Mathematical Tripos Part II: Michaelmas Term 2023

Numerical Analysis – Lecture 20

Convergence of CG The following theorem gives an important characterization of the CG method.

Theorem 4.33 *Let A be symmetric positive definite. After k iterations of the conjugate gradient method, the error $\boldsymbol{e}^{(k)} = \boldsymbol{x}^* - \boldsymbol{x}^{(k)}$ satisfies*

$$\|\boldsymbol{e}^{(k)}\|_A = \min_{P_k} \|P_k(A)\boldsymbol{e}^{(0)}\|_A$$

where the minimization is over all polynomials P_k of degree $\le k$ that satisfy $P_k(0) = 1$.

Proof. We know from Lecture 18, Theorem 4.22 that $\boldsymbol{e}^{(k)}$ is A-orthogonal to $\operatorname{span}\{\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k-1)}\}$. It is also easy to see that $\boldsymbol{e}^{(k)} - \boldsymbol{e}^{(0)}$ is in $\operatorname{span}\{\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k-1)}\}$ (see e.g., Equation (4.7) in Lecture 18, with $\boldsymbol{d} = \boldsymbol{d}^{(k)}$). Thus if we write

$$\boldsymbol{e}^{(0)} = (\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(k)}) + \boldsymbol{e}^{(k)} \tag{4.11}$$

we see that $\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(k)}$ is the A-orthogonal projection of $\boldsymbol{e}^{(0)}$ on the subspace $\operatorname{span}\{\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k-1)}\}$, and so

$$\|\boldsymbol{e}^{(k)}\|_A = \min_{\boldsymbol{v}} \|\boldsymbol{e}^{(0)} - \boldsymbol{v}\|_A$$

where the minimization is over all $\boldsymbol{v} \in \operatorname{span}(\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k-1)})$, see figure below.

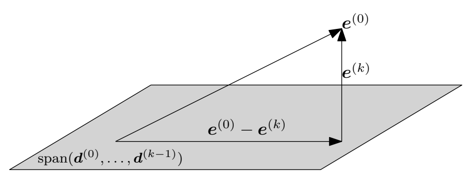

Figure 1: Geometric representation of (4.11). Orthogonality here is with respect to the A-inner product.

Since $\operatorname{span}(\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k-1)}) = \operatorname{span}(\boldsymbol{r}^{(0)}, \ldots, A^{k-1}\boldsymbol{r}^{(0)})$, and since $\boldsymbol{r}^{(0)} = A\boldsymbol{e}^{(0)}$, this means that any such $\boldsymbol{v}$ can be written as $\boldsymbol{v} = \sum_{i=1}^k c_i A^i \boldsymbol{e}^{(0)}$, i.e., $\boldsymbol{e}^{(0)} - \boldsymbol{v} = P_k(A)\boldsymbol{e}^{(0)}$ with $P_k(t) = 1 - \sum_{i=1}^k c_i t^i$ is a degree k polynomial with $P_k(0) = 1$. □

Remark 4.34 *If A has s distinct eigenvalues $\lambda_1, \ldots, \lambda_s > 0$, then with $P_s(t) = \prod_{i=1}^s (1 - t/\lambda_i)$ we have $\deg P_s = s$, $P_s(0) = 1$, and $P_s(A) = 0$. Thus this shows that the CG method terminates after s iterations, recovering the result of Corollary 4.29.*

Corollary 4.35 *Let A be symmetric positive definite, and assume that all its eigenvalues lie in $[l, L]$ where $0 < l < L$. Then after k iterations of the conjugate gradient method, the error $\boldsymbol{e}^{(k)} = \boldsymbol{x}^* - \boldsymbol{x}^{(k)}$ satisfies*

$$\|\boldsymbol{e}^{(k)}\|_A \le 2\rho^k \|\boldsymbol{e}^{(0)}\|_A \le 2(1 - \sqrt{l/L})^k \|\boldsymbol{e}^{(0)}\|_A, \qquad \rho = \frac{\sqrt{L} - \sqrt{l}}{\sqrt{L} + \sqrt{l}} < 1.$$

Proof. First note that for any polynomial P_k we have

$$\|P_k(A)\boldsymbol{e}^{(0)}\|_A \le \left(\max_{\lambda \in \operatorname{spec}(A)} |P_k(\lambda)| \right) \|\boldsymbol{e}^{(0)}\|_A$$

where $\operatorname{spec}(A)$ is the set of eigenvalues of A (its spectrum). To see why, let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ be an orthogonal basis of eigenvectors of A such that $\boldsymbol{e}^{(0)} = \sum_i \boldsymbol{w}_i$. Since the $\boldsymbol{w}_i$ are eigenvectors of A, they are also pairwise orthogonal with respect to the A-inner product, and so $\|\boldsymbol{e}^{(0)}\|_A^2 = \sum_i \|\boldsymbol{w}_i\|_A^2$. In addition $P_k(A)\boldsymbol{e}^{(0)} = \sum_i P_k(\lambda_i)\boldsymbol{w}_i$ and so

$$\begin{aligned}\|P_k(A)\boldsymbol{e}^{(0)}\|_A^2 &= \|\sum_i P_k(\lambda_i)\boldsymbol{w}_i\|_A^2 = \sum_i |P_k(\lambda_i)|^2 \|\boldsymbol{w}_i\|_A^2 \\ &\leq \left(\max_{\lambda\in\operatorname{spec}(A)} |P_k(\lambda)|^2\right) \|\boldsymbol{e}^{(0)}\|_A^2\end{aligned}$$

as desired.

