ACM/CMS 107 Linear Analysis & Applications **Fall 2016**

Assignment 2: Pseudospectra and Markov Chains **Due:** 27th October 2016

For this assignment the EigTool MATLAB software will be of use:

http://github.com/eigtool/eigtool

Additionally, the following MATLAB data will be required:

http://mdunlop.org/cms107/assignment2.zip

In this assignment we will always work with the 2-norm unless stated otherwise.

Problem 1. Pseudospectra (25 points)

Given a matrix $A \in \mathbb{C}^{n\times n}$, we define its spectrum $\Lambda(A)$ by

$$\Lambda(A) = \{\lambda \in \mathbb{C} : \lambda I - A \text{ is not invertible}\}.$$

That is, $\Lambda(A)$ is the set of eigenvalues of A. Throughout this assignment we will study the ε-pseudospectrum of A, denoted $\Lambda_\varepsilon(A)$. This is a subset of the complex plane which contains $\Lambda(A)$, and in some sense can be thought of as the set of 'approximate eigenvalues' of A.

(a) Look up at least two equivalent definitions of the ε-pseudospectrum of a matrix online and state them.

(b) Let $A \in \mathbb{C}^{n\times n}$, $z \in \mathbb{C}^n$ and $\varepsilon > 0$. Show that the following are equivalent:

(i) $\|(zI - A)^{-1}\|_2 \geq \varepsilon^{-1}$;

(ii) z is an eigenvalue of $A + E$ for some $E \in \mathbb{C}^{n\times n}$ with $\|E\|_2 \leq \varepsilon$;

(iii) there exists a vector $u \in \mathbb{C}^n$ with $\|u\|_2 = 1$ and $\|(zI - A)u\|_2 \leq \varepsilon$;

(iv) $\sigma_{\min}(zI - A) \leq \varepsilon$, where $\sigma_{\min}(zI - A)$ is the smallest singular value of $zI - A$.

Hint: A suggested order of implications is (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) and (i) $\Leftrightarrow$ (iv).

(c) Show that when A is normal, its ε-pseudospectrum is given by

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : \mathrm{dist}(z, \Lambda(A)) \leq \varepsilon\}$$

where $\mathrm{dist}(z, \Lambda(A))$ is the distance from the point z to the set $\Lambda(A)$. Describe how this pseudospectrum looks geometrically in the complex plane.

(d) Define $A \in \mathbb{C}^{2\times 2}$ by

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Show that A is not normal. Find an explicit function $f : \mathbb{C} \to \mathbb{R}$ such that[1]

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} \,|\, f(z) \leq \varepsilon\}.$$

Using MATLAB, plot the contours of f for a variety of levels $\varepsilon \in (0, 1)$. Contrast with what would be expected if A was normal.

---

[1]Do not worry about simplifying f

Problem 2. Markov Chains (25 points)

We now study properties of matrices that arise in the theory of Markov chains. Let $S = \{s_1, \ldots, s_n\}$ denote a finite state space. A Markov chain on S is a discrete time stochastic process $\{X_k\}$ taking values in S such that its future trajectory depends only upon its current state. In this assignment we will only be concerned with time homogeneous Markov chains, so that for any time $k \in \mathbb{N}$ and any states $s_i, s_j \in S$,

$$\mathbb{P}(X_{k+1} = s_j \,|\, X_k = s_i) = \mathbb{P}(X_1 = s_j \,|\, X_0 = s_i).$$

The Markov chain can then be completely characterized by its one-step transition probabilities. Denote by P_{ij} the probability of moving from state i to state j in a single step:

$$P_{ij} = \mathbb{P}(X_1 = s_j \,|\, X_0 = s_i).$$

The matrix $P = \{P_{ij}\}$ is called the transition matrix for $\{X_k\}$.

We will say that $\nu \in \mathbb{R}^n$ is a distribution if $\nu_j \in [0, 1]$ for all j and $\|\nu\|_1 = 1$. We will say that a random variable U on S is distributed according to ν if for each j,

$$\mathbb{P}(U = s_j) = \nu_j.$$

Given a Markov chain $\{X_k\}$ with transition matrix P, a distribution μ is said to be an invariant distribution for $\{X_k\}$ if

$$P^*\mu = \mu.$$

Thus μ is a right eigenvector of P^* with eigenvalue 1, or equivalently μ^* is a left eigenvector of P with eigenvalue 1. Let $F : S \to \mathbb{R}$ by a function on S. Then associated with F is a vector $f \in \mathbb{R}^n$ given by $f_j = F(s_j)$ for each j. The expected value of F under μ can then be computed as

$$\mathbb{E}\big(F(U)\big) := \sum_{j=1}^{n} F(s_j)\mathbb{P}(U = s_j) = \sum_{j=1}^{n} f_j \mu_j = \langle f, \mu \rangle_2$$

where U is distributed according to μ.

