Numerical Analysis

EE, NCKU
Tien-Hao Chang (Darby Chang)

In the previous slide

- Special matrices
  - strictly diagonally dominant matrix
  - symmetric positive definite matrix
    - Cholesky decomposition
  - tridiagonal matrix
- Iterative techniques
  - Jacobi, Gauss-Seidel and SOR methods
  - conjugate gradient method
- Nonlinear systems of equations
- (Exercise 3)

In this slide

- Eigenvalues and eigenvectors
- The power method
  - locate the dominant eigenvalue
- Inverse power method
- Deflation

Chapter 4

Eigenvalues and eigenvectors

Eigenvalues and eigenvectors

- Eigenvalue
  - λ
  - $\boldsymbol{A}\mathbf{v} = \lambda\mathbf{v} \rightarrow (\boldsymbol{A} - \lambda\boldsymbol{I})\mathbf{v} = 0$
  - $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$
    - characteristic polynomial
- Eigenvector
  - the nonzero vector $\mathbf{v}$ for which $\boldsymbol{A}\mathbf{v} = \lambda\mathbf{v}$ associated with the eigenvalue λ

In Chapter 4

- Determine the dominant eigenvalue
- Determine a specific eigenvalue
- Remove a eigenvalue
- Determine all eigenvalues

4.1

The power method

The power method

- Different problems have different requirements
  - a single, several or all of the eigenvalues
  - the corresponding eigenvectors may or may not also be required
- To handle each of these situations efficiently, different strategies are required
- The power method
  - an iterative technique
  - locate the dominant eigenvalue
  - also computes an associated eigenvector
  - can be extended to compute eigenvalues

The power method

Basics

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$, not necessarily distinct, that satisfy the relations $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$. The eigenvalue λ_1, which is largest in magnitude, is known as the *dominant* eigenvalue of the matrix A. Furthermore, assume that the associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ are linearly independent, and therefore form a basis for $\mathbf{R}^n$. It should be noted at this point that not all matrices have eigenvalues and eigenvectors which satisfy the conditions we've assumed here. At the end of the section and in the exercises, we will explore what happens when these conditions are violated.

Let $\mathbf{x}^{(0)}$ be a nonzero element of $\mathbf{R}^n$. Since the eigenvectors of A form a basis for $\mathbf{R}^n$, it follows that $\mathbf{x}^{(0)}$ can be written as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3, \ldots, \mathbf{v}_n$; that is, there exist constants $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$ such that

$$\mathbf{x}^{(0)} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 + \cdots + \alpha_n \mathbf{v}_n.$$

Next, construct the sequence of vectors $\{\mathbf{x}^{(m)}\}$ according to the rule $\mathbf{x}^{(m)} = A\mathbf{x}^{(m-1)}$ for $m \geq 1$. By direct calculation we find

$$
\begin{aligned}
\mathbf{x}^{(1)} &= A\mathbf{x}^{(0)} = \alpha_1(A\mathbf{v}_1) + \alpha_2(A\mathbf{v}_2) + \alpha_3(A\mathbf{v}_3) + \cdots + \alpha_n(A\mathbf{v}_n) \\
&= \alpha_1(\lambda_1\mathbf{v}_1) + \alpha_2(\lambda_2\mathbf{v}_2) + \alpha_3(\lambda_3\mathbf{v}_3) + \cdots + \alpha_n(\lambda_n\mathbf{v}_n), \\
\mathbf{x}^{(2)} &= A\mathbf{x}^{(1)} = A^2\mathbf{x}^{(0)} \\
&= \alpha_1(A^2\mathbf{v}_1) + \alpha_2(A^2\mathbf{v}_2) + \alpha_3(A^2\mathbf{v}_3) + \cdots + \alpha_n(A^2\mathbf{v}_n) \\
&= \alpha_1(\lambda_1^2\mathbf{v}_1) + \alpha_2(\lambda_2^2\mathbf{v}_2) + \alpha_3(\lambda_3^2\mathbf{v}_3) + \cdots + \alpha_n(\lambda_n^2\mathbf{v}_n)
\end{aligned}
$$