We know that the eigenvalues of A are all in $[l, L]$, so we consider the problem of finding the polynomial P_k of degree k, such that $P_k(0) = 1$, and that minimizes the value

$$\max_{x\in[l,L]} |P_k(x)|.$$

We take P_k to be a Chebyshev polynomial which is suitably translated and scaled, i.e.,

$$P_k(x) = T_k\left(2\frac{L-x}{L-l} - 1\right) \Big/ T_k\left(\frac{L+l}{L-l}\right)$$

where T_k is the usual Chebyshev polynomial defined by identity $T_k(\cos\theta) = \cos(k\theta)$. The polynomial P_k satisfies $P_k(0) = 1$, and since $|T_k(t)| \leq 1$ for all $t \in [-1, 1]$, we have

$$|P_k(x)| \leq \left|T_k\left(\frac{L+l}{L-l}\right)\right|^{-1},$$

for all $x \in [l, L]$. The Chebyshev polynomial satisfies the following inequality for all $|t| \geq 1$:

$$T_k(t) \geq \frac{1}{2}\left(t + \sqrt{t^2-1}\right)^k.$$

By taking $t = (L+l)/(L-l)$, we see that $t + \sqrt{t^2-1} = \frac{\sqrt{L}+\sqrt{l}}{\sqrt{L}-\sqrt{l}}$, which gives us the desired bound

$$\forall x \in [l, L],\ |P_k(x)| \leq 2\left(\frac{\sqrt{L}-\sqrt{l}}{\sqrt{L}+\sqrt{l}}\right)^k.$$

□

For a symmetric positive definite matrix A, let $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} > 1$ be its *condition number*. We saw that the convergence rate of the steepest descent method is $\approx (1 - \frac{1}{\kappa(A)})^k$, whereas the CG method achieves the better rate of $\left(1 - \frac{1}{\sqrt{\kappa(A)}}\right)^k$. When $\kappa(A) \gg 1$, note that $1 - 1/\sqrt{\kappa(A)} \ll 1 - 1/\kappa(A)$.

Remark 4.36 *The condition number defined above can be written as $\kappa(A) = \|A\|_2\|A^{-1}\|_2$ where $\|\cdot\|_2$ is the operator norm of A. This quantity measures the sensitivity of the matrix inverse operation, in a relative error sense. Let $\phi(A) = A^{-1}$ be the matrix inverse operation, and consider a perturbation $\tilde{A} = A + H$. The relative sensitivity is defined as:*

$$\frac{\|\phi(\tilde{A}) - \phi(A)\|_2/\|\phi(A)\|_2}{\|\tilde{A} - A\|_2/\|A\|_2} = \frac{\textit{output relative error}}{\textit{input relative error}}.$$

One can show that for H small, this quantity is bounded above by $\kappa(A)$.

Preconditioning Preconditioning is a technique by which we modify the linear system $A\boldsymbol{x} = \boldsymbol{b}$ in order to reduce the condition number and obtain faster convergence. The idea is to change variables, $\boldsymbol{x} = P^T\widehat{\boldsymbol{x}}$, where P is a nonsingular $n \times n$ matrix, and multiply both sides with P. Thus, instead of $A\boldsymbol{x} = \boldsymbol{b}$, we are solving the linear system

$$PAP^T\widehat{\boldsymbol{x}} = P\boldsymbol{b} \quad \Leftrightarrow \quad \widehat{A}\widehat{x} = \widehat{\boldsymbol{b}}. \tag{4.12}$$

Note that symmetry and positive definiteness of A imply that $\widehat{A} = PAP^T$ is also symmetric and positive definite since $\langle \widehat{A}\boldsymbol{y}, \boldsymbol{y}\rangle = \langle PAP^T\boldsymbol{y}, \boldsymbol{y}\rangle = \langle AP^T\boldsymbol{y}, P^T\boldsymbol{y}\rangle > 0$. Therefore, we can apply conjugate gradients to the new system. This results in the solution $\widehat{\boldsymbol{x}}$, hence $\boldsymbol{x} = P^T\widehat{\boldsymbol{x}}$. This procedure is called the *preconditioned conjugate gradient method* and the matrix P is called the *preconditioner*.

The main idea of preconditioning is to pick P in (4.12) so that $\kappa(\widehat{A})$ is much smaller than $\kappa(A)$, thus accelerating convergence. Ideally, one would like to choose P so that $PAP^T = I$, however this amounts to inverting A! Instead, we look for an approximation S of A that is easy to invert, or to factorize. If we let $S = LL^T$ be a Cholesky factorization of this approximation of A, and take $P = L^{-1}$, then $PAP^T = L^{-1}AL^{-T} \approx I$. Possible choices of S include:

1. The simplest choice of S is $D = \operatorname{diag} A$, then $P = D^{-1/2}$.
2. Another possibility is to choose S as a band matrix with small bandwidth. For example, solving the Poisson equation with the five-point formula, we may take S to be the tridiagonal part of A.

Example 4.37 Consider the tridiagonal system $A\boldsymbol{x} = \boldsymbol{b}$, and let S be defined by:

$$A = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}, \quad S = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix} = LL^T, \quad \text{with} \quad L = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$

The matrix S coincides with A except at the $(1,1)$-entry and happens to have a simple Cholesky factorization $S = LL^T$. Using $P = L^{-1}$, we note that PAP^T has only two distinct eigenvalues, and so the CG method converges in two iterations. Indeed, $PAP^T = P(S + \boldsymbol{e}_1\boldsymbol{e}_1^T)P^T = I + \boldsymbol{w}\boldsymbol{w}^T$ where $\boldsymbol{w} = L^{-1}\boldsymbol{e}_1$ is a rank-1 perturbation of the identity matrix, with all eigenvalues but one equal to 1 (the other one is equal to $1 + \|\boldsymbol{w}\|_2^2$).