CAAM 440 · APPLIED MATRIX ANALYSIS

Problem Set 5

Posted Monday 2 April 2012. Due Monday 9 April 2012.

Complete any four problems (1–8), 25 points each.

Recall the definitions of the *field of values* of $\mathbf{A} \in \mathbb{C}^{n\times n}$,

$$W(\mathbf{A}) = \left\{ \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} : \mathbf{x} \in \mathbb{C}^n \right\},$$

and the *ε-pseudospectrum* for $\varepsilon > 0$,

$$\begin{aligned} \sigma_\varepsilon(\mathbf{A}) &= \{z \in \mathbb{C} : \|(z\mathbf{I} - \mathbf{A})^{-1}\| > 1/\varepsilon\} \\ &= \{z \in \mathbb{C} : z \in \sigma(\mathbf{A} + \mathbf{E}) \text{ for some } \mathbf{E} \in \mathbb{C}^{n\times n} \text{ with } \|\mathbf{E}\| < \varepsilon\}. \end{aligned}$$

1. Recall that the derivative of a distinct eigenvalue λ_k of a generic matrix $\mathbf{A}(t) = \mathbf{A}_0 + t\mathbf{E}$ is given by

$$\lambda_k'(t)\Big|_{t=0} = \frac{\widehat{\mathbf{v}}_k^*\mathbf{E}\mathbf{v}_k}{\widehat{\mathbf{v}}_k^*\mathbf{v}_k},$$

where $\mathbf{v}_k$ and $\widehat{\mathbf{v}}_k$ are the corresponding right and left eigenvectors of $\mathbf{A}_0$.

   (a) Suppose $\mathbf{A} \in \mathbb{C}^{n\times n}$ has singular value decomposition $\mathbf{A} = \sum_{j=1}^n s_j \mathbf{u}_j \mathbf{v}_j^*$, and consider the matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^* & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{2n\times 2n}.$$

   Show that $\sigma(\mathbf{H}) = \{\pm s_1, \ldots, \pm s_n\}$. What are the corresponding right and left eigenvectors of $\mathbf{H}$?

   (b) Use the result stated at the beginning of this problem to derive a formula for the derivative of a distinct nonzero *singular value* of $\mathbf{A} + t\mathbf{E}$.

   (For your interest: This result forms the basis of *curve tracing* algorithms for computing $\sigma_\varepsilon(\mathbf{A})$. Given a point on the boundary of $\sigma_\varepsilon(\mathbf{A})$, one seeks to follow the contour where $\|(z\mathbf{I} - \mathbf{A})^{-1}\| = 1/\varepsilon$.)

2. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding right eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and left eigenvectors $\widehat{\mathbf{v}}_1, \ldots, \widehat{\mathbf{v}}_n$.

   In class we saw that the derivative of the eigenvalue $\lambda_k(t)$ of $\mathbf{A} + t\mathbf{E}$ with respect to t (for $\|\mathbf{E}\| = 1$) at $t = 0$ is bounded in magnitude by the *eigenvalue condition number*

$$\kappa(\lambda_k) = \frac{\|\widehat{\mathbf{v}}_k\|\|\mathbf{v}_k\|}{|\widehat{\mathbf{v}}_k^*\mathbf{v}_k|}.$$

   Consider the matrix

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

   where $\alpha \geq 1$ is a real constant.

   (a) Compute the eigenvalue condition numbers $\kappa(\lambda_1)$ and $\kappa(\lambda_2)$ for eigenvalues $\lambda_1 = -1$ and $\lambda_2 = -2$.

   (b) Compute the matrix exponential $\mathrm{e}^{t\mathbf{A}}$.

(c) What is $\lim_{t\to\infty} \|e^{t\mathbf{A}}\|$?
(You do not need to explicitly compute the norm to answer this question.)

(d) For fixed $\alpha > 4$, find the value of $t \geq 0$ that maximizes the largest entry in $e^{t\mathbf{A}}$.

(e) What does your solution to (d) suggest about the general behavior of $\|e^{t\mathbf{A}}\|$ for $t \in (0, \infty)$ with $\alpha > 4$? How does $\max_{t\geq 0} \|e^{t\mathbf{A}}\|$ depend on α?
(Again, you do not need to explicitly compute any matrix norms.)

3. Prove the following two results that bound the ε-pseudospectrum of $\mathbf{A}$. Roughly speaking, they can be interpreted to mean, "the ε-pseudospectrum cannot be significantly larger than the field of values or the Gerschgorin disks, beyond a factor of order ε."

