Introduction to Optimization

Jun Moon

Lecture 3 (Optimization Basics)

Overview

- Optimization Basics
- Unconstrained optimization: Necessary and Sufficient Conditions
- Convex optimization

What is Optimization?: Mathematical Formulation

$$\min_{x} f(x)$$
$$\text{subject to } x \in X \subseteq \mathbb{R}^n$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is the objective (or cost) function
- The minimization is taken over w.r.t. x
- $x \in \mathbb{R}^n$ are the decision variables
- $X \subseteq \mathbb{R}^n$ is the constraint set
- We say the optimization problem is *feasible* if X is not empty, and $x \in X$ is a *feasible point*
- If $X = \mathbb{R}^n$, the problem is an unconstrained optimization problem; otherwise it is a constrained optimization problem

What is Optimization?: Mathematical Formulation

$$\min_{x} f(x)$$
$$\text{subject to } x \in X \subseteq \mathbb{R}^n$$

- A feasible point $x^* \in X$ is an optimal point if

$$f(x^*) \leq f(x), \ \forall x \in X$$

- $f^* := f(x^*)$ is the optimal value
- Goal: Find an optimal point that minimizes the objective function over the constraint set X

What is Optimization?: Mathematical Formulation

$$\min_x f(x)$$
$$\text{subject to } x \in X \subseteq \mathbb{R}^n$$

The above formulation is abstract, since the constraint set X is not structured. In most cases, the constraint set can be broken into explicit constraints that are written in terms of other functions with equalities and inequalities, and other constraints that are not so explicit.

What is Optimization?: Mathematical Formulation

In this case, the optimization problem can be written as

$$\min_x f(x) \;\; \text{s.t.}$$
$$g_i(x) \leq 0, \;\; \forall i = 1, 2, \ldots, m$$
$$h_j(x) = 0, \;\; \forall j = 1, 2, \ldots, l$$

- Mostly, we consider $X = \mathbb{R}^n$
- $g_i : \mathbb{R}^n \to \mathbb{R}, \;\; i = 1, 2, \ldots, m$
- $h_j : \mathbb{R}^n \to \mathbb{R}, \;\; j = 1, 2, \ldots, l$
- The feasible set: $Y = \{x \in \mathbb{R}^n \; | g_i(x) \leq 0, \;\; h_j(x) = 0, \;\; \forall i, j\}$

What is Optimization?: Mathematical Formulation

$$
\begin{aligned}
&\min_{x} f(x) \\
\text{s.t. } & g_i(x) \leq 0, \ \forall i = 1, 2, \ldots, m \\
& h_j(x) = 0, \ \forall j = 1, 2, \ldots, l
\end{aligned}
$$

We consider the minimization problem only in this class, since the maximization problem can be formulated from the minimization problem by using the following fact:

$$\max_{x} f(x) = \min_{x}(-f(x))$$

Hence, if you have the maximization problem, it can easily converted into the minimization problem.

What is Optimization?: Mathematical Formulation

$$
\begin{aligned}
&\min_x f(x) \\
\text{s.t. } & g_i(x) \leq 0, \ \forall i = 1, 2, \ldots, m \\
& h_j(x) = 0, \ \forall j = 1, 2, \ldots, l
\end{aligned}
$$

Major issues of the (general) optimization problem

- Existence: When does there exist an optimal solution to the problem? (Important!!!!)
- Uniqueness: If the optimal solution exists, is it unique? (need a strong condition)
- Characterization: Are there analytical tools to characterize the optimal solution? (differentiability, simple problem)
- Algorithms: If the analytical approach is not applicable, what numerical techniques should we use to solve the problem?

Existence of an Optimal Solution

Example:

$$\min_{x \in \mathbb{R}} x \;\Rightarrow\; f(x) = x \to -\infty \text{ as } x \to -\infty$$
$$\min_{x \in \mathbb{R}} e^{-x} \;\Rightarrow\; f(x) = e^{-x} \to 0 \text{ as } x \to \infty$$

In the above examples, optimal solutions do not exist!!!!

