1 January 8

1.1 Logistics

Within the three body problem is the entirety of this course.

"In mathematics you don't understand things. You just get used to them." - John von Neumann. (Bro von Neumann is so washed for this.) For context, von Neumann was challenged that he didn't actually understand the method of characteristics. He was talking about this with his Dad, who just retired as a math prof from Ohio State, and he didn't like it. Joel's alternate form: "In mathematics it takes time for ideas and examples to sink in."

Click here for course website link.

Content: Parts of Chapters 10-12, 14-16 of textbook by Ascher and Greif (online, free). Topics: Interpolation, approximation (Ch 10 - 12); Differentiation, Integration, ODE's [PDE's] (Ch 14-16). Back in the day, 303 was meant as a followup to 302, but not anymore, so may be some repeated material (norms and condition numbers, but not thee point of this course anyway).

Discussion: please post to piazza page. If this fails, please email to jf@cs.ubc.ca with subject CPSC 303.

Grading: $(10\%)\max(h, m, f) + (35\%)\max(m, f) + (55\%)f$ where h is homework, m is midterm, f is final. So technically, can ace final and ace course. Because back in the day, Joel knew someone who couldn't attend any classes, but got 100% in the final only to get C+ in the course. There is a phenomena where if someone gets 100 on the midterm, stops doing homework. But really, these assessments are good preperation and indicators of where you stand in the course.

Please sign up for piaaza and gradescope through canvas.ubc.ca (especially gradescope).

Homework: Set Thursday 11:59pm and due Thursday 11:59pm. There is both individual and group homework. Group homework: at most 4 people, and will cover most material; only submit one. Individual homework: These are the types of things he wants to make sure everyone can do, and will be like the things on exams; you must write up your own solution even if you work with others.

1.2 Intro to ODE's

1.2 and 4.2 (norms), 14.2 (differentiation), 16.1 and 16.2 (ODE's).

He typically begins courses with the most difficult stuff the course will get. This is tough, but not the worst. Now, this course only requires two terms of calculus, but with how pertinent ML is now, most people are taking multivariable calc anyway. We will get some idea of what to expect and some intuition, but will revisit later in the course. The reason the emphasis is on ODEs and not PDEs is because the general theory for ODEs really applies to all of them, even if you have to solve differently. Families of PDEs have their own properties that have to be studied separately.

We are heading towards Ordinary Differential Equations (Ch. 16).

1.2.1 Absolute vs. Relative Error

If $v \in \mathbb{R}$ is an approximation to $u \in \mathbb{R}$, then absolute error (in v) (as an approximation to u) is $|u - v|$, and the relative error is $\frac{|u-v|}{|u|}$. The same works in $\mathbb{R}^n$ (or any normed vector space):

$$\|\vec{u}\|_2 = \|(u_1, \ldots, u_n)\|_2 = \sqrt{u_1^2 + u_2^2 + \cdots u_n^2}$$

We also use $\|\vec{u}\|_1 = |u_1| + \cdots |u_n|$ and $\|\vec{u}\|_{\max} = \|\vec{u}\|_\infty = \max_{1 \le i \le n} |u_i|$. The absolute error in $\vec{v}$ as an approximation to $\vec{u}$ is $\|\vec{u} - \vec{v}\|_p$ and the relative error is $\frac{\|\vec{u}-\vec{v}\|}{\|\vec{u}\|_p}$ where $p = 1, 2, \infty$.

1.2.2 Taylor's Theorem (p. 5)

Theorem 1. *For $f\colon (a, b) \to \mathbb{R}$ where f is $k+1$ differentiable (so $f^{(k+1)}(x)$ exists in (a, b)), for some x_0, $x_0 + h$ that lie in (a, b),*

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \cdots + \frac{h^k}{k!} f^{(k)}(x_0) + \text{error}$$

where the error is $\frac{h^{k+1}}{(k+1)!} f^{(k+1)}(\xi)$ where ξ is between x_0 and $x_0 + h$.

Remark 1. h can be negative.

We'll learn how to approximate derivatives, and then ODE solution.

2 January 10

Housekeeping:

- HW1 will be assigned on Han 11, due on gradescope on Jan 18
- Access gradscope via Canvas
- Today: Separable ODE's (see, e.g. CLP 2, Appendix D of Ascher and Grief)

Last time: Ch 1: "Reviewing" termonology. Ch 4: $\|\vec{u}\|_2 = \sqrt{u_1^2 + \cdots + u_n^2}$ for $\vec{u} \in \mathbb{R}^n$. We saw classes of functions, Taylor Series, and ODE's. An ODE is where the derivatives are a function of the same variable?? So $y' = f(t, y)$. PDEs on the other hand have partial derivatives $h = h(t, x_1, \ldots, x_n)$, and maybe something like the heat equation $\frac{\partial h}{\partial t} = -\Delta_{x_1,\ldots,x_n} h$. There might be a course on the foundations of elliptic PDEs, parabolic PDEs, etc., but won't be the focus here.

And then he talks about a sketch of proof of Taylor's, except I literally did it this morning. Essentially, we use MVT to get better approximations (not rigourously). We have

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \frac{h^3}{6} f'''(\xi_1)$$

$$hf'(x_0) = f(x_0 + h) - f(x_0) - \frac{h^2}{2} f''(\xi_2)$$

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0)}{h} - \frac{h}{2} f''(\xi_2) = \frac{f(x_0 + h) - f(x_0)}{h} + \text{Order}(h)$$

where $O(h)$ is some function (depends on f, x_0). We have $|\text{Order}(h)\| \leq \frac{h}{2} M_2$ where M_2 is bound on f'' in the interval $[x_0, x_0 + h]$. By definition, $f'(x_0) = \lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}$. All this is really taken from [A&G], Ch. 14, Sections 1,2.

We will continue our discussion of derivatives, but first, some useful notation. If $a < b \in \mathbb{R}$, $C[a, b] := \{f \colon [a, b] \to \mathbb{R}$ such that f is continuous$\}$. When $k \in \mathbb{N}$,

$$C^k(a, b) = \{f \colon (a, b) \to \mathbb{R} \text{ such that } f \text{ has } k \text{ continuous derivatives } \forall x \in (a, b)\}$$

, and similarly for $C^k[a, b]$. $C^\infty(a, b)$ is the set of f that has derivatieves of all order.

Definition 1 (Real Analytic Functions).

$$C^\omega(a, b) := \{f \colon (a, b) \to \mathbb{R} \text{ such that for all } x_0 \in (a, b),\ f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x_0) + \cdots\}$$

A function $f \colon (a, b) \to \mathbb{R}$ is real analytic when $f \in C^\omega(a, b)$.

Real analytic is a stronger condition than C^∞, so $C^\omega \subset C^\infty \subset \cdots \subset C^2 \subset C^1 \subset C^0$. [Hmm... I would like a better definition of convergence here ff]

For example, e^x ff he scrolled.

2.1 Start ODE

Simple ODE [A&G]:

$$y' = f(t, y)$$

where we use the notation $y' = \frac{dy}{dt} = \dot{y}$. (Caution: math textbooks typically use $y' = \frac{dy}{dx} = f(x, y)$.)

To solve for $y = y(t)$, we are given an "intial condition", that is we have $y_o, t_o \in \mathbb{R}$ and impose $y(t_0) = y_0$.

We expect a unique solution. Say we are given $A \in \mathbb{R}$, and find a y that satisfies

$$y'(t) = Ay(t)$$

We claim that $y(t) = e^{At}C$ is a solution. We can verify: $y'(t) = (e^{At}C)' = (Ae^{At})C = Ay(t)$. So we have solved it by shamelessly guessing.

We can plug in our t_0 and y_0 which fixes $C = y_0 e^{-At_0}$. Then

$$y(t) = y_0 e^{A(t-t_0)}$$

So when A is big enough (greater than 0), we have exponential growth.

Is this solution unique? Can we simply guess and it provides the only solution? More on Friday.

3 January 12

Today:

- ODE: $y' = y$ "isoclines"
- ODE's of the form $y' = f(y)$
- Later $\vec{y}' = \vec{f}(t, \vec{y})$, system of m ODE's where $\vec{y} = \vec{y}(t) \colon \mathbb{R} \to \mathbb{R}^m$ and $\vec{f}(t, y) \colon \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$.
- Examples; $y' = y^2$, $y' = |y|^{1/2}$.

Question, what could possibly go wrong?

Solve $y' = f(y)$, $y(t_0) = y_0$. Then

(a). If f continuous near $y = y_0$ then a solution exists locally

(b). Moreover, if f is Lipschitz (or differentiable) near $y = y_0$, the local solution is unique

(c). If f is analytic near $y = y_0$, the unique local solution is analytic

(d). If $|f(y)| \leq Ky$ for some K constant, then there is a global solution

We will start looking at Matlab next week: we will specifically be looking for how it breaks. He doesn't cover as much content as other people do in 303, but more in depth. "I just became a grandfather last week, so I get to say back in my day we always had to make our own software." So we always knew how it broke. Now we have to figure out why other people's code breaks.

Recall last time, we were looking at $y' = Ay$, $y(t_0) = y_0$ where $y \colon \mathbb{R} \to \mathbb{R}$. Isocline picture: slopes at unit points in the $y - t$ plane. And then we guessed an answer last time: $y(t) = e^{A(t-t_0)}y_0$. Now, if we had a theorem that said this was unique, we'd be done.

We can solve this with integration: $y' = y \implies \frac{dy}{y} = dt \implies \int \frac{1}{y} dy = \int dt$, and so $\ln(y) + c_1 = t + c_2$ hence $y = e^{t+c}$. Now with t_0, y_0, we can determine our constant to get $y(t) = e^{t-t_0}y$. However, this method is not foolproof. Our solution could blow up to infinity. Consider $y' = y^2$ and $y(1) = 1$. And then $\frac{1}{y^2}dy = dt \implies \frac{-1}{y} = t + c$. Then, $y = \frac{-1}{t+c}$, and plugging in the initial conditions, we get $c = -2$. So $y = \frac{1}{2-t}$. Singularities can happen, as $y(t) \to \infty$ as $t \to 2$.

What about $y' = |y|^{1/2}$? What could possible go wrong? When $y > 0$ we have $y' = y^{1/2}$. Solving in the way we did before, we can get $y^{1/2} = \frac{1}{2}(t + C)$ so $y(t) = \frac{1}{4}(t + C)^2$. Note $y(-C) = 0$... but our slope should always be increasing, yet is 0 at a point? ff idk Now when $y < 0$, we can find (with the method as before) that $y = -\frac{1}{4}(t+C)^2$. So we have piecewise function of y depending on if $y > 0$ or $y < 0$.

Is this a unique solution? Actually, let $a < b$. Then

$$y(t) = \begin{cases} \frac{1}{4}(t-b)^2 & b \leq t \\ 0 & a \leq t \leq b \\ -\frac{1}{4}(t-a)^2 & t \leq a \end{cases}$$

is also a solution. The bad situation occurs when $y = 0$. We could stay "arbitrarily" long at $y = 0$, even if to the right and left are parabola! We will continue looking at $y' = |y|^{1/2}$ next time.

4 January 17

Today's outline:

- Euler's method
- MATLAB and Euler's method
- Plugging in $y' = Ay, y(t_0) = y_0$

4.1 Euler's Method

Say we are given $y' = f(t, y)$ (or $y' = f(y)$, or $\vec{y}' = \vec{f}(t, \vec{y})$, etc.) where f is a function, and we have) initial value $y(t_0) = y_0$, $y_0, t_0 \in \mathbb{R}$. We have the approximation

$$y'(t) \approx \frac{y(t+h) - y(t)}{h}$$

for small h, since $y'(t) = \lim_{h\to 0} \frac{y(t+h)-y(t)}{h}$. But

$$y'(t) \approx \frac{y(t+h) - y(t-h)}{2h}$$

is a much better approximation (usually). Rearranging gives $y(t+h) \approx y(t) + hy'(t)$; in fact, by Taylor's theorem, $y(t+h) = y(t) + hy'(t) + \frac{h^2}{2}y''(\xi)$ for some ξ between t and $t+h$. Substituting f, we have

$$y(t+h) = y(t) + hf(t, y) + \frac{h^2}{2}y''(\xi)$$

So for small h,

$$y(t+h) \approx y(t) + hf(t, y)$$

(we will ignore the final term for now, but will be useful later for calculating error). This last approximation is the essence of Euler's method.

