

STAT 24310 Notes

Anthony Yoon

April 29, 2025

Contents

1	Lecture 1: Introduction	2
1.1	Accuracy	3
2	Lecture 2: Norms	5
2.1	Matrix Norms	5
3	3: linear Transform	6
3.1	Inner Products	7
3.2	Different kinds of multiplication	7
4	Lecture 9: Introduction to Optimization	7
4.1	Examples	8
4.2	Classes of Optimization Problems	8
4.3	Taylor's Theorem	9
4.4	Applications of convexity	10

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 $f(t) \mathcal{O}(g(t))$

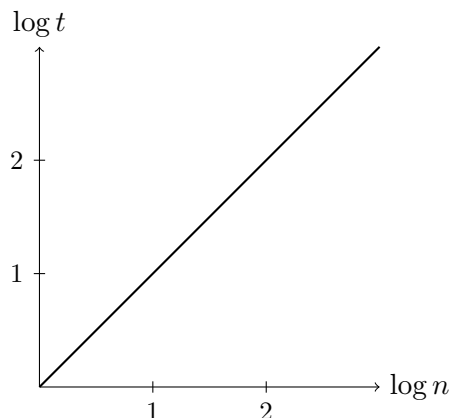
or equivalently:

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. For example, consider a problem that is $\mathcal{O}(n^3)$. If we were to increase the dimension of the problem, say a 100 times, then the algorithm would take 10000000 times longer to run. However, how do we visualize this? We do so by plotting the log-log plot, where we can note that:

$$\begin{aligned} t &= \mathcal{O}(n^3) \\ \log t &= 3 \log n + C \end{aligned}$$

where C is some constant.



where we can see that the slope gives the exponent in time complexity.

1.1 Accuracy

When we are dealing with algorithms, we can also be concerned with how accurate it is. For example, consider $f : X \rightarrow Y$, where we are interested $f(x)$ for $x \in X$. However, practically, we cannot calculate the exact values, take \sqrt{x} for example, so a computer must make an approximation, usually denoted as $\tilde{f}(x)$. We can consider the relative accuracy, defined as follows:

Definition 1.3. Given an approximation of a function $\tilde{f}(x)$, we define relative accuracy as:

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|}$$

where $\|\cdot\|$ is the norm of choosing.

A natural extension of accuracy are the following topics: **Stability and Conditional Number**. Intuively, stability is usually related to the algorithm ($\tilde{f}(x)$) used to solve the problem and the condition number is related to the actual setup of the problem ($f(x)$). We can consider the following definitions:

Definition 1.4. Backwards Stability: if $x \in X$. $\tilde{f}(x) = f(\tilde{x})$ for some $\frac{\|\tilde{x} - x\|}{\|x\|} = \mathcal{O}(\epsilon)$

Intuively, this above definition says that the problem will give the good enough answer to a slight deviation in the input. This definition is usually easier to prove.

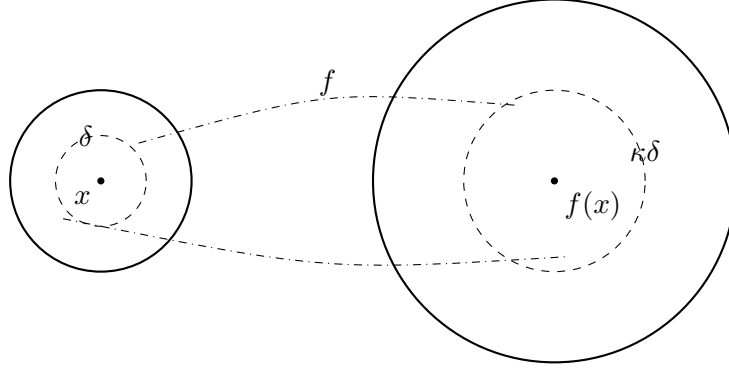
Definition 1.5. Absolute Conditional Number: for (f, x) we can consider the Absolute condition number as follows:

$$\hat{K}(f, x) = \lim_{\delta \rightarrow 0} \sup_{\|\delta x\| \leq \delta} \frac{\|f(x + \delta x) - f(x)\|}{\|\delta x\|}$$

Definition 1.6. Relative Conditional Number: for (f, x) , we define the relative condition number as:

$$\lim_{\delta \rightarrow 0} \sup_{\substack{\|\delta x\| \\ \|x\|} \leq \delta} \frac{\frac{\|f(x + \delta) - f(x)\|}{\|f(x)\|}}{\frac{\|\delta x\|}{\|x\|}} = \frac{\|f(x + \delta x) - f(x)\| \|x\|}{\|f(x)\| \|\delta x\|}$$

We can consider the diagram as follows:



We can see that the condition number measures how sensitive the actual problem is to a change in x , as seen above. The larger the problem is, the more sensitive the problem is. Think of the diameter of the circles as how values go between each set. If the condition number is large, then the diameter is larger and thus, there is chance that the small perturbation lands in a different area than original output.

