Contraction Mapping Theorem, Continuity and compactness

ROB 501

Necmiye Ozay

- Newton Raphson
- Contraction mapping
- Continuity of functions
- Compactness

Plan

- Contraction Mapping Theorem
  - This will be our main result: If you can show that a function $\boxed{P}$ is a **contraction** mapping in some set S, then the following iterations with $x_0 \in S$ converge to a unique fixed point: $\boxed{x_{k+1}=P(x_k)}$
  - Why do we care/how to use it? Assume you have an iterative algorithm where P represents your update rule, e.g.,

```
while |xk- xk_new| > 0.0001
          xk = xk_new;
          xk_new = P(xk);
end while
```

if you can show that P is a contraction mapping, you know that this loop will terminate. Moreover, as you make 0.0001 smaller or as you increase the number of iterations, it will give you a better and better result rather than oscillating arbitrarily.

  - We saw Newton-Raphson last time
  - See the handout at the end of this lecture for another nice exercise on contraction mapping theorem

$h(x) = y$

ROB 501 Handout: Grizzle

Newton Raphson Algorithm

Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable, and satisfy

$$\boxed{\det\left(\frac{\partial h}{\partial x}(x)\right) \neq 0 \quad \forall x \in \mathbb{R}^n}$$

Problem: For $y \in \mathbb{R}^n$ fixed, find a solution of $\boxed{y = h(x)}$; i.e, find $x^* \in \mathbb{R}^n$ s.t. $y = h(x^*)$. We note that this is equivalent to $h(x^*) - y = 0$. In other words, we are looking for a root of the equation $h(x) - y = 0$,

Approach: Find a convergent sequence $x_k \to x^*$ such that

$$\boxed{\lim_{k\to\infty} h(x_k) - y = h(x^*) - y = 0}$$

that is, $x^* = \lim_{k\to\infty} x_k$ is a root of $h(x) - y = 0$

Idea: Write $x_{k+1} = \boxed{x_k + \Delta x_k}$. We want

$$h(x_{k+1}) - y = h(\overbrace{x_k + \Delta x_k}^{x_{k+1}}) - y \approx 0.$$

What should Δx_k look like ?

$h(x_k + \Delta_k) \approx h(x_k) + \frac{\partial}{\partial x} h(x_k)\Delta x_k + h.o.t.$

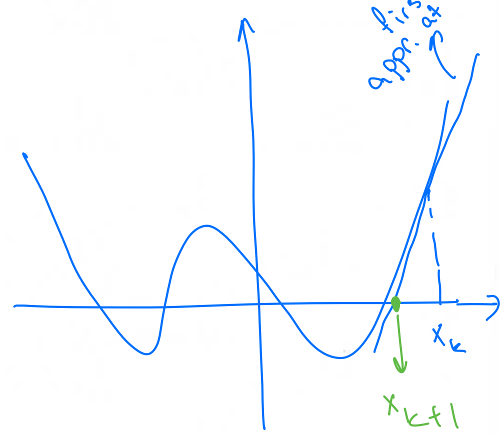
first order appr. at x_k

Apply Taylor's Theorem, to get

$$\boxed{h(x_k) + \frac{\partial h}{\partial x}(x_k)\Delta x_k} - y \approx 0$$

$$\therefore \frac{\partial h}{\partial x}(x_k)\Delta x_k \approx -(h(x_k) - y)$$

$$\Delta x_k \approx -\left(\frac{\partial h}{\partial x}(x_k)\right)^{-1} (h(x_k) - y)$$



Recalling that $x_{k+1} = x_k + \Delta x_k$, we arrive at Newton's Algorithm,

$$\boxed{x_{k+1} = x_k - \left(\frac{\partial h}{\partial x}(x_k)\right)^{-1}(h(x_k) - y)}$$

In practice, the change in x_k given by $\Delta x_k = -\left(\frac{\partial h}{\partial x}(x_k)\right)^{-1}(h(x_k) - y)$ is often too large. Hence, one uses the so-called Damped Newton Algorithm

$$\boxed{x_{k+1} = x_k - \epsilon\left(\frac{\partial h}{\partial x}(x_k)\right)^{-1}(h(x_k) - y)}$$

where $\epsilon > 0$ provides step size control!