So given t_0 = initial time and y_0 = initial value,

Step 1. Start with $y(t_0) = y_0$. [Pick a value of h, smaller the better (usually)]

Step 2. $y(t_0 + h) = y_0 + hf(t_0, y(t_0)) := y_1$

Step 3. $y(t_0 + 2h) = y_1 + hf(t_1, y_1) := y_2$ where $t_i = t_{i-1} + h = ih + t_0$.

Step n. Repeat

Actual ODE solvers (like in MATLAB) change h based off of f (especially makes h small when f is very large or f is changing quickly). When we saw the three body problem, we had error when two bodies got close because the gravitational force got so big it couldn't make h small enough to figure out what was going on.

If $y' = 2y$ and given y_0, t_0, recall that the exact solution was $y(t) = e^{2(t-t_0)}$. With the numerical approximation, we get $y_1 = y_0 + h(2y_0) = y_0(1 + 2h)$ and $y_2 = y - 1 + h(2y_1) = (1 + 2h)y_1$. So $y(t_i) = y(t_0 + ih) = (1 + 2h)^i y_0$. Now say we fix a t_{end}, perhaps N steps and so $t_{\text{end}} = t_0 + Nh$. Then $h = \frac{t_{\text{end}}-t_0}{N}$. So

$$y(t_{\text{end}}) = (1 + 2h)^N y_0 = \left(1 + \frac{2(t_{\text{end}} - t_0)}{N}\right)^N y_0 = \left(1 + \frac{\text{something}}{N}\right)^N y_0$$

Hmm... this looks a lot like a defition of e. As $N \to \infty$, we have $e^{\text{something}} y_0 = e^{2(t_{\text{end}}-t_0} y_0$. We will see this in MATLAB. On the homework, we will see MATLAB not working too well for $y' = |y|^{1/2}$.

And then he shows us his MATLAB for Euler's method. MATLAB is quirky. The first function you define in a file will always be run when you call the file, and it will assume the other functions in the file are not meant to be run globally.

5 January 24

Topics to finish:

- Look at MATLAB for a bit (Solutions to HW2 and HW1-ish)
- Vandermonde matrices and linear algebra without linear algebra
- Exponentiation and eigenpairs and norms
- Higher ordeer versions of Euler's method
- More celestial mechanics

Some quirky MATLAB: Something good to know is putting a `;` at the end of a line to supress the output (so we don't see giant vector). `1e-200` is defined as its own thing, but `1e-400` is considered `0`. `1/0` gives `Inf` and `-1/0` gives `-Inf`. But `-Inf + Inf` gives `NaN` (not a number). But all this follows the IEEE standard. MATLAB also calls the first function in a file as the name of a file when called elsewhere. We will look more at MATLAB later when we try to break it.

5.1 Vandermonde matrices

Given

$$A = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix}$$

Does this have a unique solution?

The following are equivalent:

1. $A \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}$ has a unique solution for all $y_0, \dots, y_n$.
2. A is invertible
3. $\det(A) \neq 0$
4. A is of rank $n+1$

We have actually seen Vandermonde matrices in HW1, just disguised. We were looking at 3 point schemes. We took a bunch of these formulas:

$$\begin{cases} f(x_0 + L) = \cdots \\ f(x_0) = \cdots \\ f(x_0 - h) = \cdots \\ f(x_0 + 2h) = \cdots \\ f(x_0 + \sqrt{2}h) = \cdots \\ f(x_0 + rh) = \cdots \end{cases}$$

(where r is a real number, often an integer, often $0, 1, 2, -1, -2, \dots$), and used the formula

$$\begin{aligned} c_0 \cdot f(x_0 + r_0 h) &= c_0 \left(f(x_0) + r_0 j f'(x_0) + r_0^2 \frac{h^2}{2} f''(x_0) + r_0^3 \frac{h^3}{3!} f'''(x_0) + O(h^4) \right) \\ c_1 \cdot f(x_0 + r_1 h) &= c_1(\dots) \\ c_2 \cdot f(x_0 + r_1 h) &= c_2(\dots) \\ c_3 \cdot f(x_0 + r_1 h) &= c_3(\dots) \end{aligned}$$

and we try to rearrange and change c_i and derive a scheme to get $0f(x_0)+1hf'(x_0)+0\frac{h^2}{2}f''(x_0)+0\frac{h^3}{3!}f'''(x_0)$. I.e., we are looking for c_0, c_1, c_2, c_3 such that

$$\begin{aligned}0 &= c_0 + c_1 + c_2 + c_3\\ 1 &= r_0c_0 + r_1c_1 + r_2c_2 + r_2c_3\\ 0 &= r_0^2c_0 + r_1^2c_1 + r_2^2c_2 + r_2^2c_3\\ 0 &= r_0^3c_0 + r_1^3c_1 + r_2^3c_2 + r_2^3c_3\end{aligned}$$

In homework 1, we saw this was solving a vandermonde matrix (anything that is of the form of the matrix below):

$$\begin{bmatrix}1 & 1 & 1 & 1\\ r_0 & r_1 & r_2 & r_3\\ r_0^2 & r_1^2 & r_2^2 & r_3^2\\ r_0^3 & r_1^3 & r_2^3 & r_3^3\end{bmatrix}\begin{bmatrix}c_0\\c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

(Observation: this looks like interpolation.)

Theorem 2. *The above system has a unique solution.*

Lemma 1. *Say that $p\colon \mathbb{R}\to\mathbb{R}$, $p = p(x)$ is a polynomial and p has $n+1$ distincr roots, $x_0, \dots, x_n$. Then*

1. *$p(x) = (x-x_0)(x-x_1)\dots(x-x_n)q(x)$ where $q(x)$ is a polynomial. (Bruh people in this class are actually so stupid, none of them recognize factoring).*
2. *$p'(x)$ has n distinct roots between x_0 and x_n (assuming $x_0 < x_1 < \cdots < x_n$).*

(a). $p''(x)$ has $n-1$ distinct roots between x_0 and x_n

(b). $p'''(x)$ has $n-2$ distinct roots between x_0 and x_n

To prove all of these, just done by Rolle's theorem

Theorem 3 (Rolle's Theorem)**.** *If $f \in C^0[a,b]$ (continuous on $[a,b]$), $f(a) = f(b) = 0$, and f is differentiable on (a,b), then $f'(\xi) = 0$ for some $a < \xi < b$.*

(Bro why is no one putting up their hand to haveing SEEN Rolle's theorem before, I swear there are too many bums in this course.) "How many people have seen a proof of this? Yeah, you would do this in Math 320." (Loewen could never) But this is quite intuitive anyway.

We do get a useful fact from the lemma:

Corollary 1. *If $p = p(x)$ is a polynomial of degree $\le n$, $p(x) = c_0 + c_1x + \cdots + c_nx^n$ and p has $n+1$ distinct roots, then $p = 0$*

Remark 2. $p(x) = x^2 + 1$, then p has no real roots. But $p'(x) = 2x$ has a real root, so the lemma doesn't go the other way (i.e. $q(x) = 0$ from bullet point 1.).

Claim: the system

$$\begin{bmatrix}1 & x_0 & x_0^2 & \cdots & x_0^n\\ 1 & x_1 & x_1^2 & \cdots & x_1^n\\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_n & x_n^2 & \cdots & x_n^n\end{bmatrix}\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix} = 0$$

has $c_1 = c_2 = \cdots = c_n = 0$ as its unique solution. Why? This just says $p(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$ satisfies $p(x_0) = 0, p(x_1) = 0, \dots, p(x_n) = 0$, so p has $n+1$ distinct roots, hence by our corollary, $p = 0$. (Important in the definition of a Vandermonde matrix is that $x_0, x_1, \dots, x_n$ are distinct, otherwise not invertible).

Homework: $e^{\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}t} = \begin{bmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{bmatrix}$ and $e^{\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}t} = \begin{bmatrix}\cosh t & \sinh t\\ \sinh t & \cosh t\end{bmatrix}$ (meant to get to this today).

6 January 26

Coming soon (later, we're putting a pin in it):

- More ODE's:
  - Central force: $m\vec{x}'' = -\frac{\vec{x}}{|\vec{x}|}g(|\vec{x}|)$,
  - Newton's law: $g(r) = \frac{1}{r^2}$
  - ff
- ff

Today:

- Higher Order ODE solvers. For now, Euler's method and Explicit trapezoidal
- More on e^{At}, norms, convergence, etc.s
- Where do non-diagnolizable matrices ceom from? One place: recurrences

If $y' = f(t, y), y(t_0) = y_0$, we can rewrite this as

$$y(t) - y_0 = \int_{s=t_0}^{s=t} f(s, y)ds$$

This is the integral form of "$y' = f(t, y)$". This is actually a perferred form for many applications, since we don't know if y is differentiable so we don't assume it. Rename: $f(s) = f(s, y)$ (or are we assuming??) Then $y' - f(t), y(t_0) = y_0$, then $y(t) - y(t_0) = \int_{s=t_0}^{s=t} f(s)ds$. We want to use a trapezoid to approximate the area under the curve, since it is a better approximation than a rectangle.

We have $y(t_0 + h) = y(t_0) + hy'(t_0) + O(h^2)$. Let $t_i = t_0 + ih$. We want y_1 to approximation $y(t_1)$. $y(t_1) = y(t_0 + h) \approx y(t_0) + hy'(t_0) + O(h^2)$. So $y_{i+1} = y_i + hf(t_0, y_0)$. So y_{i+1} approximates $y(t_{i+1})$ (uses y_i, approximates $y(t_i)$).

How do we improve? $y(t + h) = y(t) + h\frac{y'(t)+y'(t+h)}{2}$, which is a better approximation for y'. Trapezoidal: $\frac{y(t+h)-y(t)}{2}$ to $y'(t + \frac{h}{2})$. (I have no clue what he is doing.)

So first, we have $y(t_0), h$. Using Euler, we get $y(t_1) = y(t_0 + h) \approx y(t_0) + hy'(t_0) = y(t_0) + hf(t_0, y_0)$. $y_1 \leftarrow y(t_0) + hf(t_0, y_0)$.

Trapezoidal (Explicit) Scheme: Let $Y = y(t_0) + hf(t_0, y_0)$. And $y(t_0 + h) \approx y(t_0) + h\left(\frac{f(t_0,y_0)+f(t_1,Y)}{2}\right)$. The Y is like the otherside of the trapezoid.

6.1 Pre-interpolation

Before we get to interpolation, we need to get back at matrices, and what norms of matrices are.

6.1.1 Matrix exponential

Question: What do we mean by

$$e^A = I + A + \frac{A^2}{2} + \cdots$$

If $\vec{x} \in \mathbb{R}^n$, then $\|\vec{x}\|_2$ or $\|\vec{x}\|_{L^2}$ is $\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ (I feel like this should be ℓ^2).

Definition 2 (Matrix norm L^2). If $M \in \mathbb{R}^{n\times n}$, i.e. an $n \times n$ matrix, real entries, then we define

$$\|m\|_{L^2} = \max_{\vec{x}\neq\vec{0}} \frac{\|M\vec{x}\|_2}{\|\vec{x}\|_2}$$

So $\|M\|_{L^2}$ is the smallest real B such that $\|M\vec{x}\|_{L^2} \leq \|\vec{x}\|_{L^2} B$.

Example: $\vec{x} \in \mathbb{R}^n$, $n = 1$, $\vec{x} = x_1 \in \mathbb{R}^1$ so $A \in \mathbb{R}^1$. Say $A = \begin{bmatrix} 3 \end{bmatrix}$. $\max\left(\frac{\|A\vec{x}\|_{L^2}}{\|\vec{x}\|_{L^2}}\right)$. In this case, $\vec{x} = (x_1)$, so $\|\vec{x}\|_{L^2} = \sqrt{x_1^2} = |x_1|$. So $A \in \mathbb{R}^{|x|}$??? what is this notation, $A = \begin{bmatrix} a \end{bmatrix}$, so $\|A\|_{L^2} = |a|$. ff

We also have other norms: $\|\vec{x}\|_{\max} = \|\vec{x}\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$. $\|A\|_{L^\infty} = \max_{\vec{x} \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}$.