Remark 1.1. Note if f is a function from $\mathbb{R}^n \rightarrow \mathbb{R}^m$, we can define the following:

$$K_{abs} = J_f \quad K_{rel} = J_f \cdot \frac{\|x\|}{\|f(x)\|}$$

We can observe the following. If \tilde{f} is backwards stable with accuracy ϵ , and if f has a relative condition number k . then \tilde{f} is accurate with the following relation:

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \mathcal{O}(k\epsilon)$$

Proof. If \tilde{f} is backward stable, then there exists a point $y \in (x - \delta, x + \delta)$, with $\|\delta\| = \mathcal{O}(\epsilon\|x\|)$, such that:

$$\tilde{f}(x) = f(y)$$

This means the computed value $\tilde{f}(x)$ is the exact value of f evaluated at a slightly perturbed input y .

Now, using the definition of the relative condition number κ of f , we know that for small perturbations:

$$\frac{\|f(y) - f(x)\|}{\|f(x)\|} \leq \kappa \cdot \frac{\|y - x\|}{\|x\|} + \mathcal{O}\left(\left(\frac{\|y - x\|}{\|x\|}\right)^2\right)$$

Since $\|y - x\|/\|x\| = \mathcal{O}(\epsilon)$, it follows that:

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \frac{\|f(y) - f(x)\|}{\|f(x)\|} = \mathcal{O}(\kappa\epsilon)$$

□

2 Lecture 2: Norms

We begin with following definition of a norm in a general vector space.

Definition 2.1. Let X denote a vector space. A norm $\|\cdot\| : X \rightarrow \mathbb{R}$ satisfies

- Positivity: $\|x\| \geq 0$, $\|x\| = 0$ iff $x = 0$
- Homogeneity: $\|\alpha x\| = |\alpha| \|x\|$
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

We can also define the l_p norm

Definition 2.2. • $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$, where $0 < p < \infty$

- $\|x\|_0$ refers to number of non-zero elements of X .
- $\|x\|_\infty = \max_j \{|x_j|\}$

We also have the following definitions:

Definition 2.3. Inner product on \mathbb{C}^n : $\langle x, y \rangle = \sum_{i=1}^n \bar{x}_i y_i$

Definition 2.4. Cauchy Schwartz inequality: $|\langle x, y \rangle| \leq \|x\| \|y\|$

Definition 2.5. Holder's inequality: $|\langle x, y \rangle| \leq \|x\|_p \|y\|_q$ is $\frac{1}{p} + \frac{1}{q} = 1$

We can also define the equivalence of norms:

Definition 2.6. $\|\cdot\|_a, \|\cdot\|_b$ are equivalent if for $c_2 \geq c_1 \geq 0$:

$$C_1 \|x\|_b \leq \|x\|_a \leq C_2 \|x\|_b$$

We can prove that all norms are equivalent. This will be left as an exercise to the reader.

2.1 Matrix Norms

We can also introduce the notion of norms on matrices.

Definition 2.7. Frobenious norm: $\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2 \right)^{\frac{1}{2}}$

Definition 2.8. Induced norm: Given $A \in \mathbb{C}^{m \times n}$, $x \in \mathbb{C}^n$. We define the induced norm as follows:

$$\|A\|_{a \rightarrow b} = \sup_{x \neq 0} \frac{\|Ax\|_a}{\|x\|_b} = \sup_{\|x\|_b=1} \|Ax\|_a$$

Remark 2.1. By definition, we can see that:

$$\|Ax\|_a \leq \|A\|_{a \rightarrow b} \|x\|_b \quad \|A\|_a = \|A\|_{a \rightarrow a}$$

We also have the following properties:

Definition 2.9. • $\|A\|_1 = \max_{1 \leq j \leq n} \|A(:, j)\|_1$, which is just the maximum column sum.