(a) Prove that for all $\mathbf{A} \in \mathbb{C}^{n\times n}$,

$$\sigma_\varepsilon(\mathbf{A}) \subseteq W(\mathbf{A}) + D(\varepsilon),$$

where $D(\varepsilon) = \{z \in \mathbb{C} : |z| < \varepsilon\}$ is the open disk of radius ε.
(Addition of sets is defined in the usual way: Given sets S_1 and S_2, the sum $S_1 + S_2 = \{s_1 + s_2 : s_1 \in S_1, s_2 \in S_2\}$ is the set of all pairwise sums between elements in S_1 and S_2.)

(b) Use Gerschgorin's theorem to show that for all $\mathbf{A} \in \mathbb{C}^{n\times n}$

$$\sigma_\varepsilon(\mathbf{A}) \subseteq \bigcup_{j=1}^{n} D(a_{j,j}, r_j + \varepsilon n)$$

where $a_{j,k}$ denotes the (j,k) entry of $\mathbf{A}$, r_j denotes the absolute sum of the off-diagonal entries in the jth row of $\mathbf{A}$,

$$r_j := \sum_{k=1, k\neq j}^{n} |a_{j,k}|,$$

and $D(c, r) = \{z \in \mathbb{C} : |z - c| < r\}$ is the open disk in $\mathbb{C}$ centered at $c \in \mathbb{C}$ with radius $r > 0$.
(In fact, one can prove a stronger bound with $D(a_{j,j}, r_j + \varepsilon n)$ replaced by $D(a_{j,j}, r_j + \varepsilon\sqrt{n})$, but you are not required to obtain this sharper result.)

4. The *numerical radius* of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is the largest magnitude of any point in the field of values:

$$\mu(\mathbf{A}) := \max_{z\in W(\mathbf{A})} |z|.$$

Prove that

$$\tfrac{1}{2}\|\mathbf{A}\| \leq \mu(\mathbf{A}) \leq \|\mathbf{A}\|.$$

Hint: recall that $\mathbf{A}$ can be written as the sum of a Hermitian matrix and a skew-Hermitian matrix.

5. Throughout this semester we have considered the standard (Euclidean) inner product $\mathbf{y}^*\mathbf{x}$, where $\mathbf{y}^*$ denotes the conjugate-transpose of $\mathbf{y}$. This is but one example of a much broader class of potential inner products; these functions play an essential role in applications, where often the inner product provides a way to measure the energy in a system. This problem asks you to investigate how the usual inner product relates to these more sophisticated alternatives.

(a) Let $\mathbf{H} \in \mathbb{C}^{n\times n}$ be a Hermitian positive definite matrix, and define the $\mathbf{H}$-inner product to be

$$\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{H}} := \mathbf{y}^*\mathbf{H}\mathbf{x}.$$

Prove that the $\mathbf{H}$-inner product satisfies the fundamental axioms required of all inner products:

i. $\langle \mathbf{x}, \mathbf{x}\rangle_{\mathbf{H}} \geq 0$ for all $\mathbf{x} \in \mathbb{C}^n$ and $\langle \mathbf{x}, \mathbf{x}\rangle_{\mathbf{H}} = 0$ if and only if $\mathbf{x} = \mathbf{0}$;
ii. $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{H}} = \overline{\langle \mathbf{y}, \mathbf{x}\rangle_{\mathbf{H}}}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ (complex conjugate symmetry);
iii. $\langle \mathbf{x}+\mathbf{y}, \mathbf{z}\rangle_{\mathbf{H}} = \langle \mathbf{x}, \mathbf{z}\rangle_{\mathbf{H}} + \langle \mathbf{y}, \mathbf{z}\rangle_{\mathbf{H}}$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^n$;
iv. $\langle \alpha\mathbf{x}, \mathbf{y}\rangle_{\mathbf{H}} = \alpha\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{H}}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$.

(b) In the standard inner product, we often deal with the *conjugate transpose*, $\mathbf{A}^* = \overline{\mathbf{A}}^{\mathrm{T}}$. This definition is inherently tied to our inner product. We define the $\mathbf{H}$-*adjoint* of $\mathbf{A}$ to be the matrix $\mathbf{A}^\sharp$ such that, for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$,

$$\langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle_{\mathbf{H}} = \langle \mathbf{x}, \mathbf{A}^\sharp\mathbf{y}\rangle_{\mathbf{H}}.$$

Determine a formula for $\mathbf{A}^\sharp$ as a combination of $\mathbf{A}$, $\mathbf{A}^*$, and/or $\mathbf{H}$.

