For arbitrary $x \in \mathbb{R}^n$:

$$\|Ax-b\|_2^2 = \|\tilde{U}\tilde{\Sigma}\tilde{V}^T x - b\|^2 = \dots = \|\tilde{\Sigma}V^T x - \tilde{U}^T b\|^2 + \|U'U'^T b\|$$

This is minimal for

$$\tilde{V}^T x = \tilde{\Sigma}^{-1}\tilde{U}^T b$$

Use again decomposition from Lemma 5.25:

$$\|x\|^2 = \|\tilde{V}\tilde{V}^T x\|^2 + \|V'V'^T x\|^2 \overset{\checkmark}{=} \|\tilde{V}\tilde{\Sigma}\tilde{U}^T b\|^2 + \|V'V'^T x\|^2$$

$\Rightarrow$ x with smallest norm satisfies $V'^T x = 0$

and we get $x = \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^T b = A^+ b$ ∎

<u>Interpretation</u>

$$A^+ : b \mapsto \hat{U}\tilde{U}^T b \mapsto A_k^{-1}\tilde{U}\tilde{U}^T b = \underbrace{\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^T}_{=A^+} b$$

orthog. proj. onto range of A, ie. $\tilde{U}\tilde{U}^T$

inverse of orthogon. Kernel of A, ie.

$A_k : (\mathrm{Ker}(A))^\perp \to \mathrm{Range}\, A$

$\tilde{V}z \mapsto A\tilde{V}z = \tilde{U}\tilde{\Sigma}z$

<u>Computing SVD</u>

compute eigenvals + eigvec of $\begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix}$

6, Nonlinear equations & Newton's method

goal: find x st. $f(x) = 0$

<u>Newton method:</u> linearize f: $L(x) := f(x_n) + f'(x_n)(x - x_n)$

x_{n+1} is zero of $L(x)$

$$\Rightarrow \quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} =: \phi^{\text{Newton}}(x_n)$$

This is a fixed point iteration, as the zero of f (x^*) yields $x^* = \phi(x^*)$

<u>Definition (Contraction)</u> $\phi : \mathbb{R}^d \to \mathbb{R}^d$ is a contr. wrt. $\|\cdot\|$ if $\exists\, q \in (0,1)$ st.

$$\|\phi(x) - \phi(y)\| \le q\|x - y\| \quad \text{after some iterations}$$

__Theorem (Conv. of fixed pt. iter)__ ϕ.. contract. , $x^* = \phi(x^*)$, $q \in (0,1)$

$$\Rightarrow \quad \| x^* - x_{n+1} \| \leq q \| x^* - x_n \| \quad \forall n$$

Proof: $\| x^* - x_{n+1} \| = \| x^* - \phi(x_n) \| = \| \phi(x^*) - \phi(x_n) \| \underset{\uparrow \text{ contr. property}}{\leq} q \| x^* - x_n \|$ ∎

(if ϕ is not a contraction $\Rightarrow$ no convergence !)

__Theorem (faster conv.)__ in $\mathbb{R}$, $\phi \in C^p(\mathbb{R})$, $p \geq 2$, $x^* = \phi(x^*)$

if $\phi^{(j)}(x^*) = 0$ for $j = 1, \dots p-1$

$$\Rightarrow \quad | x^* - x_{n+1} | \leq C \cdot | x^* - x_n |^p$$

Proof: $| x^* - x_{n+1} | = | \phi^* - \phi_n | \underset{\uparrow \text{ int by parts}}{=} \left| \frac{1}{(p-1)!} \int_{x^*}^{x^n} (x_n - t)^{p-1} \phi^{(p)}(t) \, dt \right|$

$$\leq \underbrace{\dots\dots}_{=C} | x^* - x_n |^p$$ ∎

__Theorem (conv. in $\mathbb{R}^d$)__ $f \in C^2(\mathbb{R}^d)$, $f(x^*) = 0$, $f'(x^*)$ invertible

$$\Rightarrow \quad \| x^* - x_{n+1} \| \leq C \| x^* - x_n \|^2$$

Proof: ?

