STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2019
LECTURE 15

1. RATE OF CONVERGENCE AND SPECTRAL RADIUS

- we kept referring to the spectral radius of the iteration matrix $\rho(B)$ as the rate of convergence but if you think about it, since
$$\|\mathbf{e}^{(k+1)}\|_2 \le \|B\|_2 \|\mathbf{e}^{(k)}\|_2,$$
it should be $\|B\|_2$ that controls the convergence rate
- of course if the iteration matrix B is symmetric, then it makes no difference since $\|B\|_2 = \rho(B)$ but in general they are not equal, in fact is possible for $\rho(B) = 0$ and $B \ne 0$
- the reason is because we look at the *average rate of convergence*
- for any iteration matrix $B \in \mathbb{C}^{n\times n}$ (not necessarily symmetric) and any consistent norm $\|\cdot\|$ (not necessarily submultiplicative), after k steps, we get
$$\|\mathbf{e}^{(k)}\| \le \|B^k\| \|\mathbf{e}^{(0)}\| \tag{1.1}$$
- the average reduction in error per step after k steps is then
$$\left(\frac{\|\mathbf{e}^{(k)}\|}{\|\mathbf{e}^{(0)}\|}\right)^{1/k}$$
and by (1.1) is bounded by
$$\|B^k\|^{1/k}$$
- the spectral radius then drops out when we take limits
$$\lim_{k\to\infty} \|B^k\|^{1/k} = \rho(B),$$
which holds for any consistent norm (i.e., $\|\cdot\|$ satisfies $\|A\mathbf{x}\| \le \|A\|\|\mathbf{x}\|$ for any $A \in \mathbb{C}^{n\times n}$ and any $\mathbf{x} \in \mathbb{C}^n$)

2. METHOD OF STEEPEST DESCENT

- to speed up Richardson method, we consider varying the parameter α from one iteration to the next, i.e.,
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{r}^{(k)} \tag{2.1}$$
where α_k is to be chosen at the kth iteration
- again we will assume that A is symmetric positive definite
- given that
$$\mathbf{r}^{(k+1)} = \mathbf{b} - A\mathbf{x}^{(k+1)} = \mathbf{b} - A\mathbf{x}^{(k)} - \alpha_k A\mathbf{r}^{(k)} = \mathbf{r}^{(k)} - \alpha_k A\mathbf{r}^{(k)}$$
- we wish to choose α_k so that $\mathbf{r}^{(k+1)\mathsf{T}} A^{-1} \mathbf{r}^{(k+1)}$ is minimized
- note that this is just the Mahalanobis norm $\|\mathbf{r}^{(k+1)}\|_{A^{-1}}^2$ (why not minimize the 2-norm $\|\mathbf{r}^{(k+1)}\|_2^2$ instead?)

Date: November 19, 2019, version 1.0. Comments, bug reports: `lekheng@galton.uchicago.edu`.

- now

$$
\begin{aligned}
\mathbf{r}^{(k+1)\mathsf{T}} A^{-1} \mathbf{r}^{(k+1)} &= (\mathbf{r}^{(k)\mathsf{T}} - \alpha_k \mathbf{r}^{(k)\mathsf{T}} A) A^{-1} (\mathbf{r}^{(k)} - \alpha_k A \mathbf{r}^{(k)}) \\
&= \mathbf{r}^{(k)\mathsf{T}} A^{-1} \mathbf{r}^{(k)} - 2\alpha_k \mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)} + \alpha_k^2 \mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)} \qquad (2.2)
\end{aligned}
$$

- to find the minimum, we differentiate with respect to α_k and obtain

$$
\frac{d}{d\alpha_k} \mathbf{r}^{(k+1)\mathsf{T}} A^{-1} \mathbf{r}^{(k+1)} = -2\mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)} + 2\alpha_k \mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)}
$$

  which yields

$$
\hat{\alpha}_k = \frac{\mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)}}{\mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)}}
$$

