10-725/36-725: Convex Optimization **Fall 2019**

Lecture 4: September 9

Lecturer: Ryan Tibshirani *Scribes: Arish Alreja,Tara Pirnia,Yash Chandarana*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

4.1 Example of convex sets: Linear matrix inequality solution set

Given $A_1, \ldots, A_k \in \mathbb{S}^n$, a linear matrix inequality in $x \in \mathbb{R}^k$ is

$$x_1A_1 + x_2A_2 + \ldots x_kA_k \preceq B$$

The set C of points x that satisfy the above inequality is convex. This can be shown by verifying that $x, y \in C \implies tx + (1-t)y \in C$ or by using the fact that affine pre-image preserves convexity i.e. if $S \subseteq \mathbb{R}^m$ is a convex set and $A = \mathbb{R}^{m \times n}$, then $P = A^{-1}(Q)$ is convex.

4.2 Example of convex functions

4.2.1 Norms

$\|x\|$ is convex for any norm. Spectral norm ($\|X\|_{\text{op}} = \sigma_1(X)$) and trace norm ($\|X\|_{\text{tr}} = \sum_{i=1}^{r} \sigma_i(X)$) for matrices are also convex. Here σ_i are the singular values of X. Spectral norm and trace norm are analogous to L_∞ and L_1 norm.

4.2.2 Indicator function

If a set C is convex then,

$$I_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

The above indicator function is convex

4.2.3 Support function

For an set C, it's support function

$$I_C^*(x) = \max_{y \in C} x^T y$$

is convex. Note that the I^* shows that support function is conjugate to indicator function.

4.3 Operations preserving convexity

- Non-negative linear combinations
- Point-wise maximization: $\max f_i(x), \dots, f_k(x)$ is convex if $f_1, \dots, f_k$ are convex.
- Partial minimization: $g(x, y)$ is convex in x, y and C is convex then $f(x) \min_{y \in C} g(x, y)$ is convex.

4.3.1 Example: Distances to a set

C is an arbitrary set, consider the maximum distance to C under arbitrary norm, $\|\cdot\|$. The following is always convex using the convexity of $\|x - y\|$ for fixed y and pointwise maximization rule.

$$f(x) = \max_{y \in C} \|x - y\|$$

Consider minimum distance to C:

$$f(x) = \min_{y \in C} \|x - y\|$$

Using the convexity of $\|x - y\|$ for x, y jointly and assuming C to be convex, th above would be convex.

4.4 More operations preserving convexity

- Affine composition: if f is convex, so is $g(x) = f(Ax + b)$
- General composition: If $f = h \circ g$, then f is convex when
  - h is convex, g is convex and h is non-decreasing
  - h is convex, g is concave and h is non-increasing

  and f is concave when
  - h is concave, g is convex and h is non-decreasing
  - h is concave, g is concave and h is non-increasing

  To remember these consider $f : \mathbb{R} \to \mathbb{R}$ and check for sign of $f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$.
- Vector composition: Consider $f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x))$ where $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$. Then f is convex when
  - h is convex, non-decreasing for each argument and g is convex
  - h is convex, non-increasing for each argument and g is concave

  and f is concave when
  - h is concave, non-decreasing for each argument and g is convex
  - h is concave, non-increasing for each argument and g is concave

Exercise: Check that the log-sum-exponential/softmax function is convex.

4.5 Examples: Strict Convexity

$\nabla^2 f(x) \succ 0 \implies f$ is strictly convex. However, the reverse isn't guaranteed to be true, strict convexity does not imply $\nabla^2 f(x) \succ 0$.

Example: $f(x) = e^{-x}$ then $f''(x) > 0\ \exists\ m$ s.t. $f''(x) \geq m$.

If $f(x)$ is strictly convex, it is not guaranteed to be strongly convex.

Example: $f(x) = x^4$ is strictly convex but its second derivative, $f''(0) = 0$ is not strictly positive (necessary for strong convexity).

