

Convergence bounds for CG

Consider the CG residual and from that the CG error.

CG: $r_m = b - A(x_0 + z_m) = r_0 - Az_m$, where $z_m \in K^m(A, r_0)$.

This gives polynomial $r_m = r_0 - AP_{m-1}(A)r_0 = (I - AP_{m-1}(A))r_0$

Multiplying by A^{-1} gives $e_m = A^{-1}r_m = (I - AP_{m-1}(A))e_0$

Let $R_m(A) = (I - AP_{m-1}(A))$ be the residual polynomial.

Then we get a bound for the error: $\|e_m\|_A \leq \|R_m(A)\|_2 \|e_0\|_A$

Let $A = V\Lambda V^H$ then $A^k = V\Lambda^k V^H$ and so $R_m(A) = VR_m(\Lambda)V^H$

Since Λ is a diagonal matrix, we have $R_m(\Lambda) = \text{diag}(R_m(\lambda_i))$, and

$$\|R_m(A)\|_2 = \|R_m(\Lambda)\|_2 = \max_{\lambda_i \in \lambda(A)} |R_m(\lambda_i)|$$

As we do not know the eigenvalues we make a final simplification

$$\max_{\lambda_i \in \lambda(A)} |R_m(\lambda_i)| \leq \max_{a \leq \lambda \leq b} |R_m(\lambda)|, \text{ where } \lambda(A) \subset [a, b]$$

Convergence bounds for CG

Now we would like to find a bound on $\max_{a \leq \lambda \leq b} |R_m(\lambda)|$

From the optimality of CG we know that

$$\|R_m(A)A^{1/2}e_0\|_2 \leq \|\tilde{R}_m(A)A^{1/2}e_0\|_2 \leq \|\tilde{R}_m(A)\|_2 \|e_0\|_A$$

for any other residual polynomial, $\tilde{R}_m(\cdot)$, of the degree m .

One way to get a bound is to pick a particular polynomial for which we can easily compute the norm, which is known to be small.

We consider the minimax polynomial over the interval that contains the eigenvalues, the Chebyshev polynomials. We consider other choices later.

What keeps us from taking zero polynomial? Our choices must satisfy a normalization. We have $R_m(\lambda) = (1 - \lambda P_{m-1}(\lambda)) \Rightarrow R_m(0) = 1$

We consider the following problem $\min_{\substack{P \in \Pi_m \\ P(0)=1}} \max_{\lambda \in [a,b]} |P(\lambda)|$

Convergence bounds for CG

We know from approximation theory that such a polynomial must be *equioscillating*. That is obtain alternatingly maxima and minima that are equal in absolute value. We also know it is unique.

One equioscillating function is $\cos m\theta$, but is it a polynomial?

The answer is yes; it is a polynomial in $\cos \theta$.

Consider $(e^{i\theta})^m = e^{im\theta} = (\cos \theta + i \sin \theta)^m = (\cos m\theta + i \sin m\theta)$, and take the real part of the polynomial in $\cos \theta$ and $i \sin \theta$.

Let $\cos \theta = x$, for $0 \leq \theta \leq \pi$ and $\theta = \arccos x$ (pv), with $-1 \leq x \leq 1$.

Then $\sin \theta = \sqrt{1-x^2}$ and $(x + i\sqrt{1-x^2})^m = \sum_{j=0}^m \binom{m}{j} x^{m-j} (i\sqrt{1-x^2})^j$.

The real part is $x^m + \binom{m}{2} x^{m-2} (x^2 - 1) + \binom{m}{4} x^{m-4} (x^2 - 1)^2 + \dots = \cos m\theta$

This gives $T_m(x) = \cos m\theta = \cos(m \arccos x)$ for $-1 \leq x \leq 1$

Obviously $\max_{-1 \leq x \leq 1} |T_m(x)| = 1$, attained at $m - 1$ interior points and ± 1 .

Convergence bounds for CG

With some effort we see that $T_m(x) = 2^{m-1} x^m + \dots$. Let $\hat{T}_m = \frac{1}{2^{m-1}} T_m$.

Theorem:

For all monic polynomials $p_m(x)$ of degree m we have

$$\max_{-1 \leq x \leq 1} |\hat{T}_m(x)| \leq \max_{-1 \leq x \leq 1} |p_m(x)|$$

Proof:

Let $q_m(x)$ be monic, $\max_{-1 \leq x \leq 1} |q_m(x)| < \frac{1}{2^{m-1}}$, and $Q(x) = \hat{T}_m(x) - q_m(x)$.

Now \hat{T}_m takes alternating maxima and minima $\frac{(-1)^k}{2^{m-1}}$ at $m - 1$ interior points $\cos \frac{k\pi}{m}$ for $k = 1, \dots, m - 1$ and at the end points. Since $|q_m(x)| < \frac{1}{2^{m-1}}$, at those points $Q(x)$ must take alternatingly positive and negative values and hence must have m roots. However, $Q(x)$ is of degree $m - 1$, and so $Q(x) = 0$ and $q(x) = \hat{T}_m(x)$. Contradiction!

Convergence bounds for CG

So we have shown that $T_m(x)$ minimizes the maximum absolute value over the interval $[-1, 1]$ over all monic polynomials.

We also need to know how $T_m(x)$ behaves outside the interval $[-1, 1]$.

Let $x = \cos \theta = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$ with $\theta = \theta_r + i\theta_i$. So $x = \frac{1}{2}(e^{-\theta_i}e^{i\theta_r} + e^{\theta_i}e^{-i\theta_r})$.

