MATH 5330: Computational Methods of Linear Algebra

Lecture Note 12: The Eigenvalue Problem

Xianyi Zeng
Department of Mathematical Sciences, UTEP

1 Theoretical Background

The eigenvalue problem is a classical topic in mathematics, structural engineering, chemistry, and many more. Its statement is fairly simple: Let $A \in \mathbb{R}^{n\times n}$ be arbitrary, then $\lambda \in \mathbb{C}$ is *an eigenvalue* of A if and only if there exists a **non-zero** vector $\boldsymbol{v} \in \mathbb{C}^n$ such that:

$$A\boldsymbol{v} = \lambda \boldsymbol{v}\,; \tag{1.1}$$

here the "non-zero" quantifier is critical, otherwise any complex number could be an eigenvalue. The vector $\boldsymbol{v}$ is called an *eigenvector* of A associated with the eigenvalue λ.

Clearly, all the eigenvectors associated with an eigenvalue λ of A is a linear subspace of $\mathbb{C}^n$; and it is called the *eigenspace* of λ. For example, 1 is an eigenvalue of the identity matrix I, and its eigenspace is $\mathbb{R}^n$ itself. If we denote this space by $\mathcal{E}_\lambda$, it is formally defined as:

$$\mathcal{E}_\lambda = \{\boldsymbol{v} \in \mathbb{C}^n \,:\, (A-\lambda I)\boldsymbol{v} = 0\}\,. \tag{1.2}$$

Note that the definition (1.2) is valid for any complex number λ; but unless λ is an eigenvalue of A, $\mathcal{E}_\lambda$ has zero dimension and it only contains the zero vector. And by definition, if λ is an eigenvalue then $\dim\mathcal{E}_\lambda \geq 1$. In fact, we call $\dim\mathcal{E}_\lambda$ the *geometric multiplicity* of the eigenvalue λ.

Furthermore, if $\lambda_1 \neq \lambda_2$ are two eigenvalues of A, it is not difficult to show that $\mathcal{E}_{\lambda_1} \cap \mathcal{E}_{\lambda_2} = \{0\}$. Indeed, let $\boldsymbol{v} \in \mathcal{E}_{\lambda_1} \cap \mathcal{E}_{\lambda_2}$ then we have:

$$A\boldsymbol{v} = \lambda_1 \boldsymbol{v} \text{ and } A\boldsymbol{v} = \lambda_2 \boldsymbol{v} \quad \text{hence } (\lambda_1 - \lambda_2)\boldsymbol{v} = 0 \;\Rightarrow\; \boldsymbol{v} = 0\,.$$

This is a clear indication that a matrix A can only have finite distinct eigenvalues (at most n). In fact, we just showed that the sum of the geometric multiplicities of all the eigenvalues of A cannot exceed n.

This sum, however, may not be exactly n. Let us consider the following simple 2×2 matrix:

$$J = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

We'll show two facts: (1) J has only one eigenvalue 1, and (2) $\dim\mathcal{E}_1 = 1$. For the first part, we note that if λ is an eigenvalue of J then $J-\lambda I$ has an non-empty null space (i.e., $\mathcal{E}_\lambda$); hence $J-\lambda I$ is singular and $\det(J-\lambda I) = 0$. This is equivalent to $(\lambda-1)^2 = 0$ or $\lambda = 1$. Thus J has at most one eigenvalue, which is 1. Now let $\boldsymbol{v} \in \mathcal{E}_1$ be non-zero, then $J\boldsymbol{v} = \boldsymbol{v}$ leads to:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \;\Rightarrow\; v_2 = 0\,.$$

Hence $\mathcal{E}_1$ is spanned by $\boldsymbol{e}_1$ and has dimension one.

