Lecture 4: September 10

Lecturer: Ryan Tibshirani *Scribes: Priya Donti, Hima Tammineedi*

4.1 Recap

Important points we have covered till date include:

(1) Convex problems: (1) general structure, (2) local minima are global, and (3) convex combinations of minima are minima.

(2) First order optimality: For a convex problem $\min_x f(x)$ s.t. $x \in C$ and differentiable f, a feasible point x is optimal if and only if $\nabla f(x^\star)(y - x^\star) \geq 0 \ \ \forall y \in C$.

(3) Rewriting convex problems: (1) eliminating constraints, (2) partial optimization.

4.2 Relaxations

A relaxation involves turning a given problem into one with looser constraints ("enlarging the feasible set") such that the new problem is ideally easier to solve.

$$\min_{x \in C} f(x) \Rightarrow_{\text{relax}} \min_{x \in \tilde{C}} f(x), \text{ where } \tilde{C} \supseteq C.$$

The optimal value of a relaxation is always less than or equal to the optimal value of the original problem. We say that a relaxation is *tight* if the solution to the relaxed problem is still a feasible point of the original problem.

4.2.1 Relaxing non-affine equality constraints

For functions $g_i(x), \ \ i \in \{1, \ldots, d\}$ that are convex but not affine, we relax

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) = 0, \quad i \in \{1, \ldots, d\} \\ & Ax = b \\ & h_i(x) \leq 0 \end{array} \Longrightarrow_{\text{relax}} \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i \in \{1, \ldots, d\} \\ & Ax = b \\ & h_i(x) \leq 0 \end{array}$$

as the original formulation is a non-convex problem (non-affine equalities are not convex). The relaxed formulation turns affine equalities into affine *inequalities*, so thus has more feasible points and is now convex.

4.2.2 Examples

(1) Maximum utility problem:

$$\begin{aligned} \max_{x,b} \quad & \sum_{t=1}^{T} \alpha_t u(x_t) \\ \text{s.t.} \quad & b_{t+1} = b_t + f(b_t) - x_t, \ t = 1, \dots, T \\ & 0 \le x_t \le b_t, \ t = 1, \dots, T \end{aligned} \Longrightarrow_{\text{relax}} \begin{aligned} \max_{x,b} \quad & \sum_{t=1}^{T} \alpha_t u(x_t) \\ \text{s.t.} \quad & b_{t+1} \le b_t + f(b_t) - x_t, \ t = 1, \dots, T \\ & 0 \le x_t \le b_t, \ t = 1, \dots, T, \end{aligned}$$

where b_t is the budget at time t, x_t is the amount consumed at time t, α_t is the weighting of the importance of each timestep, f is the investment return function, u is the utility function (f, u are both concave and increasing).

The original formulation is not convex due to the equality constraints. The relaxation is convex and actually tight. The intuition is that you have at most the original amount of money, but you can throw some money away if you want; this is tight, since at optimum, you wouldn't throw away any money.

(2) **PCA:** Given matrix $X \in \mathbb{R}^{n \times d}$, we want to find a low-rank approximation via

$$\begin{aligned} \min_{R} \quad & ||X - R||_F^2 \\ \text{s.t.} \quad & \text{rank}(R) = k, \end{aligned}$$

where $||A||_F$ is the Frobenius norm. This is non-convex because rank is not convex.

We can compute the optimal solution by using a SVD truncated to the first k columns/elements: $R^* = U_k D_k V_k^T$

$\Rightarrow_{\text{relaxation}}$

$$\begin{aligned} \min_{Z} \quad & ||X - XZ||_F^2 \\ \text{s.t.} \quad & \text{rank}(Z) = k \end{aligned}$$

where $Z \in \mathbb{S}^d$ is a projection matrix (i.e. $Z = V_k V_k^T$).
However, this is still non-convex.

$\Rightarrow_{\text{relaxation}}$

$$\begin{aligned} \min_{Z} \quad & tr((X - XZ)^T (X - XZ)) \\ \text{s.t.} \quad & \text{rank}(Z) = k \end{aligned}$$

which is equivalent to

$$\begin{aligned} \max_{Z} \quad & tr(X^T X Z) \\ \text{s.t.} \quad & \text{rank}(Z) = k \end{aligned}$$

using the fact that a projection matrix is idempotent: $Z = ZZ^T$.

Since a projection matrix has eigenvalues either 0 or 1, we can represent the constraint set in the following form:

$$\begin{aligned} \max_{Z} \quad & tr(X^T XZ) \\ \text{s.t.} \quad & \{Z : tr(Z) = k,\ Z \in \mathbb{S},\ \lambda_i(Z) \in \{0, 1\}\} \end{aligned}$$

Where $\lambda_i(Z)$ are the eigenvalues of Z.
However, $\lambda_i(Z) \in \{0, 1\}$ is non-convex.

