Taking norms on both sides:

$$|D_j f_i(\underline{u}) - D_j f_i(\underline{v})| \leq \sup_{\substack{\underline{b} \text{ on line} \\ [\underline{u}, \underline{v}]}} |\underbrace{D(D_j f_i(\underline{b}))}_{1 \times n \text{ matrix}}| \cdot |\underline{h}|$$

$$\leq \left(\sum_{k=1}^{n} (c_{i,j,k})^2 \right)^{1/2} |\underline{h}|.$$

Kantorovich's theorem: Given $f: \mathcal{U} \to \mathbb{R}^n$ differentiable.

Initial guess $\underline{a}_0 \in \mathcal{U}$ with $Df(a_0)$ invertible. Then

$$\underline{a}_1 := \underline{a}_0 - \underbrace{[Df(\underline{a}_0)]^{-1} \cdot f(\underline{a}_0)}_{\text{call this } \underline{r}}. \quad \text{Consider } B_{|\underline{r}|}(\underline{a}_1).$$

If Df is Lipschitz on $\overline{B_{|\underline{r}|}(\underline{a}_1)} \subseteq \mathcal{U}$ with Lipschitz ratio M and if $|f(\underline{a}_0)|\,|Df(\underline{a}_0)^{-1}|^2 \cdot M \leq \frac{1}{2}$, then

$$\{a_n\} \to \text{zero of } f \text{ in } \overline{B_{|\underline{r}|}(\underline{a}_1)} \quad \text{as } n \to \infty.$$

Proof of Kantorovich's theorem requires several lemmas:

① Show $Df(\underline{a}_1)$ is invertible, so can define $\underline{a}_2$

and "radius" $\underline{r}_1 = -(Df(\underline{a}_1))^{-1} \cdot f(\underline{a}_1)$.

② Show radii shrinking $|\underline{r}_1| \leq \frac{|\underline{r}_0|}{2}$, (so Lipschitz constant M still valid.)

③ Show other two components in triple : $|f(a_1)|\,|Df(a_1)^{-1}|^2 \leq$ are ~~shrinking~~ getting no bigger $|f(\underline{a}_0)|\,|Df(\underline{a}_0)^{-1}|^2$

this guarantees we can run our algorithm, and radii shrinking ensures that $\{a_n\}$ converge to some point.

Finally remains to bound outputs $f(\underline{a}_n)$. Show: $|f(\underline{a}_1)| \leq \frac{M}{2}|\underline{r}_0|^2$

(proof is 5 pages in Appendix.)

Explaining Kantorovich's Theorem:

Why does $|f(\underline{a}_0)| \cdot |[Df(\underline{a}_0)]^{-1}|^2 \cdot M \le \frac{1}{2}$?

arise in the statement of the theorem

Try to show $[Df(\underline{a}_1)]$ is invertible.

Intuition: Df is Lipschitz, so if $\underline{a}_1$ close to $\underline{a}_0$, then knowing $Df(\underline{a}_0)$ invertible (an assumption) should imply $Df(\underline{a}_1)$ invertible.

Plan: Show $[Df(\underline{a}_0)]^{-1} \cdot [Df(\underline{a}_1)]$ invertible,

hence $[Df(a_1)]$ invertible. (if B, $B^{-1}A$ invertible then $B \cdot B^{-1}A = A$ invertible)

proof is clever: Write

$$[Df(a_0)]^{-1}[Df(a_1)] = I_n - A \quad \text{for some matrix } A.$$

then use earlier fact that if $|A| < 1$, then $(I_n - A)$ invertible with inverse $\sum_{n=0}^{\infty} A^n$

Left to show: Why is $|A| < 1$?

$$A = I_n - [Df(\underline{a}_0)]^{-1}[Df(\underline{a}_1)]$$

$$= Df(\underline{a}_0)^{-1}\left([Df(\underline{a}_0)] - [Df(\underline{a}_1)]\right) \quad \text{so}$$

$$|A| \le |Df(\underline{a}_0)^{-1}| \, M \, |\underbrace{\underline{a}_0 - \underline{a}_1}_{= Df(a_0)^{-1} \cdot f(\underline{a}_0)}|$$

M: Lipschitz ratio

so $|A| \le \frac{1}{2}$

so done.

This implies we can give well-defined sequence.

Final improvements in Newton's method.

① If ~~[illegible]~~ inequality in Kantorovich's theorem is strict, then

(i.e. $|\underline{f}(\underline{a}_0)|\,|Df(\underline{a}_0)^{-1}|^2 \cdot M < \frac{1}{2}$. Call it $= k \in [0, \frac{1}{2})$

setting $c := \frac{1-k}{1-2k} |Df(a_0)^{-1}| \cdot \frac{M}{2}$, once we get

$|\underline{r}_n| = |-Df(a_n)^{-1} \cdot f(a_n)| \leq \frac{1}{2c}$, then $\{a_n\}$ superconverges!

