10-725/36-725: Convex Optimization **Fall 2015**

Lecture 3: September 8

Lecturer: Ryan Tibshirani *Scribes: William Herlands,Maria De Arteaga, Luong Nguyen*

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

3.1 Optimization terminology

To standardize language we present a general convex optimization problem below where the objective function, f, and inequality constraint functions, g_i, are all convex. Equivilently we could study the maximization of concave functions. Note that although we do not often discuss it, the domain of this problem is $D = dom(f) \cap_{i=1}^{m} dom(g_i)$. **Example:** General form of optimization problem

$$\begin{array}{ll} \min\limits_{x \in D} & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, ..., m \\ & Ax = b \end{array} \tag{3.1}$$

Any $x \in D$ that satisfies all the constraints is feasible point. The optimal point, $x^\star$, minimizes $f(x)$ over all feasible points, yielding the optimal value, $f(x^\star) = f^\star$. A feasible point is ϵ-suboptimal when $f(x) \leq f^\star + \epsilon$. Finally, a constraint g_i is active at feasible point x when $g_i(x) = 0$. FOr example, we often discuss constraints which are active at the solution, $x^\star$.

An important nuance is that convex optimization problems need not have solutions. For example, the minimization of a linear function in R^n has no solution, but is still considered convex optimization.

3.2 Convex solution sets

In addition to minimizing an objective, we are also interested in the solution set, X_{opt} of a convex function. Instead of finding the $min_{x \in D}$ from Example , here we are interested in $argmin_{x \in D}$. We can see that X_{opt} is a convex set because given any two solutions to Example x, y, and $0 \leq \theta \leq 1$,

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta) f(y) \tag{3.2}$$

since we know that x and y are solutions,

$$\theta f(x) + (1 - \theta) f(y) = \theta f^\star + (1 - \theta) f^\star = f^\star \tag{3.3}$$

and $\theta x + (1 - \theta)y$ is also a solution!

As we mentioned in the previous section, just because X_{opt} is a convex set, does not mean that convex optimization problems have solutions. In fact a convex optimization problem may have 0, 1 or uncountably infinite solutions. X_{opt} is an empty set when no solutions are obtained (e.g. in a minimization of a linear function). Exactly 1 solution is obtained when the criterion f is strictly convex (e.g. when $f(x) = x^2$). In all other cases X_{opt} is a set of uncountably infinite solutions!

3.2.1 Lasso example

To provide some intuition for these definitions we consider the lasso, or L_1 penalized regression problem. Given $y \in R^n$ and $X \in R^{nxp}$ the lasso is, .

$$\begin{array}{rl} \min\limits_{\beta \in R^p} & \|y - X\beta\|_2^2 \\ \text{s.t.} & \|\beta\|_1 \leq s \end{array} \tag{3.4}$$

Since the objective is a least squares regression (convex) and the constraint is a shifted norm (convex) we see that this is a convex optimization problem. The feasible set are the values of $\beta \in R^p$ within the norm ball of L_1 distance s.

To understand the solution set, β_{opt}, we consider two different regimes for the input. When we have more samples than features, $n \geq p$, the input X has full rank. In this case $\bigtriangledown^2 f(\beta) = 2X^T X \succ 0$ implying that the problem is strictly convex with exactly one solution. However, in the high dimensional case when $n < p$ we cannot guarantee a unique solution. Later in the course we will revisit the lasso problem to understand under what guarantees we do have in the latter case.

3.2.2 SVM example

In the case of support vector machines, we have that the optimization problem is

$$\begin{array}{rlll} \min\limits_{\beta,\beta_0,\xi} & \frac{1}{2}\|\beta\|_2^2 + C\sum\limits_{i=1}^{n}\xi_i & & \\ \text{s.t.} & \xi_i & \geq & 0 \\ & y_i(x_i^T\beta + \beta_0) & \geq & 1 - \xi_i \quad \forall i = 1 \dots n \end{array} \tag{3.5}$$

In this case, the **criterion** is:

$$f(x) = \min_{\beta,\beta_0,\xi} \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}\xi_i$$

and the **constraints** are:

$$\begin{array}{rl} \xi_i \geq 0 & \forall i = 1 \dots n \\ y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i & \forall i = 1 \dots n \end{array}$$