Note that in the second example, although the objective function $f(x) = e^{-x}$ is bounded below from zero, the minimum does not exist. Unlike the first example where the minimum point is not achievable, in the second example, we can find x such that $f(x)$ is arbitrarily close to zero.

Existence of an Optimal Solution

We have the following question:

Question

When does there exist the optimal solution to the optimization problem of

$$\min_{x \in X} f(x)$$

$\Rightarrow$ If such an $x^* \in X$ exists, then it is called a minimizer. In this case

$$f(x^*) = \min_{x \in X} f(x) \leq f(x), \ \forall x \in X$$

Existence of an Optimal Solution

Weierstrass's theorem

If $f : X \to \mathbb{R}$ is a continuous function defined on a compact (closed and bounded) set $X \subseteq \mathbb{R}^n$, then there exist $x^*, y^* \in X$ such that

$$f(x^*) = \min_{x \in X} f(x)$$
$$f(y^*) = \max_{x \in X} f(x)$$

That is

$$f(x^*) \leq f(x) \leq f(y^*), \ \forall x \in X$$

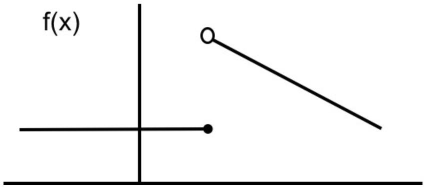
Hence, both a minimizer and a maximizer exist over the compact set X

Existence of an Optimal Solution

- It is a very strong theorem, since it guarantees existence of both minimum and maximum points.
- We won't prove this theorem (see the textbook on real analysis)
- To understand the theorem, we have to know
  - compact set, continuity, norm, bounded and closed set,

Existence of an Optimal Solution

Example 1: The function is not continuous



Example 2: $f(x) = e^{-x}$ on $X = [0, \infty)$: X is not bounded

Example 3: $f(x) = x$ on $(0, 1)$: X is not closed

Existence of an Optimal Solution

We present two more existence theorems without proofs

Theorem: Let $f : X \to \mathbb{R}$ be a continuous function. Suppose that there exists $c > 0$ such that $L_c(f) \cap X$ is a compact set, where

$$L_c(f) = \{x \in X | f(x) \leq c\}$$

then there is a minimizer to the problem

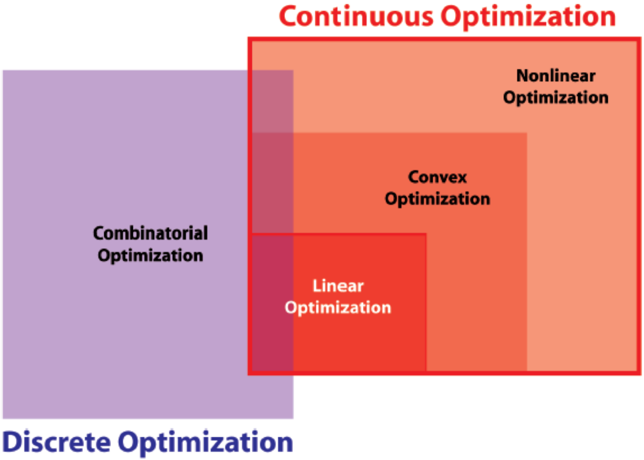
$$\min_{x \in X} f(x)$$

Theorem: If a function is continuous and coercive, i,e, $\lim_{\|x\| \to \infty} f(x) = \infty$, then a minimizer exists for the problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

First and Second-Order Necessary Conditions

Nonlinear Optimization



First and Second Order Conditions

We consider the following optimization problem:

$$\min_{x \in X} f(x)$$

Definition: local minimum

A point $x^* \in X$ is said to be a local minimum point of f over X if there exists $\delta > 0$ such that

$$f(x^*) \leq f(x) \;\; \forall x \in B_\delta(x^*) \cap X,$$

Equivalently,

$$f(x^*) \leq f(x) \;\; \forall x \in X \; \& \; |x - x^*| \leq \delta$$

First and Second Order Conditions

Definition: Global minimum

A point x^* is said to be a global minimum point if

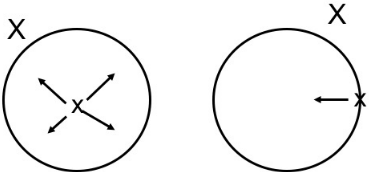
$$f(x^*) \leq f(x) \; \forall x \in X$$

- Global minimum point is a local minimum point, but the converse may not be true.