Given a Markov chain $\{X_k\}$ with an invariant distribution μ, the question arises of whether the chain converges to μ in a distributional sense. If this is the case, $\{X_k\}$ is said to be *ergodic*. Formally this means that for large enough k, the statistics of the chain coincide with those of samples from μ. When this is the case, we may approximate

$$\langle f, \mu \rangle_2 \approx \frac{1}{N} \sum_{k=1}^{N} F(X_k)$$

without the need to know or compute μ. This can be interpreted as approximating the spatial average $\mathbb{E}\big(F(U)\big)$ by the time average of $F(X_k)$.

We will give sufficient conditions for $\{X_k\}$ to be ergodic, but first we must introduce the notions of irreducibility and aperiodicity:

Definition 1. *Let $\{X_k\}$ be a Markov chain on S with transition matrix P.*

(i) $\{X_k\}$ is said to be irreducible *if there is a positive probability of moving between any two states s_i, s_j in a finite time.*

(ii) $\{X_k\}$ is said to be aperiodic *if* $\gcd\{m \geq 1 : (P^m)_{jj} > 0\} = 1$ *for each j.*

Theorem 1. *Let $\{X_k\}$ be an irreducible aperiodic Markov chain on S with transition matrix P and invariant distribution μ, then we have*

$$P^\infty := \lim_{m\to\infty} P^m = \begin{pmatrix} \mu_1 & \mu_2 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1 & \mu_2 & \cdots & \mu_n \end{pmatrix}.$$

Note that in particular the above result implies that the invariant distribution is unique. If the hypotheses of the theorem are satisfied it hence makes sense to refer to *the* invariant distribution for $\{X_k\}$.

(a) Show that the m-step transition probabilities are given by

$$\mathbb{P}(X_m = s_j \,|\, X_0 = s_i) = (P^m)_{ij}.$$

(b) Show that the left and right eigenvalues of a matrix coincide. By finding a left eigenvector for P^*, deduce that an invariant distribution for $\{X_k\}$ always exists.

(c) Let $\{X_k\}$ satisfy the hypotheses of Theorem 1.

(i) Show that given any distribution ν we have $(P^m)^*\nu \to \mu$.

(ii) Define $A = P - P^\infty$. Show that $A^m = P^m - P^\infty$ for all $m \geq 1$.

Problem 3. Ehrenfest Model (25 points)

The Ehrenfest model is a simple model for diffusion. Suppose that we have two urns, Urn 1 and Urn 2, with n balls split between them. At each time step we will choose a ball uniformly at random, move it to the other urn with probability $n/(n+1)$, or else leave it where it is. Denote by X_k the number of balls in Urn 1 at time k.

(a) Explain briefly why the transition matrix P for $\{X_k\}$ is given by

$$P = \begin{pmatrix} \frac{1}{n+1} & \frac{n}{n+1} & & & & \\ \frac{1}{n+1} & \frac{1}{n+1} & \frac{n-1}{n+1} & & & \\ & \frac{2}{n+1} & \frac{1}{n+1} & \frac{n-2}{n+1} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{n-1}{n+1} & \frac{1}{n+1} & \frac{1}{n+1} \\ & & & & \frac{n}{n+1} & \frac{1}{n+1} \end{pmatrix}.$$

(b) Show that $\{X_k\}$ is irreducible and aperiodic, and so the conclusion of Theorem 1 holds.

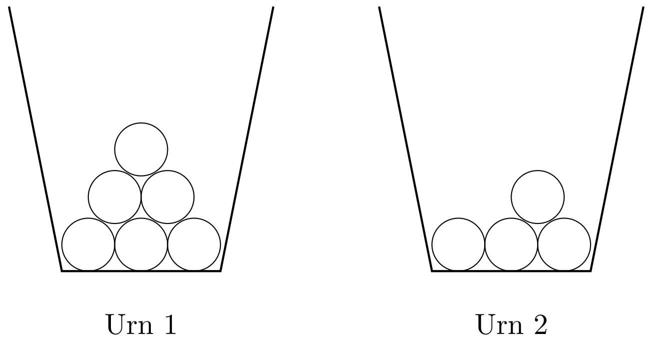

Figure 1: The Ehrenfest model with two urns and 10 balls.