Remark: Looking ahead to our discussion of contraction mappings, let's rewrite the algorithm as the iteration of a mapping $x_{k+1} = P(x_k)$

We want:

$$\boxed{P(x) := x - \epsilon\left(\frac{\partial h}{\partial x}(x)\right)^{-1}(h(x) - y)}$$

$h(x^*) = y$

$\Rightarrow P(x^*) = x^*$

A solution of $h(x) - y$ is a fixed point of $P(x)$. Indeed,

for any function P, if we have $P(x^*) = x^*$, then x^* is called a fixed point of P.

$$
\begin{gathered}
x^* = P(x^*) \\
\Updownarrow \\
x^* = x^* - \epsilon\left(\frac{\partial h}{\partial x}(x^*)\right)^{-1}(h(x^*) - y) \\
\Updownarrow \\
0 = -\epsilon\left(\frac{\partial h}{\partial x}(x^*)\right)^{-1}(h(x^*) - y) \\
\Updownarrow \\
0 = (h(x^*) - y).
\end{gathered}
$$

It can be shown that P is a local contraction on an open ball around a solution of $h(x) - y = 0$.

Example Find the solution to the coupled NL equations

$$0 = h(x) = \begin{pmatrix} x_1 + 2\,x_2 - x_1\,(x_1 + 4\,x_2) - x_2\,(4\,x_1 + 10\,x_2) + 3 \\ 3\,x_1 + 4\,x_2 - x_1\,(x_1 + 4\,x_2) - x_2\,(4\,x_1 + 10\,x_2) + 4 \\ \sin(x_3)^7 + \frac{\cos(x_1)}{2} \\ {x_4}^3 - 2\,{x_2}^2\,\sin(x_1) \end{pmatrix}$$

Initial Guess: $x_0 = \begin{bmatrix} 7 \\ 8 \\ 9 \\ 10 \end{bmatrix}$

We do 16 iterations of Newton's Algorithm (a nonlinear root finding algorithm) and we obtain:

$$x^* = \begin{pmatrix} -2.25957308738366677539068499960 \\ 1.75957308738366677539068499960 \\ 189.50954100613333978330549312824 \\ -1.68458069860197189523093013800 \end{pmatrix}$$ ✓

And the error is:

$$h(x^*) = \begin{bmatrix} 3.6734198 \times 10^{-39} \\ 2.9387359 \times 10^{-39} \\ 1.2765134 \times 10^{-38} \\ -2.5915832 \times 10^{-32} \end{bmatrix}$$ ✓

Rob 501 Handout: Grizzle
A Useful Cauchy Sequence in $(\mathbb{R}, |\cdot|)$

Proposition Let $0 \le c < 1$ and let (a_n) be a sequence of real numbers satisfying, $\forall\, n \ge 1$,

$$\underbrace{|a_{n+1} - a_n|}_{} \le c\underbrace{|a_n - a_{n-1}|}_{\Delta_n}.$$

"contracting"

Then (a_n) is Cauchy in $(\mathbb{R}, |\cdot|)$.

Proof:

Claim 1: $\forall\, n \ge 1$, $|a_{n+1} - a_n| \le c^n \underbrace{|a_1 - a_0|}_{\Delta_1}$.

Proof: First observe that $n=2$

$$|a_3 - a_2| \le c|a_2 - a_1| \le c^2|a_1 - a_0|.$$

Then complete the proof by induction. $\le c \cdot |a_1 - a_0|$ $n=1$

Claim 2: $\forall\, n \ge 1, k \ge 1$, $|a_{n+k} - a_n| \le \frac{c^n}{1-c}|a_1 - a_0|$.

Proof:

$$
\begin{aligned}
|a_{n+k} - a_n| &\le |a_{n+k} - a_{n+k-1} + a_{n+k-1} - a_{n+k-2} + \cdots a_{n+1} - a_n| \\
\text{triangular ineq.} \rightarrow &\le |a_{n+k} - a_{n+k-1}| + |a_{n+k-1} - a_{n+k-2}| + \cdots |a_{n+1} - a_n| \\
\text{claim 1} \rightarrow &\le c^{n+k-1}|a_1 - a_0| + c^{n+k-2}|a_1 - a_0| + \cdots + c^n|a_1 - a_0| \\
&\le c^n\Big(\sum_{i=0}^{k-1} c^i\Big)\,|a_1 - a_0| \\
&\le c^n\Big(\sum_{i=0}^{\infty} c^i\Big)\,|a_1 - a_0| \\
&\le c^n\Big(\frac{1}{1-c}\Big)\,|a_1 - a_0| \\
&\le \frac{c^n}{1-c}\,|a_1 - a_0|
\end{aligned}
$$

