Mathematical Optimization: Theory and Algorithms

Tokyo Institute of Technology
School of Computing
Department of Mathematical and Computing Science
Mituhiro Fukuda

2nd Quarter of 2020 *(June 25, 2020 version)*

Course description and aims

This course will cover basic notions to comprehend the gradient-based methods for convex optimization problems considered in mathematical optimization, machine learning and image processing. It starts with the basics, from the definition of convex sets in convex optimization, and will gradually focus on continuously differentiable convex functions. Along the lectures, it will also cover the characterization of solutions of optimization problems (optimality conditions), and numerical methods for general problems such as steepest descent methods, Newton method, conjugate gradient methods, and quasi-Newton methods. In the latter part, the accelerated gradient method of Nesterov for Lipschitz continuous differentiable convex functions will be detailed.

Student learning outcomes

Objectives: Learn the mathematical concepts and notions from the basics necessary for numerical methods for convex optimization problems. Definitions and proofs of theorems will be carefully explained. The objective is to understand the role of basic theorems of convex optimization in scientific articles, and to be prepared to apply them for other problems in mathematical optimization and machine learning.
Theme: In the first part, important theorems to analyze convex optimization problems will be introduced. In the second part, the Nesterov's accelerated gradient method which has received a lot of attention in the recent years will be explained from the mathematical point of view.

Keywords

Convex function, algorithm analysis, convex optimization problem, numerical methods in optimization, differentiable convex functions with Lipschitz continuous gradients, accelerated gradient method

Plan of the Lecture (tentative)

1. Convex sets and related results
2. Lipschitz continuous differentiable functions

3. Optimal conditions for differentiable functions
4. Minimization algorithms for unconstrained optimization problems
5. Steepest descent method and Newton method
6. Conjugate gradient methods, quasi-Newton methods
7. General assignment to check the comprehension
8. Convex differentiable function
9. Differentiable Convex functions with Lipschitz continuous gradients
10. Worse case analysis for gradient based methods
11. Steepest descent method for differentiable convex functions
12. Estimate sequence in accelerated gradient methods for differentiable convex functions
13. Accelerated gradient method for differentiable convex functions
14. Accelerated gradient methods for min-max problems
15. Extensions of the accelerated gradient methods

References

[Bertsekas] D. P. Bertsekas, *Nonlinear Programming*, 3rd edition, (Athena Scientific, Belmont, Massachusetts, 2016).

[Luenberger-Ye] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 4th edition, (Springer, New York, 2015).

[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).

[Nesterov18] Y. Nesterov, *Lectures on Convex Optimization, 2nd edition*, (Springer, Cham, Switzerland, 2018).

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

Prerequisites

It is preferred that the attendees have basic notions of mathematics such as linear algebra, calculus, and elementary topology, as well as understand basic methodologies to prove.

Assessment criteria and methods/Evaluation

Understand the basic theorems related to convex sets and convex functions, and the basic numerical methods to solve mathematical optimization problems. ~~Grade will be based on mid-term and final exams or on reports along the course.~~ Exceptionally this year, grade will be based on Two Reports or One Final Report. The final decision will be announced during the lectures based on the number of students.

1 Convex Sets

Definition 1.1 A subset C of $\mathbb{R}^n$ is *convex* if for $\forall \boldsymbol{x}, \boldsymbol{y} \in C$ and $\forall \alpha \in [0,1]$, $\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in C$.

Example 1.2 Examples of convex sets.

Definition 1.3 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 1.4 Given $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$, a point $\boldsymbol{y} \in \mathbb{R}^n$ is said to be a *convex combination* of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ if there exists non-negative $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}$ such that $\sum_{i=1}^m \lambda_i = 1$ and $\boldsymbol{y} = \sum_{i=1}^m \lambda_i \boldsymbol{x}_i$.

Example 1.5 Given $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$, $m+1$ distinct point of $\mathbb{R}^n$ $(m \leq n)$ such that $\boldsymbol{x}_1 - \boldsymbol{x}_0, \boldsymbol{x}_2 - \boldsymbol{x}_0, \ldots, \boldsymbol{x}_m - \boldsymbol{x}_0$ are linear independent, the set formed by all convex combination of $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ is called an *m-simplex* in $\mathbb{R}^n$.

