Expressions for the Conjugate Gradient error norms

Gérard MEURANT

Joint work with Hassane Sadok

May 6, 2008

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

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

The CG solution is defined by

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

Let

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

be the Krylov matrix

Then

$$r_k = r_0 - \sum_{i=1}^{k} a_i A^i r_0 = r_k = r_0 - AK_k a$$

with $a = (a_1, a_2, \cdots, a_k)^T$

The orthogonality condition gives

$$(K_k^T A K_k) a = K_k^T r_0$$

Then,

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

Proof: note that

$$\begin{aligned}(A\epsilon_k, \epsilon_k) &= (r_k, \epsilon_k) \\ &= (r_k, \epsilon_0) \\ &= (r_0, \epsilon_0) - (r_0, K_k (K_k^T A K_k)^{-1} K_k^T r_0)\end{aligned}$$

The right hand side is the Schur complement of $K_k^T A K_k$ in

$$K_{k+1}^T A^{-1} K_{k+1} = \begin{pmatrix} r_0^T \epsilon_0 & r_0^T K_k \\ K_k^T r_0 & K_k^T A K_k \end{pmatrix}$$

since $K_{k+1} = [r_0 \; AK_k]$

Block UL factorization

$$\begin{pmatrix} 1 & r_0^T K_k (K_k^T A K_k)^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} (A\epsilon_k, \epsilon_k) & 0 \\ K_k^T r_0 & K_k^T A K_k \end{pmatrix}$$

Taking determinants gives the first assertion, solving $K_{k+1}^T A^{-1} K_{k+1} y = e_1$ proves the other part

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

Using

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

we obtain

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

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 algorithms

In practice we never use the Krylov matrices. To solve SPD linear systems we use CG

- Init: $r_0 = b - Ax_0, \; p_0 = r_0$
- Iterations: Until convergence do,

$$\begin{aligned}
&\gamma_j = (r_j, r_j)/(Ap_j, p_j)\\
&x_{j+1} = x_j + \gamma_j p_j\\
&r_{j+1} = r_j - \gamma_j A p_j\\
&\beta_{j+1} = (r_{j+1}, r_{j+1})/(r_j, r_j)\\
&p_{j+1} = r_{j+1} + \beta_{j+1} p_j
\end{aligned}$$

$$\text{CG} \equiv \text{Lanczos}$$

- Init: $v_1 = r_0/\|r_0\|$, $v_0 \equiv 0$
- Iterations: For $j = 1, 2, \ldots$

$w_j := Av_j - \eta_j v_{j-1}$
$\alpha_j := (w_j, v_j)$
$w_j := w_j - \alpha_j v_j$
$\eta_{j+1} := \|w_j\|_2$
If $\eta_{j+1} = 0$ then Stop
$v_{j+1} := w_j/\eta_{j+1}$

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

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

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

The CG iterates are given by

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

CG and Lanczos parameters are related by

$$\alpha_k = \frac{1}{\gamma_{k-1}} + \frac{\beta_{k-1}}{\gamma_{k-2}}, \quad \beta_0 = 0, \quad \gamma_{-1} = 1$$

$$\eta_{k+1} = \frac{\sqrt{\beta_k}}{\gamma_{k-1}}$$

The Lanczos basis vectors are related to the residual vectors

$$v_{k+1} = (-1)^k \frac{r_k}{\|r_k\|}$$

Moreover

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

where $\widehat{T_k}$ is defined as

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

Then

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

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

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

Proof:

$$
\begin{aligned}
I_k &= V_k^T A^{-1} V_k T_k + \eta_{k+1} V_k^T A^{-1} v_{k+1} e_k^T \\
\widehat{T}_k &= T_k + \eta_{k+1} \widehat{T}_k V_k^T A^{-1} v_{k+1} e_k^T \\
\widehat{T}_k &= T_k + u_k e_k^T
\end{aligned}
$$

But T_k and $\widehat{T}_k$ are symmetric, therefore $u_k e_k^T$ (whose only the last column is non-zero) is also symmetric

See also Paige, Parlett, van der Vorst (1995) or Lenard (1995)