Now look at 2×2 matrices. Say

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$\text{size}(A) = \max(|a_{11}|, |a_{12}|, |a_{21}|, |a_{22}|)$. Claim: $\|A\|_\infty = \|A\|_2 = \text{size}(A)$, and so it doesn't matter what norm we pick.

What is $I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \cdots$. What does convergence mean? The rth term is $I + A + \frac{A^2}{2} + \cdots + \frac{A^r}{r!}$. So we want to measure convergence:

$$\lim_{s,r \to \infty, s > r} \left(\frac{A^{r+1}}{(r+1)!} + \cdots + \frac{A^s}{s!} \right) \to 0$$

(I'm assuming this is Cauchy). With a given metric, say 2, this means

$$\lim_{s,r \to \infty, s > r} \left\| \frac{A^{r+1}}{(r+1)!} + \cdots + \frac{A^s}{s!} \right\|_2 \to 0$$

(but it is the same with the other metrics). (I'm curious, how many of our properties from $\mathbb{R}^n$ carry over to these matrix spaces, like we always have equivalence of these L^2, L^∞ norms?). We have e^{At} where $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then computing our terms,

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
$$A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$A^3 = A^2 A = IA = A$$

So for every powers, it is I, and if it is an odd power, it is A. Then we have

$$e^{At} = I + At + \frac{(At)^2}{2!} + \frac{(At)^3}{3!} + \cdots = \begin{bmatrix} 1 + \frac{t^2}{2} + \frac{t^4}{4!} + \cdots & t + \frac{t^3}{3!} + \frac{t^5}{5!} + \cdots \\ t + \frac{t^3}{3!} + \frac{t^5}{5!} + \cdots & 1 + \frac{t^2}{2} + \frac{t^4}{4!} + \cdots \end{bmatrix} = \begin{bmatrix} \cosh(x) & \sinh(x) \\ \sinh(x) & \cosh(x) \end{bmatrix}$$

We'll do more diagonalization things next time.

7 January 29

Today:

- e^{At} where $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
- Other reasons to write $A = S \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} S^{-1}$ other than e^{At}, namely recurrences and $A^n, n = 0, 1, \ldots$.
- Examples of
  (a). Diagonlizable A
  (b). Defective A (in recurrences) ("not enough eigenvectors")... (not diagonlizable, because it does not have a complete basis of eigenvectors, some duplicates)

What is a 2×2 defective matrix? Means a (real) matrix that is not diagonliazble over $\mathbb{C}$. Diagonalizable:

$$A = S \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} S^{-1}$$

hence

$$f(A) = S \begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix}$$

The rotation matrix will have real entries, but complex (unit length) eigenvalues. So it isn't defective.

While with projection onto 2-dimensional plane, you'll have eigenvalue 1 with multiplicitiy 2.

So we have gotten some intuition of eigenvalues. Greater than 1, it is stretching it along the corresponding eigenvector, and less than 1, it is shrinking it. So iterated powers either make it go to infinity, or to 0.

7.1 Recurrence Relations and Finite Precision

7.1.1 Fibonacci numbers:

The Fibonacci numbers are defined as $F_{n+2} = F_{n+1} + F_n$ with the initial conditions $F_1 = F_2 = 1$. So our sequence is $1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$. We can also extend this backwards for $F_0, F_{-1}, F_{-2}, \ldots$ with the recurrence relation $\ldots, -8, 5, -3, 2, -1, 1, 0, 1, 1, \ldots$; as you can see, backwards follows a similar pattern.

We can use some linear algebra here. Since, we have $F_{n+2} = F_{n+1} + F_n$ and $F_{n+1} = F_{n+1}$, we can write

$$\begin{bmatrix} F_{n+2} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}$$

We can iterate backwards until we get to our initial conditions, $\begin{bmatrix} F_2 \\ F_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} F_2 \\ F_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, so

$$\begin{aligned} \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} &= A \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} \\ &= A \cdot A \cdot \begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix} \\ &\vdots \\ &= A^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} \end{aligned}$$

See how this is similar to what we did with Euler's method, except the "dyanmics" dictate that we take the power instead. There is a strong connection to Euler's method, $\vec{y}' = A\vec{y}$. We have

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = A^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}$$

How to diagonlize, at 2×2 matrix? Recipe: look for $\vec{v}$ such that there is some λ where $A\vec{v} = \lambda\vec{v} = \lambda \cdot I\vec{v}$. So $(A - \lambda I)\vec{v} = 0$. So we look for λs such that $A - \lambda I$:

(a). has non-zero vector in its nullspace

(b). $\det = 0$

(c). rank $< n$ (for $n \times n$)

(d). not invertible

(e). etc. (your favourite condition)

We can solve (obviously not the move for higher dimensions):

$$\det\left(\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix} - \begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}\right) = 0 \implies \det\begin{bmatrix}1-\lambda & 1\\ 1 & -\lambda\end{bmatrix} = 0$$

So we are solving $(1-\lambda)(-\lambda) - (1)(1) = 0$, or $\lambda^2 - \lambda - 1 = 0$. The solutions to this are the golden ratio and its "conjugate": $\lambda = \frac{1\pm\sqrt{5}}{2}$. So given that $\lambda = \frac{1+\sqrt{5}}{2}$, $\vec{v}$ such that

$$\begin{bmatrix}1-\frac{1+\sqrt{5}}{2} & 1\\ 1 & -\frac{1+\sqrt{5}}{2}\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix} = 0$$

So we want $(1-\frac{1+\sqrt{5}}{2})v_1 + v_2 = 0$. If we take $v_2 = 1$, then $\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}?\\ 1\end{bmatrix} = \frac{1+\sqrt{5}}{2}\begin{bmatrix}?\\ 1\end{bmatrix}$.

Let's look at a simpler recurrence. $F_{n+1} = 10F_n$. Then $F_n = 10^n F_0$. Just like an ODE, if we have an inital condition, reduces to e^{At} where A is a number. In the second order, our A is a matrix.

In our Fibonacci case, we can find $F_n = \left(\frac{1+\sqrt{5}}{2}\right)^n F_0$ and $F_n = \left(\frac{1-\sqrt{5}}{2}\right)^n F_0$. So we have $\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\lambda\\ 1\end{bmatrix} = \begin{bmatrix}\lambda^2\\ 1\end{bmatrix}$. So one solution is $F_n = c_1\left(\frac{1+\sqrt{5}}{2}\right)^n + c_2\left(\frac{1-\sqrt{5}}{2}\right)^n$.

This is more general than Fibonacci. If

$$\begin{bmatrix}F_1\\ F_1\end{bmatrix} = A\begin{bmatrix}\lambda\\ 1\end{bmatrix} = A\begin{bmatrix}F_1\\ F_0\end{bmatrix}$$

ff (I think using the fact that $F_1 = \lambda F_0$ or something.)

8 Feburary 2

Theorem 4. *Let $Fy' = f(t,y), y(t_0) = y_0$. Then there exists a (unique) solution if f is (Lipschitz) continuous.*

Proof is in Appedix A, but won't go into; it is quite long.

8.1 Recurrences

Theorem 5. *Let $f = f(n,x)$, i.e. $f\colon \mathbb{Z}\times\mathbb{R}\to\mathbb{R}$ and consider $x_{n+1} = f(n,x_n)$ for all $n \geq n_0$ and $x_{n_0} = x_0$. Then, there exists a unique solution $x_{n_0}, x_{n_0+1}, x_{n_0+2}, \dots$, i.e. $x\colon \{n_0, n_0+1, n_0+2, \dots\}\to\mathbb{R}$.*

Proof. Proof is trivial by induction: $x_{n_0+1} = f(n_0, x_{n_0}), x_{n_0+2} = f(n_0+1, x_{n_0+1}), \dots$ are all uniquely determined by f. □

This is one of the places where a existence/uniqueness proof is much easier. Compare this to the proof for ODEs.

Remark 3. By contrast, we can't go $x_{n_0}, x_{n_0+1}, \dots \to x_{n_0-1}$. It is not clear that $x_{n_0} = f(n_0-1, x_{n_0-1})$.

In the case of Euler's method however, this will work.

When we're solving an ODE, can intuitively think about taking little steps, and incremeneting by our f. But it is not clear what the exact relationship between ODE's and recurrences. Perhaps trapezoidal method for an ODE will give different recurrence than Euler's. But it is often that ODEs are the limit (for say Euler's method) of recurrences (depends on $h > 0$). Regardless, there is a lot that happens in common.

Reverse engineer: $(r-3)(r-2) \leftrightarrow x_n = c_1 2^n + c_2 3^n$. Then $r^2 - 5r + 6 = 0$, and so $(\sigma^2 - 5\sigma + 6)(x_n) = 0$, so $x_{n+2} - 5x_{n+1} + 6x_n = 0$.

So saw we want to solve $x_{n+2} - 5x_{n+1} + 6x_n = 0$. Guess (like with ODEs) maybe $x_n = r^n$ will be a solution for some r. Plugging in, we have $r^{n+2} - 5r^{n+1} + 6r^n = 0$. If $r \neq 0$, this holds for all $n \in \mathbb{Z}$ if and only if (???) $r^2 - 5r + 6 = 0 \implies (r-2)(r-3) = 0$. So $x_n = 2^n$ works, 3^n works. Check our work: $x_0 = 1$, $x_1 = 2$, $x_2 = 4$: $4 = 5\cdot 2 - 6\cdot 1$ works ($n = 0$), and $8 = 5\cdot 4 - 6\cdot 2$ works ($n = 1$). Similarly, for $x_n = 3^n$, so $x_n = c_1 2^n + c_2 3^n$.

Proposition 1. *For $x_{n+2} = 5x_{n+1} - 6x_n$ for any x_0, x_1, there is a unique solution.*

Proof. We have x_2 is a function of x_1, x_0, and x_3 is a function of x_2, x_1, and, etc. and $6x_n = 5x_{n+1} - x_{n+2} \implies x_n = (5x_{n+1} - x_{n+2})/6$. And so x_{-1} is a function of x_0, x_1. □

Another:

Proof. We know $x_n = c_1 2^n + c_2 3^n$ is a solution. So find c_1, c_2 that works:

$$c_1 2^0 + c_2 3^0 = x_0$$
$$c_1 2^1 + c_2 3^1 = x_1$$

or x_0, x_1 given, solve

$$\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

This is invertible, so solvable. Even more nice, this is a Vandermonde matrix! Will show up everywhere. □

Similarly, if we had $x_n = c_1 2^n + c_2 3^n + c_3 7^n$ via $(\sigma - 2)(\sigma - 3)(\sigma - 7)(x_n) = 0$ to solve for c_1, c_2, c_3 given x_0, x_1, x_2,

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 7 \\ 2^2 & 3^2 & 7^2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}$$

And get $x_{n+3} = \cdots x_{n+2} \cdots x_{n+1} \cdots x_n$.

Let's return to our recurrence $x_{n+2} = 5x_{n+1} - 6x_n$ or $x_{n+2} - 5x_{n+1} + 6x_n = 0$. So let $\vec{z}_n = \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix}$, $n \in \mathbb{Z}$.

$$\vec{z}_{n+1} = \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \vec{z}_n \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix}$$

Where we found the entries of our matrix from the relatoin above. We guessed $x_n = r^n$ for $(r = 2, 3)$ is a solution

$$\vec{z}_n = \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} r^{n+1} \\ r^n \end{bmatrix}$$
$$\vec{z}_{n+1} = \begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix} \vec{z}_n$$
$$\begin{bmatrix} r^{n+2} \\ r^{n+1} \end{bmatrix} = \begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} r^{n+1} \\ r^n \end{bmatrix}$$

Which is true if and only if $(r \neq 0)$

$$\begin{bmatrix} r \\ 1 \end{bmatrix} = \begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} r \\ 1 \end{bmatrix}$$

which is to say that $\begin{bmatrix} r \\ 1 \end{bmatrix}$ is an eigenvector. So if $r = 2, 3$ work:

$$2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$3 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Computing the eigenvalues of $\begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix}$: $\det(\lambda I - \begin{bmatrix} 5 & -6 \\ 1 & 0 \end{bmatrix})$. Which gives $0 = \lambda^2 - 5\lambda_6$.

