Nonlinear eq.

Recall fixed pt iteration.

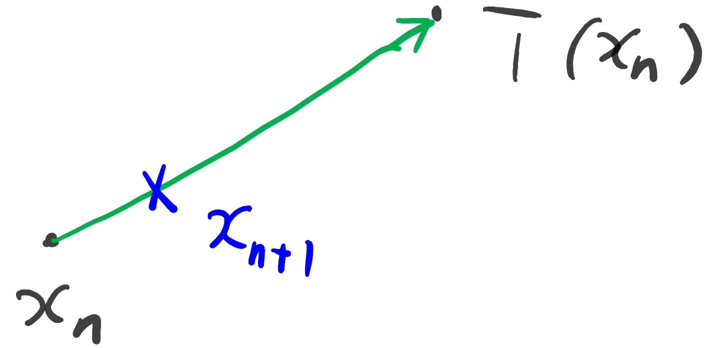
$$x = T(x) \iff F(x) = x - T(x) = 0.$$

$$x_{n+1} = T(x_n)$$

requires $T$ contraction mapping.

$$\|T(x) - T(y)\| \leq L\|x-y\|, \quad L < 1.$$

# Relaxation.




input: $x_n$

output: $T(x_n)$

$$x_{n+1} = (1-\alpha)x_n + \alpha T(x_n).$$

$\alpha = 1$. Fixed point iteration

$\alpha = 0$. stuck at $x_n$.

Intuitively $0 < \alpha < 1$.

Convergence within a small neighborhood of $x^*$.

$$\|e_n\| = \|x_n - x^*\| \ll 1.$$

$$\|e_{n+1}\| = \|(1-\alpha)x_n + \alpha T(x_n) - (1-\alpha)x^* - \alpha T(x^*)\|$$

$$= \|(1-\alpha)e_n + \alpha(T(x_n) - T(x^*))\|$$

Jacobian $J = \frac{\partial T}{\partial x}(x^*)$ invertible.

For simplicity, $J$ diagonalizable

$$J v_i = \lambda_i v_i, \quad \lambda_i \in \mathbb{R}.$$

Take $e_n = \epsilon U_i$, $\quad |\epsilon| << 1$

$$e_{n+1} \approx (1-\alpha)e_n + \lambda_i \alpha e_n + O(\|e_n\|^2)$$

$$= (1-\alpha(1-\lambda_i))e_n$$

Convergence requires

$$|1-\alpha(1-\lambda_i)| < 1 \quad \text{For any } \lambda_i$$

$$\Rightarrow \quad -1 < 1-\alpha(1-\lambda_i) < 1$$

$$\Rightarrow -2 < -\alpha(1-\lambda_i) < 0.$$

$$\Rightarrow 0 < \alpha(1-\lambda_i) < 2.$$

$$\alpha > 0 \Rightarrow \lambda_i < 1.$$

$$\Rightarrow 0 < \alpha < \frac{2}{1-\min_i \lambda_i}$$

Fixed point iteration. $|\min_i \lambda_i| < 1$

<u>Ex</u>. Prothero - Robinson.

$$\begin{cases} u' = \lambda(u-\varphi) + \varphi' . \\ u(0) = \varphi(0) \end{cases}$$

$\varphi(t)$ smooth.

$\lambda < 0$.

$\Rightarrow$ sol. $u(t) = \varphi(t)$

Back - Euler. $u_{n+1} = \underbrace{u_n + h\left(\lambda(u_{n+1} - \varphi) +\right.}_{T(u_{n+1})} \left. \varphi' \right)$

$J = \frac{\partial T}{\partial u}(u^*) = h\lambda < 0$. relaxation $\Rightarrow$ $\alpha < \frac{2}{1-h\lambda}$ small

Ex. $u' = Au$. Trapezoidal. $A \in \mathbb{R}^{n\times n}$ has only real eig.

$$u_{n+1} = \underbrace{u_n + \frac{h}{2}\left(Au_n + Au_{n+1}\right)}_{T(u_{n+1})}$$

$$J = \frac{\partial T}{\partial u} = \frac{h}{2} A. \qquad \text{eig} < 0.$$

$$0 < \alpha < \frac{2}{1 - \frac{h}{2}\min_i \lambda_i(A)}$$

General assumptions.

$F: \mathbb{R}^N \to \mathbb{R}^N$. $\quad F(u^*) = 0$.