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

Ratio of error norms

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

The last equality is true because

$$\det(\widehat{T}_k) = \det(T_k) - \tau_k \det(T_{k-1})$$

and

$$\frac{\det(T_{k-1})}{\det(T_k)} = e_k^T \, T_k^{-1} e_k$$

Ratio of residual norm and error norm

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

Moreover

$$\frac{\|r_k\|^2}{\|\epsilon_k\|_A^2} = \frac{1}{\gamma_k} - \tau_{k+1}$$

The quest for τ_k

Now, 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

Consider the LDL^T factorization of T_k with $D_{i,i} = \delta_i$

$$\delta_1 = \alpha_1, \quad \delta_l = \alpha_l - \frac{\eta_l^2}{\delta_{l-1}}, \; l = 2, \ldots, k$$

The factorization of $\widehat{T}_k$ is given by

$$\widehat{\delta}_i = \delta_i, \; i = 1, \ldots, k-1, \quad \hat{\delta}_k = \alpha_k - \tau_k - \frac{\eta_k^2}{\hat{\delta}_{k-1}} = \delta_k - \tau_k$$

and

$$\det(\hat{T}_k) = \hat{\delta}_1 \cdots \hat{\delta}_k = \delta_1 \cdots \delta_{k-1}(\delta_k - \tau_k) = \det(T_k)\left(1 - \frac{\tau_k}{\delta_k}\right)$$

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = \frac{\tau_k}{\delta_k}$$

$$\|r_k\|^2 = (\delta_{k+1} - \tau_{k+1})\|\epsilon_k\|_A^2$$

The (k,k) element of $\widehat{T}_k^{-1}$ is $1/(\delta_k - \tau_k)$, hence

$$\tau_k = \delta_k - \frac{1}{(v^k)^T A^{-1} v^k}$$

This does not seem too useful for CG!

Assuming that A has distinct eigenvalues, $AV_n = T_n V_n$, which gives

$$T_n^{-1} = V_n^T A^{-1} V_n$$

Therefore $(\hat{T}_k^{-1})_{i,j} = (T_n^{-1})_{i,j}, \; i,j = 1,\dots,k$

$$\tau_k = \delta_k - \frac{1}{(\hat{T}_k^{-1})_{k,k}} = \delta_k - \frac{1}{(T_n^{-1})_{k,k}} = \frac{1}{(T_k^{-1})_{k,k}} - \frac{1}{(T_n^{-1})_{k,k}}$$

We have also

$$\frac{\|\epsilon_{k-1}\|_A^2}{\|r_{k-1}\|^2} = (\hat{T}_k^{-1})_{k,k} = (T_n^{-1})_{k,k}$$

A recurrence relation for τ_k

Using

$$\|r_k\|^2 = (\delta_{k+1} - \tau_{k+1})\|\epsilon_k\|_A^2$$

for 2 successive iterations we have

$$\frac{\|r_k\|^2}{\|r_{k-1}\|^2} = \frac{\delta_{k+1}}{\delta_k} \left(\frac{1 - \frac{\tau_{k+1}}{\delta_{k+1}}}{1 - \frac{\tau_k}{\delta_k}} \right) \frac{\tau_k}{\delta_k}$$

After some manipulations, we obtain

$$\tau_{k+1} = \alpha_{k+1} - \frac{\eta_{k+1}^2}{\tau_k}$$

This is the same recurrence as for δ_k! But, $\delta_1 = \alpha_1$ when τ_1 is unknown

In fact the recurrence must be written as

$$\tau_k = \frac{\eta_{k+1}^2}{\alpha_{k+1} - \tau_{k+1}}$$

because we know that $\tau_n = 0$

Therefore τ_k is given by a (finite) continued fraction

This can be used to compute an approximation of τ_{k-d} by truncation

A (expensive) way to compute τ_k

We can use $\tau_k = \delta_k - 1/(v_k^T A^{-1} v_k)$
We compute $w_1 = A^{-1} v_1$ and then use the recurrence

$$\eta_{k+1} w_k = v_k - \alpha_k w_k - \eta_k w_{k-1}$$

This may not be stable!

Drawback: we have to solve a linear system

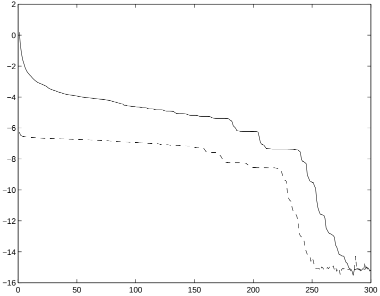
This may be circumvented by choosing $v_1 = Az/\|Az\|$ for a given z

It gives the harmonic Ritz values (without shift) but it is worst for the smallest eigenvalue

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 $\widehat{T}_k$

We use an example devised by C. Paige: 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)