The matrix J is a classical example for a non-diagonalizable matrix, and we've already encountered it before as a Jordan block. In the process of showing that J has at most one eigenvalue, we used a weaker version of the following statement: Let A be any square matrix, then λ is an eigenvalue of A **if and only if** $\det(\lambda I - A) = 0$. The determinant $\det(\lambda I - A)$ is a polynomial of λ of degree n, called the *characteristic polynomial* of A:

$$p_A(\lambda) = \det(\lambda I - A) = \lambda^n - \mathrm{tr}A\lambda^{n-1} + \cdots + (-1)^n \det A\,. \tag{1.3}$$

For reasons we will soon see, the coefficients of $p_A(\lambda)$ are called the invariants of the matrix A; and by the **fundamental theorem of algebra**, $p_A(\lambda)$ has exactly n roots (including the multiple ones). To this end we have a few conclusions:

- The set of all the distinct roots of $p_A(\lambda)$ is the same of the set of all distinct eigenvalues of A, the latter is known as the *spectrum* of A.
- Every matrix A has at least one eigenvalue, i.e., its spectrum is a non-empty finite set; and the eigenvalue with the largest modulus is known as the *spectrum radius* of A, which we denote as $\rho(A)$.

Furthermore, if λ is an eigenvalue of A, its *algebraic multiplicity* is defined as λ's multiplicity as a root of $p_A(\lambda)$. It can be shown that every eigenvalue's geometric multiplicity ($\dim\mathcal{E}_\lambda$) is no more than its algebraic multiplicity. This statement is fairly straightforward when A has only one eigenvalue λ_1, in which case $p_A(\lambda) = (\lambda - \lambda_1)^n$: λ_1 has algebraic multiplicity n and its geometric multiplicity can be at most n. For a general matrix A, we can use the Jordan canonical form (or Jordan normal form) to prove this statement.

To begin with, we say that a matrix A is *similar* to a matrix B if and only if there exists a non-singular matrix T such that:

$$A = TBT^{-1}\,. \tag{1.4}$$

It follows immediately that if A is similar to B, then their characteristic polynomials are the same:

$$p_A(\lambda) = \det(\lambda I - A) = \det(\lambda TT^{-1} - TBT^{-1}) = \det T \det(\lambda I - B)\det T^{-1} = p_B(\lambda)\,.$$

Hence the coefficients of the characteristic polynomial are invariant under the similarity relation. In Exercise 1 we will see that if A and B are similar to each other, they have the same spectrum, and all the eigenvalues have the same geometric multiplicity and the same algebraic multiplicity. Then our proof of the preceding statement follows from the fact that every matrix A is similar to its Jordan canonical form:

$$A = TJT^{-1}\,,\ J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{bmatrix}, \tag{1.5}$$

where each $J_k, k = 1,\cdots,m$ is a Jordan block; and the rest of the proof is straightforward.

Using the Jordan canonical form, one can easily prove the Cayley-Hamilton theorem:

Theorem 1.1 (Cayley-Hamilton)**.** *Let $A \in \mathbb{R}^{n\times n}$ be arbitrary and $p_A(\lambda)$ its characteristic polynomial, then $p_A(A) = 0$.*

Proof. Consider the Jordan canonical form $A = TJT^{-1}$, then we have:

$$p_A(A) = p_J(A) = p_J(TJT^{-1}) = Tp_J(J)T^{-1}\,,$$

hence we only need to show that $p_J(J) = 0$. Let J be given by (1.5), and let J_k has diagonal elements λ_k and dimension n_k (λ_k,$k = 1,\cdots,m$ are not necessarily mutually distinct). Then we have:

$$p_J(\lambda) = \prod_{k=1}^{m} (\lambda - \lambda_k)^{n_k}\,;$$

and that:

$$p_J(J) == \begin{bmatrix} p_J(J_1) & & & \\ & p_J(J_2) & & \\ & & \ddots & \\ & & & p_J(J_m) \end{bmatrix}.$$

Now it is sufficient to check $p_J(J_k) = 0$ for each $1 \le k \le m$; for example:

$$p_J(J_1) = \prod_{k=1}^{m} (J_1 - \lambda_k I_{n_1})^{n_k} = 0\,,$$

due to the fact that $(J_1 - \lambda_1 I_{n_1})^{n_1} = 0$. □

Finally, if all the eigenvalues of A have the same geometric multiplicity and algebraic multiplicity, all the Jordan block of the Jordan normal form must be of dimension one and J is diagonal. In this case, we write $J = \Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ and say that A is *diagonalizable*:

$$A = T\Lambda T^{-1} \quad \Rightarrow \quad AT = T\Lambda\,.$$

From the second equality, we see that the i-th column vector of T is an eigenvector of λ_i.

