Nonlinear equations

$f: \mathbb{R} \to \mathbb{R}$

1. Bisection
2. fixed point iteration
3. Newton's method
4. Secant method

$f: \mathbb{R}^n \to \mathbb{R}^n$

$g(x) = x \iff f(x) = 0$

pick x_0, iterate $x_{k+1} = g(x_k)$

Contraction mapping thm:

If g is a contraction on S
then g has a unique fixed ~~point~~ point on S

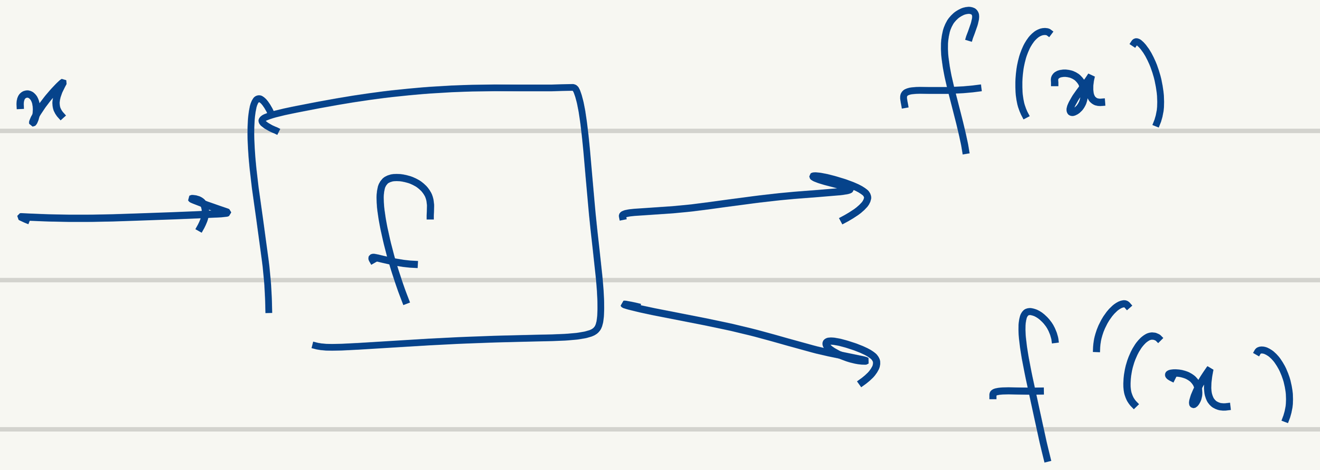
g is a contraction on $[a,b]$

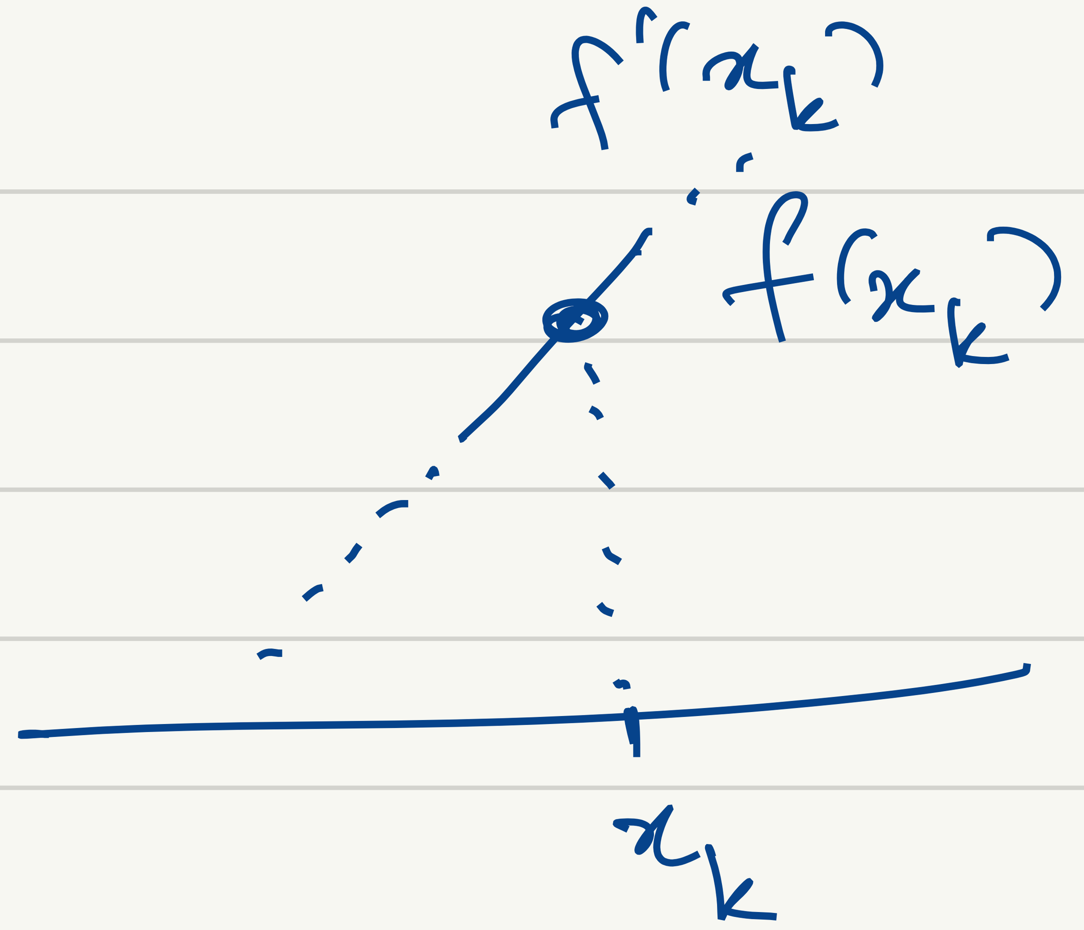
$\iff |g'(x)| < 1$ on $[a,b]$

$$e_{k+1} = \underbrace{g'(x^*)}_{\text{linear convergence}} \cdot e_k + O(|e_k|^2) \Rightarrow \text{convergence rate const.}: \lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|} = |g'(x^*)|$$