4.6 First order optimality condition

For a convex problem

$$\min_x f(x) \text{ subject to } x \in C$$

and a differentiable f, a feasible point is optimal iff

$$\nabla f(x)^T (y - x) \geq 0$$

In other words, all feasible directions from x are aligned with gradient. When the optimization is unconstrained, this reduces to $\nabla f(x) = 0$

While this is a very general condition, it is difficult to check for most problems and hence is not very useful.

4.6.1 Examples

4.6.1.1 Quadratic minimization

Consider minimizing $f(x) = \frac{1}{2}x^T Q x + b^T x + c$ where $Q \succeq 0$. Optimality condition is

$$\nabla f(x) = Qx + b = 0$$

- If $Q \succ 0, x = -Q^{-1}b$ is a unique solution
- If Q is singular and $b \notin \text{col}(Q)$, there are no solutions
- If Q is singular and $b \in \text{col}(Q)$, there are infinite solutions $x = Q^+ b + z$, where Q^+ is the pseudoinverse and $z \in \text{null}(Q)$

4.6.1.2 Projection on a convex set

Consider projection on convex set C

$$\min_x \|a - x\|_2^2 \text{ subject to } x \in C$$

According to first order optimality condition,

$$\nabla f(x)^T(y-x) = (a-x)^T(y-x) \geq 0 \ \forall \ y \in C$$

This says that $a - x \in \mathcal{N}_C(x)$ where $\mathcal{N}$ is the normal cone.

4.7 Relaxing non-affine equalities

Given an optimization problem

$$min_x f(x) \text{ subject to } x \in C$$

, when can take a larger set $\widetilde{C} \supseteq C$ and consider $x \in \widetilde{C}$.

This is called a relaxation and its optimal value is always smaller or equal to that of the original problem.

4.7.1 Special case: relaxing non-affine equality constraints

Constraints of the form $h_j(x) = 0, j = 1, \ldots, r$ where h_j are convex but not affine can be replaced with $h_j(x) \leq 0, j = 1, \ldots, r$

4.7.1.1 Example: maximum utility problem

Consider the maximum utility problem

$$\max_{x,b} \sum_{t=1}^{T} \alpha_t u(x_t)$$
$$\text{subject to } b_{t+1} = b_t + f(b_t) - x_t, t = 1, \ldots, T$$
$$0 \leq x_t \leq b_t, t = 1, \ldots, T$$

Here b_t is the budget and x_t is the amount consumed at time t; f is an investment return function, u utility function, both concave and increasing. This is not a convex optimization problem unless we relax the constraint to $b_{t+1} \leq b_t + f(b_t) - x_t, t = 1, \ldots, T$

4.7.2 Example: PCA

Given $X \in \mathbb{R}^{n \times p}$, consider the low rank approximation problem

$$\min_R \|X - R\|_F^2 \text{ subject to } \text{rank}(R) = k$$

Let $R = XZ$, subject to $\text{rank}(Z) = k$ where Z is a projection matrix. Then the problem becomes

$$\min_{Z \in \mathbb{S}^p} \|X - XZ\|_F^2 \iff \max_{Z \in \mathbb{S}^p} tr(SZ)$$

where $S = X^T X$.

The constraint set is non-convex. $C = \{Z \in \mathbb{S}^p : \lambda_i(Z) \in \{0,1\}, i = 1,\ldots,p, tr(Z) = k\}$ where λ_i are the eigenvalues of Z.

We can relax this constraint set to

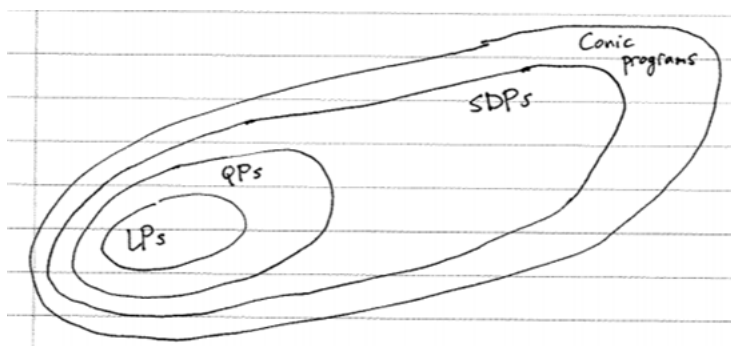
$$\begin{aligned}\mathcal{F}_k &= \{Z \in \mathbb{S}^p : \lambda_i(Z) \in [0,1], i = 1,\ldots,p, tr(Z) = k\}\\ &= \{Z \in \mathbb{S}^p : 0 \preceq Z \preceq 1, tr(Z) = k\}\end{aligned}$$