(c) We defined a matrix to be *Hermitian* (or *self-adjoint*) provided $\mathbf{A}^* = \mathbf{A}$. Similarly, we define $\mathbf{A}$ to be $\mathbf{H}$-*self-adjoint* provided $\mathbf{A}^\sharp = \mathbf{A}$. Determine a simple condition involving $\mathbf{A}$, $\mathbf{A}^*$, and/or $\mathbf{H}$ to test whether a matrix is $\mathbf{H}$-self-adjoint.

(Similar conditions can be defined to test whether $\mathbf{A}$ is $\mathbf{H}$-*unitary* and $\mathbf{H}$-*normal*, but you do not need to derive these.)

(d) The $\mathbf{H}$-inner product leads to a $\mathbf{H}$-norm for vectors,

$$\|\mathbf{x}\|_{\mathbf{H}} := \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle_{\mathbf{H}}}$$

and an induced $\mathbf{H}$-matrix norm for $\mathbf{A} \in \mathbb{C}^{n\times n}$,

$$\|\mathbf{A}\|_{\mathbf{H}} := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_{\mathbf{H}}}{\|\mathbf{x}\|_{\mathbf{H}}}.$$

Suppose you have a function for computing the usual matrix norm induced by the standard Euclidean inner product (e.g., the `norm` command in MATLAB), and you wanted to use this norm routine to compute $\mathbf{H}$-norm of a matrix.

Determine some matrix $\mathbf{B}$ (which may involve matrices like $\mathbf{A}$, $\mathbf{H}$, $\mathbf{H}^{1/2}$, etc.) such that

$$\|\mathbf{A}\|_{\mathbf{H}} = \|\mathbf{B}\|,$$

where the norm on the right is the usual norm we have been dealing with all semester.

(e) Prove that if $\langle\cdot,\cdot\rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ is any function that obeys the four axioms in part (a), then there exists a Hermitian positive definite matrix $\mathbf{H} \in \mathbb{C}^{n\times n}$ such that for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$,

$$\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{y}^*\mathbf{H}\mathbf{x}.$$

6. The exercise is designed to give you some intuition for transient growth in a dynamical system.

The *Leslie matrix* arises in models for the (female) population of a given species with fixed birth rates and survivability levels. The population is divided into n brackets of y-years each, and an average member of bracket k gives birth to $b_k \geq 0$ females in the next y years, and has probability $s_k \in [0,1]$ of surviving the next y years. Letting $p_k^{(j)}$ denote the population in the kth bracket in the jth generation, we see that the population evolves according to the matrix equation (e.g., for $n = 5$)

$$\begin{bmatrix} p_1^{(j+1)} \\ p_2^{(j+1)} \\ p_3^{(j+1)} \\ p_4^{(j+1)} \\ p_5^{(j+1)} \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 \\ s_1 & & & & \\ & s_2 & & & \\ & & s_3 & & \\ & & & s_4 & \end{bmatrix} \begin{bmatrix} p_1^{(j)} \\ p_2^{(j)} \\ p_3^{(j)} \\ p_4^{(j)} \\ p_5^{(j)} \end{bmatrix},$$

with unspecified entries equal to zero. (We presume the mortality of the last age bracket.)

Denote the matrix as $\mathbf{A}$, so that $\mathbf{p}^{(j+1)} = \mathbf{A}\mathbf{p}^{(j)}$, and hence $\mathbf{p}^{(j)} = \mathbf{A}^j\mathbf{p}^{(0)}$.

Earlier on this problem set, we considered how transient growth in matrix powers is linked to the sensitivity of eigenvalues to perturbations in the matrix entries. This problem is designed to reinforce this connection in the context of physically meaningful transient behavior.

(a) Design a set of parameters $b_1, \ldots, b_5 > 0$ and $s_1, \ldots, s_4 > 0$ (for $n = 5$) so that the population will eventually decay in size to zero ($\mathbf{A}^j \to \mathbf{0}$ as $j \to \infty$), but this will be preceded by a period of significant transient growth in the population, where $\mathbf{p}^{(j)} \gg \mathbf{p}^{(0)}$. (Think about what kind of birth and survivability values might suggest this demographic pattern.)

(b) Plot your population for a number of generations to demonstrate the transient growth and eventual decay. (You may modify `pop.m` from the class website.)