__Remarks:__

- instead of inverting f' : solve LSE
- use residual $f(x_n)$ as measure for the error

  e.g. Taylor ($f_n = f_{x^*} + f'_* \cdot (x_n - x^*)$) yields:

  $$\| x^* - x_n \| \leq \| f'^{-1}(x^*) \| \cdot \underbrace{\| f(x_n) \|}_{\text{residual}} + \mathcal{O}(\| \cdot \|^2)$$

- alternative: $\| x_n - x^* \| = \| x_n - x^* + x_{n+1} - x_{n+1} \| \underset{\uparrow \text{ triangle}}{\leq} \| x_n - x_{n+1} \| + \underbrace{\| x_{n+1} - x^* \|}_{\text{negligible due to quadr. conv.}}$

- $f'^{-1} \cdot f$: use LU-fact. and „old" values of f

Damped Newton no convergence? do this:

(Also: „globalized Newton")

$$x_{n+1} = x_n - \lambda_n f'(x_n)^{-1} \cdot f(x_n), \qquad \lambda \in (0,1)$$

Def: (Descent Method) find minimum of $g: \mathbb{R}^d \to \mathbb{R}$

1. find step length d_n
2. find search direct. λ_n, s.t. for $x_{n+1} = x_n + \lambda_n d_n$: $g_{n+1} < g_n$

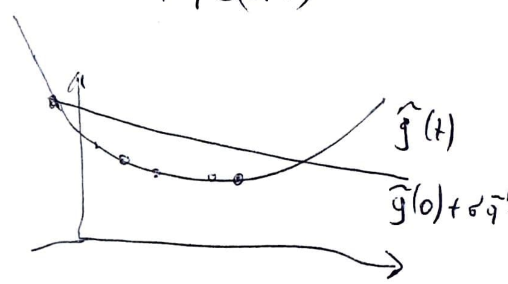
we want for $\tilde{g}(t) := g(x_n + t d_n)$: $\tilde{g}' < 0$

$$\Rightarrow \tilde{g}' = \nabla g \cdot d_n < 0 \quad \Rightarrow \quad d_n = -\nabla g$$

Armijo rule for $\lambda_n = q^k$, $k = 0, 1, \ldots$

So we look for the largest q^k s.t.

$$\tilde{g}(q^k) < \tilde{g}(0) + \sigma \tilde{g}'(0)\, q^k, \qquad \sigma, q \in (0,1)$$



Observation: zeros of f are minima of $x \mapsto \|f(x)\|_2^2 =: g(x)$

$\Rightarrow$ use descent method to find zeros

$\hookrightarrow$ Newton direction $d_n := -f_n'^{-1} f_n$

Lemma (Newton direction is a descent direc) $f \in C^2(\mathbb{R})$, $d_n = -f'(x_n)^{-1} \cdot f(x_n)$

$\Rightarrow$ $\tilde{g}(\lambda) := g(x + \lambda d)$ has Taylor exp.:

$$\tilde{g}(\lambda) = g(x) - \underbrace{2\lambda g(x)}_{>0} + \mathcal{O}(\lambda^2) \leq g(x)$$

Proof: Taylor ∎

Chose λ: $\lambda = \begin{cases} 1, & \text{if } x \text{ is close to } x^* \text{ to get quadratic conv.} \\ \text{small}, & \text{else to get descent} \end{cases}$

Non linear least squares (Gauss-Newton) find x^* st. $\|F(x^*)\| \leq \|F(x)\| \ \forall x$

minima satisfy $\nabla g(x^*) = 0$ for $g := \|F\|_2^2$

Define Newton-method for $\nabla g = F'(x)^T \cdot F(x) \doteq 0$

The Newton method is then:

$$g(x_{n+1}) = g(x_n) + g'(x_n)\,\Delta x_n \overset{!}{=} 0$$

$$\Rightarrow \quad g'(x_n)\,\Delta x_n = -g(x_n) \qquad (1)$$

with $g'(x_n) = F'^T F' + \underbrace{F''^T F}_{\to 0 \text{ for } F(x^*)=0}$ (2)

$\overset{(1),(2)}{\Rightarrow}$ $F'^T F' \Delta x_n = -F'^T F$ ... normal equ. for the linear least. squ. prob.

$$\| F' \Delta x_n + F \|_2^2 \leq \| F' y + F \|_2^2 \quad \forall y$$

(Non lin. l. squ. reduced to sequence of linear lsr. squ)

Newton - convergence: $\| x^* - x_{n+1} \|_2 \leq C \| x^* - x_n \|_2^2 \quad \forall n$

(quadratically)

if $F(x^*) \neq 0$: linear conv.

