# Solving the WF

$$m = D j \,, \quad D = [S^{-1} + R^{\dagger} N^{-1} R]^{-1}$$

we have acces to

$$D^{-1} = S^{-1} + R^{\dagger} N^{-1} R$$ as an operator

eg: $$S^{-1} = F^{\dagger} \underbrace{\hat{P}_S}_{\mathrm{diag}(P_S(k))} F$$

## harmonic series

$$D = S^{1/2} [\mathbb{1} + \underbrace{S^{1/2} \overbrace{R^{\dagger} N^{-1} R}^{M} S^{1/2}}_{Q}]^{-1} S^{1/2}$$

$$= S^{1/2} [\mathbb{1} - Q + Q^2 - Q^3 + \dots ] S^{1/2}$$

$$= S - SMS + SMSMS - \dots$$

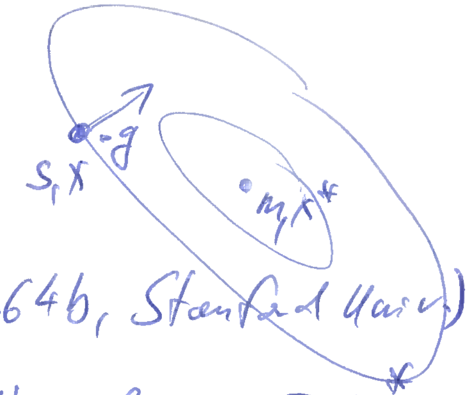
$$D j = \underbrace{S j}_{m_0} - \underbrace{S M m_0}_{m_1} + S M m_1 - \dots$$

disadvantage: converges only for $Q < 1$

requires $\infty$ iterations

original problem

$$H(d,s) \hat{=} \frac{1}{2} s^t D^{-1} s - j^t s$$




$$g(s) = \frac{\partial H(d,s)}{\partial s} = D^{-1} s - j$$

rename to match literature (EE364b, Stanford Univ.)

$$s \to x, \quad D^{-1} \to A, \quad j \to b, \quad H \to f, \quad m \to x^*$$

$$A^t = A \qquad g \to \nabla f^t$$

$$f(x) = \frac{1}{2} x^t A x - b^t x$$

$$\nabla f(x) = A x - b$$

$$f^* = f(x^*) \Leftrightarrow A x^* = b$$

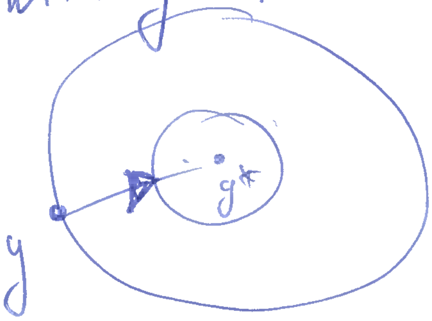
$$f(x) - f(x^*) = \frac{1}{2} x^t A x - \underbrace{b^t}_{\underbrace{A^{-1}A \; (A^{-1}b)^t A x}_{x^{*t} A x}} x \underbrace{- \frac{1}{2} x^{*t} A x^* + \overbrace{b^t x^*}^{x^{*t} A x^t}}_{+ \frac{1}{2} x^* A x^*}$$