Theorem 1.6 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 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{m+1}$ elements of the set and $\lambda_1, \lambda_2, \ldots, \lambda_{m+1} \geq 0$ 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 1.7 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 $\text{hull}(S)$. Again due to Exercise 2, $\text{hull}(S)$ is convex.

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

Theorem 1.8 The convex hull of $S \subseteq \mathbb{R}^n$, $\text{hull}(S)$, consists of all convex combinations of elements of S.

Proof:
Let $B := \{\sum_{i=1}^k \lambda_i \boldsymbol{x}_i \mid \exists k, \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1, \boldsymbol{x}_i \in S \ (i = 1, 2, \ldots, k)\}$ be the later set. If $\boldsymbol{y}_1, \boldsymbol{y}_2 \in B$, then $\exists \ell, m \in \mathbb{N}$, $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_\ell, \boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m \in S$, and non-negative $\alpha_1, \alpha_2, \ldots, \alpha_\ell, \beta_1, \beta_2, \ldots, \beta_m \in \mathbb{R}$ 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 (alternatively, note the observation at Definition 1.7). It is also clear that $S \subseteq B$, and therefore, $\text{hull}(S) \subseteq B$. From Theorem 1.6 the convex set $\text{hull}(S)$ must contain all convex combinations of elements of S. Hence $B \subseteq \text{hull}(S)$. ∎

Theorem 1.9 (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

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

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

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

Proposition 1.10 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 1.11 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. Let $C \subseteq \mathbb{R}^n$ a convex set and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ a real matrix. Show that the set $\{\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m \mid \boldsymbol{x} \in C\}$ is also convex.
7. Show that the convex hull of a set $S \subseteq \mathbb{R}^n$ is the unique and the smallest convex set containing S.
8. Prove Proposition 1.10.
9. Find an example where the sum of two closed sets is not a closed set.
10. Prove Proposition 1.11.

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 a (an orthogonal) projection onto 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 $\boldsymbol{x}_k$ onto $\bar{C}$ 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 1.10 and 1.11.

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 $\boldsymbol{d}^T \boldsymbol{x} \geq 0$, $\forall \boldsymbol{x} \in C$. 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. ∎

Remark 2.3 The Separation Theorem for Convex Sets is an essential result to show the strong duality theorem in convex optimization problems (see for example [Bertsekas]).

3 Lipschitz Continuous Differentiable Functions

Definition 3.1 Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{0} \neq \boldsymbol{s} \in \mathbb{R}^n$ be a direction (vector) in $\mathbb{R}^n$. The *directional derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ in the direction $\boldsymbol{s}$* is defined as

$$f'(\boldsymbol{x}; \boldsymbol{s}) := \lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{s}) - f(\boldsymbol{x})}{\alpha}.$$

Definition 3.2 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function on $\mathbb{R}^n$. This function is also a (Fréchet) differentiable function on $\mathbb{R}^n$ if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + o(\|\boldsymbol{y} - \boldsymbol{x}\|_2),$$

where $o(r)$ is some function of $r > 0$ such that

$$\lim_{r \to 0} \frac{1}{r} o(r) = 0, \ o(0) = 0.$$

In particular, if a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable on $\mathbb{R}^n$ and $\boldsymbol{0} \neq \boldsymbol{s} \in \mathbb{R}^n$, then,

$$f'(\boldsymbol{x}; \boldsymbol{s}) = \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{s} \rangle, \qquad \text{for } \boldsymbol{x} \in \mathbb{R}^n.$$

We say that the function is continuously differentiable if the function $\boldsymbol{\nabla f} : \mathbb{R}^n \to \mathbb{R}^n$ is continuous.

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 $\|\boldsymbol{a}\|_2 := \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}$ to it. For $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, $\|\boldsymbol{A}\|_2 := \sigma_1(\boldsymbol{A})$, where $\sigma_1(\boldsymbol{A})$ is the largest singular value. In particular, if $\boldsymbol{A} = \boldsymbol{A}^T \in \mathbb{R}^{n \times n}$, $\sigma_1(\boldsymbol{A}) = \max_{i=1,2,\ldots,n} |\lambda_i(\boldsymbol{A})|$, the largest absolute eigenvalue of $\boldsymbol{A}$.