$\Rightarrow_{\text{relaxation}}$

$$\begin{aligned} \max_{Z} \quad & tr(X^T XZ) \\ \text{s.t.} \quad & \{Z : tr(Z) = k,\ Z \in \mathbb{S},\ \lambda_i(Z) \in [0, 1]\} \end{aligned}$$

This relaxation can also be seen as the convex hull of the previous constraint set.
And this is simply the Fantope of order k.

$\mathcal{F}_k = \{Z : tr(Z) = k,\ Z \in \mathbb{S},\ \lambda_i(Z) \in [0, 1]\} = \{Z : tr(Z) = k,\ Z \in \mathbb{S},\ 0 \preceq Z \preceq I\}$ And the Fantope is convex.

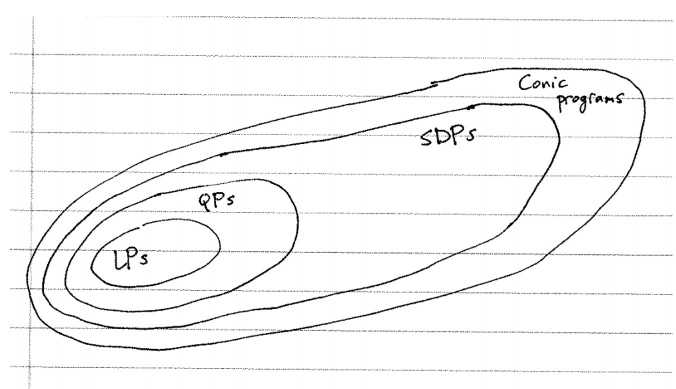
So, the following formulation is convex

$$\max_{Z \in \mathcal{F}_k} \quad tr(X^T XZ)$$

This relaxation is also tight, since $Z = V_k V_k^T$ if V_k was unique from the SVD.

4.3 Canonical Problem Types

The following graphic illustrates the relationship between the different types of convex problems. Note that conic programs are a subset of convex programs, which are a subset of non-convex programs. (Note: Second-order cone programs are in between QPs and SDPs.)



4.3.1 Linear Programs (LPs)

General form: **Standard form:**

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & Dx \leq d \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

The LP standard form eliminates all inequalities from the general form except for element-wise simple inequalities, and is equivalent to the general form. In particular, one can translate from the general form to the standard form by adding slack variables s so that the constraints $Dx \leq d$ become $Dx + s = d,\ s \geq 0$.

Each constraint defines the feasible region (which is a polytope). We note that the optimum of an LP is always at a corner of the feasible polytope.

Examples:

(1) **Diet problem.** Find the cheapest combination of foods that satisfy some nutritional requirements via $\min_x\ c^T x$ s.t. $Dx \geq d,\ x \geq 0$, where x_j is the units of food j in the diet, c_j is the per-unit cost of food j, d_i is the minimum required intake of nutrient i, and D_{ij} is the content of nutrient i per unit of food j.

(2) **Basis pursuit.** Given $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$, where $p > n$, find the sparsest solution to the underdetermined linear system $X\beta = y$. The original formulation is nonconvex, but we can form an ℓ_1 approximation ("basis pursuit") that is an LP.

Orig formulation: *Basis pursuit:* *LP formulation:*

$$\begin{aligned} \min_\beta \quad & \|\beta\|_0 \\ \text{s.t.} \quad & X\beta = y \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \min_\beta \quad & \|\beta\|_1 \\ \text{s.t.} \quad & X\beta = y \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \min_{\beta,z} \quad & 1^T z \\ \text{s.t.} \quad & z \geq \beta \\ & z \geq -\beta \\ & X\beta = y. \end{aligned}$$

The LP formulation is equivalent to basis pursuit since $z = |\beta|$ is the minimum z that satisfies the $z \geq \beta$, $z \geq -\beta$ constraints, and we have a minimization objective.

(3) **Danzig selector.** Modification of previous problem, where we allow $X\beta \approx y$: $\min_\beta \|\beta\|_1$ s.t. $\|X^T(y - X\beta)\|_\infty \leq \lambda$, where $\lambda \geq 0$ is a tuning parameter. This can be reformulated as an LP.

4.3.2 Convex quadratic programs (QPs)

General form: **Standard form:**

$$\begin{aligned} \min_x \quad & c^T x + \frac{1}{2} x^T Q x \\ \text{s.t.} \quad & Ax = b \\ & Dx \leq d \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \min_x \quad & c^T x + \frac{1}{2} x^T Q x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

The above problem is only convex when $Q \succeq 0$. (When we say "quadratic program," we implicitly assume this condition and thus that the quadratic program is convex.) As with linear programs, the general and standard forms are equivalent and can be translated between via slack variables.