$$= \frac{1}{2}\left[x^t A x - 2 x^{*t} A x + \frac{1}{2} x^* A x^*\right]$$

$$= \frac{1}{2}[x - x^*]^t A [x - x^*]$$

$$=: \frac{1}{2} \| x - x^* \|_A^2 = \frac{1}{2} \| y - y^* \|^2$$

with $y = A^{1/2} x \quad y^* = A^{1/2} y$



## Residual $r := b - Ax$

$$r = -\nabla f(x) = Ax^* - Ax = A(x^* - x)$$

$$f(x) - f^* = \frac{1}{2}(x - x^*)^t \underbrace{A}_{A^{-1}A}(x - x^*)$$

$$= \frac{1}{2} r^t A^{-1} r$$

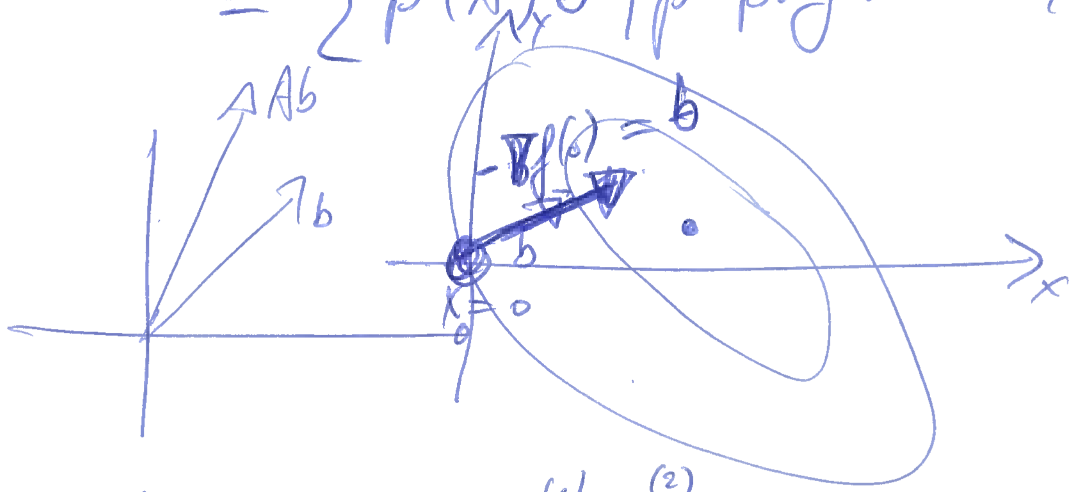
$$= \frac{1}{2} \| r \|_{A^{-1}}^2$$

accuracy measure for $x$ being close to $x^*$: $\eta = \frac{\|r\|}{\|b\|}$

## Krylov subspace

$$K_k := \operatorname{span}\{b, Ab, \dots, A^{k-1} b\}$$

$$= \{p(A)b \mid p \text{ polynomial}, \deg p < k\}$$




Krylov sequence $x^{(1)}, x^{(2)}, \dots$

$$x^{(k)} = \underset{x \in K_k}{\operatorname{argmin}} f(x) = \underset{x \in K_k}{\operatorname{argmin}} \| x - x^* \|_A^2$$

~~defines gives~~ is generated by Conjugate Gradient algorithm (CG)

## Properties

- $f(x^{(k+1)}) \le f(x^{(k)})$ (but $\|v\|$ can increase)
  since $K_k \subset K_{k+1}$
- $x^{(n)} = x^*$ (even if $K_n \neq \mathbb{R}^n$)
- $x^{(k)} = p_k(A)b$ with $p_k$ polynomial with $\deg p_k < k$
- two-term recurrence

$$x^{(k+1)} = x^{(k)} + \alpha_k r^{(k)} + \beta_k \left(x^{(k)} - x^{(k-1)}\right)$$

$$\Delta^{(k+1)} = \alpha_k r^{(k)} + \beta_k \Delta^{(k)}$$

next step — neg gradient — last step correction

k

# Cayley-Hamilton theorem

characteristic polynomial of $A$

$$\chi(s) := \det(sI - A) = s^n + \alpha_1 s^{n-1} + \dots + \alpha_n$$

$$\chi(A) = A^n + \alpha_1 A^{n-1} + \dots \alpha_n I = \det(A-A) = 0$$

multiply with $A^{-1}$

$$A^{n-1} + \alpha_1 A^{n-2} + \dots + \alpha_n A^{-1} = 0$$

$$A^{-1} = -\frac{1}{\alpha_n}\left[A^{n-1} + \alpha_1 A^{n-2} + \dots \alpha_{n-1} I\right]$$

$$= p(A) \text{ of } \deg p < n$$

$$x^* = A^{-1} b = p(A) b \in K_n \ell$$

$\Rightarrow x^{(n)} = x^*$ only $n$ steps are needed

# CG Algorithm (following C.T. Kelley)

$x^{(0)} := 0 \qquad r^{(0)} := b \qquad \rho_0 := \|r\|^2 \qquad (\rho_{-1} = \infty)$