Jacobian. $J(u) = \frac{\partial F}{\partial u}(u)$. $\quad (J(u))_{ij} = \frac{\partial F_i}{\partial u_j}(u)$

1) $u^*$ exists.

2) $J: \Omega \to \mathbb{R}^{N \times N}$ Lip. continuous. w. constant $L$.

$$\|J(u) - J(v)\|_2 < L\|u-v\|_2 \quad , u, v \in \Omega.$$

3) $J(u^*)$ non-singular $\quad$ mean diff things.

**Def** $\{u_n\} \subseteq \mathbb{R}^N$, $u^* \in \mathbb{R}^N$. $u_n \to u^*$

linear convergence rate. $\exists\, \gamma \in (0,1)$. $N > 0$,

$$\|u_{n+1} - u^*\|_2 \leq \gamma \|u_n - u^*\|_2. \quad n > N.$$

superlinear convergence rate.

$$\lim_{n\to\infty} \frac{\|u_{n+1} - u^*\|_2}{\|u_n - u^*\|_2} = 0.$$

quadratic convergence: $\exists\ K > 0,\ N$

$$\|u_{n+1} - u^*\|_2 \le K \|u_n - u^*\|_2^2, \quad n > N.$$

really fast.

quad is superlinear.

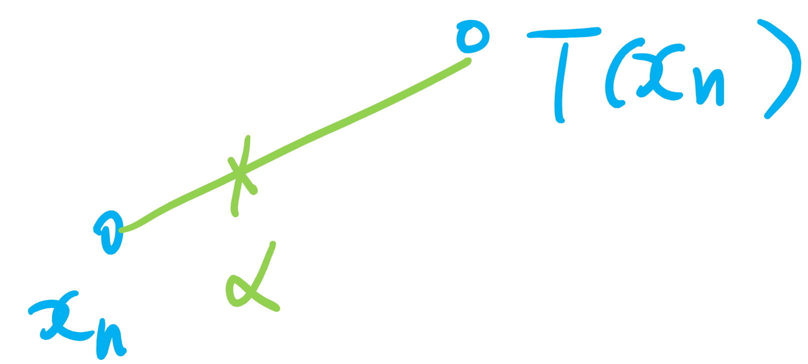
$$\frac{\|u_{n+1} - u^*\|_2}{\|u_n - u^*\|_2} = K \|u_n - u^*\| \to 0$$

Newton's method.

$$F(x) = x - T(x) = 0.$$

fixed pt

$$x_{n+1} = x_n - F(x_n)$$




relaxation

$$x_{n+1} = x_n - \alpha F(x_n) = x_n - \alpha\left(x_n - T(x_n)\right)$$

$$= (1-\alpha)x_n + \alpha T(x_n)$$

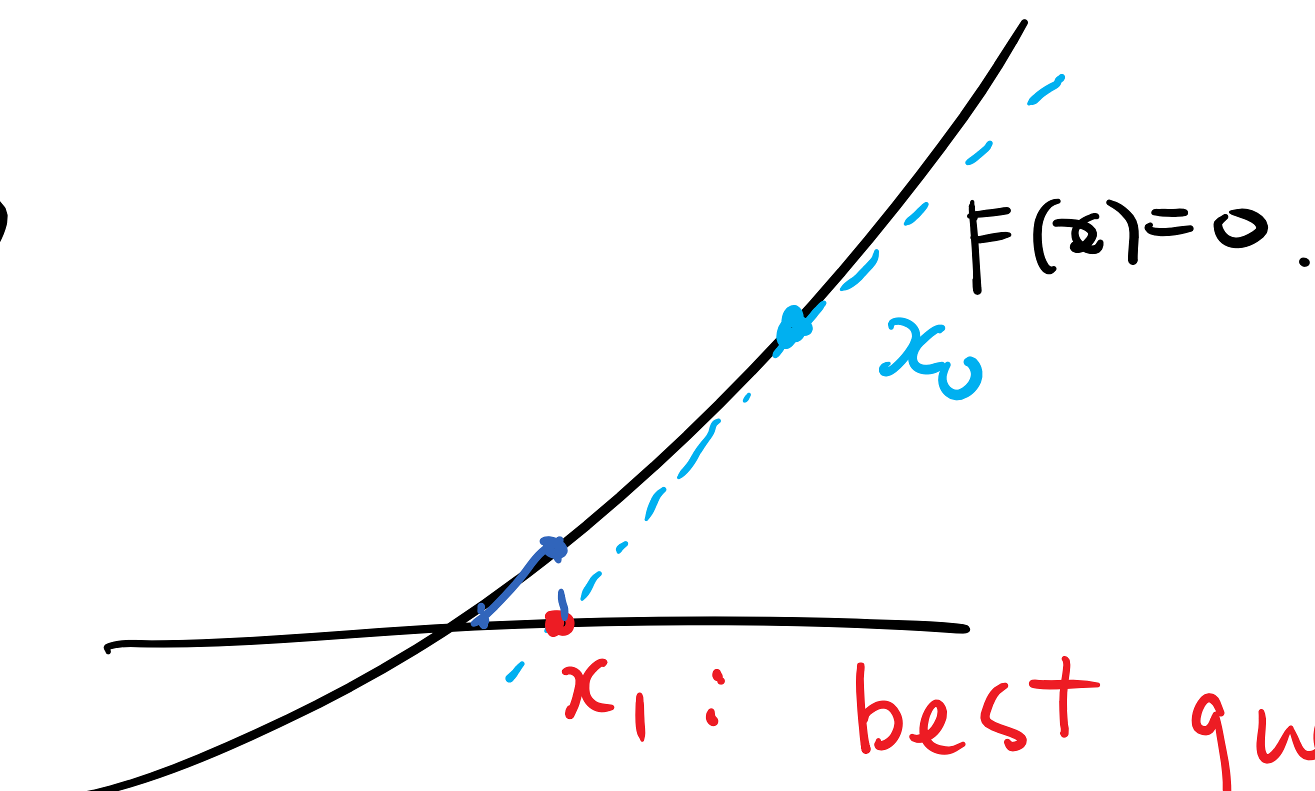
Replace $\alpha$ by a matrix.