(c) Show that the invariant distribution μ of $\{X_k\}$ is given by

$$\mu_j = 2^{-n} \binom{n}{j}, \quad j = 0, \dots, n.$$

(d) Verify that P is not normal. Implement P in MATLAB, and using EigTool or otherwise, compute a number of ε-pseudospectra of P in the cases $n = 10, 100, 1000$.

(e) For each $n = 10, 100, 1000$, compute and plot $\|P^m\|_p$ versus m for $p \in \{1, 2, \infty\}$ and a large range of m. Do the same for $\|A^m\|_p$, where $A = P - P^\infty$. How do the choice of norm and the dimension of S affect the behavior in m?

(f) Implement the Markov chain $\{X_k\}$ in MATLAB with starting point $X_0 = \lfloor n/2 \rfloor$. Fix $n = 10$, and choose a (non-constant) function $F : \{0, 1, \dots, n\} \to \mathbb{R}$. Let $f \in \mathbb{R}^{n+1}$ be the corresponding vector $f_j = F(j)$ for each j. Given $N \in \mathbb{N}$, define $\hat{Z}_N$ by

$$\hat{Z}_N = \left| \langle f, \mu \rangle_2 - \frac{1}{N} \sum_{k=1}^{N} F(X_k) \right|.$$

Plot the trace of $\hat{Z}_N$ versus N. By plotting on logarithmic axes or otherwise, estimate the rate of convergence of $\hat{Z}_N$. Compare with the rate of convergence of

$$Z_N = \left| \langle f, \mu \rangle_2 - \frac{1}{N} \sum_{k=1}^{N} F(U_k) \right|$$

where $\{U_k\}$ are independent samples from μ.

Note: Samples from μ can be generated using the `binornd` function.

In the case F is defined by $F(j) = j$, compare both $\hat{Z}_N, Z_N$ with the theoretical 95% confidence bound for Z_N:

$$Z_N \leq \frac{1.96\sqrt{n}}{2\sqrt{N}}.$$

Problem 4. State-dependent Random Walk (25 points)

We now consider a random walk on the set $S = \{0, \dots, n\}$. Let $p_0, \dots, p_n \in (0, 1)$ be a sequence of probabilities. Define the Markov chain $\{X_k\}$ by

$$X_{k+1} = \begin{cases} X_k + 1 & \text{with probability } p_{X_k} \\ X_k - 1 & \text{with probability } 1 - p_{X_k} \end{cases}$$

where arithmetic is performed modulo $n + 1$.

(a) Write down the transition matrix P for $\{X_k\}$.

(b) Verify that P is not normal if the probabilities p_j are distinct. Implement P in MATLAB, generating the sequence of probabilities $p_0, \dots, p_n \sim U(0, 1)$ i.i.d. Using EigTool or otherwise, compute a number of ε-pseudospectra of P in the case $n = 100$.

(c) We will now compare the properties of the chain when S has odd or even cardinality, specifically when $n = 50$ or $n = 51$ respectively. Load `data4.mat`. The variable `p` contains $p_0, \dots, p_{51} \sim U(0, 1)$ i.i.d. Construct $P \in \mathbb{R}^{(n+1)\times(n+1)}$ for $n = 50$ using $p_0, \dots, p_{50}$ and for $n = 51$ using $p_0, \dots, p_{51}$.

(i) For $n = 50, 51$, compute and plot $\|P^m\|_p$ versus m for $p \in \{1, 2, \infty\}$ and m up to 10^6. Compare the behavior of the the norms for the two cases.

(ii) For $n = 50, 51$, using EigTool or otherwise, compute a number of ε-pseudospectra of $(P^{10^6})^*$. How do the pseudospectra compare between the two cases?

(iii) Plot the columns of $(P^{10^6})^*$ on the same axes. What difference do you notice between the cases $n = 50$ and $n = 51$?

(iv) For each $n = 50, 51$, take a column μ from $(P^{10^6})^*$. Is it the case that $P^*\mu = \mu$?

(v) Do you think that the chain $\{X_k\}$ is ergodic in either of the cases $n = 50, 51$? Theorize why this is the case, making reference to irreducibility and aperiodicity.

Remark: To use EigTool on a matrix `A`, call `eigtool(A)` from the directory where you have extracted the software. To customize options for plotting, in the window that opens, select Extras ▷ Options Code for Printing. Copy and paste the code for the struct `opts` that appears in the console, changing it where appropriate, and then call `eigtool(A,opts)`. In particular I recommend increasing the value of `opts.npts` and the number of levels in `opts.levels`.