaces

Let $C \subseteq \mathbb{R}^n$ be a closed convex set. Then C can be written as $\bigcup_i \{x | a_i^T x + b_i \leq 0\}$, the intersection of possibly infinitely many closed halfspaces. Also, every closed convex set is the intersection of all halfspaces that contain it.

5.1.5 Covexity preserving operations

Let $C \in \mathbb{R}^n$ be a convex set. Then, the following operations preserve convexity:

- Translation C + b
- Scaling αC
- Intersection If D is a convex set, then $C \cap D$ is convex. In general, if S_{α} is a convex set $\forall \alpha \in \mathbb{A}$, then $\bigcap_{\alpha \in \mathbb{A}} S_{\alpha}$ is convex.
- Affine function Let $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^n$. Then $AC + b = \{Ax + b | x \in C\} \subseteq \mathbb{R}^m$ is convex.
- Set Sum Let $C_1 \subseteq \mathbb{R}^n$ and $C_2 \subseteq \mathbb{R}^m$ be convex sets. Then $C_1 + C_2 = \{x_1 + x_2 | x_1 \in C_1, x_2 \in C_2\}$ is convex.
- Direct Sum $C_1 \times C_2 = \{(x_1, x_2) \in \mathbb{R}^{n+m}, x_1 \in C_1, x_2 \in C_2\}$
- **Perspective projection**(pinhole camera) If $C \subset \mathbb{R}^n \times R_{++}$ is a convex set, then P(C) is also convex, where

$$P(x) = P(x_1, x_2, \dots, x_n, t) = (x_1/t, x_2/t, \dots, x_n/t) \in \mathbb{R}^n$$

• Linear fractional function Let $f(x) = \frac{Ax+b}{c^Tx+d}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}^n$, $d \in \mathbb{R}$, dom $f = \{x | c^Tx + d > 0\}$. Then $f(C) \in \mathbb{R}^m$ is convex.

An operation which does not preserve convexity is set union.

5.1.6 Separating hyperplane theorem



Figure 5.1: The hyperplane $\{x | a^T x = b\}$ separates the disjoint convex sets C and D

Theorem 5.13 For convex sets $C, D \subseteq \mathbb{R}^n, C \cap D = \emptyset, \exists a \in \mathbb{R}^n, b \in \mathbb{R}$, such that $\forall x \in C, a^T x \leq b, \forall y \in D, a^T y \geq b$.

Let C_1 and C_2 be two convex sets. Then

Definition 5.14 C_1 and C_2 are strongly separated if $a^T[C_1 + B(0,\epsilon)] > b$ and $a^T[C_2 + B(0,\epsilon)] < b$.

Definition 5.15 C_1 and C_2 are properly separated if it is not the case that both $C_1 \subseteq \{x : a^T = b\}$ and $C_2 \subseteq \{x : a^T x = b\}$

Definition 5.16 C_1 and C_2 are strictly separated if $a^T x > b \forall x \in C_1$ and $a^T x < b \forall x \in C_2$.

Theorem 5.17 If C_1, C_2 are non-empty convex sets in \mathbb{R}^n , with $\operatorname{cl} C_1 \cap \operatorname{cl} C_2 = \emptyset$, and either C_1 or C_2 bounded, there exists a hyperplane separating C_1 and C_2 strongly.

Consider $x = (x_1, x_2) \in \mathbb{R}_{++} \times \mathbb{R}$. An example where strong separation fails is when $C_1 = \{x | x_2 \ge \frac{1}{x_1}\}$, $C_2 = \{x | x_2 \le 0\}$. Here, $x_2 = 0$ separates C_1 and C_2 , but does not strongly separate them, since neither of the sets are bounded.

Theorem 5.18 If C_1, C_2 are non-empty convex sets in \mathbb{R}^n , there exists a hyperplane separating C_1 and C_2 strongly $\iff \inf_{x_1 \in C_1, x_2 \in C_2} \{|x_1 - x_2|\} > 0 \iff dist(C_1, C_2) > 0 \iff 0 \in \mathbf{cl} (C_1 - C_2).$

5.1.7 Supporting hyperplane theorem



Figure 5.1: The hyperplane $\{x | a^T x = a^T x_0\}$ supports C at $x_0/$

Theorem 5.19 For any convex set C and any boundary point x_0 , there exists a supporting hyperplane for C at x_0 . That is, for any $x_0 \in \partial C$, \exists a hyperplane $\{x | a^T x = b, a \neq 0\}$, such that $\forall x \in C, a^T x \leq a^T x_0, a^T x_0 = b$.

The partial converse of the supporting hyperplane theorem says that if a set is closed, has a non-empty interior, and has a supporting hyperplane at every point in its boundary, then it is convex.

5.1.8 Proving a set convex

To summarize, one can prove that a set is convex using any of the following:

- Definition of convexity
- Representation as a convex hull
- Representation as the intersection of halfspaces
- Partial converse of the supporting hyperplane theorem
- Using convexity-preserving operations on simple sets to build up C

References

[Boyd2004] S. BOYD and L. VANDENBERGHE, "Convex optimization", Cambridge University Press, 2004

5.2 Convex functions

Definition 5.20 A function f is convex if : 1. Domain of f is a convex set. 2. $f(\theta x + (1 - \theta)y) \le \theta f(x) + (1 - \theta)f(y)$ $\forall x, y \in Domf$ $\forall \theta \in [0, 1]$



Figure 5.1: Graphic interpretation of convex function

 $\begin{array}{l} \textbf{Definition 5.21} \quad Strictly \ convex \ function \\ f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y) \\ \forall x, y \in Domf \\ \forall \theta \in [0, 1] \end{array}$

Definition 5.22 If -f is convex, then f is concave.

