MATH 5524 · MATRIX THEORY

Pledged Problem Set 1

Posted Friday 17 February 2017. Due by 5pm on Friday 24 February 2017.

Complete any four problems, 25 points each. *You are welcome to complete more problems if you like, but specify which four you want to be graded.*

Rules: On this pledged problem set, you are welcome to use course notes (those you have taken, and those posted to the web), books (hard copies or electronic), MATLAB, Mathematica, etc. You may post questions to Piazza, and ask questions of the instructor during office hours. You are *not allowed* to more generally search the web for answers, or discuss the problems with anyone aside from the instructor. Please write out and sign the pledge ("As a Hokie, I will conduct myself with honor and integrity at all times. I will not lie, cheat or steal, nor will I acccept the actions of those who do.") on your assignment. **Pledged problem sets will not be accepted late unless you have made a previous arrangement with the instructor.**

1. Recall that the norm of a matrix is defined as

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|},$$

where $\|\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x}$.

   (a) Given any $\mathbf{A} \in \mathbb{C}^{m\times n}$, show that $\mathbf{A}^*\mathbf{A}$ is Hermitian and has nonnegative eigenvalues.

   (b) Given any $\mathbf{A} \in \mathbb{C}^{m\times n}$, show that $\|\mathbf{A}\|$ equals the square root of the largest eigenvalue of $\mathbf{A}^*\mathbf{A}$.

   (c) Show that if $\mathbf{A}$ is Hermitian, then $\|\mathbf{A}\|$ equals the maximum magnitude of the eigenvalues of $\mathbf{A}$, $\|\mathbf{A}\| = \max_{\lambda\in\sigma(\mathbf{A})} |\lambda|$.

   (d) Show by a 2×2 example that the conclusion of part (c) need not hold when $\mathbf{A}$ is non-Hermitian.

2. Suppose $\mathbf{A}, \mathbf{E} \in \mathbb{C}^{n\times n}$ are Hermitian matrices, and let $\lambda_k(\cdot)$ denote the kth eigenvalue of a Hermitian matrix ($\lambda_1(\mathbf{A}) \le \lambda_2(\mathbf{A}) \le \cdots \le \lambda_n(\mathbf{A})$). Use the Courant–Fischer minimax characterization of eigenvalues to show that

$$\lambda_k(\mathbf{A}) + \lambda_1(\mathbf{E}) \le \lambda_k(\mathbf{A}+\mathbf{E}) \le \lambda_k(\mathbf{A}) + \lambda_n(\mathbf{E}).$$

[Wilkinson; Parlett]

3. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a Hermitian matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

   (a) Suppose that $\mathbf{x} \in \mathbb{C}^n$ approximates an eigenvector up to some accuracy, $\varepsilon \ll 1$, i.e., $\mathbf{x} = \mathbf{u}_j + \mathbf{r}$ for $\varepsilon := \|\mathbf{r}\| \ll 1$. Show that the Rayleigh quotient gives an approximation to the eigenvalue that is accurate to $\mathcal{O}(\varepsilon^2)$ as $\varepsilon \to 0$:

$$\frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \lambda_j + \mathcal{O}(\varepsilon^2).$$

   (b) Illustrate this result with some numerical experiments involving the middle eigenvalue of

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

   (Use MATLAB etc. to compute the eigenvalues and eigenvectors of $\mathbf{A}$, then form $\mathbf{x} = \mathbf{u}_2 + \mathbf{r}$ for a several different $\mathbf{r}$, e.g., $\varepsilon = \|\mathbf{r}\| = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$.)

4. For real symmetric (hence Hermitian) $\mathbf{A} \in \mathbb{R}^{n\times n}$, define the function $r : \mathbb{R}^n \setminus \{\mathbf{0}\} \to \mathbb{R}$ by

$$r(\mathbf{x}) = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}}.$$

   (a) For nonzero $\mathbf{x} \in \mathbb{R}^n$, compute the gradient $\nabla r(\mathbf{x})$.

   (b) When is the gradient zero?

   (This suggests a way to computing eigenvalues: use Newton's method to compute stationary points of the Rayleigh quotient; one typically optimizes $\mathbf{x}^*\mathbf{A}\mathbf{x}$ subject to a normalization constraint, like $\|\mathbf{x}\| = 1$. For a survey of the long history of such methods, see the forthcoming *SIAM Review* paper by Tapia, Dennis, and Schäfermeyer.)

