Chapter 4
Eigenvalues and eigenvectors

Rayleigh quotient

We begin with the following theorem[1].

Theorem 1. *If A is a real symmetric matrix, then the eigenvalues λ_i are real and we can take the eigenvectors $\mathbf{q}_i$ so that they form an orthonormal basis, i.e., $\mathbf{q}_i \cdot \mathbf{q}_j = \mathbf{q}_i^T \mathbf{q}_j = \delta_{ij}$.*

Note that δ_{ij} is the Kronecker delta: $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$.

In Chapter 3, we studied that eigenvalues are given as roots of the characteristic polynomial $f_A(\lambda)$:

$$f_A(\lambda) = \det(A - \lambda I) = 0,$$

and we studied rootfinding methods in Chapter 2. So, we may think obtaining eigenvalues are not a big deal. But the following example shows that this calculation is unstable.

Example 1. Let us consider the following diagonal matrix A.

$$A = \begin{pmatrix} 1 & & & & \\ & 2 & & & \\ & & 3 & & \\ & & & 4 & \\ & & & & 5 \end{pmatrix} \quad \Rightarrow \quad \begin{aligned} f_A(\lambda) &= (1-\lambda)(2-\lambda)(3-\lambda)(4-\lambda)(5-\lambda) \\ &= -\lambda^5 + 15\lambda^4 - 85\lambda^3 + 225\lambda^2 - 274\lambda + 120. \end{aligned}$$

$$f_A(\lambda) = 0 \quad \Rightarrow \quad \lambda = 1, 2, 3, 4, 5.$$

Suppose coefficients of $f_A(\lambda)$ are slightly modified and we have

$$g_A(\lambda) = -1.01\lambda^5 + 14.98\lambda^4 - 85\lambda^3 + 225\lambda^2 - 274\lambda + 120.$$

Then,

$$g_A(\lambda) = 0 \quad \Rightarrow \quad \lambda = 0.99876,\ 2.21131,\ 2.36314,\ 4.62924 \pm 1.15532\mathrm{i}.$$

Fall 2013 Math 471 Sec 2
Introduction to Numerical Methods

Manabu Machida (University of Michigan)

[1] Math 419

Definition 1. For a given real symmetric matrix A and any $\mathbf{x} \neq \mathbf{0}$, we define

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

This $R_A(\mathbf{x})$ is called the Rayleigh quotient.

If $\mathbf{x} = \mathbf{q}_i$, then

$$R_A(\mathbf{q}_i) = \frac{\mathbf{q}_i^T A\mathbf{q}_i}{\mathbf{q}_i^T\mathbf{q}_i} = \frac{\mathbf{q}_i^T \lambda_i\mathbf{q}_i}{\mathbf{q}_i^T\mathbf{q}_i} = \lambda_i.$$

By the Taylor expansion, we have

$$R_A(\mathbf{x}) = R_A(\mathbf{q}_i) + \nabla R_A(\mathbf{q}_i)\cdot(\mathbf{x}-\mathbf{q}_i) + O\left(\|\mathbf{x}-\mathbf{q}_i\|_2^2\right).$$

Note that

$$\nabla R_A(\mathbf{x}) = \nabla\left(\frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}\right) = \frac{(\mathbf{x}^T\mathbf{x})\nabla(\mathbf{x}^T A\mathbf{x}) - (\mathbf{x}^T A\mathbf{x})\nabla(\mathbf{x}^T\mathbf{x})}{(\mathbf{x}^T\mathbf{x})^2} = \frac{(\mathbf{x}^T\mathbf{x})2A\mathbf{x} - (\mathbf{x}^T A\mathbf{x})2\mathbf{x}}{(\mathbf{x}^T\mathbf{x})^2}$$
$$= \frac{2}{\mathbf{x}^T\mathbf{x}}\Big(A\mathbf{x} - R_A(\mathbf{x})\mathbf{x}\Big),$$

and

$$\nabla R_A(\mathbf{q}_i) = \frac{2}{\mathbf{q}_i^T\mathbf{q}_i}\Big(A\mathbf{q}_i - R_A(\mathbf{q}_i)\mathbf{q}_i\Big) = 0.$$

Therefore, if $\mathbf{x} \approx \mathbf{q}_i$, then $R_A(\mathbf{x})$ is an approximation to λ_i and

$$R_A(\mathbf{x}) = \lambda_i + O\left(\|\mathbf{x}-\mathbf{q}_i\|_2^2\right).$$

The power method

Suppose we have a large $n \times n$ matrix A. We are often interested in obtaining only a few largest eigenvalues of A, or even only the largest eigenvalue.

Let λ_i ($i = 1,\ldots,n$) be the eigenvalues and $\mathbf{q}_i$ be the associated orthonormal eigenvectors. We assume

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|.$$

For a given vector $\mathbf{x}_0$ ($\|\mathbf{x}_0\|_2 = 1$), there exist constants $c_1,\ldots,c_n$ such that

$$\mathbf{x}_0 = c_1\mathbf{q}_1 + \cdots + c_n\mathbf{q}_n.$$

Let us suppose $c_1 \neq 0$, i.e., $\mathbf{q}_1 \cdot \mathbf{x}_0 \neq 0$. We construct $\mathbf{x}_k$ ($k = 1, 2, \ldots$) as

$$\mathbf{y} = A\mathbf{x}_{k-1},$$
$$\mathbf{x}_k = \frac{\mathbf{y}}{\|\mathbf{y}\|_2}.$$

The first step is written as

$$\mathbf{x}_1 = \beta_1 \left(c_1\lambda_1\mathbf{q}_1 + \cdots + c_n\lambda_n\mathbf{q}_n\right), \qquad \beta_1 = \|c_1\lambda_1\mathbf{q}_1 + \cdots + c_n\lambda_n\mathbf{q}_n\|_2^{-1}.$$