except if $g'(x^*) = 0$

Newton's method



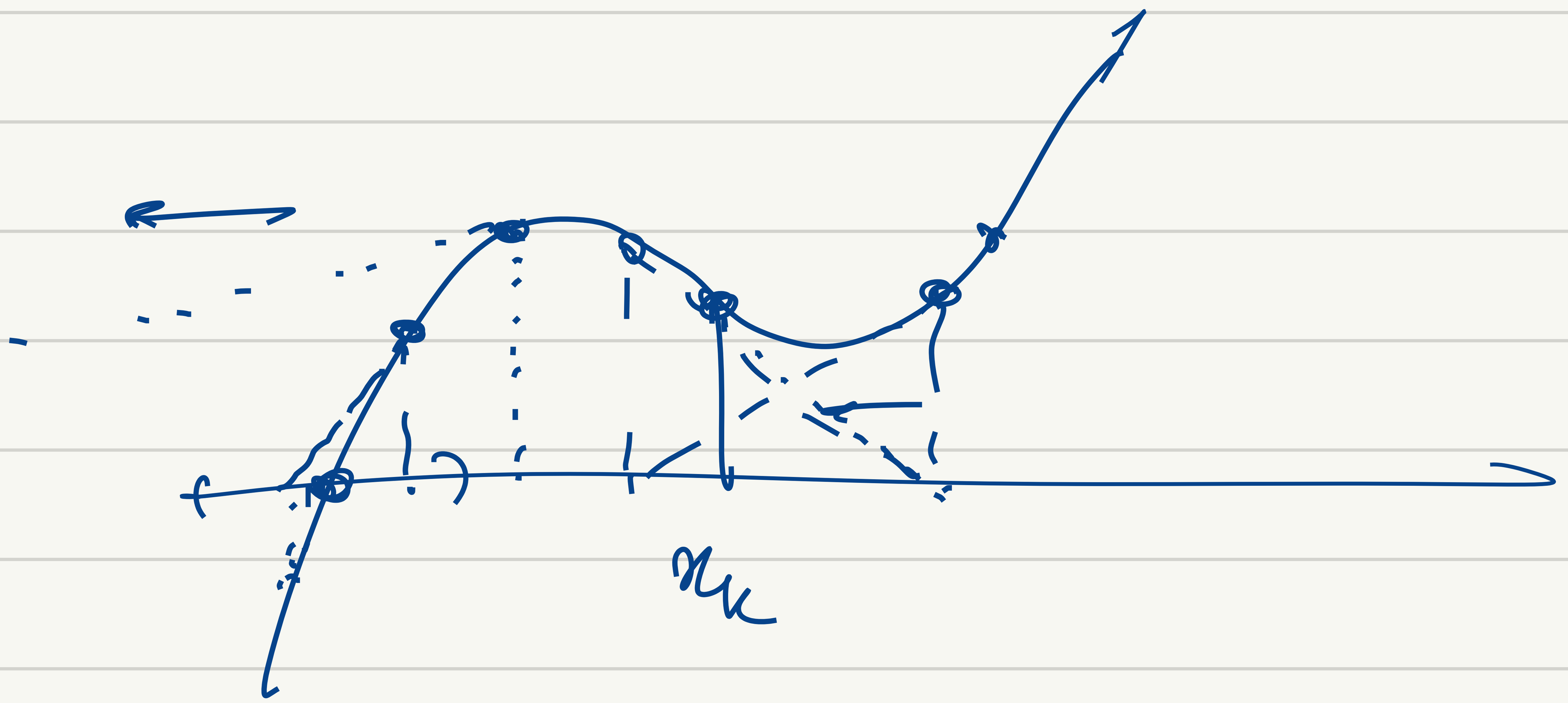


$$\tilde{f}(x) = f(x_k) + f'(x_k)\cdot(x - x_k)$$

$\boxed{\text{local linear approximation}}$ of f

let x_{k+1} = root of $\tilde{f}$

$$\tilde{f}(x_{k+1}) = f(x_k) + f'(x_k)(x_{k+1} - x_k) = 0 \Rightarrow x_{k+1} = \overbrace{\boxed{x_k - \frac{f(x_k)}{f'(x_k)}}}^{g(x_k)}$$