2 Symmetric Matrices

When A is real symmetric, it has much nicer properties related to eigenvalues. For example, any eigenvalue of a symmetric matrix A is real. To show this, let $\lambda \in \mathbb{C}$ be any eigenvalue of A and $\boldsymbol{v} \in \mathbb{C}^n$ be a non-zero eigenvector. Then:

$$A\boldsymbol{v} = \lambda \boldsymbol{v} \quad \Rightarrow \quad \overline{\boldsymbol{v}}^t A\boldsymbol{v} = \lambda \overline{\boldsymbol{v}}^t \boldsymbol{v} \quad \Rightarrow \quad \lambda = \frac{\overline{\boldsymbol{v}}^t A\boldsymbol{v}}{\overline{\boldsymbol{v}}^t \boldsymbol{v}}\,.$$

It is not difficult to see that both the denominator and the nominator are real, so λ is a real number. Furthermore, we can show that if we define the "real" eigenspace as:

$$\mathcal{E}'_\lambda = \{\boldsymbol{v} \in \mathbb{R}^n \,:\, (A - \lambda I)\boldsymbol{v} = 0\}\,, \tag{2.1}$$

then $\dim\mathcal{E}'_\lambda = \dim\mathcal{E}_\lambda$[1], the latter being given by (1.2). From this point on, we will use the same notation $\mathcal{E}_\lambda$ to denote (2.1) for symmetric matrices A and (1.2) for general matrices.

[1]Note that the first dimension is the dimension of a real-coefficient vector space whereas the latter is that of a complex-coefficient vector space

Next, we show that all symmetric matrices are diagonalizable. Indeed, this means that all the Jordan blocks in the Jordan canonical form must have dimension one. Let us suppose this is not true for some symmetric A and its first Jordan block has dimension higher than one. For simplicity, suppose $J_1 \in \mathbb{R}^{2\times 2}$ and we have:

$$J_1 = \begin{bmatrix} \lambda_1 & 1 \\ & \lambda_1 \end{bmatrix}.$$

Let the first two column vectors T of $AT = TJ$ be $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, then:

$$A\boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1 + \boldsymbol{v}_2\,, \quad A\boldsymbol{v}_2 = \lambda_1 \boldsymbol{v}_2\,.$$

Thus:

$$\begin{aligned} & \overline{\boldsymbol{v}}_2^t (A-\lambda_1 I)\boldsymbol{v}_1 = \overline{\boldsymbol{v}}_2^t \boldsymbol{v}_2 \\ \text{and} \quad & \overline{\boldsymbol{v}}_2^t (A-\lambda_1 I)\boldsymbol{v}_1 = (A\boldsymbol{v}_2 - \lambda_1 \boldsymbol{v}_2)^t \boldsymbol{v}_1 = 0\,, \end{aligned}$$

which is clearly a contradiction since $\boldsymbol{v}_2 \neq 0$. The case that J_1 has even higher dimension is similar. To this end, we showed that any symmetric matrix is diagonalizable.