Does this method always work? Consider $x_{n+2} = 4x_{n+1} - 4x_n$, so $x_{n+2} - 4x_{n+1} + 4x_n = 0$. We see then $0 = (\sigma^2 - 4\sigma + 4)(x_n) = (\sigma - 2)^2(x_n)$. Duplicate root. We guess $x_n = 2^n$: this will work, but what else will? There's no r^n for $r \neq 2$ $r^2 - 4r + 4 = (r - 2)^2 = 0$. What else might work? We will answer this on Monday.

9 February 5

3 Handouts:

- Intro to ODEs
- Recurrence Relations and Finite precision (you'll notice no mention of ODEs; from old course order)... some of the next homework problems are on here, and some have partial solutions too. We did Appendices A, B, and C
- Normal and Subnormal numbers

Today: A bit more on first two handouts, and start normal and subnormal numbers.

3-term recurrence relations: $x_{n+2} = ax_{n+1} + bx_n$, $a, b \in \mathbb{R}$, $\ldots, x_{-1}, x_0, x_1, \ldots$, $b \neq 0$ (then if $a \neq 0$, $x_{n+2} = ax_{n+1}$ which is just one term). This is actually an analog of second order ODE: $\vec{y}_{n+1} = A\vec{y}_n$ where $\vec{y}_{n+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}$. ff something I was busy on Piazza

Last time: $x_{n+2} - 4x_{n+1} + 4x_n = 0$, $(\sigma^2 - 4\sigma + 4)(x_n) = 0$. Guess $x_n = r^n$ is a solution... $r^2 - 4r + 4 = 0$, $(r-2)^2 = 0$, $r = 2, 2$. So $x_n = 2^n$ is a solution, and $x_n = c_1 2^n$ is also a solution. Think of it... $(r-2)(r-2-\varepsilon) = 0$. If $\varepsilon > 0$: $(\sigma - 2)(\sigma - 2 - \varepsilon) = 0$, then $x_n = c_1 2^n + c_2(2+\varepsilon)^n$. Fix $\varepsilon > 0$, have $x_0 = 3, x_1 1 = -40, x_2 = x_2(\varepsilon), x_3, \ldots$. So $x_{21} = (\cdot)x_0 + (\cdot)x_1$ given x_0, x_1. Taking $\varepsilon \to 0$, $x_{21}(0) = (\cdot)x_0 + (\cdot)x_1$.

What is the general solution to $x_{n+2} - 4x_{n+1} + 4x_n = 0$ if $x_n = c_1 2^n + c_2 n 2^n$ (like what we do with ODEs). Does this work? Well, $(n+2)2^n - 4(n+1)2^{n+1} - 4 \cdot 2^n = 0$. Try this: $(\sigma - 2)(\sigma - 2)(n2^n) = 0$? $(\sigma - 2)(n2^n) = (n+1)2^{n+1} \cdots 2(n2^n) = 2^n((n+1)2 - 2n) = 2^n(2)$ and $(\sigma - 2)(2^n \cdot \text{constant}) = \text{constant}(2^{n+1} - 2 \cdot 2^n) = 0$. What about $(\sigma - 2)^4(x_n) = 0$. $x_n = c_1 2^n + c_2 n 2^n + c_3 n^2 2^n + c_4 n^3 2^n 4$.

In general, we have for $p(n)$ a polynomial, $(\sigma - 2)(p(n)2^n) = 2^n \cdot$(polynomial of lower degree). We can verify for our case $(\sigma - 2)(n^k 2^n) = (n+1)^k 2^{n+1} - n^k 2^n \cdot 2 = 2^{n+1}((n+1)^k - n^k) = 2^{n+1}(kn^{k-1} + \text{lower})$. Hmm... kx^{k-1} looks like the derivative of x^k. When we interpolate a polynomial, take two nearby points and find the value in between, we are taking the derivative.

Now, the general solution to $(\sigma - r_1)^{m_1} \cdots (\sigma - r_k)^{m_k}(x_n) = 0$ is

$$x_n = p_1(n)r_1^n + \cdots + p_k(n)r_k^n$$

where $\deg(p_i) \leq m_k - 1$. E.g. $(\sigma - 2)^2(\sigma - 5)(x_n) = 0$ has the solution $x_n = c_1 2^n + c_2 n 2^n + c_3 5^n$.

Looking at $x_{n+2} - 4x_{n+1} + 4x_n = 0$, $\vec{y}_n = \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix}$, $\vec{y}_{n+1} = \begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix}$. ff

$\lambda^2 - 4\lambda + 4 = 0$, $\lambda = 2, 2$. We know then that we cannot have two linearly independent eigenvectors. When we have two eignevalues that are the same, cannot be linearly independent, otherwise $A = S \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} S^{-1} = 2SIS^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. (Why can't we do this?)

Look at $\lambda I - \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}$. ff We call these matrices deficient, defective, etc. We cannot diagonalize this matrix.

Since $2I - \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 4 \\ -1 & 2 \end{bmatrix} =: N$. We have $N \neq \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, but $N^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. We say that N is *nilpotent* if $N^k = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ for some $k \geq 1$.

So we have $2I - \big[\text{matrix of interest}\big] = N = \text{nilpotent}$. ff

10 February 16

Not Joel, the 302 guy, but he's one of the author's of the textbook.

Recall monomial interpolation: we are given two points, can easily find $c_0 + c_1 x$ such that it passes through both. If $(1,1), (2,3)$, then we let $1 = c_0 + c_1 \cdot 1$ and $3 = c_0 + c_1 \cdot 2$, and solve the linear system $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Then $p_1(x) = 2x - 1$. Similar happens if there is a third point and we use a degree 2 polynomial: $p_2(x) = c_0 + c_1 x + c_2 x^2$.

These are super-intuitive: if asked to find a polynomial, this is what you would do. But this method is not always the best. Some problems:

(1) Can't directly connect solution c_0, c_1 to the data points, not obvious what the connection is (this will be clearer when we talk about Lagrange points later today).

(2) But also, for an $n \times n$ system, our cost is $O(n^3)$ flops (floating point operations?). This is not that good, but not awful because our computers can evaluate a polynomial really fast. But it is high cost, because even though n is typically not that large, we often need to repeat many times and it becomes expensive (e.g. in machine learning, computing many polynomials). However, although the construction is expensive, the evaluation is cheap (can easily evaluate the polynomial at some other value).

(3) In MATLAB, we have the command `vander` to get the vandermonde matrix of ??? and `cond` to get that matrix's condition number. Now can compute something: the matrix is badly condition (which means inaccurate computations). Something like close to 10^{16} (particularly bad).

Recall the definition of the condition number K: $K(A) = \|A\|\|A^{-1}\|$. Suppose $Ax = b$, $\tilde{x}$ is a numerical solution. Then

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq K(A) \frac{\|b - A\tilde{x}\|}{\|b\|}$$

In our previous example, x is our solution $\begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$. In a common application, we can't compute x, and have $\tilde{x}$. And we can compute the residual $A\tilde{x}$. So we can compute $\frac{\|b-A\tilde{x}\|}{\|b\|}$. But if $K(A)$, we do not have any assurance that our $\tilde{x}$ is a good approximation of x.

Some facts about condition numbers: the smallest they can be is $K(A) = 1$, called perfectly conditioned. See $\|I\| = 1$. The worst possible is $K(A) = \infty$, this is for singular matrices.

10.1 Lagrange interpolation

Quite a nice method, very different from monomial. We are given the points $(x_j, y_j)_{j=0}^n$. Our monomial interpolation gave us

$$p_n(x) = \sum_{j=0}^{n} c_j \phi_j(x)$$

where $\phi_j(x) = x^{j-1}$. Lagrange interpolation gives

$$p_n(x) = \sum_{j=0}^{n} y_j L_j(x)$$

Notably now, we actually get the y_js showing up in the solution (remember the relation between the polynomial and the points?). How do we construct our L_js? We want

$$L_j(x_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Then $p_n(x_3) = y_3$, etc. We want to construct Lagrange polynomials that do this for us.

What is L_0 for our system from before $((1,1),(2,3),(4,3))$? We can make 2 and 3 roots, and have a coefficient out front to normalize. So $L_0(x) = a(x-2)(x-4)$, and $L_0(1) = 1 = a(1-2)(1-4) \implies a = \frac{1}{3}$. So in the arbitrary setting, we have

$$L_j(x) = \frac{1}{(x_j - x_0)\cdots(x_j - x_n)}(x - x_0)\cdots(x - x_{j-1}(x - x_{j+1})(x - x_n)$$

What is the linear system we solved? It was the identity matrix. What was the cost of construction? Well, we did n multiplications n times, and so there were $O(n^2)$ floating point operations. But the evaluation is more costly: now need to evaluate n^2.

Something about uniqueness, but didn't get to.

11 February 26

Plan:

- Review:
  - Relative error in double precision
  - Relative error and condition number

$$K_p(A) := \|A\|_p \|A^{-1}\|_p$$

- Compare
  - 10.2 Monomial Interpolation
  - 10.3 Lagrange Interpolation
- This week: 10.4 Divided Differences

Here is the basic problem of this course: CPSC 303 is not supposed to be 302. But things like condition numbers are so useful, that we want to use them, and will mainly see their application towards interpolation.

Chen is more like a numerical algebraist. When he sees condition numbers, particularly an example of a bad condition number, he gets excited. Joel really just sees it as a tool that something is going wrong.

Remark 4. Say you have scientific notation, base 10, remembering 4 digits: $1.00 \times 10^{71}, 1.001 \times 10^{71}, \ldots, 9.999 \times 10^{71}, 1.000 \times 10^{72}$. So for something with true value 3.217569×10^{71}, the scienticific notation up to 4 places is 3.218×10^{71}, which gives the relative error

$$\frac{|(\text{true value}) - (\text{sci notation})|}{|\text{true value}|} = \frac{|3.217569 - 3.218|}{|3.217569|} \leq \frac{|0.0005|}{3.217569|}$$

So the maximum size of the relative error is

$$\frac{|\frac{1}{2} \cdot 10^{-3}|}{|1.000|} = \frac{1}{2} \cdot 10^{-3}$$

Similar to in the remark, the error in double precision for a standard number:

$$\pm 1.b_1 \ldots b_{52} \cdot 2^m \quad (-1022 \leq m \leq 1023)$$

has maximum relative error

$$2^{-52} \cdot \frac{1}{2} = 2^{-53} \approx 1.1 \times 10^{-16}$$

In ODE's and recurrences, this error compounds.

Let's say we are solving $A\vec{x}_{\text{true}} = \vec{b}_{true}$, but really $\vec{b}_{\text{true}} \rightsquigarrow \vec{b}_{\text{observed}} = \vec{b}_{\text{true}} + \vec{b}_{\text{error}}$. So what we are solving is $A\vec{x}_{\text{observed}} = \vec{b}_{\text{observed}} \neq \vec{b}_{\text{true}}$ (we have the true value of A though)... I think this is a definition for $\vec{x}_{\text{observed}}$. Then $\vec{x}_{\text{observed}} = A^{-1}\vec{b}_{\text{observed}}$. The condition number answers how bad does this relative error of $\vec{x}$ differ. Specifically: $1 \leq p \leq \infty$, using $\|\cdot\|_p$. The relative error in $\vec{b}$:

$$\text{Relative error in } \vec{b} = \frac{\|\vec{b}_{\text{true}} - \vec{b}_{\text{observed}}\|}{\|\vec{b}_{\text{true}}\|}$$

Then

$$\frac{\|\vec{x}_{\text{true}} - \vec{x}_{\text{observed}}\|}{\|\vec{x}_{\text{true}}\|} \leq C \frac{\|\vec{b}_{\text{true}} - \vec{b}_{\text{observed}}\|}{\|\vec{b}_{\text{true}}\|}$$

And we call the maximum C the condition number. We might ask, what is the max C, and when is this max attained?