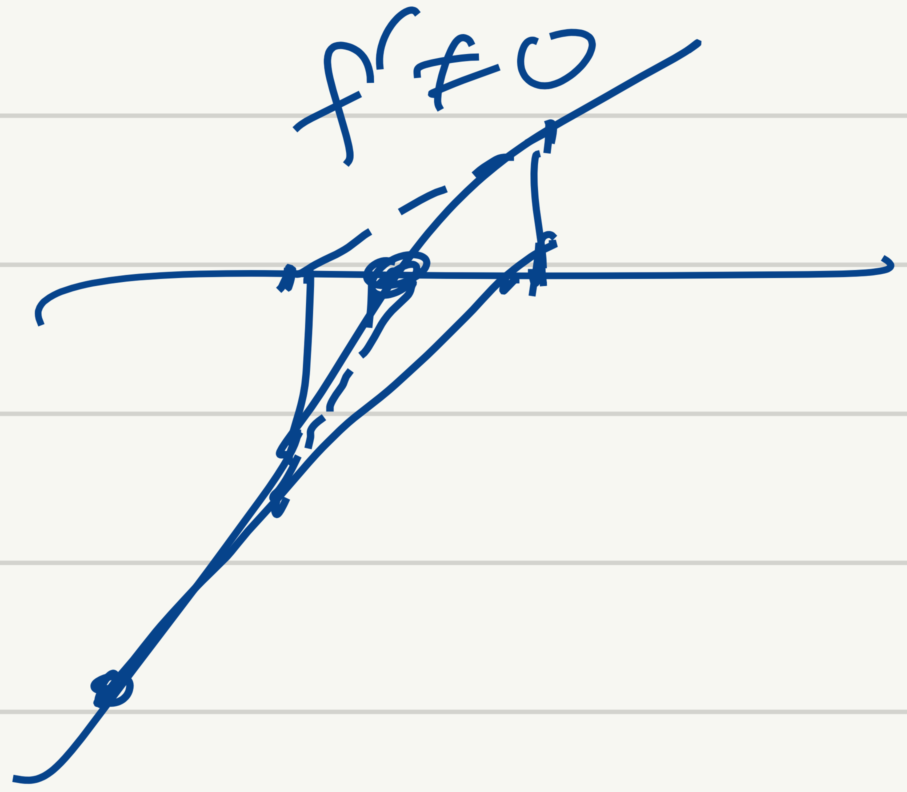


No global guarantees

But $\exists$ neighbourhood of root on which it converges

$$g(x) = x - f(x)/f'(x)$$

Convergence rate: $$g'(x^*) = \frac{\cancelto{0}{f(x^*)}\, f''(x^*)}{f'(x^*)^2}$$

