Optimization

Part 1: Steepest Descent & Conjugate Gradients

Recap

CG Iteration (Symmetric Positive Definite A)

$x_0 = 0$, $r_0 = b$, $p_0 = r_0$

for $n = 1, 2, 3, \ldots$

$\alpha_n = (r_{n-1}^T r_{n-1}) / (p_{n-1}^T A p_{n-1})$ step size

$x_n = x_{n-1} + \alpha_n p_{n-1}$ approx. soln.

$r_n = r_{n-1} - \alpha_n A p_{n-1}$ residual

$\beta_n = (r_n^T r_n) / (r_{n-1}^T r_{n-1})$ improvement

$p_n = r_n + \beta_n p_{n-1}$ search direction

Residuals $r_0, r_1, \ldots, r_{n-1}$ form ONB for $K_n(A, b)$

Directions $p_0, p_1, \ldots, p_{n-1}$ form A-ONB for $K_n(A, b)$

x_n minimizes $\|e_n\|_A = \sqrt{(x - x_n)^T A (x - x_n)}$ over $K_n(A, b)$

Distinct from GMRES (in part) because

a) builds ONBs using short recurrences (symmetry of A)

b) minimizes $\|e_n\|_A$ instead of $\|r_n\|_2$ (SPD)

CG and Polynomial Approximation

Like Arnoldi, GMRES, Lanczos: CG convergence closely linked to polynomial approximation.

$$e_n = p_n(A)e_0 \quad \text{for } \overbrace{\text{degree } n \text{ poly with } p_n(0)=1}^{\mathbb{P}_n}.$$

$$(*) \qquad \|p_n(A)e_0\|_A = \text{minimum over all } p_n \in \mathbb{P}_n$$

<u>Thm 1</u> If A is SPD and $r_{n-1} \neq 0$, then $(*)$ has a unique soln $p_n \in \mathbb{P}_n$, and

$$\frac{\|e_n\|_A}{\|e_0\|_A} = \inf_{p \in \mathbb{P}_n} \frac{\|p(A)e_0\|_A}{\|e_0\|_A} \leq \inf_{p \in \mathbb{P}_n} \max_{\lambda \in \Lambda(A)} |p(\lambda)|$$

where $\Lambda(A)$ is the spectrum of A.

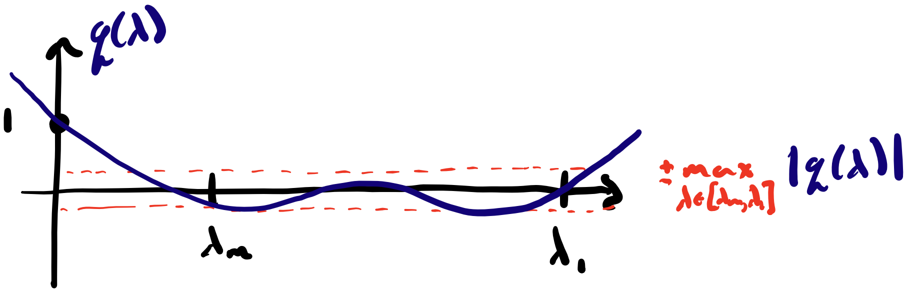
Sketch <u>Pf</u>/ SPD A has $A = V\Lambda V^*$ with $V^*V = I$.
Plug in eigendecomposition to $(*)$ and simplify

Rate of Convergence

Since $\Lambda(A) \subset [\lambda_m, \lambda_1] \subset \{x \in \mathbb{R} : x > 0\}$, we can develop explicit bounds on the rate of convergence $\|e_n\|_A / \|e_0\|_A$.

Idea: look for polynomial $q_n(x)$ that is small on $[\lambda_n, \lambda_1]$ and has $q_n(0)=1$.