Last week: $1 \leq p \leq \infty$, we said the worst C is

$$C = K_p(A) = \|A\|_p \|A^{-1}\|_p$$

Let's just believe this... Something something decoupled??? We can solve the error and true value seperately or something.

We will actually see in the homework that the condition number for a Vandermonde matrix is really bad. The homework gives a matrix where we can compute it, at least for one row (bottom righ entry???), which will help show how bad things can actually get. The monomial method relies on Vandermonde matrices, and so get a high error. We will see the Lagrange and divided differences have a much lower condition number.

Trick we will use: Recall $A\vec{x}_{\text{true}} = \vec{b}_{\text{true}}$ and $A\vec{x}_{\text{observed}} = \vec{b}_{\text{observed}}$. And so $A(\vec{x}_{\text{true}} - \vec{x}_{\text{obs}}) = \vec{b}_{\text{true}} - \vec{b}_{\text{obs}}$, and if we define $\vec{b}_{\text{error}} := \vec{b}_{\text{true}} - \vec{b}\text{obs}$, then $A\vec{x}_{\text{error}} = \vec{b}_{\text{error}}$.

So now we can express our problem in terms of the relative error in $\vec{x}$ with the relative error in $\vec{b}$ (looking for max C):

$$\frac{\|\vec{x}_{\text{error}}\|_p}{\|\vec{x}_{\text{true}}\|_p} \leq C \frac{\|\vec{b}_{\text{error}}\|_p}{\|\vec{b}_{\text{true}}\|_p}$$

We can now seperate our problem:

(1) What is the max constant C_1 such that

$$\|\vec{x}_{\text{error}}\|_p \leq C_1 \|\vec{b}_{\text{error}}\|_p$$

(2) What is the max constant C_2 such that

$$\frac{1}{\|\vec{x}_{\text{true}}\|_p} \leq C_2 \frac{1}{\|\vec{b}_{\text{true}}\|_p}$$

Where $\vec{b}_{\text{error}}$ is anything, and $\vec{b}_{\text{true}}$ is anything, so these are completely independent of each other! Probably the most remarkable thing in this course. On the homework, they will give a few 2×2 examples where this happens.

Looking at the error first, we have $\vec{x}_{\text{error}} = A^{-1}\vec{b}_{\text{error}}$, so

$$\|\vec{x}_{\text{error}}\|_p = \|A^{-1}\vec{b}_{\text{error}}\|_p \leq \|A^{-1}\|_p \|\vec{b}_{\text{error}}\|_p$$

Also, you can have $\vec{x}_{\text{error}}, \vec{b}_{\text{error}}$ non-zero, and

$$\|A^{-1}\vec{b}_{\text{error}}\|_p = \|A^{-1}\|_p \|\vec{b}_{\text{error}}\|_p$$

since

$$\|M\|_p := \max_{\vec{x} \neq 0} \frac{\|M\vec{x}\|_p}{\|\vec{x}\|_p}$$

Remark 5. To find such a pair $\vec{b}_{\text{error}}, \vec{x}_{\text{error}} = A^{-1}\vec{b}_{\text{error}}$, it is not so nice for $\|\cdot\|_2$, but not so bad for $\|\cdot\|_\infty$, since

$$\left\| \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right\|_\infty = \max(|a| + |b|, |c| + |d|)$$

So $\|\vec{x}\|_p \leq C_1 \|\vec{b}_{\text{error}}\|_p$ where C_1 can be as low as $\|A^{-1}\|_p$.

Now, we want the smallest possible C_2 such that

$$\frac{1}{\|\vec{x}_{\text{error}}\|_p} \leq C_2 \frac{1}{\|\vec{b}_{\text{error}}\|_p}$$

I.e. the smallest C_2 such that

$$\|\vec{b}_{\text{true}}\|_p \leq C_2 \|\vec{x}_{\text{true}}\|_p$$

But recall $\|\vec{b}_{\text{true}}\|_p$ ff So smallest C_2 is $\|A\|_p$, and $\|\vec{b}_{\text{true}}\|_P \leq \|A\|_p \|\vec{b}_{\text{true}}\|_P$ is attained with equality when

$$\|A\vec{x}_{\text{true}} = \|A\|_p \|\vec{x}_{\text{true}}\|_p$$

(it is stretching $\vec{x}_{\text{true}}$ out to its maximum).

So we have $C_1 = \|A^{-1}\|_p$ and $C_2 = \|A\|_p$. ff

12 February 28

For today:

- Monic interpolation (Section 10.2) versus Lagrange interpolation (Section 10.3)
- Divided differences (Sections 10.4 - 10.7) (See article: CPSC 303 Remarks on divided differences. Likely to be revised to include more comments on 10.4-10.7, since book doesn't even talk about it enough)

Homework 7 (not yet posted), will likely look at the following...

Monic interpolation: given data $(x_0, y_0), \dots, (x_n, y_n)$ and want to fit a polynomial $p(x)$ of degree $n+1$ (should be n???), specifically $p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n$, such that $p(x_i) = y_i$, $i = 0, \dots, n$. In 10.2, we use the Vandermonde matrix

$$\begin{bmatrix} 1 & x_0 & \cdots & x_0^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}$$

What could possible go wrong? Say $(x_0, y_0), (x_1, y_1)$, and $x_0 = 2, x_1 = 2 + 10^{-5}$. You just want $c_0 + c_1 x$, so to find c_0, c_1 compute halfway between: $p(2 + (10^{-5})\frac{1}{2})$, call this point x_{mid}. Our system is (??? why did he put this here)

$$\begin{bmatrix} 1 & 2 \\ 1 & 2 + 10^{-5} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \end{bmatrix}$$

but we can compute $p(x_{\text{mid}}) = \frac{y_0 + y_1}{2}$ since it is linear.

We can numerically solve our system: $A\vec{c} = \vec{y}$, where A is our matrix from above. We can solve $A^{-1} = \frac{1}{\det} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{(2+10^{-5})-2} \begin{bmatrix} 2+10^{-5} & -2 \\ -1 & 1 \end{bmatrix} = 10^5 \begin{bmatrix} 2+10^{-5} & -2 \\ -1 & 1 \end{bmatrix}$. It is clear to see that this matrix will have a very large condition number. We can compute the infinty norm (since it is easiest) to see

$$\left\| \begin{bmatrix} 1 & 2 \\ 1 & 2 + 10^{-5} \end{bmatrix} \right\|_\infty = \max(1 + 2, 1 + 2 + 10^{-5}) = 3 + 10^{-5}$$

and so ff some problem

But consider Lagrange interpolation: say fitting $(x_0, y_0), (x_1, y_1), (x_2, y_2)$. Note fixed x_0, x_1, x_2. We have

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}$$

satisfies $L_0(x_1) = (x_1 - x_1)(\text{etc.}) = 0$, $L_0(x_2) = 0$, and $L_0(x_0) = \frac{(x_0 - x_1)(x_0 - x_2)}{(x_0 - x_1)(x_0 - x_2)} = 1$ (note that this is still a polynomial, since the denominator is just a constant). So this is like the Dirac delta. Then, we have

$$p(x) = y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x)$$

Then, it is easy to compute that $p(x_0) = y_0, p(x_1) = y_1, p(x_2) = y_2$. Note that the L_i are all polynomials of degree 2.

We will see that Lagrange interpolation solves our problem. Let's look at the linear case $(x_0, y_0), (x_1, y_1)$. Then $p(x) = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0}$. This is the exact same polynomial as before with monic interpolation, and will be the same one we get with divided differences. What happens: $x_0 = 2, x_1 = 2 + 10^{-5}$, and want $p(x_{\text{mid}}) = p(2 + 10^{-5}\frac{1}{2})$. We compute it the way Lagrange suggests to do: compute each term seperately first:

$$p(x_{\text{mid}}) = y_0 \left(\frac{x_{\text{mid}} - x_1}{x_0 - x_1} \right) + y_2 \left(\frac{x_{\text{mid}} - x_0}{x_1 - x_0} \right)$$

The first term becomes $\frac{((2+10^{-5}\frac{1}{2}) - (2+10^{-5}))}{(2-(2+10^{-5}))}$. MATLAB, using double precision, has maximum relative error 10^{-16}. We can calculat $2 + 10^{-5}\frac{1}{2} = 2.000005$. then we are computing the difference $2.000005 - 2.00001$, each with at least 10^{-16} relative precision, so get $-0.000005 \pm 2 \cdot 10^{-16}$.

The point:

$$\frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)(\cdots(x_n-x_{n-1}}$$

and $x_0,\dots,x_n$ close: $x_0=2, x_1=2+\varepsilon, x_2=2+2\varepsilon,\dots$. And $x\to 2+\varepsilon/2$???

Even though the polynomials are mathematically the same, Lagrange interpolation is computed much better. Perhaps if you wanted to evaluate a ton of points, and they're not all close together, monic interpolation might be better. But this gives one siutation where Lagrange interpolatoin loses less precision than monic interpolation.

On the homework, will choose y_0, y_1 to be some irrational number, so that precision is forced to be lost. Some fractions in base 2 are exact (like 1/8), but all irrationals cannot be exact in any base.

Moving on, 10.4 - 10.7 and beyond is a much bigger story of Newton's divided differences.

13 March 4

Last time: quadratic curve fitting

$$p(x)=c_0+c_1(x-x_0)+c_1(x-x_0)(x-x_1)$$

such that

$$\begin{aligned} f(x_0)&=y_0=p(x_0)\\ f(x_1)&=y_1=p(x_1)\\ f(x_2)&=y_2=p(x_2)\end{aligned}$$

Data $(x_0,y_0),(x_1,y_1),(x_2,y_2)$. Imagine fitting $p(x)$ to $f(x)$. Instead of a monomial $p(x)=\hat{c_0}+\hat{c_1}x+\hat{c_2}x^2$, rather use $1,x,x^2$, we use $1,x-x_0,(x-x_0)(x-x_1)$.

Step 0: Adaptive: data (x_0,y_0) fit $p_0(x)$ with deg ≤ 0, i.e. a constant. Hence, $p_0(x)=y_0$.

Step 1: Use (x_1,y_1) to get a new polynomial $p_1(x)$ fit to the old points (x_0,y_0) and the new points (x_1,y_1). We want a constant c_1 for $p_1(x)=p_0(x)+c_1(x-x_0)$. At x_0, $p_1(x_0)=p_0(x_0)$. But at x_1, we require $y_1=p_1(x_1)=p_0(x_1)+c_1(x_1-x_0)$ and so

$$c_1=\frac{p_1(x_1)-p_0(x_1)}{x_1-x_0}=\frac{y_1-p_0(x_1)}{x_1-x_0}$$

Claim: $y_0=f(x_0)$, $y_1=f(x_1)$, then $c_1=f'(\xi)$, where ξ is on the interval containing x_0,x_1. Well, our formula for c_1 gives

$$c_1=\frac{y_1-p_0(x_0)}{x_1-x_0}=\frac{f(x_1)-f(x_0)}{x_1-x_0}$$

and hence by Mean Value Theorem (special case of Taylor's theorem), there's some ξ in the interval between x_0,x_1 where $c_1=f'(\xi)$.

Recall??? (Newton divided difference), we have

$$\underbrace{f[x_0]}_{\hat{c}_0}+\underbrace{f[x_0,x_1]}_{\hat{c}_1}(x-x_0)$$

Now embelish to go through $(x_2,f(x_2))=(x_2,y_2)$.

Next step: $p_1(x)=c_0+c_1(x-x_0)$. Our c_0 depends only on x_0,f, we'll call it $f[x_0]$, and c_1 depends only on x_0,x_1,f, s call it $f[x_0,x_1]$. Now we want

$$p_2(x)=p_1(x)+c_2(x-x_0)(x-x_1)$$

Since c_2 depends on x_0,x_1,x_2,f, we'll call it $f[x_0,x_1,x_2]$.