Explicitly $|\underline{r}_{n+m}| \leq \frac{1}{c} \cdot \left(\frac{1}{2}\right)^{2^m}$

radius of ball around a_{n+m} which contains the limit of the sequence

really tiny number.

$m=1$: $1/4$

$m=2$: $1/16$

$m=3$: $1/256$

$m=5$: 9 decimal places

$m=6$: 19 decimal places

---

② Second improvement. Better norm.

"Euclidean norm" to measure size

$|\underline{a}| = \sqrt{a_1^2 + \cdots + a_n^2}$

(norm : assignment of non-negative length to a vector with nice properties: scalar mult., triangle inequality sends 0 to 0.)

There are other norms we could use:

"Operator norm"

$T : \mathbb{R}^n \to \mathbb{R}^m$ linear transformation

Euclidean norm on $\mathbb{R}^m$

$\|T\|_{op} := \sup \{ |T(\underline{x})| \mid \underline{x} \in \mathbb{R}^n \text{ with } |\underline{x}| = 1 \}$

equivalently $\quad \| T \|_{op} := \sup \left\{ \dfrac{|T(\underline{x})|}{|\underline{x}|} \;\middle|\; \underline{x} \in \mathbb{R}^n \setminus \{0\} \right\}$

So can replace all ~~operator~~ Euclidean norms with operator norms, since all pfs. use triangle ineq. + scalar mult.

Good part : $\quad \| A \|_{op} \leq |A|$, for all matrices A.

Bad part : $\quad \| A \|_{op}$ is harder to compute.

For example, in our map $f : \mathbb{R}^2 \to \mathbb{R}^2$ on Monday,

$$Df(\underline{a}_0)^{-1} = \begin{bmatrix} 0 & -1/2 \\ -1/4 & 0 \end{bmatrix}. \qquad |Df(a_0)^{-1}| = \sqrt{\frac{20}{64}} = \frac{\sqrt{5}}{4}.$$

What is $\| Df(\underline{a}_0)^{-1} \|_{op}$? Unit vectors in $\mathbb{R}^2$ parametrized by unit circle coordinates $\begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$.

$$\begin{bmatrix} 0 & -1/2 \\ -1/4 & 0 \end{bmatrix} \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} = \sup_t \sqrt{\tfrac{1}{4}\sin^2 t + \tfrac{1}{16}\cos^2 t}$$

maximize interior : $\left(\frac{1}{4}\sin^2 t + \frac{1}{16}\cos^2 t \right)' = \frac{1}{2}\sin t \cos t - \frac{1}{8}\sin t \cos t$

$$= \frac{3}{8}\sin t \cos t$$

$$= \frac{3}{16}\sin 2t$$

$\sin 2t = 0$ when $t = 0, \pi/2$ ...

Better : $\frac{1}{4}\sin^2 t + \frac{1}{16}\cos^2 t$

$$= \frac{1}{16} + \frac{3}{16}\sin^2 t$$

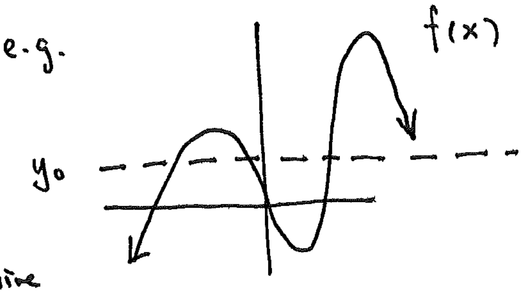
max : $1/4$. Get $\| \cdot \|_{op} = \boxed{\tfrac{1}{2}}$

When we solve linear equations $A \cdot \underline{x} = \underline{b}$, we essentially invert A, get $\underline{x} = A^{-1}\underline{b}$. (Do row reduction, which works more generally and is part way to calculating inverse)

If we have non-linear function $f : \mathbb{R}^n \to \mathbb{R}^n$, we can try to solve $f(\underline{x}) = \underline{b}$ by inverting f : $\underline{x} = f^{-1}(\underline{b})$ is our sol'n.

When is f invertible? (must be one-one, onto)

$f : \mathbb{R} \to \mathbb{R}$ e.g.



Is this invertible? "horizontal line test" to determine if f is one-one (unique input for each output)

Not invertible! multiple x's with same output yo.

But we can ask if it is invertible on smaller domain. Say between a max and a min. Then yes!

Really saying that f is invertible on sets $[a,b]$ for which $f' \neq 0$ on $[a,b]$. In terms of 1×1 linear transformations, the 1×1 matrix $[f'(c)]$ is invertible for all $c \in (a,b)$.