No proof.

Remark: Quasi Newton methods

if $f'(x_n)$ is expensive: use $f'(x_0)$ (only linear conv.)

better: Broyden

Broyden method: $x_{n+1} = x_n - \underbrace{H_n^{-1}}_{\text{approx of } f'^{-1}(x_n)} f(x_n)$

compute H_{n+1} by: $H_{n+1}(x_{n+1} - x_n) = f(x_{n+1}) - f(x_n)$ ... secant. cond.

s.t. $\| H_{n+1} - H_n \| \to \min$

Lemma: $\exists$ unique H to that problem

No proof.

Convergence: super linearly, ie: $\| \Delta x_n \| \leq \varepsilon_n \| \Delta x_{n-1} \|$ for $\varepsilon_n \to 0$

H_n^{-1} can be computed easily using the Sherman-Morrison-Woodbury-formula

Remark: minimization $\begin{cases} \text{if } \nabla f(x^*) = 0 \Rightarrow \text{Newton (Hessian required)} \\ \text{descent method} \\ \text{trust region method} \end{cases}$

Gradient descent method

1. search direction $\vec{d}_n$ s.t. $\nabla f \cdot d < 0$
   e.g. $d_n := -\nabla f$ (steepest)
2. step length λ_n s.t. $f(x_n + \lambda_n d_n) < f(x_n)$
   e.g. $\lambda := \operatorname{argmin}_t f(x_n + t \cdot d_n)$

Grad. desc. for $f \in P^2(\mathbb{R}^d)$ $\quad f(x) = \gamma + c^T x + \frac{1}{2} x^T Q x$, $\quad Q$.. SPD

$$\Rightarrow \lambda = \operatorname{argmin}_t f(x + td) = -\frac{f(x) \cdot d_n}{d^T Q d}$$

$$x_{n+1} = x_n + \lambda_n d_n$$

Lemma (conditioning) $\quad f(x_{n+1}) - f(x_*) \leq \left(\frac{K-1}{K+1}\right)^2 \cdot \left[f(x_n) - f(x^*)\right]$

$K := \frac{\lambda_{max}}{\lambda_{min}}$ ... cond. (Q)

No proof.

Mitigate by choosing $d := -H \nabla f \Rightarrow f(x_{n+1}) - f(x^*) \leq \frac{\lambda_{max}(H^{-1}Q) - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \cdot \Delta f$

$\Rightarrow$ if $H = Q$: convergence in 1 step!

(equivalent to Newton method for ∇f)

Trust region method: realize, that quadratic approx of f is only accurate close to x^*.

One tries to minimize $q_k(x)$ under constraint $\|x_{n+1} - x_n\| \leq \Delta k$

(q_k.. quadratic model of f)

model is good? check: $\rho_k := \frac{\Delta f}{\Delta q}$ (good: $\rho_k \approx 1$)

7) Eigenvalue problems

find eval and evec of $A \in \mathbb{R}^{n \times n}$

Power method

1. init guess $\vec{x}_0$
2. $\vec{x}_0 := \frac{x_0}{\|x_0\|}$, $\quad \tilde{\lambda}_0 := x_0^H A x_0$
3. $x_{\ell+1} := \frac{A x_\ell}{\|A x_\ell\|}$, $\quad \tilde{\lambda}_{\ell+1} := x_{\ell+1}^H A x_{\ell+1}$

Does not converge if $\lambda_1 \neq \lambda_2$ but $|\lambda_1| = |\lambda_2|$, slow conv if $\lambda_1 \approx \lambda_2$