Theorem 6. *If I is an interval containing x_0,x_1,x_2 and f is twice differentiable on I, then $\xi\in I$ such that*

$$f[x_0,x_1,x_2]=\frac{1}{2}f''(\xi)$$

Proof. $g(x) = f(x) - p(x)$, which has zeeros at x_0, x_1, x_2. Rolle's theorem says that there is some ξ_1 between x_0, x_1 where $g'(\xi_1) = 0$, and some ξ_2 between x_1, x_2 where $g'(\xi_2) = 0$. Then, there is a point η between ξ_1, ξ_2 where $g''(\eta) = 0$.

$g(x) = f(x) + c_2x^2 + \text{lower order}$. Then $\hat{c}_2x^2 = \hat{c}_1x + \hat{c}_0 = c_2x^2 + \text{lower}$, i.e. $c_2(x-x_0)(x-x_1) + c_1(x-x_0) + c_0 = c_2x^2 + \text{lower}$. And so $(c_2x^2 + \text{lower})'' = (c_2 2x + \text{lower}) + (2c_2 + \text{lower})'' = f''(x)$.

We have $g''(\eta) = 0$, and so $(f-p)''(\eta) = 0$, hence $f''(\eta) = p''(\eta)$, but this is equivalent to $f''(\eta) = 2c_2$. □

This is really a theorem about the highest order coefficient in polynomial interpolation (regardless of monomial, divided differences, lagrange, etc.)... I think when he said this, he was refering the start of the second paragraph, saying we could represent as such???

Hence, $p(x) = f[x_0] + f[x_0, x_1](x-x) + f[x_0, x_1, x_2](x-x_0)(x-x_1)$. We do have a slick trick to attain a formula for $f[x_0, x_1, x_2]$. The Lagrange formula: for $p(x)$, we have

$$y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}$$

which is a degree 2 polynomial, where $x_0 \mapsto 1, x_1 \mapsto 0, x_2 \mapsto 0$. We do the same for y_1, y_2. (Somehow this is like an identity matrix.) So the x^2 coefficient in L_0 is $\frac{y_0}{(x_0-x_1)(x_0-x_2)}$. Collecting these for the y_1, y_2 terms, since $p(x) = c_2x^2 + \text{lower}$ that fits $(x_0, f(x_0)), (x_1, f(x_1)), (x_2, f(x_2))$, the highest coefficient is

$$c_2 = \frac{y_0}{(x_0-x_1)(x_0-x_2)} + \frac{y_1}{(x_1-x_0)(x_1-x_2)} + \frac{y_2}{(x_2-x_0)(x_2-x_1)}$$

Recall that $f[x_0, x_1] = f[x_1, x_0]$. Claim: $f[x_0, x_1, x_2]$ also doesn't depend on the order. We could look directly at the formula:

$$c_2 = \sum_{i=0}^{2} \frac{y_i}{(x_i-x_0)\dots(x_i-x_n)}$$

(without $(x_i - x_i)$ in the denominator), and we can just rearrange the sum. But also, we are finding unique degree 2 polynomials $p(x)$ such that $p(x_0) = f(x_0)$, $p(x_1) = f(x_1)$, and $p(x_2) = f(x_2)$. This must be the same regardless of what order we choose. Since p doesn't care about the order, the coefficients can't either.

General remark: Monomial interpolation gave us $\hat{c}_0 + \hat{c}_1x + \hat{c}_2x^2$ and divided difference gave us $c_0 + c_1(x-x_0) + c_2(x-x_0)(x-x_1)$.

Remark 6. Look at $1, x, x^2$ or $1, x-x_0, (x-x_0)(x-x_1)$, the span of either set of functions is the same.

We will show this on Wednesday, maybe.

13.1 March 6

- Midterm: March 15, Location: TBA
- No new homework, instead reivew/sample problems this week.
- Next week: Part of Monday and Wednesday is fielding questions

Midterm will be 50 minutes. Similar to his 421: some true/false, based mainly on the homework; and then some short problems that are doable in 20-30 minutes, if you keeping really on top of the homework. So you should be able to do the midterm in 35 - 40 minutes, it should not be a speed test (that's his fault then). Only question 1-4 will be marked on the homework this week (out of 6), and maybe not all of 2-4: TAs will try to get feedback on this assignments up by Wednesday so can review before midterm.

Today: 2 stories about Ch. 10. The first is partially review...

Divided differences: long story, we'll get to some of the story. The connecttion with polynomial interpolation is more recent. The classical story goes like this: say you are looking at the triangular numbers $1, 3, 6, 10, \dots$ (think of a right triangle of dots, and adding a new diagonal of dots each time). We can find the pattern by looking at the difference: these are $2, 3, 4, 5, \dots$. We can take the difference of the differences, which gives 1 for all of the terms. Then, the difference of all of these terms is 0. Hence, this must be a polynomial of degree 2 or less: "third derivative is 0", or more accurately, $(\sigma - 1)(y_n) = \sigma(y_n) - (y_n) = y_{n+1} - y_n$. What we mean when we say "difference" is

applying $\sigma - 1$; taking the second difference is $(\sigma-1)^2$, etc. From what we know about finite recurrences, we know that when $(\sigma-2)^3 y_n = 0$ has the solution $y_n = 2^n p(n)$ where $\deg p \leq 2$. Similarly, $(\sigma-1)^3(y_n) = 0$ means that $y_n = a = bn + cn^2$ (from ODE/recurrences).

Say that we have $y_0 = 1$, $y_1 = 3$, $y_3 = 10$, $y_4 = 15$ etc., but no y_2 (can also remove something like y_6 as well). Then we have some not equally spaced points: taking the differences gives $2, 7, 5, 6, 15$. We have $t_0 = 1, t_1 = t_2 = 4, t_3 = 5, t_4 = 6, t_5 = 8$ with corresponding $y_0 = 1, y_1 = 3, y_2 = 10, y_3 = 15, y_4 = 21, y_5 = 36$. What do we do to fill in the missing data (the missing time steps). (One way is polynomial interpolation, but he didn't want this.) We could divide the y value by the time steps:

$$f[t_i, t_{i+1}] = \frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}$$

Which gives us $2, 7/2, 5, 6, 15/2$ as our values of $f[t_0,t_1], f[t_1,t_2], f[t_2,t_3], f[t_3,t_4], f[t_4,t_5]$. What happens if we find the differences of these? We are looking for the constant function. Remember, in interpolation

$$c_2 = f[x_0, x_1, x_2] = \frac{f[x_1,x_2] - f[x_0,x_1]}{x_2 - x_0}$$

For our purposes, we are computing $\frac{f[t_1,t_2]-f[t_0,t_1]}{t_2-t_0}$, etc. This gets us $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}$. And then regardless of the interval, the next divided difference is $0, 0, 0$. This is the story of Newton's divided difference. One could confirm that gravity is a quadratic. Whether you're computing the triangle numbers, or trying to find missing data, this works.

Theorem 7. *If given $(t_0, y_0), (t_1, y_1), \ldots$ and $y_i = p(t_i)$, where $\deg p \leq d$, then $(d+1)$st divided differences we are 0, when $f[t_i] = y_i, f[t_i, t_{i+1}] = \frac{f[t_{i+1}]-f[t_i]}{t_{i+1}-t_i}$ and*

$$f[t_i, t_{i+1}, t_{i+2}, \ldots, t_j] = \frac{f[t_{i+1}, \ldots, t_j] - f[t_i, \ldots, t_{j-1}]}{t_j - t_i}$$

The classical definition of divided differences (also in [A&G]): Given $f\colon \mathbb{R} \to \mathbb{R}$, $t_0, \ldots, t_n$ we define the Newton divided differences as above.

The following theorem is in [A&G], but not fully proven

Theorem 8. *If $p_{n-1}(x)$ with degree $n-1$ such that $p(x_0) = f(x_0), p(x_1) = f(x_1), \ldots, p(x_{n-1}) = f(x_{n-1})$ and you want $p_n(x)$ to fit $(x_0, f(x_0)), \ldots, (x_{n-1}, f(x_{n-1})), (x_n, f(x_n))$, then $p_n(x) = p_{n-1}(x) + (x-x_0)(x-x_1)\cdots(x-x_{n-1}) \cdot c_n$ and $c_n = f[x_0, x_1, \ldots, x_n]$ from the classical definition. So*

$$p_n(x) = f[x_0] + (x-x_0)f[x_0,x_1] + \cdots + (x-x_0)(x-x_1)\cdots(x-x_{n-1})f[x_0,\ldots,x_n]$$

So can either define $f[x_0, \ldots, x_n]$ to be the values such that it works, or show that the classical definition works here.

We want all the points to be distinct, but if we bring all the points together, we actually recover Taylor's theorem.

Something something, the second of these theorems can show the first: ff student remark in the notes.

14 March 8

- Remarks on exam
- Divided differences:
  (a). Finish proof that $p(x) = f[x_0] + (x-x_0)f[x_0,x_1] + \cdots$ interpolates $(x_0, f(x_0)), \ldots, (x_n, f(x_n))$
  (b). Upper triangular systems
  (c). Denegenerity

On the exam: Feb. 15, here! Sample midterm problems are on course website. Joel will take questions the last half of class Monday and Wednesday. Can bring in 2-sides of one 8.5-11 sheet oof paper.

Some tricks / alternative methods: say we have the system $m\ddot{x} = -Cmx$ where $x\colon \mathbb{R} \to \mathbb{R}$ a function of t. We can multiply by $\dot{x}$. We get $m\ddot{x}\dot{x} = -Cmx\dot{x}$. Can simplify $2\ddot{x}\dot{x} = \frac{d}{dt}(\dot{x})^2$. But note that $m\ddot{x}\dot{x} = \frac{d}{dt}\left(\frac{1}{2}m(\dot{x})^2\right) =$ kinetic energy, hence the other side must be the potential energy. If our right hand side had been x^5, now see that $\frac{d}{dt}\left(\frac{1}{2}(\dot{x})^2\right) = \frac{d}{dt}\left(\frac{1}{6}x^6\right) = \frac{1}{6}(6x^5)\dot{x}$. Thus, $m\ddot{x} = x^5 \implies m\ddot{x}\dot{x} = x^5\dot{x} \implies \frac{d}{dt}\left(\frac{1}{2}m(\dot{x})^2\right) = \frac{d}{dt}\left(\frac{1}{6}x^6\right) \implies \frac{1}{2}m(\dot{x})^2 = \frac{1}{6}x^6 + C$. This gives us the equation for the system

$$\frac{1}{2}m(\dot{x})^2 - \frac{1}{6}x^6 = C$$

Hence, $\frac{1}{6}x^6$ is the potential energy of the system, and constant total energy C.

On the exam, probably just going to stick to matrix infinity norm (p-norms for the vectors are fair game). Recall that

$$\operatorname{cond}_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty$$

and

$$\left\| \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right\| = \max\left(|a| + |b|, |c| + |d|\right)$$

So we know

$$\left\| \begin{bmatrix} 3 & 7 \\ -1 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\| \leq \underbrace{\left\| \begin{bmatrix} 3 & 7 \\ -1 & 8 \end{bmatrix} \right\|_\infty}_{10} \left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_\infty$$

But also, we know that we can get equality. Either $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} 1/3 \\ -1/3 \end{bmatrix}$. This is the maximum amount that can be stretched.

This review will be continued on Monday

14.1 Divided differences

The other story. $p_{n-1}(x)$ fits $(x_0, f(x_0)), \ldots, (x_{n-1}, f(x_{n-1}))$. We add $(x_n, f(x_n))$. See we need $p_n(x)$ with degree $\leq n$. We just make $p_n(x) = p_{n-1}(x) + c_n(x - x_0)\cdots(x - x_{n-1})$. Hence

$$c_n = \left(\frac{p_n(x) - p_{n-1}(x)}{(x - x_0)\cdots(x - x_{n-1})} \right) \Bigg|_{x = x_n} = \frac{f(x_n) - p_{n-1}(x_n)}{(x_n - x_0)\cdots(x_n - x_{n-1})}$$

We can solve for c_n. It is very nontrivial.

To fit $(x_0, y_0) = (x_0, f(x_0))$, $p_0(x) =$ const $= f(x_0)$. p_0 iss of degree ≤ 0. We first $(x_1, f(x_1))$ with $p_1(x) = p_0(x) + c_1(x - x_0)$. When $x = x_1$, $p_1(x_1) = f(x_1) = f(x_0) + c_1(x_1 - x_0)$. Then $c_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$. Then we call that this $f[x_0, x_1]$.