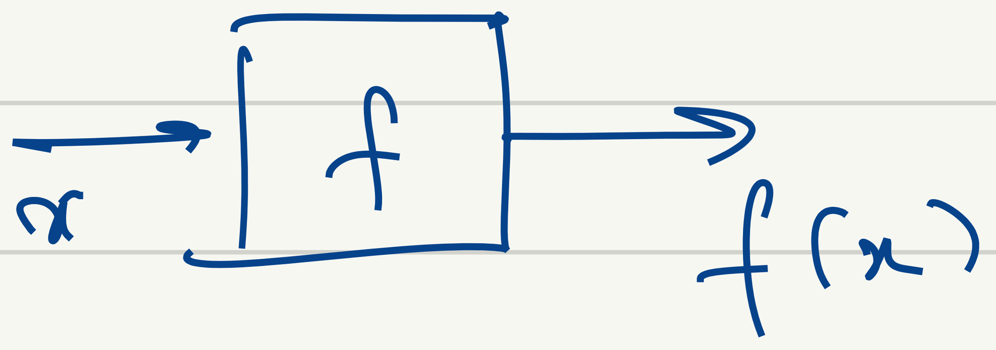


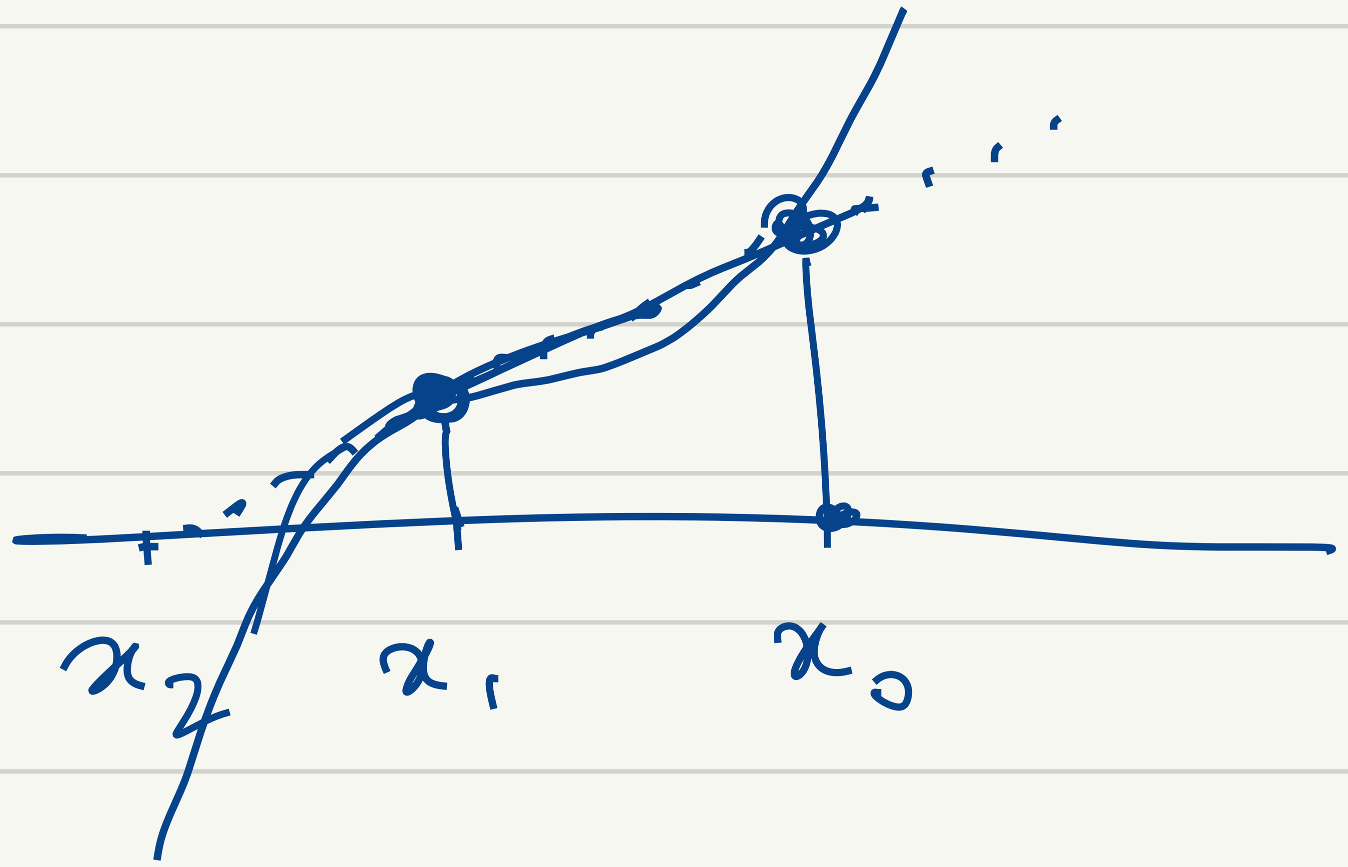
If $f'(x^*) \neq 0$, then $g'(x^*) = 0$

$\Rightarrow$ Newton's method has quadratic convergence

If $f'(x^*) = 0$ then.. linear convergence

Secant method





$$f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$$

plug into Newton's method

$$x_{k+1} = \underbrace{x_k - f(x_k)\cdot\frac{x_k - x_{k-1}}{f(x_k)-f(x_{k-1})}}_{g(x_k,\, x_{k-1})}$$

Convergence rate:

$$e_{k+1} = M e_{k-1} e_k + o(\ldots)$$

third order

Assume $e_{k+1} \approx c|e_k|^r$

$\Rightarrow r = \frac{1+\sqrt{5}}{2} \approx 1.618\ldots$ superlinear but subquad.

If evaluating f, f' is slow then secant method can converge faster!

1D :   nD :

- bisection $\not\longrightarrow$
- fixed pt iter $\longrightarrow$ - fixed point
- Newton's $\longrightarrow$ - Newton
- secant $\longrightarrow$ - secant updating

$f : \mathbb{R}^n \to \mathbb{R}^n$