Even though the objective function is convex, it is not strictly convex, since the term $C\sum_{i=1}^{n}\xi_i$ is linear on the slack variables ξ_i. A solution (β, β_0, ξ) is not necessarily unique, as there might be solutions in different directions. Changing the criterion to be

$$f(x) = \min_{\beta,\beta_0,\xi} \frac{1}{2}\|\beta\|_2^2 + \frac{1}{2}\|\beta_0\|_2^2 + C\sum_{i=1}^{n}\xi_i$$

would make solutions unique. However, this does not make sense, as it would imply we have a penalty for the separating plane going far from the origin, which is not something we are interested on doing.

3.2.3 Local minima and global minima

A point is called **locally optimal** if there exists some $R > 0$ such that

$$f(x) \leq f(y) \text{ for all feasible } y \text{ such that } ||x - y||_2 \leq R$$

One of the main results we have seen so far is that in the case of convex problems **local optima are global optima**. However, it is important to remember this results does not imply that the global optima is unique.

3.3 Properties and first-order optimality

3.3.1 Rewriting constraints

So far, we have written optimization problems in the following way:

$$\begin{array}{rrcl} \min & f(x) & & \\ \text{s.t.} & g_i(x) & \leq & 0 \quad \forall i = 1 \dots n \\ & Ax & = & 0 \end{array} \tag{3.6}$$

There are two main ways in whichh we can rewrite this.

1. **Completely general form**

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

   where we define $C = \{x : g_i(x) \leq 0, i = 1, \dots m, Ax = b\}$

2. **Unconstrained form**

   $$\min f(x) + I_C(x)$$

   where $I_C(x)$ is the indicator function.

3.3.2 First-order optimality condition

The first-order optimality condition tells us that if f is differentiable, for the convex problem

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

a feasible point is optimal if and only if the following holds:

$$\nabla f(x)^T (y - x) \geq 0 \text{ for all } y \in C$$

An important special case of this property is when $C = \mathcal{R}^n$, meaning the optimization problem is unconstrained. Since y can take any value in the $\mathbb{R}^n$ space, the above statement is equivalent to saying that:

$$\nabla f(x) = 0$$

3.3.2.1 Example: quadratic minimization

Let $f(x)$ be the quadratic function

$$f(x) = \frac{1}{2}x^T Qx + b^T x + c \text{ where } Q \succeq 0$$

Since f is differentiable, we can appply the first order condition, which indicates a point will be optimal if and only if the following holds:

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

We can solve such equality by considering three cases:

- $Q \succ 0$: In this case Q is invertible and there will be a unique solution $x = -Q^{-1}b$
- Q is singular and $b \notin col(Q)$: No solution, meaning $\min_x f(x) = -\infty$
- Q is singular and $b \in col(Q)$: Infinitely many solutions of the form $x = Q^+ b + z$ for $zinnull(Q)$.
  Note: Q^+ is the pseudoinverse of Q. This extends the notion of inverse when the matrix is not invertible. In cases where the matrix invertible the pseudoinverse is the same inverse.

3.3.2.2 Example: equality-constrained optimization

Consider the equality-constrained convex problem for a differentiable f:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax = 0 \end{array} \tag{3.7}$$

The **Lagrange multiplier optimality condition** states:

$$\nabla f(x) + A^T u = 0 \text{ for some } u$$

Here we present the proof:

According to the first-order optimality condition

$$\nabla f(x)(y - x) \geq 0 \qquad \forall y : y = Ay = b$$

Combining this with the equality constraint of the problem itself we obtain:

$$\nabla f(x)(y - x) \geq 0 \quad \forall y : A(y - x) = b$$

Setting $v = y - x$,

$$\begin{array}{rccccc} & \nabla f(x)v & \geq & 0 & \forall v \in null(A) \\ \Rightarrow & \nabla f(x)v & = & 0 & \forall v \in null(A) \\ \Rightarrow & \nabla f(x) & \in & row(A) & (\text{since } null(A)^\perp = row(A)) \\ \Rightarrow & \nabla f(x) & = & A^t u & \text{for some } u \end{array}$$