Claim 3: (a_n) is Cauchy

Proof: Consider m and n. WLOG, suppose $m \geq n$. If $m = n$, then $|a_m - a_n| = 0$. Thus assume $m = n + k$ for some $k \geq 1$. Then

$$|a_m - a_n| = |a_{n+k} - a_n| \leq \frac{c^n}{1-c}|a_1 - a_0| \xrightarrow[n\to\infty, m\to\infty]{} 0,$$

and thus it is Cauchy.

Remark: Because WLOG we could assume $m \geq n$, from $n \to \infty$, we have both $n \to \infty$ and $m \to \infty$.

$(X, \|\cdot\|)$ a normed space.

Def: Let $S \subset X$ and $T: S \to S$ a mapping (function)

(a) T is a contraction mapping if $\exists\, 0 \leq c < 1$ s.t. $\forall x, y \in S$ $\|T(x) - T(y)\| \leq c\|x - y\|$

(b) $x^* \in X$ is a fixed point of T if $T(x^*) = x^*$.

Contraction mapping theorem:

If $T: S \to S$ is a contraction mapping on a complete subset S, then $\exists$ unique $x^* \in S$ s.t. $T(x^*) = x^*$. Moreover, $\forall x_0 \in S$, the sequence $x_{k+1} = T(x_k)$, $k \geq 0$ is Cauchy and converges to x^*.

Proof: Let $x_0 \in S$, define $x_{n+1} = T(x_n) \ \forall n \geq 0$

Claim 1: (x_n) is Cauchy.

$$\|x_{n+1} - x_n\| = \| T(x_n) - T(x_{n-1}) \| \quad \text{since } T \text{ is a contraction mapping}$$
$$\leq c \| x_n - x_{n-1} \|$$

$\therefore$ From "useful Cauchy sequence" handout, we see (x_n) is Cauchy.

Because S is complete and (x_n) is Cauchy, $\exists \ x^* \in S$ s.t. $x_n \to x^*$.

Is it true that this x^* satisfies $T(x^*) = x^*$? Yes, we will prove this next.

$$\|T(x^*) - x^*\| = \|T(x^*) - x_n + x_n - x^*\|$$

$$= \|T(x^*) - T(x_{n-1}) + x_n - x^*\|$$

$$\leqslant \|T(x^*) - T(x_{n-1})\| + \|x_n - x^*\|$$

$$\leqslant \underbrace{c\|x^* - x_{n-1}\|}_{\xrightarrow[n\to\infty]{} 0} + \underbrace{\|x_n - x^*\|}_{\xrightarrow[n\to\infty]{} 0}$$

$$\leqslant 0 \qquad \therefore \; T(x^*) = x^*.$$

Is x^* the unique fixed point? Yes.

Proof: Suppose $y^* \in S$ s.t. $T(y^*) = y^*$. We will show that $y^* = x^*$.

$$\|x^* - y^*\| = \|T(x^*) - T(y^*)\|$$

$$\leqslant c\|x^* - y^*\|, \quad \text{for some } c \text{ s.t. } 0 \leqslant c < 1$$

$\therefore \|x^* - y^*\| = 0 \implies x^* = y^*$ (unique fixed point)

OPTIMIZATION:

1. How do we guarantee that functions have max and min? (Weierstrass theorem will give an answer, or rather some sufficient conditions)

2. If max or min exist, is it unique? (convexity will give an answer, or some sufficient conditions)

Rephrase Q1: $f: S \to \mathbb{R}$, can we guarantee $\exists\, x_* \in S$ and $x^* \in S$ s.t.

