Topics in Mathematical Optimization

Tokyo Institute of Technology
Department Mathematical and Computing Sciences
Mituhiro Fukuda

Spring Semester of 2015

Outline of the Lecture

The main focus of this course is on algorithms to solve convex optimization problems which have recently gained some attention in continuous optimization. The course starts with basic theoretical results and then well-known algorithms will be analyzed and discussed.

Purpose of the Lecture

Algorithms to solve large-scale convex optimization problems have been recently an important topic in continuous optimization. This lecture intends to provide basic mathematical tools to understand these algorithms focusing on computational aspects when solving large-scale problems.

Plan of the Lecture (tentative)

1. Convex sets and related results
2. Properties of Lipschitz continuous differentiable functions
3. Optimality conditions for differentiable functions
4. Complexity analysis of algorithms for minimizing unconstrained functions
5. Properties of convex and differentiable convex functions
6. Worse cases for gradient based methods
7. Steepest descent methods for differentiable convex and differentiable strongly convex functions
8. Accelerated gradient methods

References

[Bertsekas] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, (Athena Scientific, Belmont, Massachusetts, 2003).

[Luenberger-Ye] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 3rd edition, (Springer, New York, 2008).

[Mangasarian] O. L. Mangasarian, *Nonlinear Programming*, (SIAM, Philadelphia, PA, 1994).

[Nesterov03] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, (Kluwer Academic Publishers, Boston, 2004).

[Nocedal] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd edition, (Springer, New York, 2006).

Related and/or Prerequisite Courses

It is necessary to have basic knowledge of linear algebra, calculus, topology, and computational complexity.

Evaluation

Final exam and/or reports.

1 Convex Sets

Example 1.1 Examples of convex sets.

Definition 1.2 We define as a **polyhedron** the set which can be represented as an intersection of *finitely* many closed half spaces of $\mathbb{R}^n$. Due to exercise 2, polyhedra are convex sets.

Definition 1.3 A point $\boldsymbol{y} \in \mathbb{R}^n$ is said to be a **convex combination** of $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$ if there exists non-negative $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$ such that $\sum_{i=1}^m \lambda_i = 1$ and $\boldsymbol{y} = \sum_{i=1}^m \lambda_i \boldsymbol{x}_i$.

Example 1.4 Given $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$, $m+1$ distinct point of $\mathbb{R}^n$ ($m \leq n$) such that $\boldsymbol{x}_1 - \boldsymbol{x}_0, \boldsymbol{x}_2 - \boldsymbol{x}_0, \dots, \boldsymbol{x}_m - \boldsymbol{x}_0$ are linear independent, the set formed by all convex combination of $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ is called an m**-simplex** in $\mathbb{R}^n$.

Theorem 1.5 A subset of $\mathbb{R}^n$ is convex if and only if it contains all the convex combinations of its elements.

Proof:
$\boxed{\Leftarrow}$ Trivial.
$\boxed{\Rightarrow}$ Let us show by induction on the number of elements m. For $m = 2$, it follows from the definition of convexity. Let us assume that the claim is valid for any convex combination of m or fewer elements. Consider $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{m+1}$ elements of the set and $\lambda_1, \lambda_2, \dots, \lambda_{m+1} \geq 0$ such that $\sum_{i=1}^{m+1} \lambda_i = 1$. If $\lambda_{m+1} = 0$ or $\lambda_{m+1} = 1$, it falls in the previous cases. Therefore, let $0 < \lambda_{m+1} < 1$. Then $\sum_{i=1}^{m+1} \lambda_i \boldsymbol{x}_i = \left(\sum_{j=1}^m \lambda_j\right) \frac{\sum_{i=1}^m \lambda_i \boldsymbol{x}_i}{\sum_{j=1}^m \lambda_j} + \lambda_{m+1}\boldsymbol{x}_{m+1} = (1 - \lambda_{m+1}) \sum_{i=1}^m \frac{\lambda_i}{\sum_{j=1}^m \lambda_j} \boldsymbol{x}_i + \lambda_{m+1}\boldsymbol{x}_{m+1}$ belongs to the set due to the induction hypothesis and definition of convexity. ∎

Definition 1.6 The intersection of all convex sets containing a given set $S \subseteq \mathbb{R}^n$ is called **convex hull** of S and is denoted by $\text{hull}(S)$. Therefore, $\text{hull}(S)$ is convex.

The following theorem shows that a $\text{hull}(S)$ can be constructed from the convex combination consisting only by its elements.

Theorem 1.7 The convex hull of $S \subseteq \mathbb{R}^n$, $\text{hull}(S)$, consists of all convex combinations of elements of S.