First and Second Order Conditions

Definition: feasible direction

Given $x \in X$, a vector d is a feasible direction at x if there exists a constant $\bar{\alpha} > 0$ such that $x + \alpha d \in X$ for all $0 \le \alpha \le \bar{\alpha}$



First and Second Order Conditions

Recall: If a function $f : X \to \mathbb{R}$ is differentiable at $x \in X$, then we write its gradient vector as follows

$$\nabla f(x) = \begin{pmatrix} \frac{f(x)}{\partial x_1} \\ \frac{f(x)}{\partial x_2} \\ \vdots \\ \frac{f(x)}{\partial x_n} \end{pmatrix}$$

First and Second Order Conditions

We consider the following optimization problem:

$$\min_{x \in X} f(x)$$

Theorem: first-order necessary condition

Let $X \subseteq \mathbb{R}^n$, and suppose that f is differentiable on X. If x^* is a local minimum point of f over X, then x^* satisfies

$$\nabla f^T(x^*)d = \sum_{i=1}^{n} \frac{f(x)}{\partial x_i} d_i \geq 0$$

for all feasible directions d at x^*

- Note that it is a necessary condition, that is a point $\bar{x} \in X$ that satisfies the above condition is not necessarily a local minimum.

First and Second Order Conditions

Proof of the theorem:

Let d be a feasible direction, and we define

$$x(\epsilon) = x^* + \alpha d, \quad \forall \alpha \in [0, \bar{\alpha}],$$

and

$$g(\alpha) = f(x)$$

Note that $g(0) = f(x^*)$ by the definition.

First and Second Order Conditions

Then the Taylor series around $\alpha \approx 0$ implies

$$g(\alpha) = g(0) + g'(0)\alpha + o(\alpha)$$

where the term $o(\alpha)$ is negligible when α is close to zero.
We have

$$g'(\alpha) = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} \frac{dx_i}{d\alpha} = \nabla f^T(x)d,$$

and

$$g'(0) = \nabla f^T(x^*)d$$

First and Second Order Conditions

Hence, for sufficiently small α, we have

$$g(\alpha) = f(x^*) + \alpha \nabla f^T(x^*) d$$
$$\Leftrightarrow g(\alpha) - f(x^*) = \alpha \nabla f^T(x^*) d$$

If $\nabla f^T(x^*) d < 0$, then for sufficiently small value of $\alpha > 0$,

$$g(\alpha) - f(x^*) < 0$$

which contradicts the fact that $f(x^*)$ is the local minimum point. Hence,

$$\nabla f^T(x^*) d \geq 0$$

This completes the proof.

First and Second Order Conditions

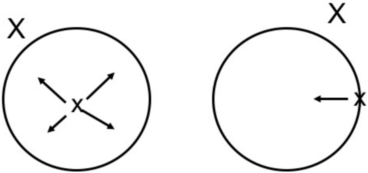
We may have the better result if x^* satisfies an additional condition

Definition: interior set

$int(X)$ means interior of X. The interior of X consists of all interior points of X.

Definition: interior point

A point $x \in X$ is an interior point of X if there exists $\delta > 0$ such that $B_\delta(x) \subseteq X$.



First and Second Order Conditions

Corollary

Corollary: If $x^* \in int(X)$ and is a local minimum point. then

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

Proof: Since d and $-d$ are both feasible directions,

$$\nabla f^T(x^*)d \geq 0, \ -\nabla f^T(x^*)d \geq 0 \Rightarrow \nabla f^T(x^*) = 0$$

First and Second Order Conditions

The previous corollary also implies the following important result:

Corollary

Suppose that $X = \mathbb{R}^n$. Let x^* be a local minimum point of the unconstrained optimization problem of $f : \mathbb{R}^n \to \mathbb{R}$, i,e,

$$f(x^*) = \min_{x \in \mathbb{R}^n} f(x)$$

Assume that f is differentiable. Then

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

First and Second Order Conditions

This corollary means that for the following problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

if x^* is the corresponding optimal solution, it must hold $\nabla f(x^*) = 0$ (high school math!!!)