- note that this is well-defined (denominator not zero) since A is symmetric positive definite
- with this choice of α_k, this method is known as the *method of steepest descent*
- note that

$$
0 < \lambda_{\min}(A) \le \frac{\mathbf{x}^{\mathsf{T}} A \mathbf{x}}{\mathbf{x}^{\mathsf{T}} \mathbf{x}} \le \lambda_{\max}(A)
$$

  and therefore

$$
\frac{1}{\lambda_{\max}(A)} \le \hat{\alpha}_k \le \frac{1}{\lambda_{\min}(A)}
$$

- substituting $\hat{\alpha}_k$ into (2.2) yields

$$
\begin{aligned}
\mathbf{r}^{(k+1)\mathsf{T}} A^{-1} \mathbf{r}^{(k+1)} &= \mathbf{r}^{(k)\mathsf{T}} A^{-1} \mathbf{r}^{(k)} - 2\mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)} \frac{\mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)}}{\mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)}} + \left(\frac{\mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)}}{\mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)}} \right)^2 \mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)} \\
&= \mathbf{r}^{(k)\mathsf{T}} A^{-1} \mathbf{r}^{(k)} - \frac{(\mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)})^2}{\mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)}}
\end{aligned}
$$

  and therefore

$$
\frac{\|\mathbf{r}^{(k+1)}\|_{A^{-1}}^2}{\|\mathbf{r}^{(k)}\|_{A^{-1}}^2} = 1 - \frac{(\mathbf{r}^{(k)\mathsf{T}} \mathbf{r}^{(k)})^2}{(\mathbf{r}^{(k)\mathsf{T}} A^{-1} \mathbf{r}^{(k)})(\mathbf{r}^{(k)\mathsf{T}} A \mathbf{r}^{(k)})}
$$

- the *Kantorovich inequality*, which comes up very often in applications such as optimization and statistics, states that for a symmetric positive definite A,

$$
\frac{\mathbf{x}^{\mathsf{T}} A \mathbf{x} \cdot \mathbf{x}^{\mathsf{T}} A^{-1} \mathbf{x}}{(\mathbf{x}^{\mathsf{T}} \mathbf{x})^2} \le \left(\frac{\sqrt{\kappa} + \sqrt{\kappa}^{-1}}{2} \right)^2, \quad \kappa = \kappa_2(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}
$$

- to prove this inequality, let $\lambda_{\max}(A) = \mu_1$ and $\lambda_{\min}(A) = \mu_n$
  - then $A - \mu_n I$ and $\mu_1 I - A$ are both symmetric positive semidefinite
  - since A^{-1} is also symmetric positive definite, the product of these three matrices

$$
(\mu_1 I - A)(A - \mu_n I) A^{-1}
$$

    is symmetric positive semidefinite and so

$$
\mathbf{x}^{\mathsf{T}} (\mu_1 I - A)(A - \mu_n I) A^{-1} \mathbf{x} \ge 0
$$

    for any $\mathbf{x} \in \mathbb{R}^n$
  - multiplying out the three matrices, we get

$$
\mathbf{x}^{\mathsf{T}} [(\mu_1 + \mu_n) I - \mu_1 \mu_n A^{-1} - A] \mathbf{x} \ge 0
$$

    or

$$
\mu_1 \mu_n \mathbf{x}^{\mathsf{T}} A^{-1} \mathbf{x} \le (\mu_1 + \mu_n) \mathbf{x}^{\mathsf{T}} \mathbf{x} - \mathbf{x}^{\mathsf{T}} A \mathbf{x} \qquad (2.3)
$$

  - now observe that for any $a, b \in \mathbb{R}$, $(a - 2b)^2 \ge 0$ which may be rewritten as

$$
a - b \le \frac{a^2}{4b}
$$

    if $b \neq 0$

- now set $a = (\mu_1 + \mu_n)\mathbf{x}^\mathsf{T}\mathbf{x}$ and $b = \mathbf{x}^\mathsf{T} A\mathbf{x}$ and we see that the right-hand side of (2.3) may be further bounded by

$$(\mu_1 + \mu_n)\mathbf{x}^\mathsf{T}\mathbf{x} - \mathbf{x}^\mathsf{T} A\mathbf{x} \leq \frac{(\mu_1 + \mu_n)^2(\mathbf{x}^\mathsf{T}\mathbf{x})^2}{4\mathbf{x}^\mathsf{T} A\mathbf{x}} \tag{2.4}$$