Examples:

(1) **Markowitz portfolio optimization.** Construct a financial portfolio, trading off performance and risk, via $\min_x \ \mu^T x - \frac{\gamma}{2} x^T Q x$ s.t. $1^T x = 1, \ x \geq 0$, where x_j is the percentage of holdings that asset j represents, μ is the assets' expected returns, Q is the covariance matrix of the assets' returns, and γ is the decisionmaker's risk aversion.

(2) **Support vector machines.** Given $y \in \{-1, 1\}^n$ and $X \in \mathbb{R}^{n\times p}$ with rows $x_1, \ldots, x_n$, the SVM problem is quadratic. This can be readily seen from the hinge form of SVM, which is equivalent to the original form.

Original form:

$$\begin{aligned} \min_{\beta,\beta_0,\xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0, \ i = 1, \ldots, n \\ & y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, \ i = 1, \ldots, n. \end{aligned}$$

$$\Longrightarrow$$

Hinge form (QP):

$$\min_{\beta,\beta_0} \quad \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n [1 - y_i(x_i^T\beta + \beta_0)]_+$$

(3) **Lasso.** Given $y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}$, the lasso problem $\min_\beta \ \|y - X\beta\|_2^2$ s.t. $\|\beta\|_1 \leq s$, where $s \geq 0$ is a tuning parameter, can be reformulated as a QP. Its alternative parameterization (Lagrange, or penalized form) $\min_\beta \ \|y - X\beta\|_2^2 + \lambda\|\beta\|_1$, where $\lambda \geq 0$ is a tuning parameter, can also be rewritten as a QP.

4.3.3 Semidefinite programs (SDPs)

SDPs are basically a generalization of linear programs, where now the decision variable X is a matrix and we generalize the $\leq$ operator to a different (partial) order. This is another convex class of problems.

To write the SDP, we first define, given $X, Y \in \mathbb{S}^n$:

- An inner product $X \bullet Y \equiv \mathrm{tr}(X^T Y)$.
- A partial ordering $X \succeq Y \iff X - Y \in \mathbb{S}^n_+$, where $\mathbb{S}^n_+$ is the set of positive semidefinite matrices. (As a special case, for $x, y \in \mathbb{R}^n$, $\mathrm{diag}(x) \succeq \mathrm{diag}(y) \iff x \geq y$.)

For $F_j \in \mathbb{S}^d$ for $j = 0, 1, \ldots, n$; $A \in \mathbb{R}^{m\times n}$; $c \in \mathbb{R}^n$; and $b \in \mathbb{R}^m$, SDPs then have the following form.

General form:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x_1F_1 + \ldots + x_nF_n \preceq F_0 \end{aligned}$$

$$\iff$$

Standard form:

$$\begin{aligned} \min_X \quad & C \bullet X \\ \text{s.t.} \quad & A_i \bullet X = b_i, \ i = 1, \ldots, m \\ & X \succeq 0. \end{aligned}$$

Examples:

(1) **Trace norm minimization/matrix completion.** Let $A : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map

$$A(X) = \begin{pmatrix} A_1 \bullet X \\ \cdots \\ A_p \bullet X \end{pmatrix}$$

for $A_1, \ldots, A_p \in \mathbb{R}^{m\times n}$. Finding the lowest rank-solution to an underdetermined system is nonconvex, but it can be shown that its trace norm approximation is an SDP (proof uses duality).

Original form:

$$\begin{aligned} \min_X \quad & \mathrm{rank}(X) \\ \text{s.t.} \quad & A(X) = b. \end{aligned}$$

$$\Longrightarrow$$

Trace norm approximation:

$$\begin{aligned} \min_X \quad & \|X\|_{\mathrm{tr}} \\ \text{s.t.} \quad & A(X) = b. \end{aligned}$$

The intuition behind this relaxation is that $\text{rank}(X) = \sum_i \mathbf{1}\{\sigma_i(x) \neq 0\}$, i.e. the ℓ_0-norm of the singular values, so as we saw with the basis pursuit problem, we use an ℓ_1 norm relaxation (where $\|X\|_{\text{tr}} = \sum_i \sigma_i(x)$).

4.3.4 Conic programs

Conic programs are of the form

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

where $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$; $D : \mathbb{R}^n \to Y$ is a linear map and $d \in Y$ for a Euclidean space Y; and $K \subseteq Y$ is a closed convex cone.

LPs and SDPs are special cases of conic programming. When $K = \mathbb{R}^n_+$, this is equivalent to an LP. For $K = \mathbb{S}^n_+$, this is equivalent to an SDP. We will talk about conic programs in more detail next class.