In general, we have

$$\mathbf{x}_k = \beta_k \left(c_1\lambda_1^k\mathbf{q}_1 + \cdots + c_n\lambda_n^k\mathbf{q}_n\right) = \beta_k c_1 \lambda_1^k \left[\mathbf{q}_1 + \frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k \mathbf{q}_2 + \cdots + \frac{c_n}{c_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k \mathbf{q}_n\right], \tag{4.1}$$

where $\beta_k = \left|c_1\lambda_1^k\right|^{-1}\left[1 + O\left(|\lambda_2/\lambda_1|\right)^k\right]$. Hence, $\mathbf{x}_k \to \pm\mathbf{q}_1$ ($\pm$ depends on the sign of $c_1\lambda_1^k$) and

$$\|\mathbf{x}_k - (\pm)\mathbf{q}_1\|_2 = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right).$$

Finally,

$$\begin{aligned}\mathbf{x}_k^T A\mathbf{x}_k &= \beta_k^2 (c_1\lambda_1^k)^2 \left[\lambda_1 + \lambda_2\left(\frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k\right)^2 + \cdots + \lambda_n\left(\frac{c_n}{c_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k\right)^2\right] \\ &= \lambda_1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right). \end{aligned} \tag{4.2}$$

The convergence is linear with asymptotic rate $|\lambda_2/\lambda_1|^2$.

Remark 1. The above discussion holds true for general nonsymmetric matrices with linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Instead of (4.2), we can look at a nonzero element $x_i^{(k)}$ in (4.1). We have

$$\frac{x_i^{(k)}}{x_i^{(k-1)}} = \frac{\lambda_1}{|\lambda_1|}\left[1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\right].$$

The convergence is linear with asymptotic rate $|\lambda_2/\lambda_1|$.

Remark 2. Recall that the matrix A for $-D_+D_-$ is tridiagonal:

$$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}.$$

In this case $\mathbf{y} = A\mathbf{x}$ can be coded as the following loop.

```
for i=1:n
    y(i)=(-x(i-1)+2*x(i)-x(i+1))/h^2;
end
```

This is more efficient than forming A and computing $\mathbf{y} = A\mathbf{x}$ by direct matrix-vector multiplication.

The power method is implemented as follows.

Step 1 Give $\mathbf{x}_0$ ($\|\mathbf{x}_0\|_2 = 1$). Set $\lambda^{(0)} = \mathbf{x}_0^T A\mathbf{x}_0$ and $k = 1$.
Step 2 $\mathbf{y} = A\mathbf{x}_{k-1}$.
Step 3 $\mathbf{x}_k = \mathbf{y}/\|\mathbf{y}\|_2$.
Step 4 $\lambda^{(k)} = \mathbf{x}_k^T A\mathbf{x}_k$.
Step 5 Set $k = k+1$ and go to Step 2

Example 2. Let us try the power method for

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}.$$

The eigenvalues of A are $\lambda_1 = 5.214320$, $\lambda_2 = 2.460811$, and $\lambda_3 = 1.324869$. For example, we can choose

$$\mathbf{x}_0 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

We obtain the following results.

k	$\lambda^{(k)}$	$\lvert\lambda^{(k)} - \lambda_1\rvert$	$\lvert\lambda^{(k)} - \lambda_1\rvert/\lvert\lambda^{(k-1)} - \lambda_1\rvert$
0	5.000000	0.214320	–
1	5.181818	0.032502	0.151650
2	5.208193	0.006127	0.188513

Note that $\lambda^{(k)} \to \lambda_1$, $|\lambda^{(k)} - \lambda_1| \to 0$, and $|\lambda^{(k)} - \lambda_1|/|\lambda^{(k-1)} - \lambda_1| \to (\lambda_2/\lambda_1)^2$ as $k \to \infty$.

The inverse power method

Here we try to find the smallest eigenvalue. We note that

$$A\mathbf{q}_i = \lambda_i \mathbf{q}_i \quad \Rightarrow \quad A^{-1}\mathbf{q}_i = \lambda_i^{-1}\mathbf{q}_i.$$

Thus the largest eigenvalue of A^{-1} is λ_n^{-1}.

The inverse power method is implemented as follows.

Step 1 Give $\mathbf{x}_0$ ($\|\mathbf{x}_0\|_2 = 1$). Set $\lambda^{(0)} = \mathbf{x}_0^T A \mathbf{x}_0$ and $k = 1$.
Step 2 Solve $A\mathbf{y} = \mathbf{x}_{k-1}$ (see Chapter 3).
Step 3 $\mathbf{x}_k = \mathbf{y}/\|\mathbf{y}\|_2$.
Step 4 $\lambda^{(k)} = \mathbf{x}_k^T A \mathbf{x}_k$.
Step 5 Set $k = k+1$ and go to Step 2

Example 3. Let us try the inverse power method for the previous example. We obtain the following results.

k	$\lambda^{(k)}$	$\lvert\lambda^{(k)} - \lambda_3\rvert$	$\lvert\lambda^{(k)} - \lambda_3\rvert / \lvert\lambda^{(k-1)} - \lambda_3\rvert$
0	5.000000	3.675131	–
1	3.816327	2.491457	0.677923
2	1.864903	0.540034	0.216754

Note that $\lambda^{(k)} \to \lambda_3$, $|\lambda^{(k)} - \lambda_3| \to 0$, and $|\lambda^{(k)} - \lambda_3|/|\lambda^{(k-1)} - \lambda_3| \to (\lambda_3/\lambda_2)^2$ as $k \to \infty$.