Proof:
Let B be the later set. If $\boldsymbol{y}_1, \boldsymbol{y}_2 \in B$, then $\exists \ell, m \in \mathbb{N}$, $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_\ell, \boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_m \in S$, and non-negative $\alpha_1, \alpha_2, \dots, \alpha_\ell, \beta_1, \beta_2, \dots, \beta_m \in \mathbb{R}$ such that $\boldsymbol{y}_1 = \sum_{i=1}^\ell \alpha_i \boldsymbol{a}_i$, $\boldsymbol{y}_2 = \sum_{j=1}^m \beta_j \boldsymbol{b}_j$, $\sum_{i=1}^\ell \alpha_i = 1$, and $\sum_{j=1}^m \beta_j = 1$. Then for $0 \leq \lambda \leq 1$, $\lambda \boldsymbol{y}_1 + (1-\lambda)\boldsymbol{y}_2 = \sum_{i=1}^\ell \lambda \alpha_i \boldsymbol{a}_i + \sum_{j=1}^m (1-\lambda)\beta_j \boldsymbol{b}_j$ with $\lambda\alpha_i, (1-\lambda)\beta_j \geq 0$, $\sum_{i=1}^\ell \lambda\alpha_i + \sum_{j=1}^m (1-\lambda)\beta_j = 1$. Therefore, B is convex. It is also clear that $S \subseteq B$, and therefore, $\text{hull}(S) \subseteq B$. From Theorem 1.5 the convex set $\text{hull}(S)$ must contain all convex combinations of elements of S. Hence $B \subseteq \text{hull}(S)$. ∎

Theorem 1.8 (Carathéodory's Theorem) Let $S \subseteq \mathbb{R}^n$. If $\boldsymbol{x}$ is a convex combination of elements of S, then $\boldsymbol{x}$ is a convex combination of $n+1$ or fewer elements of S.

Proof:
Let $\boldsymbol{x} = \sum_{i=1}^m \alpha_i \boldsymbol{x}_i$, $\boldsymbol{x}_i \in S$, $\alpha_i \geq 0$, $\sum_{i=1}^m \alpha_i = 1$. We will show that if $m > n+1$, then $\boldsymbol{x}$ can be written as a convex combination of $m-1$ elements of S. Therefore, suppose that all $0 < \alpha_i < 1$. Since $m - 1 > n$, $\exists \beta_1, \beta_2, \dots, \beta_{m-1} \in \mathbb{R}$ not all zeros such that

$$\beta_1(\boldsymbol{x}_1 - \boldsymbol{x}_m) + \beta_2(\boldsymbol{x}_2 - \boldsymbol{x}_m) + \cdots + \beta_{m-1}(\boldsymbol{x}_{m-1} - \boldsymbol{x}_m) = \mathbf{0}.$$

Define $\beta_m = -\sum_{i=1}^{m-1} \beta_i$. Then

$$\sum_{i=1}^{m} \beta_i = 0 \quad \text{and} \quad \sum_{i=1}^{m} \beta_i \boldsymbol{x}_i = \boldsymbol{0}.$$

Since $0 < \alpha_i < 1$, $\exists \gamma > 0$ such that $\delta_i := \alpha_i - \gamma\beta_i \geq 0$ $(i = 1, 2, \ldots, m)$ and only one δ_i, say $\delta_j = 0$. Then

$$\boldsymbol{x} = \sum_{i=1}^{m} \alpha_i \boldsymbol{x}_i = \sum_{i=1}^{m} \delta_i \boldsymbol{x}_i + \sum_{i=1}^{m} \gamma\beta_i \boldsymbol{x}_i = \sum_{i=1, i\neq j}^{m} \delta_i \boldsymbol{x}_i,$$

and $\delta_i \geq 0$ $(i = 1, 2, \ldots, m)$, $\sum_{i=1}^{m} \delta_i = \sum_{i=1}^{m} \alpha_i - \gamma \sum_{i=1}^{m} \beta_i = 1$.

We can do this procedure whenever $m > n + 1$. ∎

Proposition 1.9 If C_1 and C_2 are convex sets in $\mathbb{R}^n$, then so is their sum:

$$C_1 + C_2 := \{\boldsymbol{x}_1 + \boldsymbol{x}_2 \in \mathbb{R}^n \mid \boldsymbol{x}_1 \in C_1, \ \boldsymbol{x}_2 \in C_2\}.$$

Proposition 1.10 The product of a convex set in $\mathbb{R}^n$, C with a scalar $\alpha \in \mathbb{R}$:

$$\alpha C := \{\alpha \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \in C\}$$

is a convex set.

1.1 Exercises

1. Show that the set of $n \times n$ symmetric and positive definite matrices is a convex set.
2. Show that the intersection of an *arbitrary* collection of convex sets is a convex set.
3. Show that the closed ball centered at $\bar{\boldsymbol{x}} \in \mathbb{R}^n$, $\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \bar{\boldsymbol{x}}\| \leq \varepsilon\}$, with $\varepsilon > 0$ is a convex set.
4. Show that the interior of a convex set is a convex set.
5. Show that the closure of a convex set is a convex set.
6. Show that the convex hull of a set $S \subseteq \mathbb{R}^n$ is the unique and the smallest convex set containing S.
7. Prove Proposition 1.9.
8. Find an example where the sum of two closed sets is not a closed set.
9. Prove Proposition 1.10.

2 Separation Theorems for Convex Sets

The separation theorem for convex sets can be proved using the Farka's Lemma. Here, we follow [Bertsekas] and use a geometric fact of (orthogonal) projection to a convex set.

