Lecture 3

DJM

9 October 2018

Convex sets and functions

Definitions

Thanks: Much of this material is borrowed/copied from Ryan Tibshirani.

Set C is *convex* iff $\forall x, c \in C, \forall t \in [0;1] \;\; tx + (1-t)y \in C$.

So C is convex iff for any two points in C their segment is also entirely in C.

Convex combination of set of points $x_1, \ldots, x_k \in \mathbb{R}^n$ is

$$\left\{ \sum_{i=1}^{k} \Theta_i x_i : \sum_{i=1}^{k} \Theta_i = 1,\; \forall i \; \Theta_i \in [0;1] \right\}.$$

Convex hull of any $C \in \mathbb{R}^n$, denoted *conv*(C) is a union of all convex combinations of different elements of C.

Some examples

- Empty set, point, line, segment.
- Norm ball: $\{x : \|x\| < r\}$.
- Hyperplane $\left\{x : a^\top x = b\right\}$, Affine space $\{x : Ax = b\}$.
- Hyperspace: $\left\{x : a^\top x \leq b\right\}$, Polyhedron $\{x : Ax \leq b\}$.
- Cone such that if $x_1, x_2 \in C$ then $t_1 x_1 + t_2 x_2 \in C \; \forall t_1, t_2 \geq 0$.

Cones

Set C is a *cone* iff $\forall t \geq 0,\; x \in C \implies t^\top x \in C$.

Type of cones:

- Norm cone: $\{(x,t) : \|x\| \leq t\}$.
- Normal cone for some C and $x \in C$: $N_C(x) = \left\{g : g^\top x \geq g^\top y \; \forall y \in C\right\}$.
- Positive semidefinite cone $S^n_+ = \{x \in S^n : x \succeq 0\}$, S^n is Hilbert space.

Key properties of convex sets

- Separating hyperplane. A, B are convex, nonempty, disjoint. Then $\exists a, b :\; A \subseteq \left\{x : a^\top x \leq b\right\}, B \subseteq \left\{x : a^\top \geq b\right\}$.
- Supporting hyperplane. C nonempty, convex, $x_0 \in$ *boundary*(C). Then $\exists a :\; C \subseteq \left\{x : a^\top x \leq a^\top x_0\right\}$.

Operations preserving convexity

- Intersection.
- Scaling, translation. C is convex $\implies$ $aC + b$ is convex.
- Affine image and preimage. $f(x) = Ax + b$, C is convex $\implies$ $f(C), f^{-1}(C)$ are convex.
- Lots more (See (Boyd and Vandenberghe 2004), chapter 2).

$A_1, \ldots, A_k, B \in \mathbb{S}^n$ – symmetric matrices. Then $C = \left\{ x \in \mathbb{R}^k : \sum\limits_{i=1}^{k} x_i A_i \preceq B \right\}$.

$f : \mathbb{R}^k \to \mathbb{S}^n$, $f(x) = B - \sum\limits_{i=1}^{k} x_i A_i$. $C = f^{-1}(S_+^n)$ – affine preimage of convex cone.

Convex functions

Function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* iff $\mathrm{dom}\,(f) \subseteq \mathbb{R}^n$ is convex and

$$\forall x, y \in \mathrm{dom}\,(f)\,, t \in [0; 1] \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

Other definitions:

- f is *concave* iff $-f$ is convex.
- f is *strictly convex* iff $\forall t \in (0; 1)$ the inequality in definition is strict.
- f is *strongly convex* with parameter τ iff $f(x) - \frac{\tau}{2} \|x\|_2^2$ is convex.

Examples

- $f(x) = \frac{1}{x}$ is strictly convex, but not strongly.
- Univariate functions:
  - e^{ax} is convex $\forall a \in \mathbb{R}$ over $\mathbb{R}$.
  - x^a convex given $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$.
  - $\log x$ is concave over $\mathbb{R}_+$.
- Affine $a^\top x + b$ is both convex and concave.
- Quadratic $\frac{1}{2} x^\top Q x + b^\top x + c$ is convex if $Q \succeq 0$.
- $\|u - Ax\|_2^2$ convex since $A^\top A \succeq 0$.
- Norms: all vector norms and most matrix norms are convex.
- Indicator function is convex. C is a convex set, then $I_C(x) = \begin{cases} 0, & x \in C \\ \infty, & otherwise \end{cases}$.
- Support function is convex $\forall C$. $I_C^*(x) = \max\limits_{y \in C} x^\top y$.