and, in general,

$$
\begin{aligned}
\mathbf{x}^{(m)} &= A\mathbf{x}^{(m-1)} = \cdots = A^m\mathbf{x}^{(0)} \\
&= \alpha_1(A^m\mathbf{v}_1) + \alpha_2(A^m\mathbf{v}_2) + \alpha_3(A^m\mathbf{v}_3) + \cdots + \alpha_n(A^m\mathbf{v}_n) \\
&= \alpha_1(\lambda_1^m\mathbf{v}_1) + \alpha_2(\lambda_2^m\mathbf{v}_2) + \alpha_3(\lambda_3^m\mathbf{v}_3) + \cdots + \alpha_n(\lambda_n^m\mathbf{v}_n).
\end{aligned}
$$

In deriving these expressions we have made repeated use of the relation $A\mathbf{v}_j = \lambda_j\mathbf{v}_j$, which follows from the fact that $\mathbf{v}_j$ is an eigenvector associated with the eigenvalue λ_j.

Factoring λ_1^m from the right-hand side of the equation for $\mathbf{x}^{(m)}$ gives

$$\mathbf{x}^{(m)} = \lambda_1^m \left[\alpha_1 \mathbf{v}_1 + \alpha_2 \left(\frac{\lambda_2}{\lambda_1}\right)^m \mathbf{v}_2 + \alpha_3 \left(\frac{\lambda_3}{\lambda_1}\right)^m \mathbf{v}_3 + \cdots + \alpha_n \left(\frac{\lambda_n}{\lambda_1}\right)^m \mathbf{v}_n\right]. \quad (1)$$

By assumption, $|\lambda_j/\lambda_1| < 1$ for each j, so $|\lambda_j/\lambda_1|^m \to 0$ as $m \to \infty$. It therefore follows that

$$\lim_{m\to\infty} \frac{\mathbf{x}^{(m)}}{\lambda_1^m} = \alpha_1 \mathbf{v}_1.$$

Since any nonzero constant times an eigenvector is still an eigenvector associated with the same eigenvalue, we see that the scaled sequence $\{\mathbf{x}^{(m)}/\lambda_1^m\}$ converges to an eigenvector associated with the dominant eigenvalue provided $\alpha_1 \neq 0$. Furthermore, convergence toward the eigenvector is linear with asymptotic error constant $|\lambda_2/\lambda_1|$.

Approximated eigenvalue

An approximation for the dominant eigenvalue of A can be obtained from the sequence $\{\mathbf{x}^{(m)}\}$ as follows. Let i be an index for which $x_i^{(m-1)} \neq 0$, and consider the ratio of the ith element from the vector $\mathbf{x}^{(m)}$ to the ith element from $\mathbf{x}^{(m-1)}$. By equation (1),

$$\frac{x_i^{(m)}}{x_i^{(m-1)}} = \frac{\lambda_1^m \alpha_1 v_{1,i} \left[1 + O\left((\lambda_2/\lambda_1)^m\right)\right]}{\lambda_1^{m-1} \alpha_1 v_{1,i} \left[1 + O\left((\lambda_2/\lambda_1)^{m-1}\right)\right]} = \lambda_1 \left[1 + O\left((\lambda_2/\lambda_1)^{m-1}\right)\right],$$

provided $v_{1,i} \neq 0$, where $v_{1,i}$ denotes the ith element from the vector $\mathbf{v}_1$. Hence, the ratio $x_i^{(m)}/x_i^{(m-1)}$ converges toward the dominant eigenvalue, and the convergence is linear with asymptotic rate constant $|\lambda_2/\lambda_1|$.

 Any Questions?

A common practice

- Make the vector $\mathbf{x}^{(m)}$ have a unit length

To simplify the notation, let's introduce the vector $\mathbf{y}^{(m)}$ to denote the result of multiplying by the matrix A; that is, $\mathbf{y}^{(m)} = A\mathbf{x}^{(m-1)}$. $\mathbf{x}^{(m)}$ is then calculated by the formula

$$\mathbf{x}^{(m)} = \frac{\mathbf{y}^{(m)}}{y_{p_m}^{(m)}},$$

where p_m is an integer chosen so that $|y_{p_m}^{(m)}| = \|\mathbf{y}^{(m)}\|_\infty$. Note that p_m is an index into the vector $\mathbf{y}^{(m)}$.