- combining (2.3) and (2.4) gives

$$\mu_1\mu_n\mathbf{x}^\mathsf{T} A^{-1}\mathbf{x} \leq \frac{(\mu_1 + \mu_n)^2(\mathbf{x}^\mathsf{T}\mathbf{x})^2}{4\mathbf{x}^\mathsf{T} A\mathbf{x}}$$

  and thus

$$\frac{\mathbf{x}^\mathsf{T} A\mathbf{x} \cdot \mathbf{x}^\mathsf{T} A^{-1}\mathbf{x}}{(\mathbf{x}^\mathsf{T}\mathbf{x})^2} \leq \frac{(\mu_1 + \mu_n)^2}{4\mu_1\mu_n} = \left(\frac{\sqrt{\kappa} + \sqrt{\kappa}^{-1}}{2}\right)^2$$

  where $\kappa = \mu_1/\mu_n$

- it follows from the Kantorovich inequality that

$$\frac{\|\mathbf{r}^{(k+1)}\|_{A^{-1}}^2}{\|\mathbf{r}^{(k)}\|_{A^{-1}}^2} \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^2$$

- thus,

$$\frac{\|\mathbf{r}^{(1)}\|_{A^{-1}}}{\|\mathbf{r}^{(0)}\|_{A^{-1}}} \cdot \frac{\|\mathbf{r}^{(2)}\|_{A^{-1}}}{\|\mathbf{r}^{(1)}\|_{A^{-1}}} \cdot \cdots \cdot \frac{\|\mathbf{r}^{(k)}\|_{A^{-1}}}{\|\mathbf{r}^{(k-1)}\|_{A^{-1}}} \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^k$$

  which yields

$$\frac{\|\mathbf{r}^{(k)}\|_{A^{-1}}}{\|\mathbf{r}^{(0)}\|_{A^{-1}}} \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^k$$

- in other words, the rate of convergence is the same as when the parameter α_k is chosen a priori to be

$$\hat{\alpha} = \frac{2}{\mu_1 + \mu_n}$$

- so it would appear that we might as well have used Richardson's method in the first place
- but that's not quite the case — the problem is that we must know μ_1 and μ_n in order to determine the optimal $\hat{\alpha}$
- steepest descent does not require us to know μ_1 and μ_n
- however the price we pay for steepest descent is that we need to compute α_k at each step
- if you know some optimization, the steepest descent method described above is the same as applying the steepest descent method in continuous optimization to the problem

$$\min \frac{1}{2}\mathbf{x}^\mathsf{T} A\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{x}$$

3. Chebyshev iteration

- again we are interested in solving $A\mathbf{x} = \mathbf{b}$
- let us rewrite the steepest descent iteration (2.1) in the form

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k\mathbf{r}^{(k)} \\ &= (I - \alpha_k A)\mathbf{x}^{(k)} + \alpha_k\mathbf{b} \end{aligned}$$

- this time, instead of picking only α_k to minimize some quantity at the kth step, we will pick $\alpha_0, \alpha_1, \dots, \alpha_k$ simultaneously to minimize some quantity at the kth step