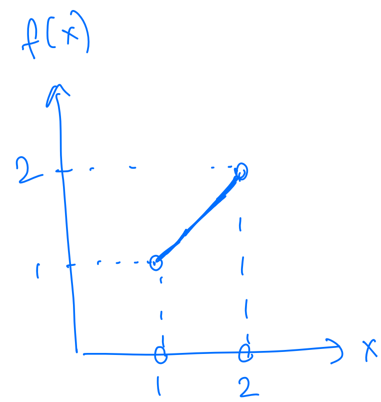
$$f(x_*) = \inf_{x \in S} f(x) \quad \text{and} \quad f(x^*) = \sup_{x \in S} f(x)$$

↳ minimum ↳ maximum

Let's see some examples!

$(X, \|\cdot\|) = (\mathbb{R}, |\cdot|)$

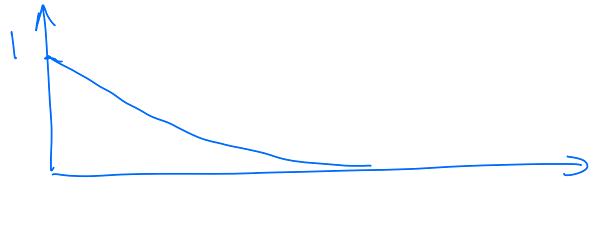
1) $S = (1, 2)$, $f: S \to \mathbb{R}$ by $f(x) = x$



→ f has neither a max nor a min in S (issue: S is not closed)

2) $S = [0, \infty)$, $f(x) = e^{-x}$

$\max_{x \in S} f(x) = 1$ with $x^* = 0$

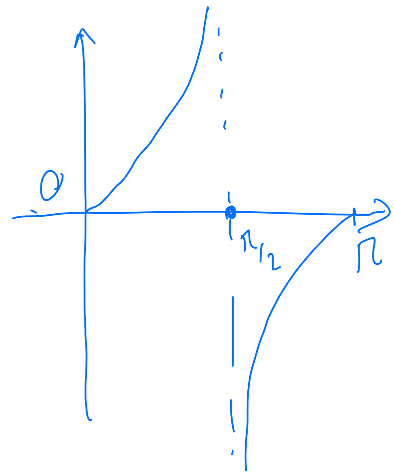


but minimum does not exist

(issue: S is unbounded)

3) $S = [0,\pi]$

$$f = \begin{cases} \tan(x), & x \neq \frac{\pi}{2} \\ 0, & x = \frac{\pi}{2} \end{cases}$$



has no max, no min

(issue: f is discontinuous)

<u>Def:</u> Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be two normed spaces. A function $f: \mathcal{X} \to \mathcal{Y}$ is <u>continuous at $x_0 \in \mathcal{X}$</u> if $(\forall \varepsilon > 0, \exists \delta(\varepsilon, x_0) > 0$ s.t. $\|x - x_0\|_{\mathcal{X}} < \delta \implies \|f(x) - f(x_0)\|_{\mathcal{Y}} < \varepsilon)$.

- f is continuous on $S \subset \mathcal{X}$ iff f is continuous $\forall x_0 \in S$.

$\forall x \in B_\delta(x_0)$, we have $\|f(x) - f(x_0)\|_{\mathcal{Y}} < \varepsilon$.

alternative way of writing this

<u>Exercise:</u> 1) Use this definition to prove $f(x) = x$, $f(x) = x^2$, $f(x) = |x|$ are continuous on $\mathbb{R}$.

2) Negate the defn. of continuous to get the defn. of discontinuous functions. (or discontinuity at a point x_0)

<u>Thm</u>: (from HW #11): Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be normed spaces and let $f : \mathcal{X} - \mathcal{Y}$ be a function:

a) If f is continuous at x_0, and (x_n) is a sequence s.t. $x_n \to x_0$, then the sequence $(f(x_n))$ in $\mathcal{Y}$ converges to $f(x_0)$. $[\, y_n = f(x_n),\ y_0 = f(x_0),\ y_n \to y_0 \,]$

b) If f is discontinuous at x_0, then $\exists\ (x_n)$ in $\mathcal{X}$ s.t. $x_n \to x_0$ but $f(x_n) \not\to f(x_0)$.

<u>Corallery</u>: f is continuous at $x_0 \iff$ every convergent sequence in $\mathcal{X}$ with limit x_0 is mapped by f in to convergent sequence in $\mathcal{Y}$.

<u>Def</u>: Let (x_n) be a sequence. Let $1 \leq n_1 < n_2 < \ldots$ be an infinite set of increasing integers. Then (x_{n_i}) is called <u>subsequence</u> of (x_n).

