IE 598: Big Data Optimization **Fall 2016**

Lecture 5: Convex Optimization – September 6

Lecturer: Niao He *Scriber: Heran Zhang*

Overview In the introductory lecture, we discussed why convex optimization is a particularly interesting family of optimization problems to consider both from theoretical and practical viewpoints. In this lecture, we formally discuss what is a convex optimization problem and special properties that make it differ from general optimization problem. Moreover, we will study the optimality conditions for both unconstrained and constrained problems.

5.1 Basics of Convex Optimization

Recall that a canonical form of an optimization problem is

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, k \\ & x \in X \end{aligned} \tag{P}$$

Definition 5.1 (Convex Program) *A program (P) is* <u>*convex*</u> *if:*

(i) *f is convex;*

(ii) *$g_i, i = 1, ..., m$ are convex.*

(iii) *there is either no equality constraints or only linear equality constraints*

(iv) *$X \subseteq dom(f) \bigcap \left(\cap_{i=1,\ldots,m} dom(g_i)\right)$ is a convex set.*

For the sake of simplicity, in the following we denote a convex program as

$$\boxed{\begin{array}{c} \min \quad f(x), \\ \text{s.t.} \quad x \in S \\ S = \{x \in X : g_i(x) \leq 0, i = 1, \ldots, m\} \end{array}}$$

Example 1 (LASSO) Recall the regularized least square regression model:

$$\min_w ||Xw - y||_2^2 + \lambda ||w||_1$$

Example 2 (Max-margin classification) Recall the classification model

$$\max_{w,b} \frac{2c}{||w||_2}, \ s.t. \ y_i(w^T x_i + b) \geq c, i = 1, \ldots, n$$

This is equivalent as

$$\min_{w,b} ||w||_2, \ s.t. \ y_i(w^T x_i + b) \geq c, i = 1, \ldots, n$$

which is a convex program.

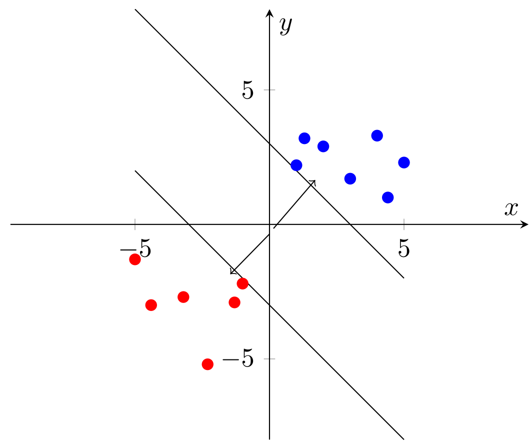

Figure 5.1: Max-margin classification

Definition 5.2 (Feasibility) *A solution x is feasible if x satisfies all constraints, i.e., $x \in S$. If $S = \emptyset$, we say program (P) is infeasible.*

Definition 5.3 (Optimality) *A solution x_* is optimal if x_* is feasible and $f(x_*) \leq f(x), \forall x \in S$. Usually, we denote as $x_* \in argmin_{x \in S} f(x)$.*

Definition 5.4 (Optimal value) *The optimal value is defined as $f_* = \inf_{x \in s} f(x)$*

- *Conventionally, if (p) is infeasible, we set $f_* = +\infty$.*
- *We say (P) is unbounded below if $f_* = -\infty$.*
- *We say (P) is solvable, if it has an optimal solution x_* and $f_* = f(x_*)$.*

Remark. The set of optimal solution, $\text{argmin}_{x \in S} f(x)$ is a convex set.

Remark. If f is strictly convex, then the optimal solution is unique.
Proof: Suppose there exist two optimal solutions x_1 and x_2, i.e. $f(x_1) = f(x_2) = f_*$. By strictly convexity, we have for $\lambda \in (0, 1)$,

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) = \lambda f_* + (1 - \lambda)f_* = f_*$$

Since $\lambda x_1 + (1 - \lambda)x_2 \in S$ is also feasible, this contradicts with the fact that f_* is the optimal value.