First and Second Order Conditions

To see the the second order necessary condition, we need the following result and definition:

If a function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, then its second derivative is as follows

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \; i = 1, 2, \ldots, n, \; j = 1, 2, \ldots, n$$

and we denote $H := \nabla^2 f(x)$. H is also called Hessian matrix

- H is a $n \times n$ dimensional square matrix
- H is a symmetric matrix, since $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$ for all $i \neq j$

First and Second Order Conditions

Positive definite matrix

A symmetric matrix H is positive definite ($H > 0$) if

$$z^T Hz > 0 \quad \forall z \in \mathbb{R}^n$$

and is positive semi-definite ($H \geq 0$) if

$$z^T Hz \geq 0 \quad \forall z \in \mathbb{R}^n$$

- $H \geq 0$ iff all the (real part) eigenvalues of H are ≥ 0
- $H > 0$ iff all the (real part) eigenvalues of H are > 0

First and Second Order Conditions

We consider the following optimization problem:

$$\min_{x \in X} f(x)$$

Theorem: Second-order necessary condition

Let $X \subseteq \mathbb{R}^n$, and let f is twice differentiable. If x^* is a local minimum point of f over X, then for any feasible direction d of x^*, we have

(i) $\nabla f(x^*)d \geq 0$

(ii) if $\nabla f(x^*)d = 0$, then $d^T \nabla^2 f(x^*)d \geq 0$

First and Second Order Conditions

Proof:

(i) is shown in the previous theorem. For the second part, first note that $\nabla f(x^*)d = 0$. Then with the same definition of $g(\alpha)$, its Taylor expansion to the second order leads to

$$g(\alpha) - g(0) = \frac{1}{2}g''(0)\alpha^2 + o(\alpha^2)$$

Note that by definition

$$g''(0) = d^T \nabla^2 f(x^*) d$$

First and Second Order Conditions

Similarly, if $g''(0) < 0$, then it contradicts the fact that x^* is a local minimum point. Hence, we need

$$d^T \nabla^2 f(x^*) d \geq 0,$$

This completes the proof.

First and Second Order Conditions

Corollary

Let $X \subseteq \mathbb{R}^n$, and let f is twice differentiable. If $x^* \in int(X)$ is a local minimum point of f over X, then

(i) $\nabla f(x^*) = 0$

(ii) for all d, $d^T \nabla^2 f(x^*) d \geq 0 \Leftrightarrow H \geq 0$

- The above corollary implies that if $\bar{x} \in int(X)$ but H is not positive semi-definite at $\bar{x}$, then $\bar{x}$ is not a local optimal point.

First and Second Order Conditions

The above corollary also implies

Let x^* be a local minimum point of the unconstrained optimization problem of $f : \mathbb{R}^n \to \mathbb{R}$, and assume that f is twice differentiable. Then

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) = H \geq 0$$

- This means that the Hessian H is positive semi-definite.

First and Second Order Conditions

So far, we have need necessary conditions for optimality. We now see the sufficient condition for a local minimum.

Theorem: Let $X \subseteq \mathbb{R}^n$, and let f is twice differentiable. Then $x^* \in int(X)$ is a local minimum point if

(i) $\nabla f(x^*) = 0$

(ii) $H = \nabla^2 f(x^*) > 0$

First and Second Order Conditions

In this case, there exist $\gamma > 0$ and $\epsilon > 0$ such that

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x - x^*\|^2 \quad \forall x \ \|x - x^*\| \leq \epsilon$$

- Note that from (ii), we need $H > 0$ (compare with the previous corollary, where $H \geq 0$.
- Also, we have $f(x) > f(x^*)$, in which case x^* is a strict local minimum point

End of Unconstrained Optimization Problems

- What we have seen so far is local analysis, since we did not say anything about global optimality.
- But mostly, we want global optimality
- If we have some convexity conditions on f and X, then local optimal becomes global; hence, the previous results can be extended to the global analysis.
- To do this, we need an additional convexity assumption for f and X

Convex Optimization

Convex Optimization?

