Expressions for the Conjugate Gradient error norms

Gérard MEURANT

Joint work with Hassane Sadok

June 1–6, 2008

1 Summary of results

Introduction

We use CG to solve

$$Ax = b, \quad A \in \mathbb{R}^{n\times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^n$$

with A symmetric positive definite (SPD)

Let $\epsilon_k = x_k - x$ be the error vector, our goal is to find expressions for

$$\|\epsilon_k\|_A^2 = (A\epsilon_k, \epsilon_k)$$

and also $\|r_k\|^2$ where

$$r_k = b - Ax_k$$

We assume exact arithmetic. For finite precision arithmetic, see Strakoš and Tichý or M. and Strakoš

Krylov matrices

Krylov subspace

$$\mathcal{K}_k(A, v) \equiv span\{v, Av, \dots, A^{k-1}v\}$$

The CG solution is defined by

$$x_k - x_0 \in \mathcal{K}_k(A, r_0) \equiv \mathcal{K}_k, \quad r_k = b - Ax_k = A(x - x_k) = A\epsilon_k \perp \mathcal{K}_k$$

Let

$$K_k = [r_0, Ar_0, \dots, A^{k-1}r_0]$$

be the Krylov matrix

First result

$$\|\epsilon_k\|_A^2 = \frac{\det(K_{k+1}^T A^{-1} K_{k+1})}{\det(K_k^T A K_k)} = \frac{1}{e_1^T (K_{k+1}^T A^{-1} K_{k+1})^{-1} e_1}$$

This also gives

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_0\|_A^2} = \frac{1}{e_1^T(K_{k+1}^T\, A^{-1}\, K_{k+1})e_1\;\, e_1^T(K_{k+1}^T A^{-1} K_{k+1})^{-1} e_1}$$

Using the Kantorovich inequality we obtain

$$1 > \frac{\|\epsilon_k\|_A}{\|\epsilon_0\|_A} \geq \frac{2\,\sqrt{\kappa(K_{k+1}^T A^{-1} K_{k+1})}}{\kappa(K_{k+1}^T A^{-1} K_{k+1}) + 1}$$

The norm of the error is large as long as $K_{k+1}^T A^{-1} K_{k+1}$ is well conditioned

Second result

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{\det(K_k^T K_k)^2}{\det(K_k^T A^{-1} K_k)\det(K_k^T A K_k)}$$

Concerning the residual we have

$$\|r_k\|^2 = \frac{\det(K_k^T K_k)\det(K_{k+1}^T K_{k+1})}{\det(K_k^T A K_k)^2}$$

CG and Lanczos

$$V_k^T V_k = I_k, \quad \text{where} \quad V_k \equiv [v_1, \dots, v_k]$$

$$A\, V_k = V_k\, T_k + \eta_{k+1}\, v_{k+1}\, e_k^T$$

T_k is the tridiagonal matrix $(\eta_k, \alpha_k, \eta_{k+1})$

The CG iterates are given by

$$x_k = x_0 + V_k y_k, \quad T_k y_k = \|r_0\| e_1$$

We are interested in

$$\widehat{T_k} = (V_k^T\, A^{-1}\, V_k)^{-1}$$

Third result

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{1}{\det(V_k^T A V_k)\det(V_k^T A^{-1} V_k)} = 1 - \frac{\det(\widehat{T}_k)}{\det(T_k)}$$

Hence it is worth studying $\widehat{T}_k$

Properties of $\widehat{T}_k$

$\widehat{T}_k$ is a tridiagonal matrix and

$$\widehat{T}_k = T_k - \tau_k e_k \, e_k^T \, ,$$

where τ_k is a positive real element

$$\widehat{T}_k = (V_k^T \, A^{-1} \, V_k)^{-1} = \begin{pmatrix} \alpha_1 & \eta_2 & & & \\ \eta_2 & \alpha_2 & \eta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \eta_{k-1} & \alpha_{k-1} & \eta_k \\ & & & \eta_k & \alpha_k - \tau_k \end{pmatrix}$$

Eigenvalue properties

Let $\theta_i^{(k)}$ (resp. $\widehat{\theta}_i^{(k)}$) be the eigenvalues of T_k (resp. $\widehat{T}_k$) s.t.

$$\theta_k^{(k)} \leq \ldots \leq \theta_2^{(k)} \leq \theta_1^{(k)}$$

and

$$\widehat{\theta}_k^{(k)} \leq \ldots \leq \widehat{\theta}_2^{(k)} \leq \widehat{\theta}_1^{(k)}$$

Then we have interlacing properties

$$\begin{aligned}
&1)\ \theta_{i+1}^{(k)} \leq \widehat{\theta}_i^{(k)} \leq \theta_i^{(k)}, \quad i \in \{1, \ldots, k-1\} \\
&2)\ \widehat{\theta}_i^{(k)} \leq \theta_i^{(k)} \leq \widehat{\theta}_{i-1}^{(k)}, \quad i \in \{2, \ldots, k\} \\
&3)\ \widehat{\theta}_i^{(k)} \leq \theta_{i-1}^{(k-1)} \leq \widehat{\theta}_{i-1}^{(k)}, \quad i \in \{2, \ldots, k\} \\
&4)\ \lambda_{i+n-k} \leq \widehat{\theta}_i^{(k)} \leq \theta_i^{(k)} \leq \lambda_i, \quad i \in \{1, \ldots, k\}
\end{aligned}$$

Note that the smallest eigenvalue $\widehat{\theta}_k^{(k)}$ is closer to λ_{min} than the smallest Ritz value

Main result

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{\det(\widehat{T}_k)}{\det(T_k)} = 1 - \prod_{i=1}^{k} \frac{\widehat{\theta}_i^{(k)}}{\theta_i^{(k)}} = \tau_k \, e_k^T \, T_k^{-1} e_k$$

Ratio of residual norm and error norm

$$\frac{\|r_k\|^2}{\|\epsilon_k\|_A^2} = \frac{\det(\widehat{T}_{k+1})}{\det(T_k)} = \frac{1}{e_{k+1}^T\ \widehat{T}_{k+1}^{-1}\ e_{k+1}}$$

The quest for τ_k

The main question is:

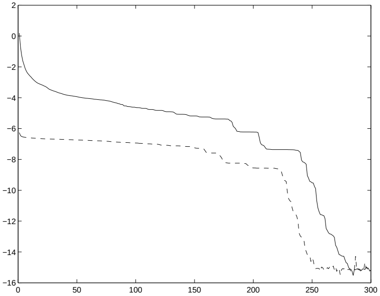
Can we compute (or estimate) τ_k?

This will give estimates of the speed of CG convergence and/or give better estimates of the smallest eigenvalues

Results for the smallest eigenvalue

To motivate you for joining the quest for τ_k, let us compare the smallest Ritz value and the smallest eigenvalue of $\widehat{T}_k$

As an example we use a 1D biharmonic pb: A has 5 diagonals $(1, -4, 6, -4, 1)$ except $A_{1,1} = A_{n,n} = 5$ with $n = 100$ and the "exact" value of τ_k



log_{10} of distance to λ_{min}, T_k (plain), $\widehat{T}_k$ (dashed)