3.4 Equivalent transformation

3.4.1 Partial optimization

Reminder: **If** f is convex in (x, y) and C is convex **then** $g(x) = \min_{y \in C} f(x, y)$ is convex in x.
Because of this property, we can partially optimize a convex problem and retain convexity.
Example: Let $x = x1, x2 \in \mathbb{R}^{n_1+n_2}$, then the following two convex optimization problems are equivalent:

1. $\min_{x_1,x_2} f(x_1, x_2)$ s.t. $g(x_1) \leq 0$ and $g(x_2) \leq 0$

2. $\min_{x_1} \tilde{f}(x_1)$ s.t. $g_1(x_1) \leq 0$

where $\tilde{f}(x_1) = \min\{f(x_1, x_2) : g_2(x_2) \leq 0\}$. If the first problem is convex then so is the second problem.

Example: Hinge form of SVMs

$$\begin{array}{rll} \min\limits_{\beta,\beta_0,\xi} & \frac{1}{2}\|\beta\|_2^2 + C\sum\limits_{i=1}^{n}\xi_i & \\ \text{s.t.} & \xi_i \geq 0 & \\ & y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i & \forall i = 1 \dots n \end{array} \tag{3.8}$$

The constraint can be rewritten as $\xi_i \geq \max 0, 1 - y_i(x_i^T\beta + \beta_0)$ and we achieve equality at the optimal ξ. Therefore, the hinge form of SVMs for optimal ξ is:

$$\min_{\beta,\beta_0} \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}[1 - y_i(x_i^T\beta + \beta_0)]_+ \tag{3.9}$$

where $a_+ = \max a, 0$ is called the hinge function.

3.4.2 Transformation and change of variables

Lemma 3.1 *If h: $\mathbb{R} \to \mathbb{R}$ is a monotone increasing transformation, then:*

$$\begin{array}{rl} & \min\limits_{x} \quad f(x) \text{ subject to } x \in C \\ \iff & \min\limits_{x} \quad h(f(x)) \text{ subject to } x \in C \end{array} \tag{3.10}$$

Inequality and equality constraints can be transformed and yield equivalent optimization problems.

Lemma 3.2 *If $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-on, and its image covers feasible set C, then we can change variables in an optimization problem:*

$$\begin{array}{rl} & \min\limits_{x} \quad f(x) \text{ subject to } x \in C \\ \iff & \min\limits_{y} \quad f(\phi(y)) \text{ subject to } \phi(y) \in C \end{array} \tag{3.11}$$

3.4.3 Eliminiating equality constraints

Equality constraints could be eliminated by incorporating them into inequality constraints. Given the problem:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, i = 1 \dots m \\ & Ax = b \end{aligned} \tag{3.12}$$

Any feasible point can be expressed as $x = My + x_0$, where $Ax_0 = b$ and $\text{col}(M) = \text{null}(A)$. Convex optimization problem 3.12 is equivalent to:

$$\begin{aligned} \min_{x} \quad & f(My + x_0) \\ \text{subject to} \quad & g_i(My + x_0) \leq 0, i = 1 \dots m \end{aligned} \tag{3.13}$$

Note that eliminating equality constraints is fully general but not always a good idea. There are two main reasons:

- Computing M might be expensive
- If A is a sparse matrix, there are neat tricks to solve problem 3.12. On the other hand, M will be dense and make it difficult to solve problem 3.13.

3.4.4 Introducing slack variables

Slack variables can be introduced to the optimization problem by decomposing the inequality constraints into affine equalities and slack variables. By introducing slack variables, the standard convex optimization problem 3.12 is transformed into:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & s_i \geq 0, i = 1 \dots m \\ & g_i(x) + s_i = 0, i = 1 \dots m \\ & Ax = b \end{aligned} \tag{3.14}$$

Note that problem 3.14 is no longer convex unless each $g_i(x)$ is an affine function. In that case, problem 3.14 becomes linear programming, which can be solved using simplex algorithms.

References

[BL] S. Boyd and L. Vandenberghe, "Convex Optimization," Chapter 4

[G] Guler (2010), "Foundations of Optimization," Chapter 4