Note: $n_i \geq i$

<u>Lemma (exercise)</u>: Suppose $x_n \to x_0$ and (x_{n_i}) is a subsequence of (x_n). Then, $x_{n_i} \xrightarrow[i \to \infty]{} x_0$

Def: Let $(X, \|\cdot\|)$ be a normed space. Then, $C \subset X$ is compact if every sequence in C has a convergent subsequence with limit in C.

Remark: Often called sequential compactness.

Ex 1: A non-convergent sequence with no convergent subsequence:

(x_n) s.t. $x_n = n \Rightarrow (x_n) = 1, 2, 3, 4, 5, \dots$

$\rightarrow$ no convergent subsequence.

Ex 2: A non-convergent sequence with a convergent subsequence:

(x_n) s.t. $x_n = (-1)^n \Rightarrow (x_n) = -1, 1, -1, 1, -1, 1, \dots$

Take $n_i = 2i - 1$ $(x_{n_i}) = -1, -1, -1, \dots \rightarrow -1$.

Def: A set $S \subset X$ is bounded if $\exists r < \infty$ s.t. $S \subset B_r(0)$.

Fact 1) (Exercise) S is bounded $\Leftrightarrow \sup_{x \in S} \|x\| < \infty$

Fact 2) (Exercise) S is unbounded $\Leftrightarrow$ $\exists (x_k)$ s.t. $\forall k \geq 1$, $x_k \in S$, and $\|x_{k+1}\| \geq \|x_k\| + 1$, and $\forall p \geq 1$ $\|x_{k+p} - x_k\| \geq p$

Hint: $\|x_{n+p}\| \geq \|x_n\| + p$ and "reverse triangular inequality" $\left(\forall x, y \in \mathcal{X} \quad \|x-y\| \geq \big| \|x\| - \|y\| \big| \right)$

<u>Bolzano - Weierstrass Theorem</u>: (simple characterization of compactness in finite dimensional spaces).

In a finite-dim normed space $\mathcal{X}$, TFAE for a set $C \subset \mathcal{X}$:

(a) C is closed and bounded.

(b) C is sequentially compact.

$$\|A\|_2 = \sup_{x \in \mathbb{R}^n} \frac{\|Ax\|_2}{\|x\|_2} = \sup_{\|x\|_2 = 1} \|Ax\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$$

we can do this because norm is cont. and $\|x\|_2 = 1$ is compact

Rob 501 Handout: Grizzle
Cauchy Sequence Example and Contraction Mapping Theorem

Suggested Exercise: Suppose A is a square invertible matrix and we want to solve $Ax = b$. You know a few ways to do this, such as inverting A or using QR-factorization. Here, I will let you investigate another method via Contraction Mappings! Recall in the following that we assume A is invertible.

- Let's first note that the solution to $A^\top Ax = A^\top b$ is the same as that of $Ax = b$.
- We recall that $A^\top A > 0$ hence its e-values are all positive.
- Find the range of $\alpha > 0$ such that $-1 < \lambda_{\max}(I - \alpha A^\top A) < 1$. **Hint**: For any square real matrix M, e-values of $I+M$ satisfy: $\lambda_i(I+M) = 1+\lambda_i(M)$.
- **Exercise:** Recall from the SVD Handout, $\sqrt{\lambda_{\max}(M^\top M)}$ is the *induced 2-norm* of the matrix M. Prove that if M is real and symmetric, then $\sqrt{\lambda_{\max}(M^\top M)} = |\lambda_{\max}(M)|$.
- Define $P(x) := x - \alpha\big(A^\top Ax - A^\top b\big)$, for an α you found above.
- Check that $x^* = P(x^*) \Leftrightarrow A^\top Ax^* - A^\top b = 0$
- Choose random A and b with A invertible. Choose a random initial condition x_0. Define
$$x_{k+1} = P(x_k)$$
and check that the resulting sequence approaches a solution to $Ax = b$.
- Choose different values of α and see what you get.
- **Remark:** $||P(x) - P(y)||_2 \leq |\lambda_{\max}(I - \alpha A^\top A)|||x - y||_2$. Hence, you will see in Thursday's lecture that you are building a Cauchy Sequence when you choose α such that $0 \leq |\lambda_{\max}(I - \alpha A^\top A)| < 1$.