For x to be real we need either real θ or $\theta_r = k\pi$. (real case already seen)

For $x = \pm \cos i\theta$ (θ real) we get $x = \pm \frac{1}{2}(e^{-\theta} + e^{\theta}) = \pm \cosh \theta$.

So we get $T_m(x) = \pm \cosh(m \cosh^{-1}x) = \pm \frac{1}{2}(e^{-m\theta} + e^{m\theta})$ for $x \notin [-1, 1]$.

In order to get the minimax polynomial over the required interval, containing the eigenvalues) we use the (linear polynomial) function

$$w(x) = \frac{2x-b-a}{b-a},$$

which maps $[a, b]$ to $[-1, 1]$ where the Tschchebyshev polynomial is defined.

Convergence bounds for CG

This gives the function $T_m(w(x))$, which is equioscillating but not normalized over the interval $[a, b]$. We normalize this function by dividing by $T_m(w(0))$ (a scalar constant):

$$\hat{T}_m(x) = \frac{T_m(w(x))}{T_m(w(0))} = \frac{T_m\left(\frac{2x-b-a}{b-a}\right)}{T_m\left(-\frac{b+a}{b-a}\right)}.$$

Since $\max_{x \in [a, b]} |T_m(x)| = 1$ we get $\max_{x \in [a, b]} |\hat{T}_m(x)| = |T_m(w(0))|^{-1}$

How large is $|T_m(w(0))|$?

Outside interval $-1 \leq x \leq 1$, we have $T_m(x) = \cosh(m \cosh^{-1}x)$.

Let $y = e^\theta$ and $x = \frac{1}{2}(y + y^{-1})$. Then $T_m(x) = \frac{1}{2}(y^m + y^{-m})$.

Now y defined by $y^2 - 2xy + 1 = 0$. Take solution $|y| \geq 1$.

$$y = x \pm \sqrt{x^2 - 1} \quad (\text{note since } b > a > 0, -\frac{b+a}{b-a} < -1)$$

Convergence bounds for CG

We have $x = w(0) = -\frac{b+a}{b-a}$. Take $b = \lambda_{\max}$ and $a = \lambda_{\min}$.

Since $|T_m(x)| = |T_m(-x)|$, we can take $x = \frac{b+a}{b-a} = \frac{b/a+1}{b/a-1} = \frac{\kappa+1}{\kappa-1}$,

where $\kappa = \frac{b}{a}$ is the condition number of A . From $y = x + \sqrt{x^2 - 1}$ we get

$$y = \frac{\kappa+1}{\kappa-1} + \left(\frac{\kappa^2+2\kappa+1}{\kappa^2-2\kappa+1} - \frac{\kappa^2-2\kappa+1}{\kappa^2-2\kappa+1} \right)^{1/2} = \frac{\kappa+1}{\kappa-1} + \left(\frac{4\sqrt{\kappa}^2}{(\kappa-1)^2} \right)^{1/2}$$

$$y = \frac{\kappa+1}{\kappa-1} + \frac{2\sqrt{\kappa}}{\kappa-1} = \frac{\sqrt{\kappa}^2+2\sqrt{\kappa}+1}{(\sqrt{\kappa}+1)(\sqrt{\kappa}-1)} = \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}, \text{ and so}$$

$$\left| T_m\left(\frac{b+a}{b-a}\right) \right| = \frac{1}{2} \left(\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \right)^m + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \right)^{-m} \right) \geq \frac{1}{2} \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \right)^m$$

$$\text{Result: } \max_{x \in [a,b]} \left| \hat{T}_m(x) \right| = |T_m(w(0))|^{-1} \leq 2 \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right)^m$$

$$\|e_m\|_A \leq \|R_m(A)e_0\|_A \leq \|\hat{T}_m(A)e_0\|_A \leq 2 \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right)^m \|e_0\|_A$$

Convergence bounds for CG

Consider case where one eigenvalue λ_n much larger than others.

Construct better polynomial than $T_m\left(\frac{2\lambda-\lambda_n-\lambda_1}{\lambda_n-\lambda_1}\right)/T_m\left(\frac{-\lambda_n-\lambda_1}{\lambda_n-\lambda_1}\right)$ using this information.

For example, “polynomial that is zero at extreme eigenvalue and lower degree Chebyshev over other eigenvalues”.

$$p_m(z) = \left[T_{m-1}\left(\frac{2\lambda-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right) / T_{m-1}\left(\frac{-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right) \right] \left(\frac{\lambda_n-\lambda}{\lambda_n} \right)$$

Clearly $p_m(\lambda_n) = 0$ and $|p_m(\lambda_i)| < \left| T_{m-1}\left(\frac{2\lambda_i-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right) / T_{m-1}\left(\frac{-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right) \right|$, $i < n$

So, new bound $\frac{\|e_m\|_A}{\|e_0\|_A} \leq 2 \left(\frac{\sqrt{\kappa_{n-1}}-1}{\sqrt{\kappa_{n-1}}+1} \right)^{m-1}$, where $\kappa_{n-1} = \frac{\lambda_{n-1}}{\lambda_1}$,

versus old bound: $\frac{\|e_m\|_A}{\|e_0\|_A} \leq 2 \left(\frac{\sqrt{\kappa_n}-1}{\sqrt{\kappa_n}+1} \right)^m$, where $\kappa_n = \frac{\lambda_n}{\lambda_1}$.