The next step is $p_2(x) = p_1(x) + c_2(x - x_0)(x - x_1)$ fit with $(x_2, f(x_2))$. Then

$$\begin{aligned} c_2 &= \frac{p_2(x_2) - p_1(x_2)}{(x_2 - x_0)(x_2 - x_1)} \\ &= \frac{f(x_2) - p_1(x_2)}{(x_2 - x_0)(x_1 - x_0)} \\ &= \frac{f(x_2) - (f(x_0) + f[x_0, x_1](x_2 - x_1))}{(x_2 - x_0)(x_1 - x_0)} \\ &= \frac{f(x_2) - \left(f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x_2 - x_1)\right)}{(x_2 - x_0)(x_2 - x_1)} \end{aligned}$$

Where are we going with this? Will it even be nice?

We can try isolating $f(x_0) = y_0, f(x_1) = y_1, f(x_2) = y_2$. We have

$$= \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)} + \frac{-f(x_1)}{(x_1 - x_0)(x_2 - x_1)} + f(x_0) \text{ term } = \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_0)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + f(x_0) \text{ term}$$

There is a symmetry in our first two terms. It looks like we could write it as $\sum_{i=0,1,2} \frac{f(x_i)}{\prod_{j \neq i}(x_i - x_j)}$. This looks like it can be inferred from Lagrange interpolation. Here is the key insight, probably from Newton:

$$= \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = f[x_0, x_1, x_2]$$

This is a long calculation, and so won't prove the full formula. But this is our formula from Newton's divided differences. Will show a bit more in the notes. As mentioned before, we have as well

$$f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$$

On Monday, we will assume that this is true, and use it to prove Taylor's theorem. Apparently, Taylor originally proved using Newton's formula.

15 March 11

Today: topics related to divided differences

- $f[x_0, \ldots, x_n] = \frac{f[x_1, \ldots, x_{n-1}], f[x_2, \ldots, x_n]}{x_n - x_0}$ proven in "Remarks on Divided Differences (2024)"
- Change of basis: $1, x, x^2 \leftrightarrow 1, x - 1, (x - 1)(x - 2)$ and lower/upper triangular systems
- Error in interpolation, Chebyshev Interpolation (10.6)
- Degenerate Interpolation: Taylor interpolation, Hermite interpolation

Reminder; Midterm will be here, ESB 1012, on Friday

Recall last time, we showed the nontrivial fact

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$$

We can do this using Lagrange interpolation formulas. To compute $f[x_i, x_{i+1}, x_{i+2}]$, etc., for $x_0, \ldots, x_n$, the total time is roughly n^2 (can see: we are have a branching operation... isn't this log time??? in the textbook). We have

$$\sum_i f(y_i) \left(\frac{(x - x_0) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_n)} \right)$$

where the $\frac{x - x_i}{x_i - x_i}$ term is omitted. We have $O(n^2)$ floating point operations ("flops").

Using monomial, we use a matrix, and this is roughly $O(n^3)$ operations. Note that we get the exact same polynomial, but Lagrange is faster, and is also more accurate (as we saw directly on the homework).

One point about upper/lower traingular matrices and change of bases: Polynomials of deg ≤ 2 is $\{\alpha_0 + \alpha_1 x + \alpha_2 x^2\}$. Newton divided difference said: $x_0 = 1, x_1 = 3$, then the polynomials that satisfy this this with deg ≤ 2 are those polynomials with deg ≤ 2 such that $\{c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1)\} = \{c_0 + c_1(x - 1) + c_2(x - 1)(x - 3)\}$. Note that

$$1 = 1$$
$$x - 1 = x(1) + (1)(-1)$$
$$(x - 1)(x - 3) = x^2(1) + x(-4) + (1)(3)$$

Hence

$$\begin{bmatrix} 1 \\ x-1 \\ (x-1)(x-3) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & -4 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}$$

Since this is a lower triangular matrix, it is invertible. So we also can find

$$\begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & -4 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ x-1 \\ (x-1)(x-3) \end{bmatrix}$$

More generally, ff no he erased shit because people don't know matrices. Basically, the fact that the matrix is invertible means we can backwards solve our system to be in terms of the other basis. This helps us find

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & -4 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 4 & 1 \end{bmatrix}$$

We can multiply our matrices together to verify that they are inverses:

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The Lagrange basis for $x_0 = 1, x_1 = 3, x_2 = 5$ are the polynomials

$$\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}, \dots$$

ff Note that we can convert between a space of functions to another space of functions using matrices, if we consider them as vector spaces.

15.1 Interpolation Error

The next topic he wants to talk about with divided differences is how to get Taylor's theorem from divided differences. This is apparently how he originally derived Taylor's theorem. Here is the cute idea: divided differences give you an error in approximating a function.

Here is the idea: $f(x)$ agrees with polynomial $p_n(x)$ on data points $x_0, \dots, x_n$:

$$p_n(x) = c_0 + c_1(x-x_0) + \cdots + c_n(x-x_0)\cdots(x-x_n)$$

and on $x_0, \dots, x_n, x_{n+1}$ for

$$p_{n+1}(x) = p_n(x) + c_{n+1}(x-x_0)\cdots(x-x_n)$$

where $c_{n+1} = f[x_0, \dots, x_{n+1}]$. Now consider x_{n+1} to be a variable. We claim that this leads to Taylor's.

Newton's formula gives us that if $f(x_{n+1}) = p_{n+1}(x_{n+1})$, $f(x) = p(x)$, and ??? $f(x) = p_{n+1}(x) = c_0 + c_1(x - x_0) + \cdots = c_n(x-x_0)\cdots(x-x_{n-1}) + f[x_0, \dots, x_n, x](x-x_0)\cdots(x_0-x_n)$. We know

$$f[x_0, \dots, x_n, x] = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

where ξ is in the interval containing $x_0, \dots, x_n, x$.

Theorem 9. $f(x) = c_0 + c_1(x-x_0) + \cdots + c_n(x-x_0)\cdots(x-x_{n-1}) + \frac{f^{(n+1)}(\xi)}{n!}(x-x_0)\cdots(x-x_n)$

Tayor's theorem says that as $x_1, x_2, \dots, x_n \to x_0$, we have

$$f(x) = c_0 + c_1(x-x_0) + c_2(x-x_0)^2 + \cdots + c_n(x-x_0)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}$$

where each $c_i = \frac{f^{(i)}(\text{point in interval with } x_0, x_1, \ldots, x_i)}{(i+1)!}$ and as $x_1, \ldots, x_n \to x_0$, this givs $\frac{f^{(i)}(x_0)}{(i+1)!}$.

Something about how we can use this interpolate, and get Hermite interpolation. This leads into Chapter 11 with splines.

Then some stuff about practice question for exam on course website. Recall that for double precision, the smallest possible number is the subnormal number 2^{-1074}. The next smallest would be $2 \cdot 2^{-1074}, 3 \cdot 2^{-1074}$. Recall in homework we could find some crazy behaviour that cycles.

16 March 13

Today: 10.7 Hermite Interpolation. This will lead us to Ch. 11 on Splines (Piecewise Hermite Interpolation).

16.1 Hermite Interpolation

ff beginning

Write $p(2) = f(2)$ and $p(2+\varepsilon) = f(2+\varepsilon)$. Monomoial interpolation gives $p(x)$ of the form $p(x) = c_0 + c_1 x$, where $c_0 + 2c_1 = f(2)$ and $c_0 + (2+\varepsilon)c_1 = f(2+\varepsilon)$, which is given by solving

$$\begin{bmatrix} 1 & 2 \\ 1 & 2+\varepsilon \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} f(2) \\ f(2+\varepsilon) \end{bmatrix}$$

and the condition number of this matrix goes to ∞ as $\varepsilon \to 0$. But the linear calculation is: $1 + 2c_0 = f(2)$ and $1 + (2+\varepsilon)c_0 = f(2+\varepsilon)$. Then $c_0 = \frac{f(2+\varepsilon) - f(2)}{\varepsilon}$. So as $\varepsilon \to 0$, we actually get $c_0 \to f'(2)$.

Now, what if we used Hermite interpolation at 2 and 4 for the system $x_0 = 2, x_1 = 2+\varepsilon, x_2 = 4, x_3 = 4+\varepsilon$. Monomial interpolation would give us $p(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$ by solving the system

$$\begin{bmatrix} 1 & 2 & 2^2 & 2^3 \\ 1 & 2+\varepsilon & (2+\varepsilon)^2 & (2+\varepsilon)^3 \\ 1 & 4 & 4^2 & 4^3 \\ 1 & 4+\varepsilon & (4+\varepsilon)^2 & (4+\varepsilon)^3 \end{bmatrix} \vec{c} = \begin{bmatrix} f(2) \\ f(2+\varepsilon) \\ f(4) \\ f(4+\varepsilon) \end{bmatrix}$$

As $\varepsilon \to 0$, not clear would happen. We can probably try to simplify this like before, but we're looking for a more general principal.

Hermite interpolation is just regular interpolation, but we allow derivative agreement. In the case above, we would guess $p(2) = f(2), p'(2) = f'(2), p(4) = f(4), p'(4) = f(4)$. This gives us that

$$\begin{aligned} f(2) &= c_0 + c_1(2) + c_2(2)^2 + c_3(2)^3 \\ f'(2) &= c_1 + 2c_2(2) + 3c_3(2)^2 \\ f(4) &= c_0 + c_1(4) + c_2(4)^2 + c_3(4)^3 \\ f'(4) &= c_1 + 2c_2(4) + 3c_3(4)^2 \end{aligned}$$

To show that we always get a unique solution, we could do the matrix algebra method (we subtract row 1 to row 2 and divide by ε, etc.). But we can do our "linear algebra without doing linear algebra". Let's look at the homogeneous form of the linear system of the matrix:

$$\begin{aligned} 0 &= c_0 + c_1(2) + c_2(2)^2 + c_3(2)^2 \\ 0 &= c_1 + 2c_2(2) + 3c_3(2)^2 \\ 0 &= c_0 + c_1(4) + c_2(4)^2 + c_3(4)^2 \\ 0 &= c_1 + 2c_2(4) + 3c_3(4)^2 \end{aligned}$$

What would this mean for $p(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$? This would mean that $p(2) = p'(2) = p(4) = p'(4) = 0$. And so we get to use Rolle's theorem.

Theorem 10. *If $p(x)$ is a polynomial of degree ≤ 3, and $p(2) = p'(2) = p(4) = p'(4) = 0$, then p is the zero polynomial.*

Proof. Rolle's gives us some ξ such that $2 < \xi < 4$ and $p'(\xi) = 0$. But Rolle's again gives us that p'' has zeros between $2, \xi$ and $\xi, 4$. Similarly for p'''.

Note that $p'''(x) = 6c_3$. But $p'''(x)$ has a zero, hence, $c_3 = 0$. But $p''(x) = 2c_2$ has a zero, so $c_2 = 0$. Similarly, $c_1 = c_0 = 0$. □

So without doing any linear algebra at all, we know that our system will have a unique solution (just assume two distinct solutions, subtract to get the homogenous case).

We could also show this with divided differences: $f[2, 2+\varepsilon] = f'(\xi)$ where $2 \leq \xi \leq 2+\varepsilon$. Then $f[2,2] = \lim_{\varepsilon\to 0} f[2, 2+\varepsilon] = \lim_{\varepsilon\to 0} f'(\xi) = f'(2)$. Likewise, $f[4,4] = \lim_{\substack{x_0\to 4\\ x_1\to 4}} f[x_0, x_1] = f'(4)$. So $p(x) = f[2] + f[2, 2+\varepsilon](x-2) + f[2, 2+\varepsilon, 4](x-2)(x-(2+\varepsilon)) + f[2, 2+\varepsilon, 4, 4+\varepsilon](x-2)(x-2-\varepsilon)(x-4)$. As $\varepsilon \to 0$, if f' exists and is continuous, have the limit

$$p(x) = f[2] + f[2,2](x-2) + f[2,2,4](x-2)^2 + f[2,2,4,4](x-2)^2(x-4)$$

where

$$f[2, 2+\varepsilon, 4] = \frac{f[2+\varepsilon, 4] - f[2, 2+\varepsilon]}{4-2} \to \frac{f[2,4] - f[2,2]}{2}$$

as $\varepsilon \to 0$. And that's it for chapter 10.