Key properties

- f is convex iff its epigraph is convex, where $epi(f) = \{(x, t) \in \mathrm{dom}\,(f) \times \mathbb{R} : f(x) \leq t\}$.
- f is convex $\implies$ all its sublevel sets are convex. $C_t = \{x \in \mathrm{dom}\,(f) : f(x) \leq t\}$. The converse is false.
- Assume f is differentiable. Then f is convex iff $\mathrm{dom}\,(f)$ is convex and $\forall x, y \in \mathrm{dom}\,(f) \quad f(y) \geq f(x) + \nabla f(x)^\top (y - x)$. Essentially, it means that f's graph is above any tangent plain.
- Assume f is twice differentiable. f is convex iff $\mathrm{dom}\,(f)$ is convex and $\forall x \in \mathrm{dom}\,(f) \quad \nabla^2 f(x) \succeq 0$.

Operations preserving function convexity

- Nonnegative linear combination.
- Pointwise maximum. $\forall s \in S \quad f_s$ is convex $\implies f(x) = \max\limits_{S} f_s(x)$ is also convex.
- Partial minimum. $g(x, y)$ convex over variables x, y; C convex. Then $f(x) = \min\limits_{y \in C} g(x, y)$ is also convex. E.g., $f(x) = \max\limits_{y \in C} \|x - y\|$ or $f(x) = \min\limits_{y \in C} \|x - y\|$.

Terminology

Optimization problem

A convex optimization problem (program)

$$\begin{aligned} \min_{x \in D} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0 \quad \forall i \in [1 : m] \\ & Ax = b \end{aligned} \tag{1}$$

where f, g_i are convex and $D = \mathrm{dom}\,(f) \cap \mathrm{dom}\,(g_i)$.

Terms

$$\begin{aligned} \min_{x \in D} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0 \quad \forall i \in [1 : m] \\ & Ax = b \end{aligned} \tag{2}$$

- f – criteria or objective function.
- g_i – inequality constraints.
- x is a *feasible point* if it satisfies the conditions, namely $x \in D$, $g_i(x) \leq 0$, and $Ax = b$.
- $\min f$ over feasible points points – *optimal value* f^*.
- If x is feasible and $f(x) = f^*$ then x is an *optimum* (solution, minimizer).
- Feasible x is a *local optimum* if $\exists R > 0$ such that $\forall y \in B_R(x) \quad f(x) \leq f(y)$.
- If x is feasible and $f(x) \leq f^* + \varepsilon$ then x is *ε-suboptimal*.
- If x is feasible and $g_k(x) = 0$ then g_k is *active* at x (otherwise inactive).

Properties

- Solution set X_{opt} is convex.
- If f is strictly convex then the solution is unique.
- For convex optimization problems all local optima are global.
- The set of feasible points is convex.

Example: Lasso.

$\min_{\beta} \|y - X\beta\|_2^2$ subject to $\|\beta\|_1 \leq s$.

- $g_1(\beta) = \|\beta\|_1 - s$ – convex, no equality constraints.
- X is $n \times p$ matrix
  - If $n \geq p$ and X is full rank then $\nabla^2 f(\cdot) = 2X^\mathsf{T} X$ is positive definite matrix. The function is strictly convex, therefore the solution is unique.
  - If $p > n$ then $\exists \beta \neq 0$ such that $X\beta = 0 \implies$ multiple solutions.