- Why we need this step?

question

The power method

A common practice

- Make the vector $\mathbf{x}^{(m)}$ have a unit length

To simplify the notation, let's introduce the vector $\mathbf{y}^{(m)}$ to denote the result of multiplying by the matrix A; that is, $\mathbf{y}^{(m)} = A\mathbf{x}^{(m-1)}$. $\mathbf{x}^{(m)}$ is then calculated by the formula

$$\mathbf{x}^{(m)} = \frac{\mathbf{y}^{(m)}}{y_{p_m}^{(m)}},$$

where p_m is an integer chosen so that $|y_{p_m}^{(m)}| = \|\mathbf{y}^{(m)}\|_\infty$. Note that p_m is an index into the vector $\mathbf{y}^{(m)}$.

- Why we need this step?

- Make the vector $\mathbf{x}^{(m)}$ have a unit length
  - to avoid overflow and underflow

To avoid overflow and underflow problems when calculating the sequence $\{\mathbf{x}^{(m)}\}$ (note that $\lim_{m\to\infty}\lambda_1^m \to \pm\infty$ when $|\lambda_1| > 1$, whereas $\lim_{m\to\infty}\lambda_1^m \to 0$ when $|\lambda_1| < 1$), it is common practice to scale the vectors $\mathbf{x}^{(m)}$ so that they are all of unit length. Here, we will use the l_∞-norm to measure vector length. Thus, in a practical implementation of the power method, the vector $\mathbf{x}^{(m)}$ would be computed in two steps: First multiply the previous vector by the matrix A and then scale the resulting vector to unit length.

The power method

Complete procedure

To simplify the notation, let's introduce the vector $\mathbf{y}^{(m)}$ to denote the result of multiplying by the matrix A; that is, $\mathbf{y}^{(m)} = A\mathbf{x}^{(m-1)}$. $\mathbf{x}^{(m)}$ is then calculated by the formula

$$\mathbf{x}^{(m)} = \frac{\mathbf{y}^{(m)}}{y_{p_m}^{(m)}},$$

where p_m is an integer chosen so that $|y_{p_m}^{(m)}| = \|\mathbf{y}^{(m)}\|_\infty$. Note that p_m is an index into the vector $\mathbf{y}^{(m)}$. Whenever there is more than one possible choice for the index p_m, we will adopt the convention of always selecting the smallest value. The vector $\mathbf{x}^{(m)}$ now converges specifically to that multiple of $\mathbf{v}_1$ which has unit length measured in the infinity norm. As for the eigenvalue, since $\mathbf{x}^{(m-1)}$ is approximately an eigenvector associated with λ_1, $\mathbf{y}^{(m)} = A\mathbf{x}^{(m-1)} \approx \lambda_1\mathbf{x}^{(m-1)}$. By construction $x_{p_{m-1}}^{(m-1)} = 1$, so it follows that $y_{p_{m-1}}^{(m)}$ converges to λ_1.

 Any Questions?

http://thomashawk.com/hello/209/1017/1024/Jackson%20Running.jpg

In action

EXAMPLE 4.2 A Demonstration of the Power Method

Consider the matrix

$$A = \begin{bmatrix} -2 & -2 & 3 \\ -10 & -1 & 6 \\ 10 & -2 & -9 \end{bmatrix},$$

whose eigenvalues are $\lambda_1 = -12$, $\lambda_2 = -3$, and $\lambda_3 = 3$. Let's start with the vector $\mathbf{x}^{(0)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$, which already has an infinity norm of 1. Since the first element in $\mathbf{x}^{(0)}$ is the only element that has an absolute value of one, we set $p_0 = 1$.