16.2 Extraneous comments because of midterm

Chen made a comment that Lagrange interpolation is like the identity matrix. For monomial interpolation, given $\begin{bmatrix} 1 & 2 \\ 1 & 2+\varepsilon \end{bmatrix}$ the condition number is about $\frac{1}{\varepsilon} \cdot c$, since the determinant has an ε. And so the condition number is not good with close points. Lagrange interpolation is as good as it gets. We have $y_0 \frac{x-x_1}{x_0-x_1} + y_1 \frac{x-x_0}{x_1-x_0}$. Waving some hands, this is like $\alpha_0 \frac{x-x_1}{x_0-x_1} + \alpha_1 \frac{x-x_0}{x_1-x_0}$, which kinda gives

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$$

But... when $x_1 \to x_0$, this starts being weird, like... idk ff. We won't rigourosly prove sad.

17 March 18

Homework 8 to be assigned by Thursday, due Thursday March 28.

17.1 Chapter 11: Splines (Cubic and related)

Say we have $(x_i, f_i(x))$, we want $v(x)$ near $f(x)$. Say we are modelling the side profile of a car with with points in a plane. Note that the two parts at the hood and at the trunk have almost nothing to do with each other. We have an outline $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$ for n quite large, then $(x_0, y_0), (x_1, y_1)$ has little to do with (x_n, y_n). We observe: interpolation is a bad idea globally. Instead, we take cubic splines, C^2, bounded variation, etc. all these other ways to do it.

17.1.1 Approach

Have $f = f(x)$, and at points $x_0, \ldots, x_n$, $y_i = f(x_i)$, f pretty smooth, designed "locally", and you want $v = v(x)$ that "looks like" what f should look like.

Assume that f is measured exactly. We want $v = v(x)$ such that

(a). $v(x_i) = y_i$ for all $i = 0, 1, \ldots, n$ (why we don't use least-squares, we can't gaurantee it will go through all the points)

(b). v is as smooth as is reasonable

Answer (cubic spline): choose v so that

$$\mathcal{U} = \{u \in C^k(A, B) \colon u(x_i) = y_i, i = 0, \ldots, n\}$$

where k is some number of derivatives, and the energy for the second derivative

$$E_2(u) = \int_{A=x_0}^{B=x_n} (u''(x))^2 dx$$

is minimized at $v \in \mathcal{U}$ over all functions in $\mathcal{U}$.

There are some other ways to require the function be smooth (that's what minimizing energy is really doing):

$$E_1(u) = \int_{x_0}^{x_n} (u'(x))^2 dx$$

(this is called the Dirichlet integral) and also

$$\text{Length}(u) = \int_{x_0}^{x_n} \sqrt{1 + (u'(x))^2} dx$$

There are other measures that we don't mention. So why did we choose E_2? We'll do this trying not to go into calculus of variations. If we had chosen our length as the thing to minimize, we get something that looks like just connecting the dots (we do have restrictions that u is differentiable, but there is a way to smooth it at the points so infinitely differentiable in some limit, and looks similar to connecting the dots... this is what would be in a more thorough course in things like PDEs).

If we had chosen the Dirichlet integral as the thing to minimize, we claim we get u = piecewise linear, which is not good (again, locally smooth it so infinitely differentiable). This follows from the Calculus of Variations: say $v \in \mathcal{U}$ such that v goes through $(x_i, y_i), i = 0, \ldots, n$ and $E_1(u)$ is minimal. (We need to show that this exists...) take any $g \in C^\infty[x_0, x_1]$ such that $g(x_0) = 0, g(x_1) = 0$. Then $v + g$ also goes through $(x_0, y_0), (x_1, y_1), \ldots$. Now look at $v_\varepsilon(x) = v(x) + \varepsilon g(x)$. If v really minimizes the energy, this pertubation v_ε has a larger energy. So $E_1 \colon \mathcal{U} \to \mathbb{R}$ has $E_1(v) \leq E_1(v_\varepsilon)$. Perhaps we can get $E_1(v + \varepsilon g) \leq E_1(v) + \varepsilon$, then we get that our energy is going to be 0. Expanding:

$$\begin{aligned}
E_1(v + \varepsilon g) &= \int_{x_0}^{x_1} (v'(x) + \varepsilon g'(x))^2 dx \\
&= \int_{x_0}^{x_1} \left[(v')^2 + 2\varepsilon (v')(g') + \varepsilon^2 (g')^2\right] dx \\
&= E_1(v) + \varepsilon \int_{x_0}^{x_1} 2(v')(g') dx + \varepsilon^2 \int_{x_0}^{x_1} (g'(x))^2 dx \\
&= E_1(v) + \varepsilon \int_{x_0}^{x_1} 2(v')(g') dx
\end{aligned}$$

(the integral on the right evaluates to 0). So for the "best lowest energy" v, $\int_{x_0}^{x_1} 2(v')(g')dx = 0$. And this is true for any g such that $g(x_0) = g(x_1) = 0$. We can integrate this by parts to get

$$0 = \int_{x_0}^{x_1} 2(v')(g')dx = v'(x)g(x)\big|_{x_0}^{x_1} - \int_{x_0}^{x_1} v''(x)g(x)dx = \int_{x_0}^{x_1} v''(x)g(x)dx$$

And hence, $v''(x) = 0$ for all $x_0 < x < x_1$ (otherwise, we have some g bump function at a point but this integral is still 0, contradiction). And so this means that v is linear from x_0 to x_1. Existence is fine now: we just plug in the linear.

Next time: we will see that if $E_2(u)$ is minimized at $u = v$, then we will see that $\int v''g'' = 0$, and integration by parts gives $\int v^{(4)} g = 0$, which implies that v is a cubic polynomial. If you don't want to read [A&G] why cubic splines are what we are looking for, this is a classical motivation for it.

18 March 20

Today:

- Cubic splines, minimize $E_2(u) = \int_A^B (u''(x))^2 dx$ over $\mathcal{U} = \{u = \mathcal{C}^2[A,B] \colon u(x_i) = y_i, i = 0, \dots, n\}$.
- Boundary conditions: $\mathcal{U}^{\text{free}}, \mathcal{U}^{\text{clamped}}$, etc.

How is this different from interpolation from chapter 10? ff

Last time: $u \colon [A,B] \to \mathbb{R}$, $E_1(u) = \int_A^B (u'(x))^2 dx$ and fix $(x_0, y_0), \dots, (x_n, y_n)$. $\mathcal{U}^1_{A,B} = \{u \in \mathcal{C}^1[A,B] \colon u(x_i) = y_i, i = 0, \dots, n\}$. We essentially proved there is no $v \in \mathcal{U}^1_{A,B}$ such that $E_1(v) \le E_1(u)$ for all $u \in \mathcal{U}^1_{A,B}$ (what??? I thought we proved piecewise linear... unless this is a remark about how these functions are not in $\mathcal{C}^1$).

What we did was that if $u \colon [x_0, x_1] \to \mathbb{R}$ continuous differentiable, $u \in \mathcal{C}^1[x_0, x_1]$ and $u(x_0), u(x_1)$ given, and u minimizes $E_1(u)$ or $\text{Length}(u) = \int_{x_0}^{x_1} \sqrt{1 + (u'(x))^2} dx$, then $u''(x) = 0$ for all $x \in (x_0, x_1)$. So we have piecewise linear functions, and in general, it is not differentiable at each x_i.

The advantage with splines is that we get the desired property.

Beyond this course: consider the functions where the following integral exists

$$\int_A^B (u'(x))^2 dx$$

We need our function to be square integrable. This goes into Lebesgue theory. There is truly a space of functions $w^{1,2}[A,B]$ "Sobolev space" of the functions where $\text{Energy}_{1,A,B}, \text{Length}_{A,B}, \dots$ makes sense. We call $w^{k,p}[A,B] = \{u \in L^p[A,B] \colon u \text{ is weakly } k\text{-times differentiable}\}$. This is just a digression, this course isn't about functional analysis and measure theory, but might put some extra credit problems on the homework that hint towards this. For now, just forget this.

Definition 3. Let $(x_0, y_0), \dots, (x_n, y_n)$ be fixed, $A = x_0 < x_1 < \cdots < x_n = B$. Let

$$\mathcal{U} = \{u \in \mathcal{C}^2[A,B] \colon u(x_i) = y_i, i = 0, \dots, n\}$$

Let $E_2(u) = \int_A^B (u''(x))^2 dx$.

Then

(a). There is a $v \in \mathcal{U}$ on which $E_2 \colon \mathcal{U} \to \mathbb{R}$ is minimized

(b). For this v, $v^{(4)}(x) = 0$ for $x_i < x < x_{i+1}$

(c). Also (not in [A&G]), $v''(A) = 0, v''(B) = 0$

In other words, $v^{(4)}(x)$, $x_i < x < x_{i+1}$ means v is piecewise a cubic polynomial, i.e.

$$s_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

where $a_i, b_i, c_i, d_i \in \mathbb{R}$ for $i = 0, \dots, n$, and

$$v(x) = \Big\{ s_i(x) \quad \text{if } x_i \le x \le x_{i+1}$$

Let's believe the theorem, i.e. that twice differentiable functions are the biggest space we want to look at (ignoring the Sobolev space), then we can do linear algebra without linear algebra: We have $\{a_i, b_i, c_i, d_i\}$ for each of the n patches from A to B, and so there are $4n$ variables (parameters). To solve for $v(x)$, how many equations do we get such that

$$s_0(x_1) = y_1 = s_1(x_1)$$

and $s_0'(x_1) = s_1'(x_1)$ and $s_0''(x_1) = s_1''(x_1)$. To have the values of s_i match, i.e. $s_i(x_i) = y_i$ and $s_i(x_{i+1}) = y_{i+1}$, this imposes $2(n+1)$ conditions. Matching derivatives gives $2(n-1)$ conditions. The total number of linear conditions is then $2n + 2(n-1) = 4n - 2$??? what happened to second derivative condition

Theorem 11. *E_2 minimized when $v''(x_0) = 0$, $v''(x_n) = 0$. ff*

ff

Theorem 12. *If you fix $v'(x_0) = y_0'$, $v'(x_n) = y_n'$, you also look at*

$$\mathcal{U}_{2,A,B,y_0',y_n'} = \left\{u \in \mathcal{C}^2[A,B] \colon U(x_i) = y_i, i = 0, \dots, n, u'(x_0) = y_0', u'(x_n) = y_n'\right\}$$

Then there exists a unique $v \in \mathcal{U}_{2,A,B,y_0',y_n'}$ that minimizes E_2.

There are $4n$ variables a_i, b_i, c_i, d_i, $i = 0, \dots, n-1$, and so there are $4n$ equations, so we get a unique solution.

18.1 Questions after class

So the idea behind the Sobolev space stuff is that even though we can't ask for a $C^\infty[A,B]$ to minimize E_2, we can get a sequence of functions in $C^\infty[A,B]$ that converge weakly (weaker than pointwise, like equal almost everywhere? I think he said the integral of their difference is 0) to something that minimizes E_2... I should have asked, does this converge to something in C^∞? Probably not, right?

Also, we were talking about what conditions we place on u depending on our energy function. Apparently, the conditions only depend on what highest derivative order we have, so something like $\in e^{(f')^9}|f'|dx$ would still minimize to piecewise linear (do calculus of variations or something, $a + b\varepsilon + c\varepsilon^2$), because the conditions it places is not enough to ensure that we get derivatives at the points equal each other. But when f'' is involved, this gives us enough conditions to require some things out of u to get a nice solution that's continuous and differentiable.

Also lol, he was thinkking of making a bonus question on homework about taking a spline, and convolving it with something like an approximate identity (not his words) that is inn C^∞, and the convolution would be in C^∞ as well. And I think this would be done at each point?