

STAT 24310 Notes

Anthony Yoon

April 24, 2025

Contents

1	Lecture 1: Introduction	2
2	Lecture 9: Introduction to Optimization	2
2.1	Examples	2
2.2	Classes of Optimization Problems	2
2.3	Taylor's Theorem	3
2.4	Applications of convexity	5

Abstract

Notes for STAT24310 taught by Professor Yuehaw Khoo. Prior Linear Algebra knowledge will be assumed.

1 Lecture 1: Introduction

This class is heavily based on Trefethen and Bau's Textbook. Take a look at it if you have the time.

When we run an algorithm, we often are interested in how long it takes to run the algorithm. For example, if we have a matrix $A \in \mathbb{R}^{n \times n}$ an $x, b \in \mathbb{R}^n$, and we are interested in solving

$$Ax = b$$

and n is very large, say $n = 100,000$, a concern is that the algorithm takes forever to run. In this case, we may be worried about the worst case time complexity, denoted as *big O notation*. In this case, solving $x = A^{-1}b$ is $\mathcal{O}(n^3)$. But what is this notation?

Definition 1.1. If there exists a $C \in \mathbb{R}$ such that for fg , where $g \geq 0$, where for all sufficiently t large enough, such that $|f(t)| \leq C \cdot g(t)$, then $f(t) = \mathcal{O}(g(t))$

or equivalently:

Definition 1.2. If there exists a constant $C \in \mathbb{R}$ and a real number t_0 such that for all $t \geq t_0$, $|f(t)| \leq C \cdot g(t)$, where $g(t) \geq 0$, then we write $f(t) = \mathcal{O}(g(t))$ as $t \rightarrow \infty$.

Both are equivalent.

2 Lecture 9: Introduction to Optimization

Last class, we showed that there existed a solution to the Least Squares problem, or rather $\min \|Ax - b\|_2^2$ through a combination of the QR and LU decomposition. We can consider the following general optimization problem, where we define $f : \Omega \rightarrow \mathbb{R}$:

$$x^* = \arg \min_{x \in \Omega} f(x)$$

where we say that Ω is the objective domain. The intuition is that we want to find the input that would minimize the value of the function.

2.1 Examples

Consider the nonlinear inverse problem used within physics.

Explain
this better

2.2 Classes of Optimization Problems

Given a domain Ω that we want to optimize over, we have the following scenarios:

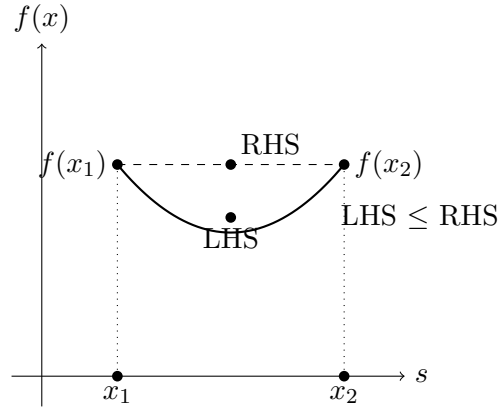
- a set of discrete points, which is discrete optimization

- $\Omega \subseteq \mathbb{R}^n, \subseteq \mathbb{R}^n$, a constrained optimization problem. An example of this would be induced norms.
- $\Omega = \mathbb{R}^n$, a unconstrained optimization problem. An example of this would be the Least Squares Regression Problem.

We can often times utilize properties of the function, which is where we can introduce the idea of convexity.

Definition 2.1. f is convex over S if for any $x_1, x_2 \in S$ and $t \in [0, 1]$, $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$

This concept can be illustrated in the following diagram:



We can then expand the definition of convexity as follows:

Definition 2.2. f is strongly convex over S if for any $x_1, x_2 \in S$ and $t \in [0, 1]$, $f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2)$

A natural expansion of this is when f is convex, then $f(\sum_{i=1}^n t_i x_i) \leq \sum_{i=1}^n t_i f(x_i)$ where $\sum_{i=1}^n t_i = 1$ such that $t_i \geq 0$. We can also consider the following definition of convexity:

Definition 2.3. Locally Convex: $f : \Omega \rightarrow \mathbb{R}$, where f is convex over some $S \subseteq \Omega$

2.3 Taylor's Theorem

We begin with an introduction of smoothness.

Definition 2.4. A function is said to be C^n if it is continuous and differentiable n number of times.

Also a refresher on what the Taylor's Theorem in 1D is.

Definition 2.5. if $f : \mathbb{R} \rightarrow \mathbb{R}$ is $(k + 1)$ times differentiable at $x = a$, then

$$f(x) = f(a) + \frac{\partial f(a)}{\partial x}(x - a) + \cdots + \frac{1}{k!} \frac{\partial^k f(x)}{\partial x^k}(x - a)^k + \frac{L}{(k + 1)!}(x - a)^{k+1}$$

where

$$L = \sup_{\xi \in [x, a]} \left| \frac{\partial^{k+1} f(x)}{\partial x^{k+1}} \right|$$

note that if $|x - a|$ is sufficiently small, we can discard the remainder term (the one with the L in it). Note that also that if the $(k + 1)$ derivative is bounded, then we can approximate f with the k th order polynomial. We can also consider the case of the special case of the third order Taylor polynomial expansion.

Definition 2.6. For $f : \mathbb{R}^n \rightarrow \mathbb{R}$, and f has a bounded 3rd derivative, the expansion is as follows

$$f(x) = f(a) + \frac{\partial f(a)}{\partial x}(x - a) + \frac{1}{2}(x - a)^T H(a)(x - a) + \mathcal{O}(\|x - a\|^3)$$

where we define

$$H(a)_{ij} = \frac{\partial^2 f(a)}{\partial x_i \partial x_j}$$

where $H(a)$ is $n \times n$

To illustrate the nature of the Hessian Matrix, we can visualize f locally to see that we can use a contour map. The Hessian usually shows the directions of the biggest changes in direction, given by the Eigenvalues.

However, we can see that if our function is convex and smooth, then *optimality is guaranteed based on the local derivative*. We can see this in the following proof:

Need to
fix

Proof. if f is convex and differentiable, we can see that if we let a, b be in the domain of f . we can see that we are left with the following inequality.

$$f(b) \geq f(a) + \frac{\partial f(a)}{\partial x}(b - a)$$

by the definition of convexity. If we choose a such that $f'(a) = 0$, we see that we have guaranteed an optimal solution. \square

However, we can see that if we are working with a strongly convex differentiable solution, we can use the same assumptions as above to see in the following proof.

Proof. We are given that f is strongly convex and differentiable. Thus, we can see that for the following Taylor expansions

$$f(b) \geq f(a) + \frac{\partial f(a)}{\partial x}(b-a) + \frac{a}{2}\|b-a\|_2^2$$

We see that we can choose a point such that $\frac{\partial f(a)}{\partial x} = 0$, which implies that

$$\begin{aligned} f(b) &\geq f(a) + \frac{a}{2}\|b-a\|^2 \\ \implies f(b) &> f(a), \forall b, b \neq a \end{aligned}$$

Since we know that the function is strongly convex and thus, we know that for any point b and a . the last inequality holds, we can see that we have indeed found the unique optimizer. \square

Remark 2.1. Strongly convex implies strictly convex, but not the other way around.

2.4 Applications of convexity

Consider the following:

$$f(x) = \frac{1}{2}ax^2 - bx, a > 0$$

and consider the following:

$$\begin{aligned} f(x) &= \frac{1}{2}ax^2 - bx \\ f(y) - f(x) &= \frac{1}{2}a(y^2 - x^2) - b(y - x) \end{aligned}$$