For the first iteration of the power method, we compute

$$\mathbf{y}^{(1)} = A\mathbf{x}^{(0)} = \begin{bmatrix} -2 & -10 & 10 \end{bmatrix}^T,$$

from which we obtain our first estimate for the dominant eigenvalue: $\mathbf{x}\lambda^{(1)} = y_{p_0}^{(1)} = y_1^{(1)} = -2$. Note that the infinity norm of the vector $\mathbf{y}^{(1)}$ is 10. Sticking with our convention of selecting the smallest index for which the magnitude of the vector element is equal to the infinity norm of the vector, we take $p_1 = 2$. Therefore, for the second iteration, we have

what is the first estimate?

$$\mathbf{x}^{(1)} = \frac{\mathbf{y}^{(1)}}{-10} = \begin{bmatrix} 1/5 & 1 & -1 \end{bmatrix}^T.$$

EXAMPLE 4.2 A Demonstration of the Power Method

Consider the matrix

$$A = \begin{bmatrix} -2 & -2 & 3 \\ -10 & -1 & 6 \\ 10 & -2 & -9 \end{bmatrix},$$

whose eigenvalues are $\lambda_1 = -12$, $\lambda_2 = -3$, and $\lambda_3 = 3$. Let's start with the vector $\mathbf{x}^{(0)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$, which already has an infinity norm of 1. Since the first element in $\mathbf{x}^{(0)}$ is the only element that has an absolute value of one, we set $p_0 = 1$.

For the first iteration of the power method, we compute

$$\mathbf{y}^{(1)} = A\mathbf{x}^{(0)} = \begin{bmatrix} -2 & -10 & 10 \end{bmatrix}^T,$$

from which we obtain our first estimate for the dominant eigenvalue: x$\lambda^{(1)} = y_{p_0}^{(1)} = y_1^{(1)} = -2$. Note that the infinity norm of the vector $\mathbf{y}^{(1)}$ is 10. Sticking with our convention of selecting the smallest index for which the magnitude of the vector element is equal to the infinity norm of the vector, we take $p_1 = 2$. Therefore, for the second iteration, we have

$$\mathbf{x}^{(1)} = \frac{\mathbf{y}^{(1)}}{-10} = \begin{bmatrix} 1/5 & 1 & -1 \end{bmatrix}^T.$$

The calculations for the second iteration produce the results

$$\mathbf{y}^{(2)} = A\mathbf{x}^{(1)} = \begin{bmatrix} -27/5 & -9 & 9 \end{bmatrix}^T,$$
$$\lambda^{(2)} = y_{p_1}^{(2)} = y_2^{(2)} = -9,$$
$$p_2 = 2$$

and

$$\mathbf{x}^{(2)} = \frac{\mathbf{y}^{(2)}}{-9} = \begin{bmatrix} 3/5 & 1 & -1 \end{bmatrix}^T.$$

The third iteration then produces

$$\mathbf{y}^{(3)} = A\mathbf{x}^{(2)} = \begin{bmatrix} -31/5 & -13 & 13 \end{bmatrix}^T,$$
$$\lambda^{(3)} = y_{p_2}^{(3)} = y_2^{(3)} = -13,$$
$$p_3 = 2$$

and

$$\mathbf{x}^{(3)} = \frac{\mathbf{y}^{(3)}}{-13} = \begin{bmatrix} 31/65 & 1 & -1 \end{bmatrix}^T.$$

The following table displays the output from the 11 iterations of the power method needed for the eigenvector to converge to within a tolerance of 5×10^{-6}. The final estimates are

$$\lambda_1 \approx -12.000014 \quad \text{and} \quad \mathbf{v}_1 \approx \begin{bmatrix} 0.500000 & 1.000000 & -1.000000 \end{bmatrix}^T.$$

The values in the column headed "Convergence" were computed according to the formula

$$\left| \frac{\lambda^{(j)} - \lambda^{(j-1)}}{\lambda^{(j-1)} - \lambda^{(j-2)}} \right| .$$

This quantity is an estimate for the asymptotic rate of linear convergence of the sequence $\{\lambda^{(j)}\}$ toward the value $\lambda_1 = -12$. Note the values in this column approach the value predicted by theory: $|\lambda_2/\lambda_1| = 3/12 = 0.25$.

j	$\mathbf{x}^{(j)^T}$			$\lambda^{(j)}$	Convergence
0	1.000000	0.000000	0.000000		
1	0.200000	1.000000	−1.000000	-2.000000	
2	0.600000	1.000000	−1.000000	-9.000000	
3	0.476923	1.000000	−1.000000	-13.000000	0.571429
4	0.505882	1.000000	−1.000000	-11.769231	0.307692
5	0.498537	1.000000	−1.000000	-12.058824	0.235294
6	0.500366	1.000000	−1.000000	-11.985366	0.253659
7	0.499908	1.000000	−1.000000	-12.003663	0.249084
8	0.500023	1.000000	−1.000000	-11.999085	0.250229
9	0.499994	1.000000	−1.000000	-12.000229	0.249943
10	0.500001	1.000000	−1.000000	-11.999943	0.250014
11	0.500000	1.000000	−1.000000	-12.000014	0.249996

Any Questions?

