Taking norms on both sides:

\[ |D_j f_i(y) - D_j f_i(x)| \leq \sup_{b \text{ on line}} |b(D_j f_i(b))| \cdot |y - x| \]

\[ = \left( \sum_{k=1}^{n} (c_{ij,k})^2 \right)^{1/2} |b|_1. \]

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

Initial guess \( a_0 \in U \) with \( Df(a_0) \) invertible. Then

\[ a_1 := a_0 - [Df(a_0)]^{-1} f(a_0). \]

Consider \( B_{\|a_1\|} \).

If \( Df \) is Lipschitz on \( \overline{B}_{\|a_1\|} \), \( U \) with Lipschitz ratio \( M \) and if \( \|f(a_0)\| \cdot \|Df(a_0)^{-1}\|^2 \cdot M \leq \frac{1}{2} \), then

\[ \frac{g_n}{2^n} \to \text{zero of } f \quad \text{as } n \to \infty, \]

in \( \overline{B}_{\|a_1\|} \).
Proof of Kantrovich's theorem requires several lemmas:

1. Show \( Df(a_1) \) is invertible, so can define \( a_2 \)
   and "radius" \( r_1 = - \left( Df(a_1) \right)^{-1} \cdot f(a_1) \).

2. Show radii shrinking \( r_1 \leq \frac{r_0}{2} \), (so Lipschitz constant \( M \) still valid.)

3. Show other components in triple \( \frac{|f(a_1)|}{|Df(a_1)|^{-1}} \leq \frac{|f(a_0)|}{|Df(a_0)|^{-1}} \)
   are shrinking getting no bigger.

This guarantees we can run our algorithm, and radii shrinking ensures
that \( \{a_n\} \) converges to some point.

Finally remains to bound outputs \( f(a_n) \). Show: \( |f(a_n)| \leq \frac{M}{2} \left| r_0 \right|^2 \)

(proof is 5 pages in Appendix.)
Explaining Kantorovich's Theorem:

Why does \(|f(a_0)| \cdot |[Df(a_0)]^{-1}|^2 \cdot M \leq \frac{1}{2}\) arise in the statement of the theorem?

Try to show \([Df(a_1)]\) is invertible.

Intuition: \(Df\) is Lipschitz, so if \(a_1\) close to \(a_0\), then knowing \([Df(a_0)]\) invertible (an assumption) should imply \([Df(a_1)]\) invertible.

Plan: Show \([Df(a_0)]^{-1} \cdot [Df(a_1)]\) invertible, hence \([Df(a_1)]\) invertible. (If \(B, B^{-1}A\) invertible, then \(B \cdot B^{-1}A = A\) invertible)

Proof is clever: Write

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

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

Left to show: Why is \(|A| < 1\)?

\[A = I_n - [Df(a_0)]^{-1} [Df(a_1)]\]

\[= Df(a_0)^{-1} \left( [Df(a_0)] - [Df(a_1)] \right)\]

\[|A| \leq |Df(a_0)^{-1}| M |a_0 - a_1| \quad M: \text{Lipschitz rate}\]

\[= Df(a_0)^{-1} \cdot f(a_0)\]

So \(|A| \leq \frac{1}{2}\)
Final improvements in Newton's method.

1. If the inequality in Kantorovich's theorem is strict, then

\[ |f'(a_0)| \cdot |Df(a_0)^{-1}|^2 \cdot M < \frac{1}{2} \cdot \text{Call it } k \in \left(0, \frac{1}{2}\right). \]

Setting \( c := \frac{1 - k}{1 - 2k} \cdot \frac{M}{2} \), once we get

\[ |r_n| = \left| -Df(a_n)^{-1} \cdot f(a_n) \right| \leq \frac{1}{2c} \]

then \( \{a_n\} \) superconverges!

Explicitly

\[ |r_{n+m}| \leq \frac{1}{c} \cdot \left(\frac{1}{2}\right)^{2^m} \]

radius of ball around \( a_{n+m} \) which contains the limit of the sequence.


"Euclidean norm" to measure size

\[ |a| = \sqrt{a_1^2 + \cdots + a_n^2} \]

There are other norms we could use:

"Operator norm"

\[ T: \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ linear transformation} \]

\[ \|T\|_{op} := \sup_{x \in \mathbb{R}^n} \left| \frac{T(x)}{\|x\|} \right| \quad x \in \mathbb{R}^n \text{ with } \|x\| = 1 \]

(Euclidean norm on \( \mathbb{R}^m \))

(norm: assignment of non-negative length to a vector with nice properties:

scalar mult., triangle inequality)

such \( 0 \) to \( 0 \).
equivalently \( \| T \|_{\text{op}} := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|T(x)|}{|x|} \) 

So can replace all \( \ell_2 \) norms with operator norms, since all pos. use triangle ineq. + scalar mult.

Good part: \( \| A \|_{\text{op}} \leq |A| \), for all matrices \( A \).

Bad part: \( \| A \|_{\text{op}} \) is harder to compute.

For example, in our map \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) on Monday,

\[
Df(A_0)^{-1} = \begin{bmatrix} 0 & -\frac{1}{2} \\ -\frac{1}{4} & 0 \end{bmatrix}, \quad |Df(A_0)^{-1}| = \sqrt{\frac{20}{64}} = \frac{\sqrt{5}}{4}.
\]

What is \( \| Df(A_0)^{-1} \|_{\text{op}} \)? Unit vectors in \( \mathbb{R}^2 \) parametrized by unit circle coordinates \( \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} \).

\[
\begin{bmatrix} 0 & -\frac{1}{2} \\ -\frac{1}{4} & 0 \end{bmatrix} \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} = \sup_t \sqrt{\frac{1}{4} \sin^2 t + \frac{1}{16} \cos^2 t}
\]

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

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

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

\[= \frac{3}{16} \sin 2t \]

\[\max: \frac{1}{4}. \text{ Get } \| f \|_{\text{op}} = \frac{1}{2} \]

\[\sin 2t = 0 \text{ when } t = 0, \frac{\pi}{2}, \ldots\]
When we solve linear equations $A \cdot x = b$, we essentially invert $A$, get $x = A^{-1} b$.

If we have non-linear function $f: \mathbb{R}^n \to \mathbb{R}^n$, we can try to solve $f(x) = b$ by inverting $f$: $x = f^{-1}(b)$, which works more generally and is a good way to calculating inverse.

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)

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 $1 \times 1$ linear transformations, the $1 \times 1$ matrix $[f'(c)]$ is invertible for all $c \in (a, b)$. 