$$x_{n+1} = x_n - \underbrace{J^{-1}(x_n)}_{\uparrow} F(x_n)$$

Newton's method.

$\alpha$ is a matrix!

1D




$x_1$ : best guess within lin. approx.

Linear: $F(x) = Ax - b$. Newton's method converges within one step.

Start from any $x_0$.

$$x_1 = x_0 - A^{-1}(Ax_0 - b) = A^{-1}b.$$

Jacobian

sol!

<u>Thm</u>. Initial condition is sufficiently close to $u^*$. then Newton's method converges quadratically.

sketch. ① Most importantly.

$$\|e_{k+1}\| \le C_k \|e_k\| \, \|e_k\|$$

operator level

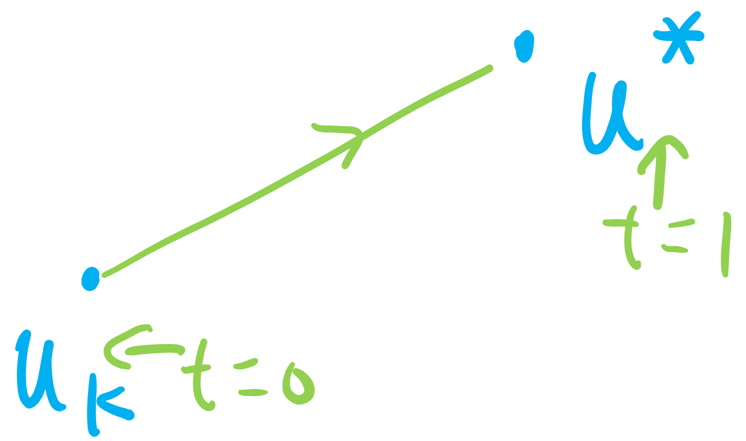
usual.

② (technical) $|C_k|$ won't be large.

---

Pf: $e_k = u^* - u_k$. $\quad F(u^*) = 0$

$$\underbrace{F(u^*)}_{=0} - F(u_k) = \int_0^1 J(u_k + t e_k)\, e_k \, dt$$




$$u_{k+1} = u_k - J(u_k)^{-1} F(u_k)$$
$$= u_k + J(u_k)^{-1} \int_0^1 J(u_k + t e_k)\, e_k \, dt$$