Furthermore, let λ_1,λ_2 be two distinct eigenvalues of A; and let $\boldsymbol{v}_1 \in \mathcal{E}_{\lambda_1}$ and $\boldsymbol{v}_2 \in \mathcal{E}_{\lambda_2}$ be two non-zero real eigenvectors, then:

$$\boldsymbol{v}_1^t A \boldsymbol{v}_2 = \boldsymbol{v}_1^t (\lambda_2 \boldsymbol{v}_2) = \lambda_2 (\boldsymbol{v}_1 \cdot \boldsymbol{v}_2) \quad \text{and} \quad \boldsymbol{v}_2^t A \boldsymbol{v}_1 = \boldsymbol{v}_2^t (\lambda_1 \boldsymbol{v}_1) = \lambda_1 (\boldsymbol{v}_1 \cdot \boldsymbol{v}_2)\,.$$

Consequently $(\lambda_1 - \lambda_2)(\boldsymbol{v}_1 \cdot \boldsymbol{v}_2) = 0$ and we must have $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$. This implies that the eigenspaces of two distinct eigenvalues are mutually orthogonal to each other. Suppose A has m distinct eigenvalues $\lambda_1,\cdots,\lambda_m$ with multiplicities $n_1,\cdots,n_m$, respectively; we can choose an orthonormal basis $\{\boldsymbol{v}_1^k,\cdots,\boldsymbol{v}_{n_k}^k\}$ of $\mathcal{E}_{\lambda_k}$, then the vectors in all these bases form an orthonormal basis of $\mathbb{R}^n$. Denote:

$$T = [\boldsymbol{v}_1^1\ \boldsymbol{v}_2^1\ \cdots\ \boldsymbol{v}_{n_1}^1\ \boldsymbol{v}_1^2\ \cdots\ \boldsymbol{v}_{n_m}^m] \in \mathbb{R}^n\,,$$

we obtain the eigenvalue decomposition $A = T\Lambda T^{-1}$, where

$$\Lambda = \begin{bmatrix} \lambda_1 & & & & & \\ & \cdots & & & & \\ & & \lambda_1 & & & \\ & & & \lambda_2 & & \\ & & & & \ddots & \\ & & & & & \lambda_m \end{bmatrix},$$

where λ_k is repeated n_k times.

Sometimes for convenience we use an alternative notation. Let A be symmetric and $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$ be its n eigenvalues (they are not necessarily different from each other); then we may choose an unit eigenvector $\boldsymbol{v}_i$ of λ_i for all $1 \leq i \leq n$ such that $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$ if $i \neq j$. So that the eigenvalue decomposition of A is given by:

$$A = Q\Lambda Q^{-1}\,, \quad Q = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ \cdots\ \boldsymbol{v}_n] \text{ is orthogonal } \text{, and } \Lambda = \operatorname{diag}(\lambda_1,\cdots,\lambda_n)\,. \tag{2.2}$$

As an application of the eigenvalue decomposition (2.2), we can write any vector $\boldsymbol{v} \in \mathbb{R}^n$ as linear combination of the column vectors of Q:

$$\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i \,,$$

then:

$$A\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \lambda_i \boldsymbol{v}_i \,.$$

If we compute the last term of the first equation in this section, we have:

$$\frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} = \frac{\sum_{i=1}^{n} \alpha_i^2 \lambda_i}{\sum_{i=1}^{n} \alpha_i^2} \in [\lambda_{\min},\, \lambda_{\max}] \,.$$

Here $\lambda_{\min} = \min_{i=1}^{n} \lambda_i$ and $\lambda_{\max} = \max_{i=1}^{n} \lambda_i$. The quotient $(\boldsymbol{v}^t A \boldsymbol{v})/(\boldsymbol{v}^t \boldsymbol{v})$ is called the *Rayleigh quotient*, and it is used in the *Courant-Fischer theorem* (also known as the *min-max theorem*) to characterize all eigenvalues of symmetric matrices.