- since the exact solution $\mathbf{x}$ satisfies
$$\mathbf{x} = (I - \alpha_k A)\mathbf{x} + \alpha_k \mathbf{b},$$
it follows that
$$\mathbf{e}^{(k+1)} = (I - \alpha_k A)\mathbf{e}^{(k)}$$
where $\mathbf{e}^{(k)} = \mathbf{x} - \mathbf{x}^{(k)}$
- so we have
$$\begin{aligned} \mathbf{e}^{(1)} &= (I - \alpha_0 A)\mathbf{e}^{(0)} \\ &\vdots \\ \mathbf{e}^{(k)} &= (I - \alpha_{k-1} A)(I - \alpha_{k-2} A)\cdots(I - \alpha_0 A)\mathbf{e}^{(0)} \end{aligned}$$
- in other words,
$$\mathbf{e}^{(k)} = P_k(A)\mathbf{e}^{(0)}$$
where
$$P_k(A) = (I - \alpha_{k-1} A)(I - \alpha_{k-2} A)\cdots(I - \alpha_0 A).$$
is a polynomial of degree k
- by the Cayley–Hamilton theorem, the minimal polynomial $\psi(x)$ of A has the following property:
$$\psi(A) = \prod_{k=0}^{d-1} (A - \mu_k I) = 0$$
where d is the number of distinct eigenvalues μ_k of A, by Homework **1**, Problem **5**(a)
- in other words
$$\prod_{k=0}^{d-1} \left(I - \frac{1}{\mu_k} A \right) = (-1)^d \mu_0 \cdots \mu_{d-1} \psi(A) = 0$$
so we could choose $\alpha_k = 1/\mu_k$ and
$$P_d(A) = \prod_{k=0}^{d-1} \left(I - \frac{1}{\mu_k} A \right)$$
but this is a bad choice because we almost never know the eigenvalues of A and even if we do, this choice is unstable because μ_k can vary immensely in magnitude
- however, no matter what P_k we have, it is always true that
$$\frac{\|\mathbf{e}^{(k)}\|}{\|\mathbf{e}^{(0)}\|} \le \|P_k(A)\|_2,$$
which allows us to use approximation theory to find a suitable P_k as follows
- if $A = Q\Lambda Q^\mathsf{T}$ where $\Lambda = \operatorname{diag}(\mu_1, \dots, \mu_n)$, then $P_k(A) = QP_k(\Lambda)Q^\mathsf{T}$, and therefore
$$\frac{\|\mathbf{e}^{(k)}\|}{\|\mathbf{e}^{(0)}\|} \le \|P_k(\Lambda)\|_2$$
- since
$$P_k(\Lambda) = \begin{bmatrix} P_k(\mu_1) & & \\ & \ddots & \\ & & P_k(\mu_n) \end{bmatrix},$$
it follows that
$$\|P_k(\Lambda)\|_2 = \max_{1 \le i \le n} |P_k(\mu_i)|$$

- so we want a polynomial P_k that minimizes

$$\max_{1\le i\le n} |P_k(\mu_i)|$$

but this is trivial since we can pick $P_k(x) = 0$
- so we need to restrict the coefficients of P_k and we will impose the condition

$$P_k(1) = 1$$

- if $P_k(x) = a_k x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0$, then $P_k(1) = 1$ means

$$a_k + a_{k-1} + \cdots + a_1 + a_0 = 1$$

- so we want to solve the problem

$$\min_{P_k(1)=1} \max_{1\le i\le n} |P_k(\mu_i)|$$

but this is still too difficult
- so we are content with an upper bound

$$\min_{P_k(1)=1} \max_{1\le i\le n} |P_k(\mu_i)| \le \min_{P_k(1)=1} \max_{\mu_n\le\mu\le\mu_1} |P_k(\mu)| \le \min_{P_k(1)=1} \max_{\alpha\le\mu\le\beta} |P_k(\mu)|$$

where $\alpha \le \mu_n \le \mu_1 \le \beta$
- ideally we want $\alpha = \mu_n$ and $\beta = \mu_1$ but often we don't have the eigenvalues but only lower and upper bounds (e.g., from Gerschgorin theorem)
- the solution to this problem is given by the *Chebyshev polynomials*, suitably normalized so that it is monic
- the Chebyshev polynomial of degree k is most easily defined to be the polynomial expression that gives the expansion of $\cos(k\theta)$ in terms of $\cos(\theta)$, formally,

$$C_k(\cos x) = \cos kx$$

or

$$C_k(x) = \cos(k \cos^{-1}(x))$$

- for example,

$$C_0(x) = 1, \quad C_1(x) = x, \quad C_2(x) = 2x^2 - 1, \quad C_3(x) = 4x^3 - 3x, \quad C_4(x) = 8x^4 - 8x^2 + 1$$