$$e_{k+1} = J(u_k)^{-1}\left(J(u_k)e_k - \int_0^1 J(u_k+te_k)\,e_k\,dt\right)$$

$$= J(u_k)^{-1}\int_0^1 \left[J(u_k) - J(u_k+te_k)\right]e_k\,dt.$$

$$\|e_{k+1}\| \le \|J(u_k)^{-1}\|\cdot L \int_0^1 t\|e_k\|^2\,dt$$

$$= \frac{L}{2}\|J(u_k)^{-1}\|\,\|e_k\|^2$$

Now prove $\|J(u_k)^{-1}\| \le 2\,\|J(u^*)^{-1}\|.$

$$\left\| \left( J(u_k) - J(u^*) + J(u^*) \right)^{-1} \right\|$$

$$= \left\| J(u^*)^{-1} \cdot \left( I - \underbrace{\left( I - J(u_k) J(u^*)^{-1} \right)}_{\text{small}} \right)^{-1} \right\| \qquad (*)$$

$$\| I - J(u_k) J(u^*)^{-1} \| \le \| J(u^*)^{-1} \| \cdot \| J(u^*) - J(u_k) \|$$

$$\le L \, \| J(u^*)^{-1} \| \, \| e_k \| \quad < \frac{1}{2}$$

choose $\delta < \dfrac{1}{2L \| J(u^*)^{-1} \|}$, $\| u_0 - u^* \| < \delta$.

initial condition.

$$(*) \leq \|J(u^*)^{-1}\| \cdot \frac{1}{1 - \|I - J(u_k) J(u^*)^{-1}\|}$$

$$\leq 2 \|J(u^*)^{-1}\| .$$

show $\|u_{k+1} - u^*\| < \delta$ .

$$\|e_{k+1}\| \leq \frac{L}{2} \|J(u^*)^{-1}\| \, \delta \, \|e_k\|$$

$$\leq \|e_k\| \leq \delta . \quad \text{assumption} \qquad \square .$$

Lem. $A \in \mathbb{R}^{N \times N}$. $\|A\| < 1$. Then $I - A$ is invertible.

$$\|(I-A)^{-1}\| \leq \frac{1}{1-\|A\|}$$

Pf: $(I-A)^{-1} = \sum_{n=0}^{\infty} A^n$ (multiply $(I-A)$ on both sides)

$$\|(I-A)^{-1}\| \leq \sum_{n=0}^{\infty} \|A\|^n = \frac{1}{1-\|A\|}$$

Quantification of quality of update from inexact Newton.

$$u_{k+1} = u_k + d_k$$

Generate $d_k$:

① reduce frequency for constructing $J$

② Use iterative method to solve

$J^{-1}F$ Newton-Krylov.

③ $J(u_k)\,x \approx \dfrac{F(u_k + \delta x) - F(u_k)}{\delta}$ finite diff

Jacobian free Newton Krylov (JFNK)

want $d_k \approx -J(u_k)^{-1} F(u_k)$

$$\|J(u_k)\,d_k + F(u_k)\| \le \eta_k \|F(u_k)\| \qquad \boxed{*}$$

inexact Newton cond. implementable. sharp.

Thm. $\eta_k < \eta$ (small enough).

1) If $\boxed{*}$ is satisfied. Inexact Newton. conv. linearly.

2) If $\lim_{k\to\infty} \eta_k = 0$. conv. superlinearly.

Pf: $u_{k+1} = u_k + d_k = \underbrace{u_k - J(u_k)^{-1}F(u_k)}_{\text{Newton}} + \underbrace{d_k + J(u_k)^{-1}F(u_k)}_{\text{residual}}$.

$$\|e_{k+1}\| \le \underbrace{\|e_k - J(u_k)^{-1}F(u_k)\|}_{\text{Newton.}} + \underbrace{\|J(u_k)^{-1}\| \cdot \|F(u_k) + J(u_k)d_k\|}_{\text{inexact Cond}}$$

$$\le \frac{L}{2}\|J(u^*)^{-1}\|\|e_k\|^2 + 2\|J(u^*)^{-1}\|\eta_k\|F(u_k)\|$$

$$\|F(u_k)\| \le 2\|J(u^*)\|\|e_k\| \rightarrow \text{integral formulation.}$$

$$\Rightarrow \|e_{k+1}\| \le \left(\frac{L}{2}\|J(u^*)^{-1}\|\|e_k\| + 4\underbrace{\|J(u^*)\|\|J(u^*)^{-1}\|}_{\text{cond. number.}}\eta_k\right)\|e_k\|$$

$\square$

Broyden's method (quasi-Newton method)

$$u_{k+1} = u_k - B_k F(u_k)$$

Newton : $B_k = J(u_k)^{-1}$

$$s_k = u_k - u_{k-1}$$

$$y_k = F(u_k) - F(u_{k-1}) \approx J(u_k)(u_k - u_{k-1}) = J(u_k)\, s_k$$

$\boxed{s_k = B_k y_k}$ Broyden cond.