Theorem 2.1 (Courant-Fischer)**.** *Let $A \in \mathbb{R}^{n \times n}$ be real symmetric with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, then:*

$$\lambda_k = \min_{\{\mathcal{U} \subseteq \mathbb{R}^n:\ \dim \mathcal{U} = k\}} \max_{\{\boldsymbol{v} \in \mathcal{U}:\ \boldsymbol{v} \neq 0\}} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} \,, \tag{2.3}$$

and

$$\lambda_k = \max_{\{\mathcal{U} \subseteq \mathbb{R}^n:\ \dim \mathcal{U} = n-k+1\}} \min_{\{\boldsymbol{v} \in \mathcal{U}:\ \boldsymbol{v} \neq 0\}} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} \,. \tag{2.4}$$

The theorem seems daunting but it has two simple special cases:

$$\lambda_{\min} = \lambda_1 = \min_{\boldsymbol{v} \in \mathbb{R}^n:\ \boldsymbol{v} \neq 0} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} \quad \text{and} \quad \lambda_{\max} = \lambda_n = \max_{\boldsymbol{v} \in \mathbb{R}^n:\ \boldsymbol{v} \neq 0} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} \,. \tag{2.5}$$

Proof. First we look at (2.3), and notice that:

$$\mathcal{V}_k = \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k)$$

is a k-dimensional subspace of $\mathbb{R}^n$ and it is not difficult to show that:

$$\max_{\boldsymbol{v} \in \mathcal{V}_k:\ \boldsymbol{v} \neq 0} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} = \max_{\alpha_1, \cdots, \alpha_k \in \mathbb{R}:\ \alpha_1^2 + \cdots + \alpha_k^2 \neq 0} \frac{\alpha_1^2 \lambda_1 + \cdots + \alpha_k^2 \lambda_k}{\alpha_1^2 + \cdots + \alpha_k^2} \leq \lambda_k \,,$$

and the equality is attainable with $\boldsymbol{v} = \boldsymbol{v}_k$. Hence:

$$\lambda_k = \max_{\boldsymbol{v} \in \mathcal{V}_k:\ \boldsymbol{v} \neq 0} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} \geq \min_{\{\mathcal{U} \subseteq \mathbb{R}^n:\ \dim \mathcal{U} = k\}} \max_{\{\boldsymbol{v} \in \mathcal{U}:\ \boldsymbol{v} \neq 0\}} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} \,.$$

To show the other direction, we consider any k-dimensional subspace $\mathcal{U}$ of $\mathbb{R}^n$, and notice that:

$$\mathcal{U} \cap \operatorname{span}(\boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \cdots, \boldsymbol{v}_n) \neq \{0\} \,.$$

Thus there must exists a non-zero $\boldsymbol{w} \in \mathcal{U}$, such that:

$$\boldsymbol{w} = \alpha_k \boldsymbol{v}_k + \alpha_{k+1} \boldsymbol{v}_{k+1} + \cdots + \alpha_n \boldsymbol{v}_n \,.$$

For this particular $\boldsymbol{w}$, we have:

$$\frac{\boldsymbol{w}^t A \boldsymbol{w}}{\boldsymbol{w}^t \boldsymbol{w}} = \frac{\alpha_k^2 \lambda_k + \alpha_{k+1}^2 \lambda_{k+1} + \cdots + \alpha_n^2 \lambda_n}{\alpha_k^2 + \cdots + \alpha_n^2} \geq \lambda_k \,,$$

hence:

$$\min_{\boldsymbol{v} \in \mathcal{U}:\ \boldsymbol{v} \neq 0} \frac{\boldsymbol{v}^t A \boldsymbol{v}}{\boldsymbol{v}^t \boldsymbol{v}} \geq \frac{\boldsymbol{w}^t A \boldsymbol{w}}{\boldsymbol{w}^t \boldsymbol{w}} \geq \lambda_k \,.$$

Because the choice $\mathcal{U}$ is arbitrary, we obtain that the left hand side of (2.3) is no more than the right hand side. Combining with the previous inequality, (2.3) is proved.