We consider the following optimization problem:

$$\min_{x \in X} f(x)$$

- We will impose some conditions of f and X
- This guarantees that the first-order necessary condition becomes sufficient

Convex Optimization

We consider the following optimization problem:

$$\min_{x \in X} f(x)$$

Assume that

- $X \subset \mathbb{R}^n$ is a convex set
- f is a convex function

Why Do We Study Convex Optimization?

We consider the following optimization problem:

$$\min_{x \in X} f(x)$$

Fact 1: local minimum is global minimum in convex optimization

If x^* is a local minimum to the above problem, then it is also global minimum point for a convex optimization problem.

Why Do We Study Convex Optimization?

Proof: Assume that we have $y \in X$ such that

$$f(y) < f(x^*)$$

Since X is convex, for any $\alpha \in [0, 1]$, we have

$$\alpha x^* + (1 - \alpha)y \in X$$

By the convexity of f, we have

$$f(\alpha x^* + (1 - \alpha)y) \leq \alpha f(x^*) + (1 - \alpha)f(y^*) < f(x^*)$$

Choosing α close to 1, we can obtain a contradiction to the fact that $f(x^*)$ is a local minimum.

Useful Facts

Recall: First order necessary condition

Let $X \subseteq \mathbb{R}^n$, and suppose that f is differentiable on X. If x^* is a local minimum point of f over X, then x^* satisfies

$$\nabla f^T(x^*)d = \sum_{i=1}^{n} \frac{f(x)}{\partial x_i} d_i \geq 0$$

for all feasible directions d at x^*

Useful Facts

Fact 2: first order necessary optimality condition becomes sufficient for the convex optimization problem

If x^* satisfies the first-order necessary condition, then x^* is a global optimal point of the convex optimization problem.

Useful Facts

Proof:
Recall that f is convex if and only if its domain is convex and

$$f(x) + \nabla f(x)^\top (z - x) \leq f(z) \quad \forall x, z \in X$$

Then for the convex function that is differentiable, we have

$$f(y) \geq f(x^*) + \nabla f(x^*)(y - x^*) \;\; \forall y \in X$$

Note that $(y - x^*)$ is a feasible direction, and by the first-order optimality condition, we have

$$\nabla f(x^*)(y - x^*) \geq 0.$$

Hence,

$$f(y) \geq f(x), \; \forall y \in X$$

This completes the proof.

Useful Facts

Fact 3: Uniqueness

Assume that f is strictly convex, i.e.,

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2),\ \forall \alpha \in (0,1),\ x_1, x_2 \in X.$$

Then There exists a unique minimizer.

Proof: Assume that there are two minimizers, x^* and y^*. Then consider for any $\theta \in (0,1)$

$$f(\theta x^* + (1-\theta) y^*) < \theta f(x^*) + (1-\theta) f(y^*) = f(x^*) = f(y^*)$$

since $f(x^*)$ and $f(y^*)$ are both the minimum point. This contradicts the fact that $f(x^*)$ and $f(y^*)$ are the same minimum point.

Useful Facts

Summary

We consider the following optimization problem:

$$\min_{x \in X} f(x)$$

Assume that X is a convex set and f is a convex function

- Any local optimal solution becomes the global optimal solution
- The first-order necessary condition is also the sufficient condition
- If f is strictly convex, then there is a unique optimal solution

Example: Least Square Problem

Consider

$$f(x) = \|Ax - b\|^2 = x^T A^T Ax - 2x^T A^T b + b^T b$$

Since $A^T A \geq 0$, the problem is convex, and we have the global optimal solution. This optimization problem is also known as the quadratic programming.

- When $A^T A > 0$, i.e. $A^T A$ is positive definite, there is a unique minimizer, which can be written as

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

  Hence,

  $$x^* = (A^T A)^{-1} A^T b$$

Conclusions

- Optimization theory