5. (a) Derive a formula for the (j,k) entry of the pth power of the Jordan block

$$\mathbf{J} = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix} \in \mathbb{C}^{n\times n}.$$

   (b) Use the Jordan form to prove that for any $\mathbf{A} \in \mathbb{C}^{n\times n}$,

$$\lim_{k\to\infty} \|\mathbf{A}^k\|^{1/k} = \max_{\lambda\in\sigma(\mathbf{A})} |\lambda|.$$

   The quantity on the right, often denoted $\rho(\mathbf{A})$, is called the *spectral radius* of $\mathbf{A}$. In particular, you have just proved that $\mathbf{A}^k \to 0$ if and only if $\rho(\mathbf{A}) < 1$.
   Hint: You might find it helpful to first prove that, for any matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$,

$$\|\mathbf{B}\| \le C(n) \max_{j,k} |b_{j,k}|,$$

   from some scalar value $C(n)$ that depends on n but not on $\mathbf{B}$.

6. Show that the Cauchy Interlacing Theorem is sharp when $m = n-1$, in the sense that for any distinct real numbers

$$\lambda_1 < \lambda_2 < \cdots < \lambda_n$$

   and

$$\theta_1 < \theta_2 < \cdots < \theta_{n-1}$$

   that satisfy the (sharp) interlacing conditions

$$\lambda_k < \theta_k < \lambda_{k+1}$$

   for $k = 1, \ldots, n-1$, one can construct a Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ of the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{H} & \mathbf{b} \\ \mathbf{b}^* & \delta \end{bmatrix}$$

   $\mathbf{b} \in \mathbb{C}^{n-1}$, $\delta \in \mathbb{C}$ such that

$$\sigma(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$$

   and

$$\sigma(\mathbf{H}) = \{\theta_1, \ldots, \theta_{n-1}\}.$$

   (Specify how to construct $\mathbf{H}$, $\mathbf{b}$, and δ.)

   [Parlett]

7. The Courant–Fischer minimax characterization describes the mth eigenvalue of the Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ as

$$\lambda_m = \min_{\dim(\mathcal{S})=m} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}.$$

It follows that for *any* m dimensional subspace $\mathcal{S}$ we choose,

$$\lambda_m \le \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}.$$

This provides a natural upper bound on λ_m. Let $\mathbf{q}_1, \ldots, \mathbf{q}_m$ be an orthonormal basis for $\mathcal{S}$, and write

$$\mathbf{Q} = [\,\mathbf{q}_1 \quad \cdots \quad \mathbf{q}_m\,] \in \mathbb{C}^{n\times m}.$$

Then any $\mathbf{x} \in \mathcal{S}$ can be written as $\mathbf{Q}\mathbf{s}$ for some $\mathbf{s} \in \mathbb{C}^m$, and so

$$\lambda_m \le \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \max_{\mathbf{s}\in\mathbb{C}^m} \frac{\mathbf{s}^*\mathbf{Q}^*\mathbf{A}\mathbf{Q}\mathbf{s}}{\mathbf{s}^*\mathbf{Q}^*\mathbf{Q}\mathbf{s}} = \max_{\mathbf{s}\in\mathbb{C}^m} \frac{\mathbf{s}^*(\mathbf{Q}^*\mathbf{A}\mathbf{Q})\mathbf{s}}{\mathbf{s}^*\mathbf{s}} = \theta_m,$$

where θ_m is the *largest eigenvalue* of $\mathbf{H} := \mathbf{Q}^*\mathbf{A}\mathbf{Q} \in \mathbb{C}^{m\times m}$. (Here we have used the orthonormality of $\mathbf{q}_1, \ldots, \mathbf{q}_m$ to conclude that $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}$.) Notice that the (j,k) entry of $\mathbf{H}$ is given by

$$h_{j,k} = \mathbf{q}_j^*\mathbf{A}\mathbf{q}_k.$$

(a) Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ have the form

$$\mathbf{A} = \begin{bmatrix} 1 & 0.1 & & \\ 0.1 & 2 & \ddots & \\ & \ddots & \ddots & 0.1 \\ & & 0.1 & n \end{bmatrix}.$$

Let $\mathbf{q}_1, \ldots, \mathbf{q}_m$ denote the first m columns of the $n \times n$ identity matrix. Take $n = 32$, and use the above procedure to produce (in MATLAB) estimates θ_m to λ_m for various values of m. (That is, for a range of m values, build $\mathbf{H} \in \mathbb{C}^{m\times m}$, find its largest eigenvalue θ_m, and compare that to the eigenvalue λ_m of $\mathbf{A}$.)

(b) Repeat part (a) (using the same vectors $\mathbf{q}_1, \ldots, \mathbf{q}_m$) for the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix} \in \mathbb{C}^{32\times 32}.$$

(c) Why do you think this procedure gave good estimates in part (a) but not in part (b)? (Full credit will be given for any plausible explanation.)

(d) Parts (a) and (b) of this problem were rather artificial. This part gives a better indication of how one could use this bounding technique in practice – but this requires a little background.