■

5.2 Local = Global

Definition 5.5 (Local minimum) *We say $x_* \in S$ is a local minimum if it has the smallest objective around its neighborhood, i.e. $\exists r > 0,$ s.t. $f(x_*) \leq f(x), \ \forall x \in B(x_*, r) \cap S$.*

Proposition 5.6 *A local minimum is a global minimum.*

Proof: Let x_* be a local minimum. For any $x \in S$, when ϵ is small enough, we have $y = x_* + \epsilon(x - x_*) \in B(x_*, r) \cap S$ and $0 \le \frac{f(y)-f(x_*)}{\|y-x_*\|} \le \frac{f(x)-f(x_*)}{||x-x_*||} \Rightarrow f(x) \ge f(x_*), \forall x \in S$. ■

Note that in general, for any optimization (possibly non-convex) problems, this is not true. A local minimum is not necessarily global optimum. This is a key property that makes convex optimization problem different from general non-convex problem.

5.3 Optimality Conditions for Simple Constrained Problems

We consider the simple constrained optimization problem (no inequality constraints)

$$\boxed{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X \end{array}}$$

5.3.1 Differentiable Case

Theorem 5.7 (Sufficient and Necessary Condition) *Assume f is convex and differentiable on $x_* \in X$,*

$$x_* \textit{ is an optimal solution } \iff (x - x_*)^T \nabla f(x_*) \ge 0, \forall x \in X.$$

Proof:

- ($\Leftarrow$) Since f is convex, we have $\forall x \in X,\ f(x) \ge f(x_*) + \nabla f(x_*)^T (x - x_*) \ge f(x_*)$.
- ($\Rightarrow$) We have $(x - x_*)^T \nabla f(x_*) = \lim_{\epsilon \to 0} \frac{f(x_* + \epsilon(x - x_*)) - f(x_*)}{\epsilon} \ge 0$.

■

Corollary 5.8 (Unconstrained Case) *If $x_* \in int(X)$, then x_* is optimal if and only if $\nabla f(x_*) = 0$. In particular, if $X = \mathbb{R}^n$, x_* is optimal if and only if $\nabla f(x_*) = 0$.*

Remark. Note that for unconstrained convex problems, $\nabla f(x_*) = 0$ is both sufficient and necessary for x_* to be optimal. However, for general (non-convex) optimization problems, $\nabla f(x_*)$ is only a necessary condition and is not sufficient to guarantee the global optimality. In fact, if $\nabla f(x_*) = 0$, we call x_* a stationary point of f, which could be a local minimum, a local maximum, or a saddle point.

Example 3 (Quadratic Problem) $f(x) = \frac{1}{2} x^T Q x + b^T x + c,\ Q \succeq 0$.
The optimal solution satisfies $\nabla f(x_*) = Qx_* + b = 0$.

$$\underset{x \in \mathbb{R}^n}{\text{argmin}} f(x) = \begin{cases} -Q^{-1}b & \text{Q is nonsingular} \\ \emptyset & \text{Q is singular, } b \notin col(Q) \\ x_* + null(Q) & \text{Q is singular, } b \in col(Q) \end{cases}$$

where x_* is such that $Qx_* + b = 0$ and $null(Q) = \{d : Qd = 0\}$.
Note that the set of optimal solution might be an empty set, or a unique solution, or a convex set.

5.3.2 Non-differentiable Case

Let us consider the situation when f is not necessarily differentiable.

Proposition 5.9 (Sufficient and Necessary Condition) *Assume f is convex on X, then*

(a) $x_* \in X$ is optimal if and only if $\exists g \in \partial f(x_*)$, s.t. $g^T(x - x_*) \geq 0, \forall x \in X$.

(b) Suppose $X = \mathbb{R}^n$, x_* is optimal if and only if $0 \in \partial f(x_*)$.