$$\vec{f}(\vec{x}) = 0 \iff \begin{matrix} f_1(x_1, x_2 \ldots) = 0 \\ f(x_1, x_2, \ldots) = 0 \\ \vdots \end{matrix}$$

$$J(\vec{x}) = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$\vec{f}(\vec{x} + \Delta\vec{x}) = \vec{f}(\vec{x}) + J(\vec{x}) \cdot \Delta\vec{x} + O(\|\Delta x\|^2)$$

$$G(\vec{x}) = [\partial g_i/\partial x_j]$$

<u>fixed point iterations</u>:

$$\vec{x}_{k+1} = \vec{g}(\vec{x}_k)$$

$$\vec{x}_k = \vec{x}_* + \vec{e}_k \quad . \quad g(\vec{x}_k) = \vec{x}_* + J(\vec{x}_*)\vec{e}_k$$

$$\vec{e}_{k+1} = G(\vec{x}_*)\vec{e}_k + O(\|\vec{e}_k\|^2)$$

Converges if $\lim_{k \to 0} G^k \vec{e} = 0$ for all $\vec{e}$ $\iff$ $\rho(G(\vec{x}^*)) < 1$

$= \max_j |\lambda_j|$

$\boxed{\text{spectral radius}}$

If $G = V \Lambda V^{-1} \Rightarrow G^k = V \Lambda^k V^{-1} \to 0$

$\Rightarrow \lambda_j^k \to 0$

What does this have to do with contraction mappings?

