Remember $Df(x)$ is in $\mathrm{Mat}_{n\times n} \simeq \mathbb{R}^{n^2}$, so

$|Df(x) - Df(y)|$ is size of $n\times n$ matrix : $\sqrt{a_{11}^2 + \dots + a_{1n}^2 + \dots + a_{nn}^2}$

if $(a_{ij}) = [Df(x) - Df(y)]$

Easiest examples : Quadratic functions

$f: \mathbb{R}^2 \to \mathbb{R}^2$ then $Df(\underline{x}) = \begin{bmatrix} 6x_1 & -1 \\ -1 & 2x_2 \end{bmatrix}$

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} 3x_1^2 - x_2 \\ x_2^2 - x_1 \end{bmatrix}$

So $Df(\underline{x}) - Df(\underline{y}) = \begin{bmatrix} 6(x_1-y_1) & 0 \\ 0 & 2(x_2-y_2) \end{bmatrix}$ and

$$|Df(\underline{x}) - Df(\underline{y})| = \sqrt{36(x_1-y_1)^2 + 4(x_2-y_2)^2}$$

While $|\underline{x} - \underline{y}| = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$. Want M s.t.

$$\frac{|Df(\underline{x}) - Df(\underline{y})|}{|\underline{x} - \underline{y}|} \le M$$

What should M be?

One answer : $M = 6$ since

$|Df(\underline{x}) - Df(\underline{y})| \le 6|\underline{x} - \underline{y}|$.

Generally, only prove such an inequality holds on some restricted set.

Definition : We say Df satisfies the "Lipschitz condition" on a set V in $\mathbb{R}^n$ if $\exists M$ s.t. for all $\underline{x}, \underline{y} \in V$

M : Lipschitz ratio.

$$|Df(\underline{x}) - Df(\underline{y})| \le M|\underline{x} - \underline{y}|$$

Another method for finding Lipschitz ratios:

Bound all second partial derivatives. $\mathcal{U}$: open ball in $\mathbb{R}^n$

$f: \mathcal{U} \to \mathbb{R}^n$ twice differentiable ("C2")

If $\left| D_k D_j f_i(x) \right| \le c_{i,j,k}$ $\quad i,j,k \in [1,\dots,n]$

$\forall\, x \in \mathcal{U}$ $\quad$ (n^3 inequalities)

then

$$\frac{\left| Df(\underline{u}) - Df(\underline{v}) \right|}{\left| \underline{u} - \underline{v} \right|} \le \left(\sum_{i,j,k} (c_{i,j,k})^2 \right)^{1/2}$$

We expect partial derivatives to greatly simplify calculation of Lipschitz ratio. In our earlier example, $Df(\underline{X}) = \begin{bmatrix} 6X_1 & -1 \\ -1 & 2X_2 \end{bmatrix}$

so only non-vanishing second partials are

$$\underbrace{D_1 D_1 (f_1)}_{=6}, \quad \underbrace{D_2 (D_2 (f_2))}_{=2}$$

Lipschitz ratio $\sqrt{6^2 + 2^2} = \sqrt{40}$.

(we got slightly better constant, but with more thinking)

---

proof of above criterion:

(Mean Value Thm) For any $\underline{u}, \underline{v} \in \mathcal{U}$, consider the line joining them.

Write $\underline{u} = \underline{v} + \underline{h}$. Then

$$\left| Df(\underline{u}) - Df(\underline{v}) \right| = \left(\sum_{i,j} \left(D_j f_i(\underline{v}+\underline{h}) - D_j f_i(\underline{v}) \right)^2 \right)^{1/2}$$

But $D_j f_i(\underline{v}+\underline{h}) - D_j f_i(\underline{v}) \overset{\text{MVT}}{=} D D_j f_i(\underline{b}) \, |\underline{h}|$ for some $\underline{b}$ along $[\underline{v}, \underline{v}+\underline{h}]$

functions from $\mathbb{R}^n \to \mathbb{R}$

Taking norms on both sides:

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

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

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

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

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

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

Example: $f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^3 - x_2^2 + 4 \\ x_1^2 - x_1 x_2 + 1 \end{pmatrix}$

then $Df = \begin{bmatrix} 3x_1^2 & -2x_2 \\ 2x_1 - x_2 & -x_1 \end{bmatrix}$. If we pick $\underline{a_0} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

then $f\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ so $|f(a_0)| = \boxed{1}$

while $Df(\underline{a_0}) = \begin{pmatrix} 0 & -4 \\ -2 & 0 \end{pmatrix}$ with inverse $-\frac{1}{8}\begin{pmatrix} 0 & 4 \\ 2 & 0 \end{pmatrix}$

so $\left|Df(a_0)^{-1}\right|^2 = \frac{1}{64}(4^2 + 2^2) = \frac{20}{64} = \boxed{\frac{5}{16}}$

To find Lipschitz ratio for Df on $B_{|\underline{r}|}(\underline{a_1})$, we compute:

$$\underline{a_1} = \underline{a_0} - Df(a_0)^{-1} \cdot \left(f(a_0)\right) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \frac{1}{8}\begin{pmatrix} 0 & 4 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 2 \end{pmatrix}$$

and $\underline{r} = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$ with $|\underline{r}| = 1/2$. So need Lipschitz ratio on $B_{1/2}\begin{pmatrix} 1/2 \\ 2 \end{pmatrix}$

To find it compute second ~~partial~~ partial derivatives: Non-zero ones are:

$D_1 D_1 f_1 = 6x_1 \leftarrow$ bounded by 6 on $B_{1/2}\begin{pmatrix} 1/2 \\ 2 \end{pmatrix}$ where maximum of x_1 coord. is 1.

(and min of x_1 is 0)

$D_2 D_2 f_1 = -2$

$D_1 D_1 f_2 = 2$

$D_2 D_1 f_2 = -1$

$D_1 D_2 f_2 = -1$

Get ratio

$$M = \sqrt{6^2 + 2^2 + 2^2 + 1^2 + 1^2} = \sqrt{46}.$$

So we fail Kantorovich criterion.

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 $|\underline{r}_1| \leq \frac{|\underline{r}_0|}{2}$.

③ Show other components in triple : $|f(a_1)|\,|Df(a_1)^{-1}|^2 \leq$
are shrinking

$$|f(\underline{a}_0)|\,|Df(\underline{a}_0)^{-1}|^2$$

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.)