Definition 3.3 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 $L \geq 0$:

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

In particular, $\boldsymbol{f}^{(1)}(\boldsymbol{x}) = \boldsymbol{\nabla f}(\boldsymbol{x})$ and $\boldsymbol{f}^{(2)}(\boldsymbol{x}) = \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x})$. Observe that if $f_1 \in \mathcal{C}_{L_1}^{k,p}(Q)$, $f_2 \in \mathcal{C}_{L_2}^{k,p}(Q)$, 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 3.4 Let $f \in \mathcal{C}^2(\mathbb{R}^n)$. Then $f \in \mathcal{C}_L^{2,1}(\mathbb{R}^n)$ if and only if $\|\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x})\|_2 \leq L, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$.

Proof:
For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\begin{aligned} \boldsymbol{\nabla f}(\boldsymbol{y}) &= \boldsymbol{\nabla f}(\boldsymbol{x}) + \int_0^1 \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}) d\tau \\ &= \boldsymbol{\nabla f}(\boldsymbol{x}) + \left(\int_0^1 \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) d\tau \right) (\boldsymbol{y} - \boldsymbol{x}). \end{aligned}$$

Since $\|\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x})\|_2 \leq L$,

$$\begin{aligned} \|\boldsymbol{\nabla f}(\boldsymbol{y}) - \boldsymbol{\nabla f}(\boldsymbol{x})\|_2 &\leq \left\| \int_0^1 \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) d\tau \right\|_2 \|\boldsymbol{y} - \boldsymbol{x}\|_2 \\ &\leq \int_0^1 \|\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))\|_2 d\tau \|\boldsymbol{y} - \boldsymbol{x}\|_2 \\ &\leq L \|\boldsymbol{y} - \boldsymbol{x}\|_2. \end{aligned}$$

On the other hand, for $\boldsymbol{s} \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$, $\alpha \neq 0$,

$$\|\boldsymbol{\nabla f}(\boldsymbol{x} + \alpha \boldsymbol{s}) - \boldsymbol{\nabla f}(\boldsymbol{x})\|_2 \leq |\alpha| L \|\boldsymbol{s}\|_2.$$

Dividing both sides by $|\alpha|$ and taking the limit to zero,

$$\|\nabla^2 \boldsymbol{f}(\boldsymbol{x})\boldsymbol{s}\|_2 \le L\|\boldsymbol{s}\|_2, \quad \boldsymbol{s} \in \mathbb{R}^n.$$

Therefore, $\|\nabla^2 \boldsymbol{f}(\boldsymbol{x})\|_2 \le L$. ∎

Example 3.5

1. The linear function $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \boldsymbol{x}\rangle \in \mathcal{C}_0^{2,1}(\mathbb{R}^n)$ since
$$\nabla \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{a}, \quad \nabla^2 \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{O}.$$
2. The quadratic function $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \boldsymbol{x}\rangle + 1/2\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}\rangle$ with $\boldsymbol{A} = \boldsymbol{A}^T$ belongs to $\mathcal{C}_L^{2,1}(\mathbb{R}^n)$ where
$$\nabla \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{A}\boldsymbol{x}, \quad \nabla^2 \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}, \quad L = \|\boldsymbol{A}\|_2.$$
3. The function $f(x) = \sqrt{1+x^2} \in \mathcal{C}_1^{2,1}(\mathbb{R})$ since
$$f'(x) = \frac{x}{\sqrt{1+x^2}}, \quad f''(x) = \frac{1}{(1+x^2)^{3/2}} \le 1.$$

Lemma 3.6 Let $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$. Then for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| \le \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2.$$

Proof:
For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \int_0^1 \langle \nabla \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x}\rangle d\tau \\ &= f(\boldsymbol{x}) + \langle \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle + \int_0^1 \langle \nabla \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle d\tau. \end{aligned}$$

Therefore,

$$\begin{aligned} |f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| &= \left|\int_0^1 \langle \nabla \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle d\tau\right| \\ &\le \int_0^1 |\langle \nabla \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| d\tau \\ &\le \int_0^1 \|\nabla \boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \nabla \boldsymbol{f}(\boldsymbol{x})\|_2 \|\boldsymbol{y} - \boldsymbol{x}\|_2 d\tau \\ &\le \int_0^1 \tau L \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 d\tau = \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2. \end{aligned}$$

∎

Consider a function $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$. Let us fix $\boldsymbol{x}_0 \in \mathbb{R}^n$, and define two quadratic functions:

$$\begin{aligned} \phi_1(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \langle \nabla \boldsymbol{f}(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0\rangle - \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2, \\ \phi_2(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \langle \nabla \boldsymbol{f}(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0\rangle + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2. \end{aligned}$$

Then the graph of the function f is located between the graphs of ϕ_1 and ϕ_2:

$$\phi_1(\boldsymbol{x}) \le f(\boldsymbol{x}) \le \phi_2(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Lemma 3.7 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$. Then for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\|\nabla \boldsymbol{f}(\boldsymbol{y}) - \nabla \boldsymbol{f}(\boldsymbol{x}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})\|_2 \leq \frac{M}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2,$$

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle - \frac{1}{2}\langle \nabla^2 \boldsymbol{f}(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle| \leq \frac{M}{6}\|\boldsymbol{y}-\boldsymbol{x}\|_2^3.$$

Lemma 3.8 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, with $\|\nabla^2 \boldsymbol{f}(\boldsymbol{x}) - \nabla^2 \boldsymbol{f}(\boldsymbol{y})\|_2 \leq M\|\boldsymbol{x}-\boldsymbol{y}\|_2$. Then

$$\nabla^2 \boldsymbol{f}(\boldsymbol{x}) - M\|\boldsymbol{y}-\boldsymbol{x}\|_2 \boldsymbol{I} \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{y}) \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{x}) + M\|\boldsymbol{y}-\boldsymbol{x}\|_2 \boldsymbol{I}.$$

Proof:
Since $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, $\|\nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x})\|_2 \leq M\|\boldsymbol{y}-\boldsymbol{x}\|_2$. This means that the eigenvalues of the symmetric matrix $\nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x})$ satisfy:

$$|\lambda_i(\nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x}))| \leq M\|\boldsymbol{y}-\boldsymbol{x}\|_2, \quad i = 1, 2, \ldots, n.$$

Therefore,

$$-M\|\boldsymbol{y}-\boldsymbol{x}\|_2 \boldsymbol{I} \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x}) \preceq M\|\boldsymbol{y}-\boldsymbol{x}\|_2 \boldsymbol{I}.$$

∎

3.1 Exercises

1. Prove Lemma 3.7.

4 Optimality Conditions and Algorithms for Minimizing Functions

4.1 General Minimization Problem and Terminologies

Definition 4.1 We define the *general minimization problem* as follows

$$\begin{cases} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & f_j(\boldsymbol{x}) \ \& \ 0, \quad j = 1, 2, \ldots, m \\ & \boldsymbol{x} \in S, \end{cases} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $f_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, 2, \ldots, m)$, the symbol & could be $=$, $\geq$, or $\leq$, and $S \subseteq \mathbb{R}^n$.

Definition 4.2 The *feasible set* Q of (1) is

$$Q = \{\boldsymbol{x} \in S \mid f_j(\boldsymbol{x}) \ \& \ 0, \ (j = 1, 2, \ldots, m)\}.$$

In the following, we assume $S \equiv \mathbb{R}^n$.