$p^0 = 0$

$x_n$ for $k = 1, \dots, n$ (or $n_{max}$)

quit if $\sqrt{\rho_{k-1}} \le \varepsilon \|b\|$

1) search direction $p^{(k)} \Leftarrow r^{(k-1)} + \underbrace{\frac{\rho_{k-1}}{\rho_{k-2}}}_{\text{if } k>1} p^{(k-1)}$

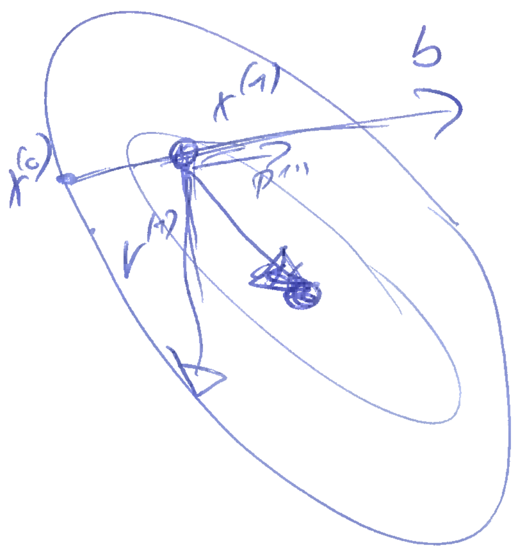
$w^{(k)} \Leftarrow A p^{(k)}$

$\alpha^{(k)} \Leftarrow \rho_{k-1} / p^{(k)} w^{(k)}$

$x^{(k)} \Leftarrow x^{(k-1)} + \alpha^{(k)} p^{(k)}$

$r^{(k)} \Leftarrow r^{(k-1)} - \alpha^{(k)} w^{(k)}$

$\rho_k \Leftarrow \|r\|^2$




First step:

$x^{(0)} = 0, \; r = b, \; \rho_0 = b^2, \; k = 1$

$p := r = b, \; w = Ap = Ar = Ab$

$x^{(1)} = \underbrace{x^{(0)}}_{0} + \alpha p = \alpha b$

$f(\alpha) := f(x^{(1)}(\alpha)) = f(\alpha b) = \frac{1}{2}\alpha^2 b^t A b - \alpha b^t b$

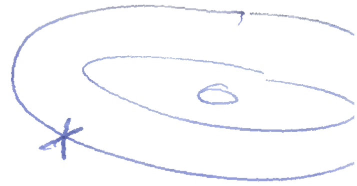
$0 = \frac{\partial f}{\partial \alpha} = \alpha \, b^t A b - b^t b \Rightarrow \alpha = \frac{b^t b}{b^t A b}$

$\alpha^{(1)} \overset{?}{=} \frac{\rho_{k-1}}{p^t w} = \frac{\|r^{(0)}\|^2}{b^t A b} = \frac{b^t b}{b^t A b}$ ✓

$p^{(2)} = r^{(1)} + \frac{\rho_1}{\;} p^{(1)}$

# Convergence

## Spectral analysis



$A = Q \Lambda Q^T$, $Q$ orthogonal

$\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$

$y := Q^T x \quad \bar b = Q^T b, \quad y^* = Q^T x^*$

$$f(x) =: \bar f(y) = \frac12 \underbrace{x^T Q}_{y^T} \Lambda \underbrace{Q^T x}_{y} - b^T Q Q^T x$$

$$= \frac12 y^T \Lambda y - \bar b^T y$$

$$= \frac12 \sum_{i=1}^{n} \lambda_i y_i^2 - \bar b_i y_i$$

$$\Rightarrow y_i^* = \frac{\bar b_i}{\lambda_i} \qquad f^* = \bar f(y^*) = -\frac12 \sum_{i=1}^{n} \frac{\bar b_i^2}{\lambda_i}$$