Definition 5.23 Strongly convexity $(\bigtriangledown f(x) - \bigtriangledown f(y))^T (x - y) \ge m \parallel y - x \parallel_2^2$ or equivantly, $f(y) \ge f(x) + \bigtriangledown f(x)^T (y - x) + \frac{m}{2} \parallel y - x \parallel_2^2$ If without gradient: $\exists t \in [0, 1]$ $f(tx + (1 - t)y) \le tf(x) + (1 - t)f(y) - \frac{1}{2}mt(1 - t) \parallel x - y \parallel_2^2$

If with Hessian: $\bigtriangledown^2 f(x) \succeq mI$

Note: A strongly convex function is also strictly convex, but not vice versa.

Example: $f(x) = x^4$ is convex, strictly convex, not strongly convex. f(x) = |x| is convex, not strictly convex. Convex functions: $|x|^p, (p \ge 1), f(x) = max(x_1, ..., x_n)$ Concave functions: $f(x) = (\sum_{i=1}^n x_i)^{1/n}, \log(x)$ **Theorem 5.24** Extended reals Let

$$\tilde{f} = \begin{cases} f(x) & x \in Dom f \\ \infty & x \notin Dom f \end{cases}$$

Then f is convex means \tilde{f} is also convex.

Definition 5.25 Epigraph:

 $Epi(f) = \{(x,t) : x \in Domf, t \ge f(x)\}$ If f is convex, then Epi(f) is a convex set



Figure 5.2: Illustration of epigraph

5.2.1 Convex function properties

5.2.1.1 0th order characterization

f is convexity, iff, g(t) = f(x + tv) is convex.

5.2.1.2 1st order characterization

Let f be differentiable, f is convex, iff $f(y) \leq f(x) + \nabla f(x)^T (y - x), \forall y$

5.2.1.3 2nd order characterization

Let f be twice differentiable, f is convex iff $\nabla^2 f(x) \leq 0, \forall x \in domain$

5.2.1.4 Jensen's inequality

f is convex, so $f(Ex) \leq Ef(x)$



Figure 5.3: Illustration of Jensen's inequality

5.2.2 Proving a function convex

Use definition directly Prove that epigraph is convex via set methods 0th, 1st, 2nd order convexity properties Construct f from simpler convex functions using convexity preserving operations

5.2.3 Convexity-preserving function operations

Nonnegative weighted sum Pointwise max/sup Extension of pointwise max/sup Affine map Composition Perspecive map

5.3 Gradient Descent

5.3.1 Choose step size

Recall that we have $f : \mathbb{R}^n \to \mathbb{R}$, convex and differentiable. We want to solve

 $\min_{x\in\mathbb{R}^n}f(x)$

i.e, to find x^* such that $f(x^*) = \min f(x)$.

Gradient descent: choose initial $x^{(0)} \in \mathbb{R}^n$, repeat :

$$x^{(k)} = x^{(k-1)} - t_k \cdot \nabla f(x^{(k-1)}), k = 1, 2, 3, \dots$$

Stop at some point (When to stop is quite dependent on what problems you are looking at).

Figure 5.1 is shows a example that we cannot always continue and it depends where we start. i.e. If we start at a spot somewhere between the purple and orange, it would stay there and go nowhere.



At each iteration, consider the expansion

$$f(y) \approx f(x) + \nabla f(x)^{T}(y-x) + \frac{1}{2t} ||y-x||^{2}$$

We can use quadratic approximation, replacing usual $\nabla^2 f(x)$ by $\frac{1}{t}I$, then we have

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

which is a linear combination to f, and

$$\frac{1}{2t}\|y-x\|^2,$$

which is a proximity term to x, with weight $\frac{1}{2t}$.

Then, choose next point $y = x^+$ to minimize quadratic approximation

$$x^+ = x - t\nabla f(x)$$

as shown in Figure 5.2.

5.3.2 Fixed step size

Simply take $t_k = t$ for all k = 1, 2, 3, . can diverge if t is too big. Consider $f(x) = (10x_1^2 + x_2^2/2)$, Figure 5.3 shows the gradient descent after 8 steps. It can be slow if t is too small. As for the same example, gradient descent after 100 steps in Figure 5.4, and gradient descent after 40 appropriately sized steps in Figure 5.5.



blue point is x, red point is x^+

Convergence analysis will give us a better idea which one is just right.





5.3.2.1 Backtracking line search

Adaptively choose the step size:

First, fix a parameter $0 < \beta < 1$, then at each iteration, start with t = 1, and while

$$f(x - \nabla f(x)) > f(x) - t\alpha \|\nabla f(x)\|^2,$$

update $t = \beta t$, as shown in Figure 5.6 (from B & V page 465), for us $\Delta x = -\nabla f(x)$, $\alpha = 1/2$.

Backtracking line search is simple and work pretty well in practice. Figure 5.7 shows that backpacking picks up roughly the right step size(13 steps) for the same example, with $\beta = 0.8$ (B & V recommend $\beta \in (0.1, 0.8)$).





5.3.3 Exact line search

At each iteration, do the best we can along the direction of the gradient,

$$t = \underset{s \ge 0}{\operatorname{argmin}} f(x - s\nabla f(x)).$$

Usually, it is not possible to do this minimization exactly.

Approximations to exact line search are often not much more efficient than backtracking, and it's not worth it.



Figure 5.9:



Figure 5.10: