The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

Lecture 1

Caramanis & Mokhtari Wednesday, January 19, 2022

Goal: In this lecture, we discuss the general optimization formulation as well as basics of optimization problem classes. Then, we talk about the general global minimization problem with ℓ_∞-Lipschitz functions, its lower complexity bound, and the uniform grid method for solving this class of problems.

1 References and Resources

In this course, we closely follow the textbook on Optimization Algorithms by Yurii Nesterov:

- Book: "Lectures on Convex Optimization" (Second Edition) by Yurii Nesterov.

Other resources that you might find useful are

- Book: "Numerical Optimization" (Second Edition) by Jorge Nocedal and Steve Wright.
- Book: First-Order Methods in Optimization" MOS-SIAM Series on Optimization by Amir Beck.
- Book: "Convex Optimization Algorithms" by Dimitri Bertsekas.
- Book: "Convex Analysis and Optimization" by Dimitri Bertsekas, Angelia Nedic, and Asuman Ozdaglar.
- Book: "Nonlinear Programming" (Third Edition) by Dimitri Bertsekas.
- Book: "Convex Optimization Algorithms and Complexity" Foundations and Trends in Machine Learning, by Sebastian Bubeck.

2 Important Remark Regarding Notation

In all lecture notes, we use

- lowercase normal font for scalars: x, y, a, ...
- lowercase boldface for vectors: $\mathbf{x}$, $\mathbf{y}$, $\mathbf{a}$, ...
- uppercase boldface for matrices: $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{A}$, ...

3 General Form of Optimization Problems

Consider the decision variable $\mathbf{x} \in \mathbb{R}^n$ and the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 0, \ldots, m$ and the functions $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, p$, and the set $\mathcal{Q} \subseteq \mathbb{R}^n$. The general form of an optimization problem is given by

$$\begin{array}{rll} \text{minimize} & f_0(\mathbf{x}) & \\ \text{subject to} & f_i(\mathbf{x}) \leq 0 & \text{for } i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1, \ldots, p \\ & \mathbf{x} \in \mathcal{Q} & \end{array} \tag{1}$$

Note that the **feasible set** of the above problem is defined as

$$\mathcal{S} = \{\mathbf{x} \in \mathcal{Q} \mid f_i(\mathbf{x}) \leq 0 \text{ for } i = 1, \ldots, m, \;\; h_i(\mathbf{x}) = 0, \text{ for } i = 1, \ldots, p\}.$$

There are different types of minimization problems that we consider in this class

- Constrained problems: $\mathcal{S} \subsetneq \mathbb{R}^n$.
- Unconstrained problems: $\mathcal{S} = \mathbb{R}^n$.
- Smooth problems: f_0, f_i for $i = 1, \ldots, m$ and h_i for $i = 1, \ldots, p$ are all differentiable.
- Nonsmooth problems: subset of the above functions are not differentiable.

A few remarks:

- $\mathbf{x} \in \mathbb{R}^n$ is called the optimization (decision) variable
- f_0 is the objective function
- $f_i(\mathbf{x}) \leq 0 \quad \text{for } i = 1, \ldots, m$ are inequality constraints
- $h_i(\mathbf{x}) = 0 \quad \text{for } i = 1, \ldots, p$ are equality constraints
- If a vector $\hat{\mathbf{x}}$ satisfies all the constraints it is called a feasible point, i.e., $\mathbf{x} \in \mathcal{S}$
- The set of all feasible points $\mathbf{x}$ is called feasible set
- A problem is called infeasible if there is no point in $\mathbf{x} \in \mathbb{R}^n$ that satisfies all the constraints.
- $\mathbf{x}^*$ is called a (global) optimal solution if $\mathbf{x} \in \mathcal{S}$ and $f_0(\mathbf{x}^*) \leq f_0(\mathbf{y})$ for all $\mathbf{y} \in \mathcal{S}$. In this case, $f_0(\mathbf{x}^*)$ is called the (global) optimal value.
- The set of all optimal solutions is called the optimal set
- $\hat{\mathbf{x}}^*$ is called a local solution if $\mathbf{x} \in \hat{\mathcal{S}} \subseteq \mathcal{S}$ and $f_0(\hat{\mathbf{x}}^*) \leq f_0(\mathbf{y})$ for all $\mathbf{y} \in \hat{\mathcal{S}}$.

4 Performance of Numerical Algorithms

A few important remarks:

- There is no algorithm that works well for any problem P, but at least for a class of problems $\mathcal{P} \in P$ we might be able to come up with *proper* algorithms. Thus, a natural choice is to report the performance of an algorithm $\mathcal{A}$ on the whole class $\mathcal{P}$. For instance, we state that the performance of gradient descent algorithm on smooth convex problems is such.
- We often do not have access to the full information about the problem, but we have access to some **oracle** $\mathcal{O}$ that we can call. For instance, calls to the objective function or its gradient.
- When we state that we solve a problem, it means that we find a solution with accuracy $\epsilon > 0$.
- **Analytical complexity** of an algorithm $\mathcal{A}$ is the number of calls to the oracles that $\mathcal{A}$ needs for solving problem P up to accuracy ϵ.
- **Arithmetical complexity** of an algorithm $\mathcal{A}$ is the number of arithmetic operations that $\mathcal{A}$ needs for solving problem P up to accuracy ϵ.

4.1 General Iterative Scheme

Initialization Phase

Input: Starting point $\mathbf{x}_0$ and the required accuracy $\epsilon > 0$

Initialization: Set $k = 0$ and $\mathcal{I}_{-1} = \emptyset$

(k is the time index and $\mathcal{I}_k$ is the accumulated information set up to time k)

Main Loop

$\boxed{1.}$ Call oracle $\mathcal{O}$ at input x_k.

$\boxed{2.}$ Update the information set: $\mathcal{I}_k = \mathcal{I}_{k-1} \cup (\mathbf{x}_k, \mathcal{O}(\mathbf{x}_k))$.

$\boxed{3.}$ Apply algorithm $\mathcal{A}$ to $\mathcal{I}_k$ and generate a new point $\mathbf{x}_{k+1}$.

$\boxed{4.}$ If the accuracy ϵ is achieved return the new point.
Otherwise, set $k \leftarrow k + 1$ and go to Step 1.

4.2 Oracle Definition

In this course, we focus on the **Black Box** assumption which states:

- The only information available for the numerical scheme is the answer of the oracle.

Different types of oracle that we consider in this course:

- **Zeroth-order**: For a given $\mathbf{x}$, returns the function value $f(\mathbf{x})$
- **First-order**: For a given $\mathbf{x}$, returns the function value $f(\mathbf{x})$ and the gradient $\nabla f(\mathbf{x})$
- **Second-order**: For a given $\mathbf{x}$, returns the function value $f(\mathbf{x})$, the gradient $\nabla f(\mathbf{x})$, and the Hessian $\nabla^2 f(\mathbf{x})$

5 Complexity Bounds for Global Optimization

Consider the following optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ & \mathbf{x} \in \mathcal{B}_n \end{array}$$

where $\mathcal{B}_n = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}\}$. Further, assume that the objective function is L-Lipschitz with respect to the infinity norm:

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|_\infty.$$

Therefore, we formally have:

Model: $\min_{\mathbf{x}\in\mathcal{B}_n} f(\mathbf{x})$, where f is ℓ_∞-Lipschitz continuous with constant L on $\mathcal{B}_n$.

Oracle: Zeroth-order Black Box.

Approximation Solution: Find $\hat{\mathbf{x}} \in \mathcal{B}_n$ such that $f(\hat{\mathbf{x}}) - f^* < \epsilon$

Definition 1. *We call the above class of problems $\mathcal{P}_\infty$.*

Now consider the following algorithm called the **Uniform Grid Method**:

- 0. **Input:** p
- 1. Form p^n points

$$\mathbf{x}_\alpha = \left[\frac{2i_1 - 1}{2p}, \frac{2i_1 - 1}{2p}, \ldots, \frac{2i_n - 1}{2p}\right]$$

  where $\alpha = (i_1, \ldots, i_n) \in \{1, \ldots, p\}^n$

- 2. Among all points $\mathbf{x}_\alpha$, find the one $\hat{\mathbf{x}}$ with the minimal objective function value.
- 3. Return $\hat{\mathbf{x}}$ and $f(\hat{\mathbf{x}})$ as the output.