The equality (2.4) can be proved in a similar way, which is left as an exercise. □

3 The Eigenvalue Problem

The classical eigenvalue problem concerns finding some information about the eigenvalues of a given matrix A. Our first example is the linear evolution model in general and the population growth in particular. Let $\boldsymbol{p} : \mathbb{R}^+ \to \mathbb{R}^n$ be a time-dependent non-negative vector describing the population of n species; and we use a dot to denote the time derivative. Then a simple model to describe the species' evolution is:

$$\dot{\boldsymbol{p}} = G\boldsymbol{p} \,,$$

where each entry of $G = [g_{ij}]$ is endowed with an ecological meaning. For example, g_{ii} denotes the natural growth rate of the i-th specie; and if it is a prey of the j-th specie we expect $g_{ij} < 0$ and $g_{ij} > 0$ (a simplified version of the predator-prey model).

A basic question to ask is whether this colony will blow up (the population grows without bound) or become extinct (all the species will disappear as $t \to \infty$). This problem is tightly related to the sign of the real parts of the eigenvalues of the matrix G, which in turn, can be determined once we manage to solve the eigenvalue problem for G.

The second example comes from structural engineering, or specifically the natural frequency analysis. Say we have a one-dimensional beam spanning the distance l, whose two end points are clamped. Denoting the vertical displacement of the beam by u, which is a function of $x \in [0,\, l]$ and time $t \geq 0$, we have the following equation for the beam deflection:

$$\rho \frac{\partial^2 u}{\partial t^2} = -\frac{\partial^2}{\partial x^2}\left(EI \frac{\partial^2 u}{\partial x^2} \right) + f(x) \,,$$

where ρ is mass per unit length, E is the elasticity modulus and I is the second moment of the cross section area. $f(x)$ is the external force excited on the beam; if we consider the beam as a simple bridge model, $f(x)$ can be thought as the weight of the traffic on the bridge or the force from transverse wind (search for the Tacoma bridge accident!). The phenomenon of resonance happens when some component of f has the frequency coinciding with one of the natural frequency

of the beam. To determine this natural frequency, people need to solve for the eigenvalue of the differential operator:

$$-\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 u}{\partial x^2}\right), \quad u(0)=u(l)=0\,.$$

Note that this is a different problem from what we've been considering so far – the eigenvalue is that of a general linear operator. Nevertheless, the idea is similar: We want to find a pair of real number λ and a function ϕ, such that:

$$-\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 \phi}{\partial x^2}\right)=\lambda\phi\,, \quad \phi(0)=\phi(1)=0\,;$$

once this pair is available, we can obtain a natural frequency of the dynamic equation by solving:

$$\rho\frac{\partial^2\phi}{\partial t^2}=\lambda\phi\,.$$

Hence the important question in design is: Given a beam model what is the range of its eigenvalues, such that the resulting natural frequencies will be far away from expected frequencies in the external forces.

Analytical solutions to the preceding eigenvalue problem do not always exist when the parameters varies in space and when the model becomes more complicated. But a good approximation can be obtained by discretizing the system into a collection of nodes that are connected by springs. The result is a linear system:

$$K\boldsymbol{\phi}=\lambda\boldsymbol{\phi}\,,$$

where $\boldsymbol{\phi}=[\phi_0,\phi_1,\cdots,\phi_{n+1}]\in\mathbb{R}^{n+1}$ describes the displacement at the locations $0/n,l/n,\cdots,(n-1)l/n,l$, and $K\in\mathbb{R}^{(n+1)\times(n+1)}$ is known as the stiffness matrix. Usually, the first row of K is [1, 0, $\cdots$, 0] and the last row of K is [0, $\cdots$, 0, 1] – these two rows correspond to the two boundary conditions $\phi(0)=\phi(l)=0$; the middle $n-1$ rows of K correspond to the numerical approximation to the differential operator at the inner nodes $l/n,2l/n,\cdots,(n-1)l/n$. For engineering purpose, people are interested in the range of these eigenvalues, i.e., both $\lambda_{\min}$ and $\lambda_{\max}$ need to be evaluated.