Assume $B_{k-1}$ is known

Find $B_k$ closest to $B_{k-1}$. while satisfying

Broyden cond.

$\hookrightarrow$ Frobenius norm

$A \in \mathbb{R}^{m \times n}$. $\|A\|_F^2 = \sum_{ij} A_{ij}^2 = \mathrm{Tr}[A^T A]$, $B \in \mathbb{R}^{m \times n}$

$$\frac{\partial}{\partial A} \mathrm{Tr}[A^T A] = 2A \qquad \frac{\partial}{\partial A} \mathrm{Tr}[A^T B] = B$$

$\Downarrow$

$$\frac{\partial}{\partial A_{ij}} \sum_{i'j'} A_{i'j'}^2 = 2A_{ij}$$

$$\min_{B_k \in \mathbb{R}^{N \times N}} \| B_k - B_{k-1} \|_F^2$$

$s_k, y_k, B_{k-1}$ known

$$\text{s.t.} \quad B_k y_k = s_k$$

Constrained optimization → Lagrange multiplier.

$$\mathcal{L}[B_k, \Lambda] = \frac{1}{2} \| B_k - B_{k-1} \|_F^2 - \Lambda^T [ s_k - B_k y_k ]$$

$\mathbb{R}^N$

$$\frac{\partial \mathcal{L}}{\partial B_k} = B_k - B_{k-1} + \Lambda y_k^T = 0$$

Figure out $\Lambda$.

$$\underbrace{B_k y_k}_{= s_k} - B_{k-1} y_k + \Lambda y_k^T y_k = 0 .$$

$$\Rightarrow \Lambda = -(s_k - B_{k-1} y_k)(y_k^T y_k)^{-1}$$

$$\Rightarrow \boxed{B_k = B_{k-1} + (s_k - B_{k-1} y_k)\, y_k (y_k^T y_k)^{-1}}$$

Usually start with $B_0 = \alpha I$.

Variant:

① Broyden that is exact along more than 1 dir?

$$S_k = [\, s_{k-l+1}, \; s_{k-l+2}, \; \cdots, \; s_k \,] \longrightarrow l \text{ steps of hist}$$

$$Y_k = [\, y_{k-l+1}, \; \cdots \; y_k \,]$$

$$\min_{B_k} \; \| B_k - B_{k-1} \|_F^2$$

$$\text{s.t.} \quad B_k Y_k = S_k .$$

$$\mathcal{L}(B_k, \Lambda) = \frac{1}{2}\|B_k - B_{k-1}\|_F^2 - \mathrm{Tr}\left[\Lambda^T (S_k - B_k Y_k)\right]$$

$$\Lambda \in \mathbb{R}^{n \times l}$$

$$\frac{\partial \mathcal{L}}{\partial B_k} = B_k - B_{k-1} + \Lambda Y_k^T = 0.$$

$$\underbrace{B_k Y_k}_{= S_k} - B_{k-1} Y_k + \Lambda \underbrace{(Y_k^T Y_k)}_{} = 0.$$

often ill-conditioned.

use pseudo-inverse.

$$S_k \left(Y_k^T Y_k\right)^{\dagger} - B_{k-1} Y_k \left(Y_k^T Y_k\right)^{\dagger} + \Lambda = 0 .$$

$$\Rightarrow \quad B_k = B_{k-1} + \left(S_k - B_{k-1} Y_k\right) \underbrace{\left(Y_k^T Y_k\right)^{\dagger} Y_k^T}_{Y_k^{\dagger}}$$

requires storing $B_k$ as dense matrix

Small - medium size ✓ large size ✗

large problem. Fix $B_{k-1} \equiv B_0$

suitable

plays the role of a preconditioner.

best if $B_0 = J(u^*)^{-1}$

$B_k$ = simple + low rank.

Anderson's method.

Anderson's update

$$u_{k+1} = u_k - \left[ B_0 + (S_k - B_0 Y_k) Y_k^{\dagger} \right] F(u_k)$$

$$= \left[ u_k - S_k Y_k^{\dagger} F(u_k) \right] - \left[ B_0 \left( I - Y_k Y_k^{\dagger} \right) F(u_k) \right]$$

Thm. (Basic version of) Broyden's method converges locally & superlinearly.

Idea: $\| B_k^{-1} d_k + F(u_k) \| \le \eta_k \|F(u_k)\|$.

$\eta_k \to 0$