- If $Q \equiv \mathbb{R}^n$, (1) is a *unconstrained optimization problem.*
- If $Q \subsetneq \mathbb{R}^n$, (1) is a *constrained optimization problem.*
- If all functionals $f(\boldsymbol{x}), f_j(\boldsymbol{x})$ are differentiable, (1) is a *smooth optimization problem.*
- If one of functionals $f(\boldsymbol{x})$, $f_j(\boldsymbol{x})$ is non-differentiable, (1) is a *nonsmooth optimization problem.*
- If all constraints are linear $f_j(\boldsymbol{x}) = \langle \boldsymbol{a}_j, \boldsymbol{x}\rangle + b_j$ $(j = 1, 2, \ldots, m)$, (1) is a *linear constrained optimization problem.*

- In addition, if $f(\boldsymbol{x})$ is linear, (1) is a *linear programming problem.*
- In addition, if $f(\boldsymbol{x})$ is quadratic, (1) is a *quadratic programming problem.*

- If $f(\boldsymbol{x})$, $f_j(\boldsymbol{x})$ $(j = 1, 2, \ldots, m)$ are quadratic, (1) is a *quadratically constrained quadratic programming problem.*

Definition 4.3 $\boldsymbol{x}^*$ is called a *global optimal solution* of (1) if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in Q$. Moreover, $f(\boldsymbol{x}^*)$ is called the *global optimal value.* $\boldsymbol{x}^*$ is called a *local optimal solution* of (1) if there exists an open ball $B(\boldsymbol{x}^*, \varepsilon) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}^*\|_2 < \varepsilon\}$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in B(\boldsymbol{x}^*, \varepsilon) \cap Q$. Moreover, $f(\boldsymbol{x}^*)$ is called a *local optimal value.*

4.2 Complexity Bound for a Global Optimization Problem on the Unit Box

Consider one of the simplest problems in optimization, that is, minimizing a function on the n-dimensional box.

$$\begin{cases} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in B_n := \{\boldsymbol{x} \in \mathbb{R}^n \mid 0 \leq [\boldsymbol{x}]_i \leq 1, \ i = 1, 2, \ldots, n\}. \end{cases} \tag{2}$$

To be coherent, we use the ℓ_∞-norm:

$$\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq n} |[\boldsymbol{x}]_i|.$$

Let us also assume that $f(\boldsymbol{x})$ is *Lipschitz continuous on* B_n:

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_\infty, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in B_n.$$

Let us define a very simple method to solve (2), the **uniform grid method**.

Given a positive integer $p > 0$,

1. Form $(p+1)^n$ points
$$\boldsymbol{x}_{i_1, i_2, \ldots, i_n} = \left(\frac{i_1}{p}, \frac{i_2}{p}, \ldots, \frac{i_n}{p}\right)^T$$
where $(i_1, i_2, \ldots, i_n) \in \{0, 1, \ldots, p\}^n$.

2. Among all points $\boldsymbol{x}_{i_1, i_2, \ldots, i_n}$, find a point $\bar{\boldsymbol{x}}$ which has the minimal value for the objective function.

3. Return the pair $(\bar{\boldsymbol{x}}, f(\bar{\boldsymbol{x}}))$ as the result.

Theorem 4.4 Let $f(\boldsymbol{x}^*)$ be the global optimal value for (2). Then the uniform grid method yields

$$f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq \frac{L}{2p}.$$

Proof:

Let $\boldsymbol{x}^*$ be a global optimal solution. Then there are coordinates $(i_1, i_2, \ldots, i_n)$ such that $\boldsymbol{x} := \boldsymbol{x}_{i_1, i_2, \ldots, i_n} \leq \boldsymbol{x}^* \leq \boldsymbol{x}_{i_1+1, i_2+1, \ldots, i_n+1} =: \boldsymbol{y}$. Observe that $[\boldsymbol{y}]_i - [\boldsymbol{x}]_i = 1/p$ for $i = 1, 2, \ldots, n$ and $[\boldsymbol{x}^*]_i \in [[\boldsymbol{x}]_i, [\boldsymbol{y}]_i]$ $(i = 1, 2, \ldots, n)$.