First Order Condition

- convex problem with differentiable f
- a feasible x is optimal iff $\nabla f(x)^T (x - y) \geq 0$, $\forall$ feasible y
- if unconstrained, the condition reduces to $\nabla f(x) = 0$

$$\min_x \frac{1}{2} x^T Q x + b^T x + c, \qquad Q \succeq 0$$

- FOC: $\nabla f(x) = Q^T x + b = 0$
- if $Q \succ 0 \to x^* = -Q^{-1} b$
- if Q singular, $b \notin Col[Q] \to$ no solution
- if Q singular, $b \in Col[Q] \to x^* = -Q^* b + z$ with $z \in null[Q]$

Projection onto convex C :

$$\min_x \|a - x\|_2^2 \qquad \text{s.t.} \qquad x \in C$$

-FOC : $\nabla f(x)^T (y - x) = (x - a)^T (y - x) \geq 0$, $\forall y \in C \Leftrightarrow a - x \in N_2(x)$

Useful operations

Partial optimization

Recall: $h(x) = \min_{y \in C} f(x, y)$ is convex if f is convex, and C is convex.

$$\begin{aligned} \min_{x_1, x_2} \quad & f(x_1, x_2) & & \min_{x_1} \quad \tilde{f}(x_1) \\ \text{s.t.} \quad & g_1(x_1) \leq 0 & \iff & \text{s.t.} \quad g_1(x_1) \leq 0 \\ & g_2(x_2) \leq 0 & & \end{aligned} \tag{3}$$

where $\tilde{f}(x_1) = \min \{ f(x_1, x_2) : g_2(x_2) \leq 0 \}$.

- The right problem is convex if the left is (and vice versa)

Transformations

- We can use a monotone increasing transformation $h : \mathbb{R} \to \mathbb{R}$:

$$\min_{x \in C} f(x) \Rightarrow \min_{x \in C} h(f(x))$$

- We can use a change of variable transformation $\phi : \mathbb{R}^n \Rightarrow \mathbb{R}^m$:

$$\min_{x \in C} f(x) \Leftrightarrow \min_{\phi(y) \in C} f(\phi(y))$$

Example: Geometric Program

$$\min_{x \in C} f(x) = \sum_{k=1}^{p} \gamma_k x_1^{a_{k_1}} x_2^{a_{k_2}} .. x_n^{a_{k_n}} \qquad \text{(posynomial)}$$

- C : involves (convex) inequalities in some form and equalities are affine.

- We can change above non-convex problem to the following convex problem by letting $y_i = \log x_i$

Eliminate equality constraints

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0 \\ & Ax = b \end{aligned} \tag{4}$$

- x feasible means $\exists M : col[M] = null[A]$ and $x_0 s.t. A x_0 = b$
- Let $x = My + x_0$

Then the following is an equivalent problem:

$$\begin{aligned} \min_{y} \quad & f(My + x_0) \\ \text{s.t.} \quad & g_i(My + x_0) \le 0 \end{aligned} \tag{5}$$

Introduce slack variables

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0 \\ & Ax = b \end{aligned} \tag{6}$$

- Can add equality constraints:

$$\begin{aligned} \min_{x,s} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) + s_i = 0 \\ & s_i \ge 0 \\ & Ax = b \end{aligned} \tag{7}$$

- But this is nonconvex unless g_i are affine

Relaxation

We can relax nonaffine constraints

$$\min_{x \in C} f(x) \Rightarrow \min_{x \in \tilde{C}} f(x)$$

with $C \subset \tilde{C}$

- In this case optimum of new problem is smaller or equal to the optimum of the original problem.

Standard problems (with examples)

LP (Linear Programs)

$$\min_x c^T x$$

with affine constraints

- Basis Pursuit
  $\min_\beta \|\beta_0\|$ s.t. $X\beta = y$
- Above problem can be relaxed to :
  $\min_\beta \|\beta\|_1$ s.t. $X\beta = y$.
- This relaxation can be reformulated to a LP problem:
  $\min_{\beta, z} 1^T z$ s.t. $z \geq \beta, z \geq -\beta, X\beta = y$
- Dantzig selector
  $\min_\beta \|\beta\|_1$ s.t. $\|x^T(y - X\beta)\|_\infty \leq \lambda$

QP (Quadratic program)

Lasso, ridge regression, OLS, Portfolio Optimization

SDP (Semi-Definite program)

$$\begin{aligned} \min_{X \in S_n} \quad & tr(C^T X) \\ \text{s.t.} \quad & \operatorname{tr}(A_i^\top X) = b_i \\ & X \succeq 0 \end{aligned} \tag{8}$$

Conic program

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & D(x) + d \in K \end{aligned} \tag{9}$$

D a linear mapping, K a closed convex cone.