If g is a contraction on $S \subseteq \mathbb{R}^n$ $\Rightarrow$ $\|g(x+\Delta x) - g(x)\| < \|\Delta x\|$

$\Rightarrow \|G(x)\| < 1$ for all $\vec{x} \in S$

$\|G(x)\| < 1 \Rightarrow \rho(G(x)) < 1$

$$G = \begin{bmatrix} 0.2 & 100 \\ & 0.3 \end{bmatrix}$$

g is a contraction for some norm $\implies$ fixed point iter converges

$\Downarrow$

$\exists$ induced norm $\|G(x^*)\| < 1 \iff \rho(G(x^*)) < 1$

$\rho(G) \leq \|G\|$ for any induced norm

Sufficient condition: $\|G(x^*)\| < 1$ for some norm

Newton's method

Choose x_{k+1} s.t. $\tilde{f}(x_{k+1}) = \underbrace{f(x_k)}_{} + \overbrace{\underbrace{f'(x_k)}_{}}^{J(x_k)}\underbrace{(x_{k+1} - x_k)}_{} = 0$

$$J(x_k)\underbrace{\Delta x}_{} = \underbrace{-f(x_k)}_{}$$

$x_{k+1} = x_k + \Delta x$

Then $x_{k+1} = x_k + \Delta x$

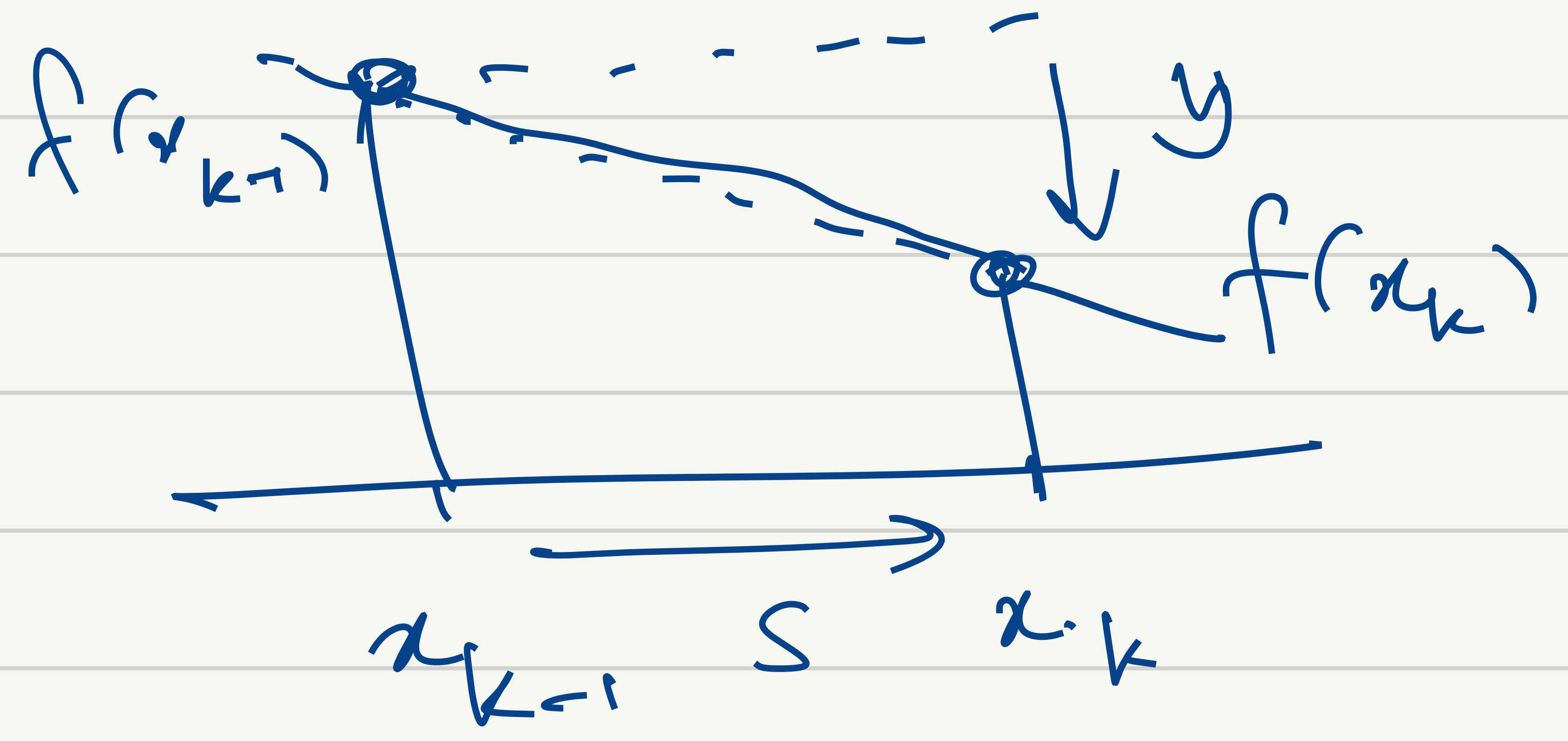
<u>Convergence</u>: quadratic unless $J(x^*)$ is singular