Consider $\hat{\boldsymbol{x}} = (\boldsymbol{x} + \boldsymbol{y})/2$ and form a new point $\tilde{\boldsymbol{x}}$ as:

$$[\tilde{\boldsymbol{x}}]_i := \begin{cases} [\boldsymbol{y}]_i, & \text{if } [\boldsymbol{x}^*]_i \geq [\hat{\boldsymbol{x}}]_i \\ [\boldsymbol{x}]_i, & \text{otherwise.} \end{cases}$$

It is clear that $|[\tilde{\boldsymbol{x}}]_i - [\boldsymbol{x}^*]_i| \leq 1/(2p)$ for $i = 1, 2, \ldots, n$. Then $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}^*\|_\infty = \max_{1 \leq i \leq n} |[\tilde{\boldsymbol{x}}]_i - [\boldsymbol{x}^*]_i| \leq 1/(2p)$. Since $\tilde{\boldsymbol{x}}$ belongs to the grid,

$$f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq f(\tilde{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq L\|\tilde{\boldsymbol{x}} - \boldsymbol{x}^*\|_\infty \leq L/(2p).$$

∎

Let us define our goal

$$\boxed{\text{Find } \boldsymbol{x} \in B_n \text{ such that } f(\boldsymbol{x}) - f(\boldsymbol{x}^*) < \varepsilon.}$$

Corollary 4.5 The number of iterations necessary for the problem (2) to achieve the above goal using the uniform grid method is at most

$$\left(\left\lfloor \frac{L}{2\varepsilon} \right\rfloor + 2 \right)^n.$$

Proof:

Take $p = \lfloor L/(2\varepsilon) \rfloor + 1$. Then, $p > L/(2\varepsilon)$ and from the previous theorem, $f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq L/(2p) < \varepsilon$. Observe that we constructed $(p+1)^n$ points. ∎

Consider the class of problems $\mathcal{P}$ defined as follows:

Model:	$\min_{\boldsymbol{x} \in B_n} f(\boldsymbol{x}),$ $f(\boldsymbol{x})$ is ℓ_∞-Lipschitz continuous on B_n.
Oracle:	Only function values are available
Approximate solution:	Find $\bar{\boldsymbol{x}} \in B_n$ such that $f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) < \varepsilon$

Theorem 4.6 For $\varepsilon < \frac{L}{2}$, the number of iterations necessary for the class of problems $\mathcal{P}$ using any method which uses only function evaluations is always at least $(\lfloor \frac{L}{2\varepsilon} \rfloor)^n$.

Proof:

Let $p = \lfloor \frac{L}{2\varepsilon} \rfloor$ (which is ≥ 1 from the hypothesis).

Suppose that there is a method which requires $N < p^n$ calls of the oracle to solve the problem in $\mathcal{P}$.

Then, there is a point $\hat{\boldsymbol{x}} \in B_n = \{\boldsymbol{x} \in \mathbb{R}^n \mid 0 \leq [\boldsymbol{x}]_i \leq 1, \ i = 1, 2, \ldots, n\}$ where there is no test points in the interior of $B := \{\boldsymbol{x} \mid \hat{\boldsymbol{x}} \leq \boldsymbol{x} \leq \hat{\boldsymbol{x}} + \boldsymbol{e}/p\}$ where $\boldsymbol{e} = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$.

Let $\boldsymbol{x}^* := \hat{\boldsymbol{x}} + \boldsymbol{e}/(2p)$ and consider the function $\bar{f}(\boldsymbol{x}) := \min\{0, L\|\boldsymbol{x} - \boldsymbol{x}^*\|_\infty - \varepsilon\}$. Clearly, $\bar{f}$ is ℓ_∞-Lipschitz continuous with constant L and its global minimum is $-\varepsilon$. Moreover, $\bar{f}(\boldsymbol{x})$ is non-zero valued only inside the box $B' := \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\|_\infty \leq \varepsilon/L\}$.

Since $2p \leq L/\varepsilon$, $B' \subseteq \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\|_\infty \leq 1/(2p)\} = B$.

Therefore, $\bar{f}(\boldsymbol{x})$ is equal to zero to all test points of our method and the accuracy of the method is ε.

If the number of calls of the oracle is less than p^n, the accuracy can not be better than ε. ∎

Theorem 4.6 supports the claim that the *general optimization problem is unsolvable.*

Example 4.7 Consider a problem defined by the following parameters. $L = 2$, $n = 10$, and $\varepsilon = 0.01$.