- Very similar to an LP

Relations

$LP \subset QP \subset SOCP \subset SDP \subset CP(ConicProgramming)$

Duality

Introduction

- Suppose we want to *Lower bound* our convex program
- Find $B \leq \min_x f(x), \quad x \in C$.

Example:

$$\begin{aligned} \min_{x,y} \quad & x + y \\ \text{s.t.} \quad & x + y \geq 2 \\ & x, y \geq 0 \end{aligned} \tag{10}$$

- What is B?

Harder

Example:

$$\begin{aligned} \min_{x,y} \quad & x + 3y \\ \text{s.t.} \quad & x + y \geq 2 \\ & x, y \geq 0 \end{aligned} \tag{11}$$

- What is B?

Why?

Example:

$$\begin{aligned} \min_{x,y} \quad & x + 3y \\ \text{s.t.} \quad & x + y \geq 2 \\ & x, y \geq 0 \end{aligned} \iff \begin{aligned} \min_{x,y} \quad & (x + y) + 2y \\ \text{s.t.} \quad & x + y \geq 2 \\ & x, y \geq 0 \end{aligned} \tag{12}$$

- What is B?

Harderer

Example:

$$\begin{aligned} \min_{x,y} \quad & px + qy \\ \text{s.t.} \quad & x + y \geq 2 \\ & x, y \geq 0 \end{aligned} \tag{13}$$

- What is B?

Solution

$$
\begin{aligned}
\min_{x,y} \quad & px + qy \\
\text{s.t.} \quad & x + y \geq 2 \\
& x, y \geq 0
\end{aligned}
\iff
\begin{aligned}
\min_{x,y} \quad & px + qy \\
\text{s.t.} \quad & ax + ay \geq 2a \\
& bx, cy \geq 0 \\
& a, b, c \geq 0
\end{aligned}
\tag{14}
$$

- Adding implies
$$(a+b)x + (a+c)y \geq 2a$$
- Set $p = (a+b)$ and $q = (a+c)$ we get that $B = 2a$

Better

- We can make this lower bound bigger by maximizing:

$$
\begin{aligned}
\max_{a,b,c} \quad & 2a \\
\text{s.t.} \quad & a + b = p \\
& a + c = q \\
& a, b, c \geq 0
\end{aligned}
\tag{15}
$$

- This is the **Dual** of the **Primal** LP
$$
\begin{aligned}
\min_{x,y} \quad & px + qy \\
\text{s.t.} \quad & x + y \geq 2 \\
& x, y \geq 0
\end{aligned}
\tag{16}
$$

- Note that the number of Dual variables (3) is the number of Primal constraints

General LP

$$
\begin{aligned}
& \text{(P)} \\
\min_{x} \quad & c^\top x \\
\text{s.t} \quad & Ax = b \\
& Gx \leq h
\end{aligned}
\qquad
\begin{aligned}
& \text{(D)} \\
\max_{u,v} \quad & -b^\top u - h^\top v \\
\text{s.t} \quad & -A^\top u - G^\top v = c \\
& v \geq 0
\end{aligned}
\tag{17}
$$

Exercise

Alternate derivation

$$
\begin{aligned}
\min_{x} \quad & c^\top x \\
\text{s.t} \quad & Ax = b \\
& Gx \leq h
\end{aligned}
\tag{18}
$$

- Suppose that some x is feasible.
- Then, for that x,
$$c^\top x \geq c^\top x + u^\top (Ax - b) + v^\top (Gx - h) =: L(x, u, v).$$
as long as $v \geq 0$ and u is anything.
- We call $L(x, u, v)$ the **Lagrangian**.