Krylov sequence in terms of $y$

$$y^{(k)} = \operatorname*{argmin}_{y \in K_k} \bar f(y) \qquad K_k = \mathrm{span}\{\bar b, \Lambda \bar b, \dots, \Lambda^{(k-1)} \bar b\}$$

$$y_i^{(k)} = p_k(\lambda_i)\, \bar b_i \qquad \deg p_k < k$$

$$p_k = \operatorname*{argmin}_{\deg p < k} \sum_{i=1}^{n} \bar b_i^2 \left[\frac{\lambda_i}{2} [p(\lambda_i)]^2 - p(\lambda_i)\right]$$

$$f(x^{(k)}) - f^* = \bar{f}(y^{(k)}) - f^*$$

$$= \min_{\deg p < k} \Big[ \frac{1}{2} y^{(k)T} \Lambda y^{(k)} - \underbrace{\bar{b}^T}_{\Lambda^{-1}\Lambda} y^{(k)} + \underbrace{\frac{1}{2} \bar{b}^{(k)T} \bar{b}^{(k)}}_{\text{const}} \Big]$$

$$\boxed{y_i^{(k)} = p_k(\lambda_i)\bar{b}_i}$$

$$\cancel{= \min_{\deg p < k} \Big[ \sum_{i} \frac{1}{2} (p_k(\lambda_i)\bar{b}_i) \lambda_i \Big]}$$

$$= \min_{\deg p < k} \frac{1}{2} (y^{(k)} - \Lambda^{-1}\bar{b})^T \Lambda (y^{(k)} - \Lambda^{-1}\bar{b})$$

$$= \min_{\deg p < k} \frac{1}{2} \sum_{i=1}^{n} (p_k(\lambda_i)\bar{b}_i - \lambda^{-1}\bar{b}_i)^2 \lambda_i$$

$$= \min_{\deg p < k} \frac{1}{2} \sum_{i=1}^{n} \frac{\bar{b}_i^2}{\lambda_i} (\lambda_i p_k(\lambda_i) - 1)^2$$