<u>Cost</u> : $O(n^2)$ to form $J(x_k)$ + $O(n^3)$ to solve $J(x_k)\Delta x = -f(x_k)$

(unless J is sparse)

(Also verify $f(\bar{x}+\Delta x) - (f(x) + J(x)\Delta x)$
$= O(\|\Delta x\|^2)$)

Secant updating methods



x_{k-1}, x_k , $s = x_k - x_{k-1}$

$f(x_{k-1}), f(x_k)$, $y = f(x_k) - f(x_{k-1})$

$f'(x_k) \approx y/s$

In $\mathbb{R}^n$, $\quad \vec{s} = x_k - x_{k-1} \in \mathbb{R}^n$

$$\vec{y} = f(x_k) - f(x_{k-1}) \in \mathbb{R}^n$$

$$J(\vec{x}_k)\,\vec{s} \approx \vec{y}$$

Approximation of J : B_k

$$\left\{ \begin{array}{l} B_k = B_{k-1} + \underbrace{\Delta B} \\ B_k \vec{s} = \vec{y} \\ \text{Make } \Delta B \text{ ~~as~~ small as possible :} \end{array} \right.$$

$\Delta B = u v^T$, $\quad \min \|\Delta B\|_F \implies \boxed{B_k = B_{k-1} + \frac{\vec{r}\,\vec{s}^T}{\vec{s}^T \vec{s}}}$ where $\vec{r} = \underbrace{\vec{y} - B_{k-1}\vec{s}}$

Broyden's first method

Then $\vec{x}_{k+1} = \vec{x}_k - B_k^{-1} f(x_k)$

$B_k = B_{k-1} + \Delta B$ where ...

$x_{k+1} = x_k - \underbrace{B_k^{-1} f(x_k)}$

~~$O(n^3)$ time?~~

$B_{k-1} = QR$ or LU

$B_k = B_{k-1} + \underbrace{uv^T}_{\text{rank-1 update}} \longrightarrow$ QR or LU of B_k in $O(n^2)$ time

or... Instead of $B_k \approx J$, let's update $H_k \approx J^{-1}$

$\boxed{\text{Broyden's second method}}$

$$J\vec{s} \approx \vec{y} \iff J^{-1}\vec{y} \approx \vec{s} \Rightarrow H_k\vec{y} = \vec{s} \Rightarrow H_k = H_{k-1} + \frac{\overbrace{(s - H_{k-1}\vec{y})\vec{y}^T}}{\vec{y}^T\vec{y}}$$

$$x_{k+1} = x_k - H_k f(x_k)$$