To solve the differential equation $-u''(x) = f(x)$ for $x \in [0, \pi]$ with $u(0) = u(\pi) = 0$, one can write the solution as a linear combination of eigenfunctions q_k that satisfy $-q_k''(x) = \lambda_k q_k(x)$ for $x \in [0, \pi]$ and $q_k(0) = q_k(\pi) = 0$. As one learns in a basic PDE class, these eigenfunctions are simply

$$q_k(x) = \frac{2}{\pi}\sin(kx).$$

One can use the same approach to solve more complicated equations. For example, consider

$$-u''(x) + (x\sin(x) - 1)u(x) = f(x)$$

again with $u(0) = u(\pi) = 0$. To solve this equation, one could use the eigenvalues λ_k and eigenfunctions $u_k(x)$ that satisfy

$$-u_k''(x) + (x\sin(x) - 1)u_k(x) = \lambda_k u_k(x)$$

for $x \in [0, \pi]$ and $u_k(0) = u_k(\pi) = 0$. In general, the variable coefficient $(x\sin(x)-1)$ makes these eigenvalues and eigenfunctions difficult to compute. We will use the Courant–Fischer minimax characterization to *estimate* λ_k.

We are working with infinite dimensional problems (in the space $L^2(0,\pi)$, rather than $\mathbb{C}^n$), so we need a different definition of inner product. The inner product of functions $f, g \in L^2(0,\pi)$ is defined as

$$(f, g) = \int_0^\pi f(x)g(x)\,\mathrm{d}x.$$

We wish to approximate the eigenvalues of the linear operator T defined (for u sufficiently differentiable) by

$$Tu := -u'' + (x\sin(x) - 1)u.$$

In this setting, the Courant–Fischer characterization gives

$$\lambda_m = \min_{\substack{\mathcal{S}\subset \mathrm{Dom}(T)\\ \dim(\mathcal{S})=m}} \max_{u\in\mathcal{S}} \frac{(Tu, u)}{(u, u)}.$$

(Here $\mathrm{Dom}(T)$ is the domain of T, a subset of $L^2(0,\pi)$.) For a particular subspace, say $\mathcal{S} = \mathrm{span}\{q_1, \ldots, q_m\}$ for orthonormal $q_1, \ldots, q_m$ (that is, $(q_j, q_k) = 0$ if $j \neq k$, and $(q_j, q_j) = 1$),

$$\lambda_m \leq \max_{u\in\mathcal{S}} \frac{(Tu, u)}{(u, u)} = \theta_m,$$

where θ_m is the largest eigenvalue of the $m \times m$ *matrix* $\mathbf{H}$ with

$$h_{j,k} = (Tu_j, u_k) = \int_0^\pi \big(-u_j''(x) + (x\sin(x) - 1)u_j(x)\big)u_k(x)\,\mathrm{d}x.$$

One particularly interesting choice for the functions $q_1, \ldots, q_m$ is the set of orthonormal eigenfunctions associated with the easier problem $-u'' = f$ (which doesn't have the complicating variable coefficient term),

$$q_k(x) = \frac{2}{\pi}\sin(kx), \qquad k = 1, \ldots, m.$$

With this choice, one has

$$h_{j,k} = \begin{cases} \dfrac{(1 - k^2 + 4k^4)}{4k^2 - 1}, & j = k; \\[2ex] \dfrac{-(1 + k + k^2)}{4k^2 + 4k}, & j = k + 1; \\[2ex] \dfrac{-(1 - k + k^2)}{4k^2 - 4k}, & j = k - 1; \\[2ex] \dfrac{4(-1)^{j+k+1}jk}{(j^2 - 1)^2 - 2(j^2 + 1)k^2 + k^4}, & \text{otherwise.} \end{cases}$$

For $m = 1, \ldots, 20$, construct $\mathbf{H} \in \mathbb{C}^{m\times m}$ and compare its largest eigenvalue θ_m to the true eigenvalue λ_m of T given in the table below. (The "true" eigenvalues were computed using a different high-precision strategy.)

k	λ_k
1	1.2708040093...
2	4.0947832505...
3	9.0367214840...
4	16.0211554874...
5	25.0136704727...
6	36.0095471867...
7	49.0070396015...
8	64.0054028151...
9	81.0042762020...
10	100.0034680458...
11	121.0028688359...
12	144.0024123528...
13	169.0020566555...
14	196.0017741396...
15	225.0015460372...
16	256.0013592296...
17	289.0012043240...
18	324.0010744527...
19	361.0009645003...
20	400.0008705941...

(e) Did you notice that in all of these problems, as m got bigger the smallest eigenvalue, θ_1, of $\mathbf{H} \in \mathbb{C}^{m\times m}$ became a better and better upper bound for the eigenvalue λ_1 of $\mathbf{A}$? Use the Courant–Fischer minimax characterization to explain why this is the case.