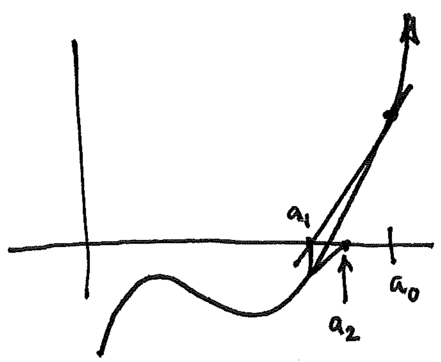
Newton's method: Finding roots of functions $f$.

In pictures: pick initial point $a_0$. Approximate $f$ by tangent line at $a_0$ : $f(x) \approx f(a_0) + (x-a_0) f'(a_0)$

Find where tangent line $= 0$ : $x = -\frac{f(a_0)}{f'(a_0)} + a_0$

Should be close to 0. Let above sol'n be called $a_1 = a_0 - \frac{f(a_0)}{f'(a_0)}$. Repeat.

$a_2 = a_1 - \frac{f(a_1)}{f'(a_1)}$, etc... Hope: $\{a_n\} \to$ root of $f$ as $n \to \infty$.




Big questions:

(1) What to pick for $a_0$?

(2) When can we certify that $\{a_n\} \to$ root ?

(3) How to generalize to multivariable function? ⟵ if $f: \mathbb{R}^n \to \mathbb{R}^n$ or from open set $U \subseteq \mathbb{R}^n \to \mathbb{R}^n$

(4) Possible bad situations. - initial choices $a_0$ which don't converge to a root.

then tangent hyperplane:

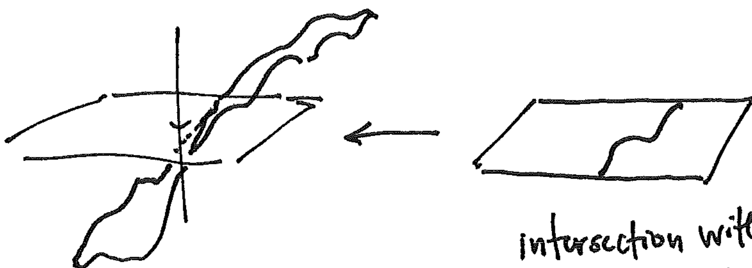
$$f(\underline{a_0}) + [Df(\underline{a_0})](\underline{x} - \underline{a_0}) \stackrel{?}{=} \underline{0}$$

Solve linear system. for $\underline{x} - \underline{a_0}$, then add $\underline{a_0}$ to find $\underline{x}$

For most $Df(a_0)$, its row reduction is $I_n$, so sol'n is unique.

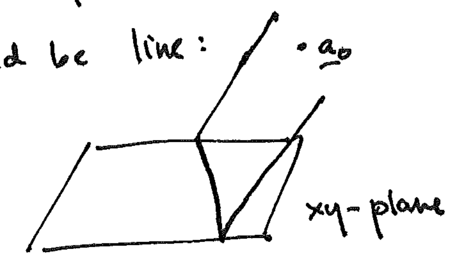
Why $\mathbb{R}^n \to \mathbb{R}^n$? Zeros (i.e. places where $f(\underline{x}) = \underline{0}$) should be sparse (special)

If $\mathbb{R}^2 \to \mathbb{R}$, then geometrically, asking where graph $(x, y, f(x,y))$ meets the $xy$-plane.



intersection with $xy$-plane should be curve.

Indeed if we tried Newton's method we'd get tangent plane. Zeros would be line:




Iterations always giving line, not a curve.

If you think about $\mathbb{R} \to \mathbb{R}^2$ get tangent line ~~equation~~ required to pass through $(0,0)$

(linear system with 2 equations, 1 unknown)

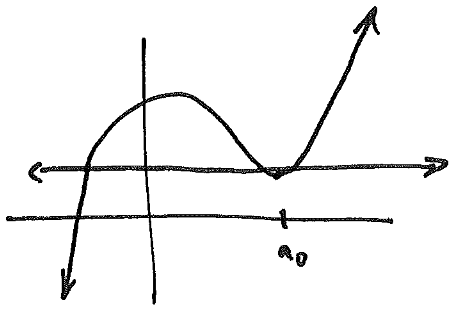
$$\begin{bmatrix} f_1(t_0) \\ f_2(t_0) \end{bmatrix} + \begin{bmatrix} f_1'(t_0) \\ f_2'(t_0) \end{bmatrix} (t - t_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

---

(1) What to pick for $a_0$? Often best to get as close to 0 with $f(a_0)$ as possible.
(use any extra info about where a 0 might be.)

(4) Pitfalls. – illustrate these in 1-var case.

(4a)



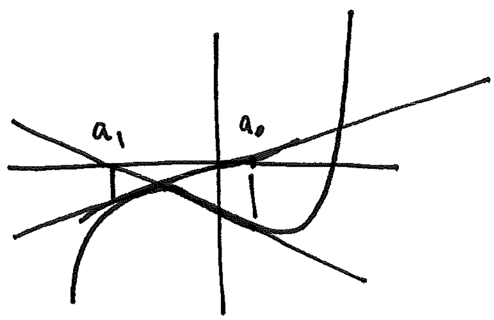

No sol'ns to tangent line.

i.e. $1\times 1$ matrix $f'(a_0)$ is not invertible (i.e. not reducing to identity in echelon form.)

Resolution: Pick another $a_0$.

(4b) Periodicity  $a_0 = a_2 = \dots$  $a_1 = a_3 = \dots$

never converge to 0 of function.




Book gives well chosen example of this.

$$f(x) = x^3 - x + \frac{\sqrt{2}}{2} \quad \text{with } a_0 = 0,\ a_1 = \frac{\sqrt{2}}{2}$$

in fact $a_0 = \epsilon$, small pos. $\epsilon > 0$ no better.

$a_2, a_4, \dots$ sequence tending to ~~$a_0$~~ 0

$a_1, a_3, \dots$ sequence tending to ~~$a_1$~~ $\frac{\sqrt{2}}{2}$

(i.e. heading toward "death spiral" of periodicity)

(2) When can we certify that $\{a_n\} \longrightarrow$ root of $f$ as $n \to \infty$?

Book has nice metaphor: if I show you pictures of a plane in flight, can you tell me when it will crash (i.e. height is function. When = 0)

Ans: Yes. Know position. First derivative from angle of plane in picture.

And: second derivative — plane is limited in how fast it can reverse ~~direction~~ change in height.

In multivariable setting, don't have second derivative as #. Stand in:

Bound (size of) secants in all directions: $\dfrac{|Df(x) - Df(y)|}{|x-y|} \le M$ : max.

Remember $Df(x)$ is in $Mat_{n\times n} \simeq \mathbb{R}^{n^2}$, 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

$$f : \mathbb{R}^2 \to \mathbb{R}^2 \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} 3x_1^2 - x_2 \\ x_2^2 - x_1 \end{bmatrix}$$

then $Df(\underline{x}) = \begin{bmatrix} 6x_1 & -1 \\ -1 & 2x_2 \end{bmatrix}$

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 $V \subset \mathbb{R}^n$ if $\exists M$ s.t. for all $\underline{x}, \underline{y} \in V$

$M$ : Lipschitz ratio.

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