(c) Show that this transient growth coincides with sensitivity of the eigenvalues of your matrix. You may choose to do this in several different ways: compute the condition numbers of the eigenvalues; show that the numerical range of $\mathbf{A}$ contains points of significantly larger magnitude than the spectral radius; or download EigTool for MATLAB (see the link on the class website) and use it to plot the pseudospectra of $\mathbf{A}$.

7. We have often discussed the linear, constant-coefficient dynamical system $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, whose solutions $\mathbf{x}(t)$ decay to zero as $t \to \infty$, provided all eigenvalues of $\mathbf{A}$ have negative real part.

Is the same true for *variable-coefficient* problems? Suppose $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$, and that all eigenvalues of the matrix $\mathbf{A}(t) \in \mathbb{C}^{n\times n}$ have negative real part for all $t \geq 0$. Is this enough to guarantee that $\mathbf{x}(t) \to \mathbf{0}$ as $t \to \infty$? This problem asks you to explore this possibility.

(a) Consider the matrix

$$\mathbf{U}(t) = \begin{bmatrix} \cos(\gamma t) & \sin(\gamma t) \\ -\sin(\gamma t) & \cos(\gamma t) \end{bmatrix}.$$

Show that $\mathbf{U}(t)$ is unitary for any fixed real values of γ and t.

(b) Now consider the matrix $\mathbf{A}(t) \in \mathbb{C}^{n\times n}$ defined by

$$\mathbf{A}(t) = \mathbf{U}(t)\mathbf{A}_0\mathbf{U}(t)^*, \qquad \mathbf{A}_0 = \begin{bmatrix} -1 & \alpha \\ 0 & -2 \end{bmatrix}.$$

(Notice that $\mathbf{A}_0$ is the matrix that featured in Problem 2.)

Explain why $\sigma(\mathbf{A}(t)) = \sigma(\mathbf{A}_0)$, $W(\mathbf{A}(t)) = W(\mathbf{A}_0)$, and $\sigma_\varepsilon(\mathbf{A}(t)) = \sigma_\varepsilon(\mathbf{A}_0)$ for all real t.

(In other words, show the spectrum, field of values, and ε-pseudospectra are identical for all t.)

(c) Now we wish to investigate the behavior of the dynamical system

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t). \qquad (*)$$

Define $\mathbf{y}(t) = \mathbf{U}(t)^*\mathbf{x}(t)$. Explain why equation $(*)$ implies that

$$\mathbf{y}'(t) = \big(\mathbf{A}_0 + (\mathbf{U}(t)^*)'\mathbf{U}(t)\big)\mathbf{y}(t). \qquad (**)$$

(Here $(\mathbf{U}(t)^*)' \in \mathbb{C}^{n\times n}$ denotes the t-derivative of the conjugate-transpose of $\mathbf{U}(t)$.)

(d) Compute $(\mathbf{U}(t)^*)'\mathbf{U}(t)$. Does this matrix vary with t?

(e) Define the matrix

$$\widehat{\mathbf{A}} = \mathbf{A}_0 + (\mathbf{U}(t)^*)'\mathbf{U}(t).$$

Fix $\alpha = 7$. Plot the (real) eigenvalues of $\widehat{\mathbf{A}}$ (e.g., in MATLAB) as a function of $\gamma \in [0, 7]$. Do the eigenvalues of $\widehat{\mathbf{A}}$ fall in the left half of the complex plane for all γ?

(f) Calculate the eigenvalues of $\widehat{\mathbf{A}}$ for $\gamma = 1$ and $\alpha = 7$.
What can be said of solutions $\mathbf{y}(t)$ to the system $(**)$ as $t \to \infty$ for these α, γ values?
What then can be said of $\|\mathbf{x}(t)\| = \|\mathbf{U}(t)\mathbf{y}(t)\|$, where $\mathbf{x}(t)$ solves $(*)$, as $t \to \infty$?
How does this compare to the similar constant coefficient problem $\mathbf{x}'(t) = \mathbf{A}_0\mathbf{x}(t)$ (where we have seen that $\mathbf{A}_0$ has the same spectrum, field of values, and pseudospectra as $\mathbf{A}(t)$ for all t)?

[Examples of this sort were perhaps first constructed by Vinograd; see Dekker and Verwer, or Lambert.]

8. Prove or disprove Crouzeix's conjecture: For any matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ and any function f analytic on $W(\mathbf{A})$,

$$\|f(\mathbf{A})\| \leq 2 \max_{z\in W(\mathbf{A})} |f(z)|.$$

[This is a challenge problem: a solution has not yet been discovered!]