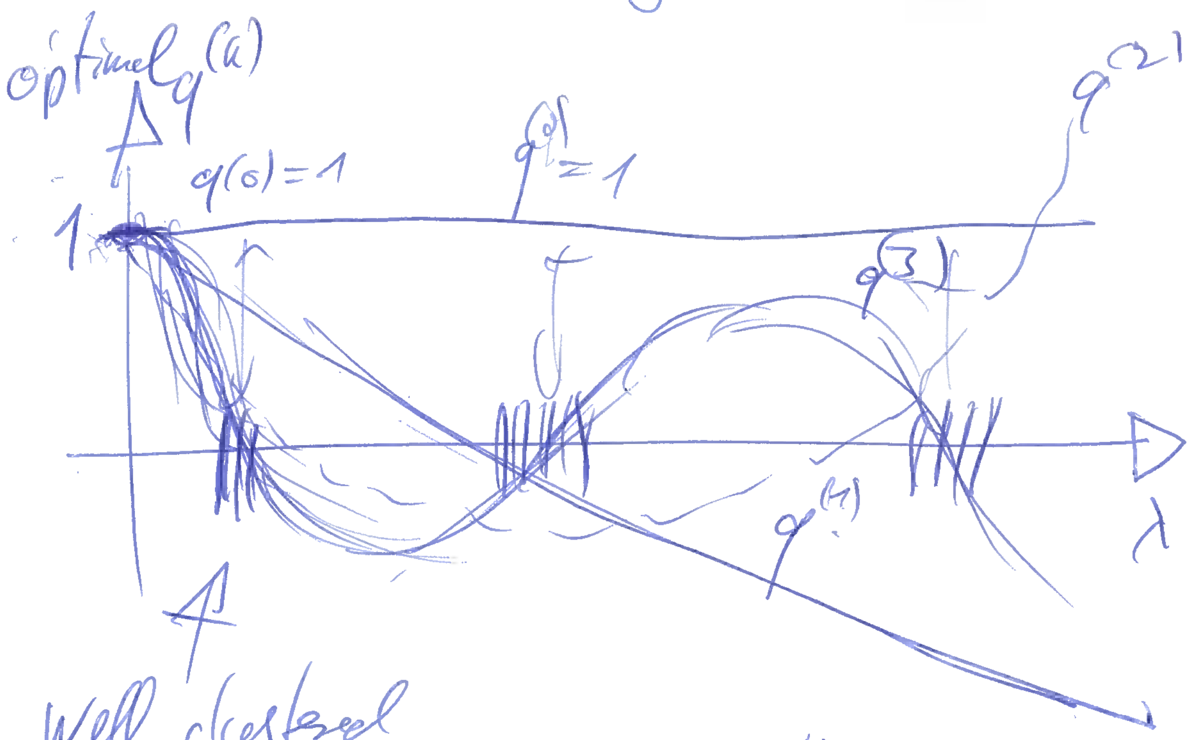
$$= \min_{\deg q < k,\ q(0)=1} \frac{1}{2} \sum_{i=1}^{n} \bar{y}_i^{*2} \lambda_i (q(\lambda_i))^2$$

$$= \min_{\deg q < k,\ q(0)=1} \frac{1}{2} \sum_{i=1}^{n} \frac{\bar{b}_i^2}{\lambda_i} [q(\lambda_i)]^2$$

relative measure of error small if polynomial exists:

$$\tau = \frac{f(x) - f^*}{f(0) - f^*} = \frac{\|x - x^*\|_A^2}{\|x^*\|_A^2} = \min_q \frac{\sum_i y_i^{*2} \lambda_i (q(\lambda_i))^2}{\sum_i y_i^{*2} \lambda_i}$$

that is small at eigenvalues of $A$




well clustered eigenvalues make algorithm small will be dealt with together! $\kappa = \lambda_{max}/\lambda_{min}$

$$\tilde{\varepsilon}_k \le \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$$

improve spectral properties via

<u>preconditioning</u>

$M \approx A^{-1}$

$M = T T^\dagger \quad \det T \neq 0$

$x = T y, \quad x^* = T y^* \qquad \operatorname{cond} \bar{A} = \frac{\bar{\lambda}_{max}}{\bar{\lambda}_{min}} < \operatorname{cond} A$

$Ax = b$

$\to \underbrace{T^\dagger A T}_{\bar{A}}\, y = T^\dagger b \quad \to$ solve via CG

example $A = D^{-1} = S^{-1} + \widehat{R^\dagger N^{-1} R}$

use $T = S^{1/2} = \mathcal{F} P_{(S)}^{1/2} \mathcal{F}$

Finite data $d \in \mathbb{R}^m$ $\quad Q = S^{1/2} R^t N^{-1} R S^{1/2} = S^{1/2} M S^{1/2}$

$\operatorname{rang} Q = \operatorname{rang} M = \operatorname{rang} R^t N^{-1} R \leq \operatorname{rang} N^{-1} = m$

$\cancel{A}\; T^t A T = \mathbb{1} + Q$

$m$ Eigenvalues $> 0$

$n-m$ Eigenvalues $= 0$

$\operatorname{spec}(TAT^t) = \{\lambda_1 \ldots, \lambda_m, 0 \ldots 0\}$

KC converges exactly after $m \ll n$ iterations!