Here $\mathcal{F}_k$ is called a fantope which is analogous to polytopes but for matrices.

4.8 Canonical Problems

The 4 canonical problems described in these notes relate to each other according to the high level picture shown below (from lecture slides).

1. Linear Programs (LPs)
2. Quadratic Programs (QPs)
3. Semidefinite Programs (SDPs)
4. Conic Programs (CPs)



CVX solves conic programs generically using interior point methods. However, depending upon the problem at hand, other methods can be faster for a given problem.

Convex problems are a small part of the broader soup of optimization problems which include non-convex problems. However, non-convex is not a useful characterization for optimization problems because of how overbroad the term is. For example, neural networks are continuous non-convex problems, while integer programming is a combinatorial non-convex problem, both are very different from each other.

4.9 Linear Programs

Fundamental problem in convex optimization, which can be solved using the simplex algorithm (non-iteratively) and interior point methods (iteratively). Any LP described in the Basic Form below can be re-written in the Standard Form.

Basic Form

$$\min_x c^T x, \quad s.t \;\; Dx \le d \text{ and } Ax = b$$

Standard Form

$$\min_x c^T x, \quad s.t \;\; Ax = b \text{ and } x \ge 0$$

4.9.1 Example Problems

4.9.1.1 Diet: Find cheapest combination of foods that satisfies nutritional requirements

$$\min_x c^T x, \quad s.t \;\; Dx \geq d \text{ and } x \geq 0$$

Interpretation of variables

- c_j : per-unit cost of food j
- d_i : minimum required intake of nutrient i
- D_{ij} : content of nutrient i per unit of food j
- x_j : units of food j in the diet

4.9.1.2 Transportation: Ship commodities from given sources to destinations at min cost

$$\min_x \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \quad s.t. \quad \sum_{j=1}^{n} x_{ij} \leq s_i \ , \quad i = 1, \ldots, m \quad \& \quad \sum_{i=1}^{m} x_{ij} \geq \ d_j \ , \quad j = 1, \ldots, n \ \ \& \ \ x \geq 0$$

Interpretation of variables

- s_i : supply at source i
- d_j : demand at destination j
- c_{ij} : per-unit shipping cost from i to j
- x_{ij} : units shipped from i to j

4.9.1.3 Basis Pursuit

Given $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$, where $p > n$. Suppose that we seek the sparsest solution to under-determined linear system $X\beta = y$.

We can draft a non-convex formulation (recall $||\beta||_0 = \sum_{j=1}^{p} \mathbb{I}_{\beta_j \neq 0}$ is the L0 norm)

$$\min_\beta ||\beta||_0 \quad s.t \;\; X\beta = y$$

The L1 approximation of the non-convex problem above is referred to as **Basis Pursuit**

$$\min_\beta ||\beta||_1 \quad s.t \;\; X\beta = y$$

and can be cast as a linear program as shown below

$$\min_{\beta, z} 1^T z \quad s.t \;\; X\beta = y \;\; \& \;\; z \geq \beta \;\; \& \;\; z \geq -\beta$$

4.9.1.4 Dantzig Selector

The **Dantzig Selector** is a modification of the **Basis Pursuit** problem, where we allow $X\beta \approx y$ instead of seeking an exact solution. This can be formulated as

$$\min_{\beta} ||\beta||_1 \quad s.t \quad ||X^T(y - X\beta)||_\infty \leq \lambda$$

with $\lambda \geq 0$ as a tuning parameter.