Proof:

(a) "If" part : by definition of subgradient, we have $\forall x \in X, f(x) \geq f(x_*) + g^T(x - x_*) \geq f(x_*)$.
"Only if" part: this uses the fact that

$$\lim_{\epsilon \to 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon} = \sup_{g \in \partial f(x)} g^T d.$$

We therefore have $\sup_{g \in \partial f(x_*)} g^T(x - x_*) \geq 0$. Since $\partial f(x_*)$ is closed, this implies that there exists $g \in \partial f(x_*)$ such that $g^T(x - x_*) \geq 0$.

(b) This is because

$$f(x) \geq f(x_*) \Longleftrightarrow f(x) \geq f(x_*) + 0^T(x - x_*) \Leftrightarrow 0 \in \partial f(x_*)$$

■

5.4 Optimality Condition for Constrained Problem

We now consider the constrained optimization problem

$$\boxed{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & x \in X \end{array}}$$

5.4.0 Motivation

A solution x_* is optimal if and if the following holds true:

1. The system $\{x \in X : f(x) \leq f(x_*), g_i(x) \leq 0, i = 1, ..., m\}$ has a solution
2. The system $\{x \in X : f(x) < f(x_*), g_i(x) \leq 0, i = 1, ..., m\}$ has no solution

One verifiable condition to show that the second system has no solution is when there exists nonnegative coefficients $\lambda_i \geq 0, i = 1, \ldots, m$, such that the system $\{x \in X : f(x) + \sum \lambda_i g_i(x) < f(x_*)\}$ has no solution. This is equivalent as $\inf_{x \in X} f(x) + \sum \lambda_i g_i(x) \leq f(x_*)$. In fact, under some condition, the opposite direction is also true. This is due to the useful convex theorem of alternative, which we present below without detailing the proof. The proof follows from the Separation theorem.

Definition 5.10 (Slater Condition) *Problem (P) is said to satisfy the Slater condition if there exists $x \in X$ such that $g_i(x) < 0,\ \forall i = 1, ..., m$.*

Theorem 5.11 (Convex Theorem of Alternatives) *Consider the two systems*

$$(I) : \{x \in X : f(x) < c, g_i(x) \leq 0, i = 1, ..., m\}$$

$$(II) : \{\lambda \geq 0 : \inf_{x \in X} f(x) + \sum \lambda_i g_i(x) \geq c\}$$

Then (I) is insolvable if and only if (II) is solvable.

5.4.1 Lagrangian Duality

We define

- $L(x, \lambda) = f(x) + \sum \lambda_i g_i(x)$ (Lagrangian function)
- $\underline{L}(\lambda) = \inf_{x \in X} L(x, \lambda)$ (Lagragian dual function)

Primal:

$$\begin{aligned} \mathrm{Opt}(P) = \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & x \in X \end{aligned}$$

Lagrangian dual:

$$\mathrm{Opt}(D) = \sup_{\lambda \geq 0} \underline{L}(\lambda)$$

$$\text{where } \underline{L}(\lambda) = \inf_{x \in X} L(x, \lambda)$$

Theorem 5.12 (Duality)

1. **(Weak duality)** Opt*(D)* $\leq$ Opt*(P)*
2. **(Strong duality)** *If (P) is solvable and satisfies slater condition, then (D) is solvable and* $\mathrm{Opt}(D) = \mathrm{Opt}(P)$.

Proof:

1. $\forall \lambda \geq 0$, we have

$$\underline{L}(\lambda) = \inf_{x \in X} L(x, \lambda) \leq \inf_{x \text{ feasible}} L(x, \lambda) \leq \inf_{x \text{ feasible}} f(x) = \mathrm{Opt}(P)$$

   Hence, $\mathrm{Opt}(D) = \sup_{\lambda \geq 0} \underline{L}(\lambda) \leq \mathrm{Opt}(P)$.

2. If x_* is optimal for (P), then the system $\{x \in X : f(x) < f(x_*), g_i(x) \leq 0, i = 1, ..., m\}$ has no solution. From the convex theorem of alternative, this implies that there exists $\lambda \geq$, such that $\underline{L}(\lambda) \geq \mathrm{Opt}(P)$. Hence, $\mathrm{Opt}(D) \geq \mathrm{Opt}(P)$.

∎