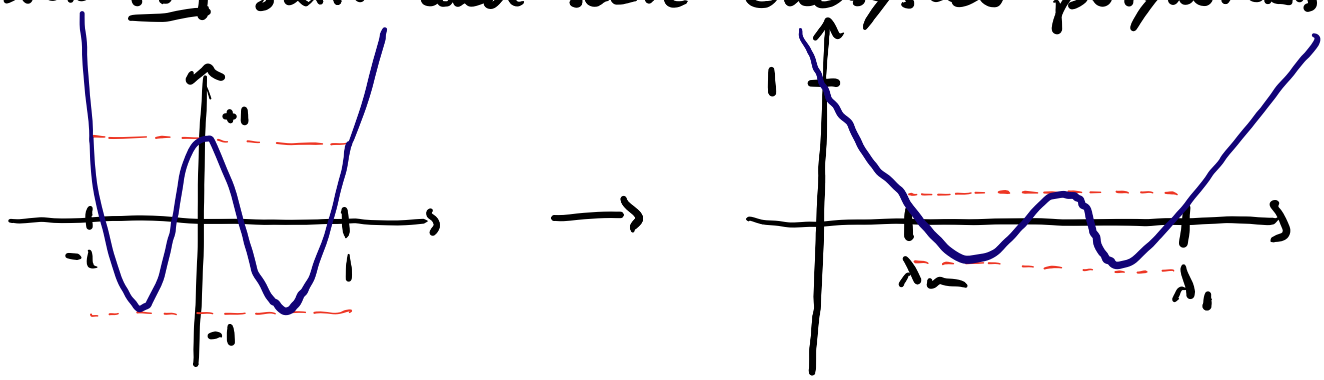
$$\Rightarrow \quad \frac{\|e_n\|_A}{\|e_0\|_A} \le \inf_{p\in\mathbb{P}_n} \max_{\lambda\in\Lambda(A)} |p(\lambda)| \le \max_{\lambda\in[\lambda_n,\lambda_1]} |q(\lambda)|$$



Thm 2 Let SPD A have 2-norm condition number $\kappa = \lambda_1/\lambda_n$. Then CG A-norm error after n iterations satisfy

$$\frac{\|e_n\|_A}{\|e_0\|_A} \le \frac{2}{\left[\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^n + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{-n}\right]} \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n$$

Sketch Pf Shift and scale Chebyshev polynomials



Note: for moderately ill-conditioned matrices

$$\#\text{ iterations} = \mathcal{O}(\sqrt{\kappa}\log \varepsilon_{mach})$$

to reach $\|e_n\|_A / \|e_0\|_A \leq \varepsilon_{mach}$.

CG & Optimization

CG Iterations can also be understood as an improved first-order (gradient based) optimization routine. CG minimizes

$$\|e_n\|_A^2 = (x-x_n)^* A (x-x_n)$$

$$= x_n^* A x_n - 2x_n^* A x + x^* A x$$

$$= \underbrace{x_n^* A x_n - 2x_n^* b}_{\varphi(x_n)} + \underbrace{x^* b}_{\text{const}}$$

CG minimizes $\varphi(x) = \frac{1}{2} x^* A x - x^* b$

over $K_n(A,b)$, by taking "steps" in

the search direction p_{n-1}, of length α_n.

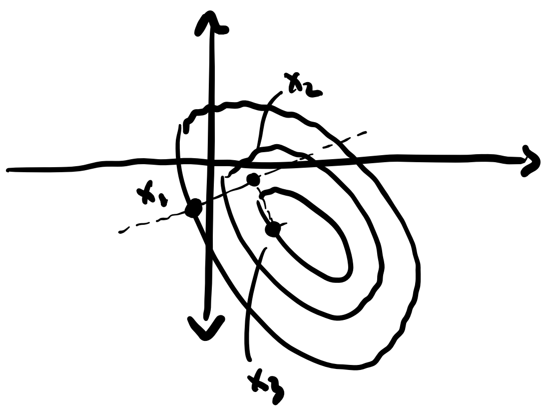
Steepest Descent (Gradient Descent)

Task: minimize smooth function $\varphi : \mathbb{R}^d \to \mathbb{R}$.

Idea: starting from point x, walk "downhill" along the steepest slope.

The direction of steepest slope at point $x_* \in \mathbb{R}^d$ is given by $-\nabla \varphi \big|_{x = x_*}$

E.g. $\quad \varphi(x) = \frac{1}{2} x^* A x - x^* b$



$$\nabla \varphi(x) = \frac{1}{2} A^T x + \frac{1}{2} A x - b$$

$$= Ax - b$$

$$= r$$

To minimize $\varphi(x)$ along line $x - \alpha r$

need $\quad \alpha = \dfrac{r^T r}{r^T A r}$.

Steepest Descent for $\phi(x)$

1. Compute $r_n = b - Ax_n$ $\quad$ Gradient
   $= r_{n-1} - \alpha_{n-1} A r_{n-1}$
2. $\alpha_n = r_n^T r_n / r_n^T A r_n$ $\quad$ step length
3. $x_{n+1} = x_n + \alpha_n r_n$ $\quad$ update