Theorem 1. *Let f^* be the global optimal value of the problem. Then,*

$$f(\hat{\mathbf{x}}) - f^* \leq \frac{L}{2p}$$

Proof. Consider the set

$$\mathcal{X}_\alpha = \left\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}_\alpha\|_\infty \leq \frac{1}{2p}\right\}$$

Obvisously we have $\mathcal{B}_n = \cup_{\alpha \in \{1,\ldots,p\}^n} \mathcal{X}_\alpha$. Suppose $\mathbf{x}^*$ is a global optimal solution of the problem. Then it belongs to one of the sets which we denote by $\mathcal{X}_\alpha^*$. Then, indeed

$$\|\mathbf{x}^* - \mathbf{x}_\alpha^*\|_\infty \leq \frac{1}{2p}$$

Then, we can conclude that

$$f(\hat{\mathbf{x}}) - f(\mathbf{x}^*) \leq f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) \leq L\|\mathbf{x}^* - \mathbf{x}_\alpha^*\|_\infty \leq \frac{L}{2p},$$

where the first inequality follows from the fact that $\hat{\mathbf{x}}$ has the smallest objective function value among all the selected points and the second inequality follows from the Lipschitz continuity of the function. $\square$

The above theorem implies that to achieve an ϵ accurate solution we need to ensure that $\frac{L}{2p} < \epsilon$. Hence, we need to select p as

$$p := \left\lfloor \frac{L}{2\epsilon} \right\rfloor + 1,$$

and therfore the analytical complexity of the uniform grid algorithm is

$$\left(\left\lfloor \frac{L}{2\epsilon} \right\rfloor + 1\right)^n$$

Remark 1. *Note that this result $\left(\left\lfloor \frac{L}{2\epsilon} \right\rfloor + 1\right)^n$ provides an **upper complexity bound** for the considered class of problems.*

Now to understand if our algorithm is efficient or our analysis is tight we need to come up with a lower complexity bound. To do so, we first need to formally identify the class of problems that we study and then show that for any algorithm the number of calls to the oracle should be at least a specific number.

In the following we establish a **lower complexity bound** for the class of problems $\mathcal{P}_\infty$.

Theorem 2. *For $\epsilon < \frac{L}{2}$, the analytical complexity of the problem class $\mathcal{P}_\infty$ is at least $\left(\left\lfloor \frac{L}{2\epsilon} \right\rfloor\right)^n$ calls of the oracle.*

Proof. Let $p = \left\lfloor \frac{L}{2\epsilon} \right\rfloor$. Suppose there exits a method that requires $N < p^n$ calls to the oracle to solve any problem from the class of problems $\mathcal{P}_\infty$. Let's apply this method to the following resisting strategy

$$\text{Return } f(\mathbf{x}) = 0 \text{ at any test point } \mathbf{x}.$$

Therefore, the returned point $\hat{\mathbf{x}}$ satisfies $f(\hat{\mathbf{x}}) = 0$. Now since $N < p^n$ one region $\mathcal{X}_{\hat{\alpha}}$ is indeed not sampled from. Set $\mathbf{x}^*$ as the center of that region, i.e., $\mathbf{x}^* = \mathbf{x}_{\hat{\alpha}}$. Now consider the function

$$\bar{f}(\mathbf{x}) = \min\{0, L\|\mathbf{x} - \mathbf{x}^*\|_\infty - \epsilon\}.$$

This function is indeed ℓ_∞-Lipschitz with L and its optimal value is $-\epsilon$. Further outside the region $\mathcal{X}_{\hat{\alpha}}$ its value is 0. Thus, at all test points its value is 0. Therefore,

$$\bar{f}(\hat{\mathbf{x}}) - \bar{f}^* = 0 - (-\epsilon) = \epsilon,$$

and the proof is complete. □

The above theorem shows the following result:

Remark 2. $\left(\left\lfloor \frac{L}{2\epsilon} \right\rfloor\right)^n$ *is a* ***lower complexity bound*** *for the considered class of problems $\mathcal{P}_\infty$.*