Now, suppose C is the feasible set, and f^* is the optimum

$$f^* \geq \min_{x \in C} L(x,u,v) \geq \min_{x} L(x,u,v) =: g(u,v)$$

- We call $g(u,v)$ the **Lagrange Dual Function**

Weak duality

Consider the generic (primal) convex program

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t} \quad & l_i(x) = 0 \\ & h_i(x) \leq 0 \end{aligned} \tag{19}$$

- For feasible x, $v \geq 0$

$$f(x) \geq f(x) + u^\top h(x) + v^\top l(x) \geq \min_{x} L(x,u,v) = g(u,v).$$

- Therefore,

$$f^* \geq \max_{\forall u, v \geq 0} g(u,v) = g^*.$$

- This is **weak duality**.
- Note that the Dual Program is always convex even if P is not (pointwise max of linear functions)

Strong duality

$$f^* = g^*$$

- Sufficient conditions for strong duality: **Slater's conditions**
- If P is convex and there exists x such that for all i, $h_i(x) < 0$ (strictly feasible), then we have strong duality. (Extension: strict inequalities for i when h_i not affine.)
- Sufficient conditions for strong duality of an LP: strong duality if either P or D is feasible. (Dual of D $= P$)

Example

Dual for Support Vector Machine

$$\begin{array}{llcll} & \text{(P)} & & & \text{(D)} \\ \min\limits_{\xi,\beta,\beta_0} & \frac{1}{2}\|\beta\|_2^2 + C\mathbb{1}^\top \xi & & \max\limits_{w} & -\frac{1}{2} w^\top \tilde{X}^\top \tilde{X} w + \mathbb{1}^\top w \\ \text{s.t} & \xi_i \geq 0 & & \text{s.t} & 0 \leq w \leq C\mathbb{1} \\ & y_i(x_i^\top \beta + \beta_0) \geq 1 - \xi_i & & & w^\top y = 0 \end{array} \tag{20}$$

where $\tilde{X} = \text{diag}(y)X$.

- $w = 0$ is Dual feasible.
- Rewrite $0 \leq w \leq C\mathbb{1}$ as $w > 0, \quad w_i - C \leq 0$.
- Thus, $w = 0$ is strictly feasible.
- We have strong duality by Slater's conditions.

KKT conditions

1. Stationarity: $0 \in \partial(f(x) + u^T h(x) + v^T l(x))$: For some pair (u, v), x minimizes the Lagrangian.
2. Complementary slackness: $u_i h_i(x) = 0 \ \ , \forall i$
3. Primal feasibility: $h_i(x) \leq 0 \ , \ l_i(x) = 0$
4. Dual feasibility: $u \geq 0$

Theorem: Solutions x^* and (u^*, v^*) Primal-Dual optimal and $f^* = g^*$, then they satisfy the KKT conditions.

Theorem: Solutions x^* and (u^*, v^*) that satisfy the KKT conditions are Primal-Dual optimal.

Example (SVM cont.)

1. Stationarity: $w^\top y = 0 \ , \ \beta = w^\top \tilde{X} \ , \ w = C1 - v$
2. CS: $v_i \zeta_i = 0 \ , \ w_i(1 - \zeta_i - y_i(x_i^\top \beta + \beta_0)) = 0$
3. Clear.
4. Clear.

Constraints and Lagrangians

When are the two following forms equivalent?

constrained form (C):

$$\min f(x)$$
$$s.t. \ h(x) \leq t$$

Lagrangian form (L):

$$\min f(x) + \lambda h(x)$$

When C is strictly feasible, strong duality holds. So there exists λ such that for each x that solves C those x minimize L.

Now, if x^* solves L, then KKT condition for C hold by taking $t = h(x^*)$ and so x^* is a solution of C.

Algorithms

References

Boyd, S.P., and L. Vandenberghe. 2004. *Convex Optimization.* Cambridge, UK: Cambridge Univ Press.