Proposition 2.1 Let $C \subseteq \mathbb{R}^n$ a convex set and $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ be a point that does not belong to the interior of C. Then there exists a vector $\boldsymbol{d} \neq \boldsymbol{0}$ such that

$$\boldsymbol{d}^T \boldsymbol{x} \geq \boldsymbol{d}^T \hat{\boldsymbol{x}}, \quad \forall \boldsymbol{x} \in C.$$

Proof:

Since $\hat{\boldsymbol{x}} \notin \text{int}(C)$, there is a sequence $\{\boldsymbol{x}_k\}$ which does not belong to the closure of C, $\bar{C}$, and converges to $\hat{\boldsymbol{x}}$. Now, denote by $p(\boldsymbol{x}_k)$ the orthogonal projection of $\boldsymbol{x}_k$ into $\bar{C}$ by a standard norm. One can see that by the convexity of $\bar{C}$ [Bertsekas]

$$(p(\boldsymbol{x}_k) - \boldsymbol{x}_k)^T(\boldsymbol{x} - p(\boldsymbol{x}_k)) \geq 0, \quad \forall \boldsymbol{x} \in \bar{C}.$$

Hence,

$$(p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T\boldsymbol{x} \geq (p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T p(\boldsymbol{x}_k) = (p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T(p(\boldsymbol{x}_k)-\boldsymbol{x}_k)+(p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T\boldsymbol{x}_k \geq (p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T\boldsymbol{x}_k.$$

Now, since $\boldsymbol{x}_k \notin \bar{C}$, calling $\boldsymbol{d}_k = \frac{p(\boldsymbol{x}_k)-\boldsymbol{x}_k}{\|p(\boldsymbol{x}_k)-\boldsymbol{x}_k\|}$,

$$\boldsymbol{d}_k^T\boldsymbol{x} \geq \boldsymbol{d}_k^T\boldsymbol{x}_k, \quad \forall \boldsymbol{x} \in \bar{C}.$$

Since $\|\boldsymbol{d}_k\| = 1$, it has a converging subsequence which will converge to let us say $\boldsymbol{d}$. Taking the same indices for this subsequence for $\boldsymbol{x}_k$, we have the desired result. ∎

Theorem 2.2 (Separation Theorem for Convex Sets) Let C_1 and C_2 nonempty non-intersecting convex subsets of $\mathbb{R}^n$. Then, $\exists \boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d} \neq 0$ such that

$$\sup_{\boldsymbol{x}_1 \in C_1} \boldsymbol{d}^T\boldsymbol{x}_1 \leq \inf_{\boldsymbol{x}_2 \in C_2} \boldsymbol{d}^T\boldsymbol{x}_2.$$

Proof:

Consider the set

$$C := \{\boldsymbol{x}_2 - \boldsymbol{x}_1 \in \mathbb{R}^n \mid \boldsymbol{x}_2 \in C_2, \quad \boldsymbol{x}_1 \in C_1\}$$

which is convex by Propositions 1.9 and 1.10.

Since C_1 and C_2 are disjoint, the origin $\boldsymbol{0}$ does not belong to the interior of C. From Proposition 2.1, there is $\boldsymbol{d} \neq \boldsymbol{0}$ such that $\boldsymbol{d}^T\boldsymbol{x} \geq \boldsymbol{0}$, $\forall \boldsymbol{x} \in C$. Therefore

$$\boldsymbol{d}^T\boldsymbol{x}_1 \leq \boldsymbol{d}^T\boldsymbol{x}_2, \quad \forall \boldsymbol{x}_1 \in C_1 \text{ and } \boldsymbol{x}_2 \in C_2.$$

Finally, since both C_1 and C_2 are nonempty, it follows the result. ∎

3 Lipschitz Continuous Differentiable Functions

Hereafter, we define for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, the standard inner product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle := \sum_{i=1}^n a_i b_i$, and the associated norm to it $\|\boldsymbol{a}\|_2 := \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}$.

Definition 3.1 Let Q be a subset of $\mathbb{R}^n$. We denote by $\mathcal{C}_L^{k,p}(Q)$ the class of functions with the following properties:

- Any $f \in \mathcal{C}_L^{k,p}(Q)$ is k times continuously differentiable on Q;
- Its pth derivative is Lipschitz continuous on Q with the constant $L \geq 0$:

$$\|f^{(p)}(\boldsymbol{x}) - f^{(p)}(\boldsymbol{y})\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in Q.$$

Observe that if $f_1 \in \mathcal{C}_L^{k,p}(Q)$, $f_2 \in \mathcal{C}_L^{k,p}(Q)$, and $\alpha, \beta \in \mathbb{R}$, then for $L_3 = |\alpha|L_1 + |\beta|L_2$ we have $\alpha f_1 + \beta f_2 \in \mathcal{C}_{L_3}^{k,p}(Q)$.

Lemma 3.2 Let $f \in \mathcal{C}^2(\mathbb{R}^n)$. Then $f \in \mathcal{C}_L^{2,1}(\mathbb{R}^n)$ if and only if $\|f''(\boldsymbol{x})\|_2 \leq L, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$.