• $\|A\|_\infty = \max_{1 \leq i \leq m} \|A(i, :)\|_1$ Which is the max of the row sum.

We also have the Cauchy Schwartz inequality for the matrix norm:

Definition 2.10. $\|AB\|_2 \leq \|A\|_2 \|B\|_2$ and $\|AB\|_f \leq \|A\|_F \|B\|_f$ and $\|AB\|_f \leq \|A\|_2 \|B\|_F$

We can prove the fact $\|AB\|_2 \leq \|A\|_2 \|B\|_2$, which is as follows:

Proof. $\|AB\|_2 \leq \|A\|_2 \|B\|_2$. To show this note,

$$\sup_{\|x\|=1} \|ABx\|_2 \leq \|Ay\|_2 \leq \|A\|_2 \|y\|_2 \leq \|A\|_2$$

□

3 3: linear Transform

In numerical linear algebra, most problems and algorithms consist of linear maps.

Definition 3.1. Linear Transform is a function from $T : X \rightarrow Y$ where

- $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$
- $T(v) = Av$ where $A \in \mathbb{C}^{m \times n}$, $v \in \mathbb{C}^n$

For example consider the following example.

Example

Let $P_k(x) = x^k$, where it is monomial, where $k \in \mathbb{N} \cup \{0\}$, where $X = \text{span}\{P_0, P_1, P_2\}$ and $Y = \text{span}\{P_0, P_1, P_2, P_3\}$. Note that if $q \in X$. this means that q is equivalent to a linear combination of the polynomials, where we can see that $Tq = (q+1)q$ It is trivial to show that T is a linear transform. Thus, we aim to find the matrix for this. Consider the following

$$Tp_1 = p_2 + p_1 \quad Tp_2 = p_3 + p_2$$

where we can construct the matrix as follows¹

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

where columns vector is $\{p_0, p_1, p_2, p_3\}$ and row vectors are $\{p_0, p_1, p_2\}$.

3.1 Inner Products

We can introduce the inner product as follos.

Definition 3.2. Given a inner product $\langle \cdot, \cdot \rangle$ where $v \in \mathbb{C}^n$, we can see that

- It sastifies bilinearity
- $\langle v, w \rangle = \langle w, v \rangle$
- $\langle v, v \rangle \geq 0$ if and only if $v = 0$

Note that $\langle v, v \rangle = \|v\|^2$. We can also define $\langle v, v \rangle_M = v^T M v$ where M is a positive definite matrix.

Definition 3.3. A matrix is positive definite if and only if all the eigenvalues are positive.

Definition 3.4. A matrix is negative definite if and only if all the eigenvalues are negative.

Definition 3.5. Hermitian matrix is the complex analgue to the transpose of a matrix. Symmmetric matrix is $A^T = A$ and Hermitian matrix is $A = A^*$ ²

3.2 Different kinds of multiplication

4 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.

¹Professor Khoo did some weird proof about the change of basis here, and frankly I don't think its worth putting here as an FYI

²The * superscript denotes the complex conjugate, where every entry of the matrix is conjugated.

4.1 Examples

Consider the nonlinear inverse problem used within physics.

Explain
this better

4.2 Classes of Optimization Problems

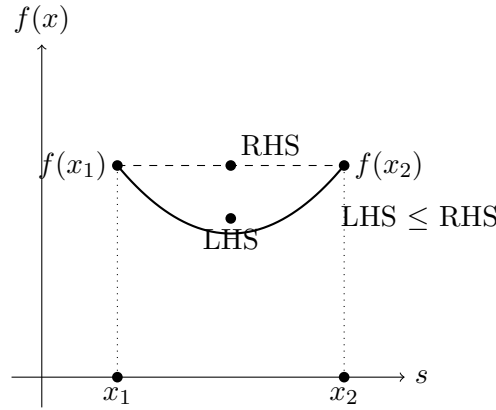
Given a domain Ω that we want to optimize over, we have the following scenarios:

- 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 4.1. 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 4.2. 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 4.3. Locally Convex: $f : \Omega \rightarrow \mathbb{R}$, where f is convex over some $S \subseteq \Omega$

4.3 Taylor's Theorem

We begin with an introduction of smoothness.

Definition 4.4. 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 4.5. 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 4.6. 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 4.1. Strongly convex implies strictly convex, but not the other way around.

4.4 Applications of convexity

Consider the following:

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

and consider the following:

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