The third example is about chemical reactions. If we use $\boldsymbol{c}$ to denote the concentration of n species of chemicals that may react with each other, the dynamics of their concentrations can be described by the linear ODE:

$$\dot{\boldsymbol{c}}=R\boldsymbol{c}\,,$$

In combustions, there can be thousands of species and people are interested in short-to-mid term behavior of this dynamical system. The solution to the preceding equation is given by:

$$\boldsymbol{c}(t)=T\exp(Jt)T^{-1}\boldsymbol{c}(0)\,,$$

where J is the Jordan canonical form of R. For the range of time that is of engineering interests, usually the largest one-third of eigenvalues are significant; so for a large system like this example, people focus on methods that can provide accurate estimates to at least one-third of the largest eigenvalues and their eigenvectors.

The last example is in the Markov chain models, which is widely used in pattern recognition and machine learning. In a simple Markov chain model, there are n possible states labeled by the integers 1 through n. Let X_k denote the state at the step k, then in this model:

$$P(X_k=j\mid X_1=j_1,\,X_2=j_2,\,\cdots,\,X_{k-1}=j_{k-1})=P(X_k=j\mid X_{k-1}=j_{k-1})=p_{j_{k-1}j}\,.$$

If we use $\boldsymbol{p}_k$ to denote the possibilities of each state at step k, then:

$$\boldsymbol{p}_{k+1} = P\boldsymbol{p}_k \,,$$

where $P = [p_{ij}]$ is the transition probability matrix. A stationary state $\boldsymbol{p}$ is defined by:

$$\boldsymbol{p} = P\boldsymbol{p} \,,$$

that is, an eigenvector of the eigenvalue 1 of P.

So far, except for the last example, we're interested in the eigenvalues themselves but not their eigenvectors. We've seen some theoretical results that may help in finding them. First, we can try to find the roots of the characteristic polynomial $p_A(\lambda)$. However, for a general polynomial of degree n, there is no algebraic form of its roots if $n > 4$. Finding numerical approximations to the roots of a polynomial, however, is usually a more difficult task than the eigenvalue problem itself; and some of the best root finding algorithms convert the problem to an eigenvalue problem. In fact, any polynomial of degree n with leading coefficient being unity is the characteristic polynomial of some matrix (Exercise 5).

In the next lectures, we'll talk about two classes of methods to find the eigenvalues. The first class is based on factorization, such that the matrix A is transformed into a form whose characteristic polynomial has roots that are easier to find. The second one is iterative methods, which typically make use of the min-max theorem.

Exercises

Exercise 1. *Let $A = TBT^{-1}$ where T is a non-singular matrix. Show that λ_1 is an eigenvalue of A if and only if λ_1 is also an eigenvalue of B; furthermore, it has the same geometric multiplicity and the same algebraic multiplicity for the two matrices.*

Exercise 2. *Let A be real symmetric, show that the (2.3) alone leads to (2.5), then show that (2.4) alone also leads to (2.5).*

Exercise 3. *Prove (2.4) of Theorem 2.1. You can use the proof of (2.3) as an example.*

Exercise 4. *Let $P = [p_{ij}] \in \mathbb{R}^{n\times n}$ be a transition probability matrix, it has the following property:*

- *Each p_{ij} is the probability $P(X_{k+1} = j \mid X_k = i)$ for any k, hence it is non-negative: $p_{ij} \geq 0$.*
- *Every row sum of P is one:*

$$\sum_{j=1}^{n} p_{ij} = \sum_{j=1}^{n} P(X_{k+1} = j \mid X_k = i) = P(X_{k+1} \in \{1,2,\cdots,n\} \mid X_k = i) = 1 \,.$$

Show that 1 is an eigenvalue of P.

Exercise 5. *Let α_0, α_1, $\cdots$, α_{n-1} be arbitrary real numbers, and we construct a matrix A as:*

$$A = \begin{bmatrix} -\alpha_{n-1} & -\alpha_{n-2} & \cdots & -\alpha_1 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Show that the characteristic polynomial of A is given by:

$$p_A(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0 \,.$$