The power method for generic matrices

The power method for symmetric matrices

- When A is symmetric
  - more rapid convergence
    - still linear, but smaller asymptotic error
  - different scaling scheme (norm)
  - based on the theorem

Theorem. If A is an $n \times n$ symmetric matrix, then there exists a set of n eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ that are orthogonal with respect to the standard inner product on $\mathbf{R}^n$; that is, whenever $i \neq j$

$$\mathbf{v}_i^T \mathbf{v}_j = 0.$$

Recall that for arbitrary vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, the standard inner product is the scalar quantity $\mathbf{x}^T\mathbf{y}$ (or $\mathbf{y}^T\mathbf{x}$), and associated with this inner product is the norm $\sqrt{\mathbf{x}^T\mathbf{x}}$. When written in component form,

$$\sqrt{\mathbf{x}^T\mathbf{x}} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2},$$

we can readily see that this is just the l_2, or Euclidean, norm.

The power method variation

To exploit the orthogonality of the eigenvectors of a symmetric matrix within the power method, we will measure vector length and scale the vectors $\mathbf{x}^{(m)}$ to unit length using the Euclidean norm. Furthermore, we will compute an estimate for the dominant eigenvalue using the standard inner product as follows. Premultiplying both sides of the relation $\mathbf{y}^{(m)} \approx \lambda_1 \mathbf{x}^{(m-1)}$ by $\mathbf{x}^{(m-1)^T}$ yields $\mathbf{x}^{(m-1)^T}\mathbf{y}^{(m)} \approx \lambda_1 \mathbf{x}^{(m-1)^T}\mathbf{x}^{(m-1)} = \lambda_1$, since $\mathbf{x}^{(m-1)^T}\mathbf{x}^{(m-1)} = 1$ by construction. Putting these changes together, we arrive at the power method for symmetric matrices:

> let $\mathbf{x}^{(0)}$ be a nonzero element of $\mathbf{R}^n$ with $\mathbf{x}^{(0)^T}\mathbf{x}^{(0)} = 1$. For $m = 1, 2, 3, \ldots$, calculate

$$\begin{aligned} \mathbf{y}^{(m)} &= A\mathbf{x}^{(m-1)} \\ \lambda^{(m)} &= \mathbf{x}^{(m-1)^T}\mathbf{y}^{(m)} \quad \text{and} \\ \mathbf{x}^{(m)} &= \mathbf{y}^{(m)} / \sqrt{\mathbf{y}^{(m)^T}\mathbf{y}^{(m)}}. \end{aligned}$$

Then $\lambda^{(m)} \to \lambda_1$ and $\mathbf{x}^{(m)}$ converges to an eigenvector associated with λ_1 that has unit length in the Euclidean norm.

What about the convergence rate for this version of the power method? Using equation (1) and the orthogonality of the eigenvectors, it follows that

$$\mathbf{x}^{(m-1)} = \frac{\lambda_1^{m-1}\left[\alpha_1\mathbf{v}_1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{m-1}\mathbf{v}_i\right]}{\sqrt{\lambda_1^{2m-2}\left[\alpha_1\mathbf{v}_1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{m-1}\mathbf{v}_i\right]^T\left[\alpha_1\mathbf{v}_1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{m-1}\mathbf{v}_i\right]}}$$

$$= \frac{\alpha_1\mathbf{v}_1 + \sum_{i=2}^n \alpha_i(\lambda_i/\lambda_1)^{m-1}\mathbf{v}_i}{\sqrt{\alpha_1^2\mathbf{v}_1^T\mathbf{v}_1 + \sum_{i=2}^n \alpha_i^2(\lambda_i/\lambda_1)^{2m-2}\mathbf{v}_i^T\mathbf{v}_i}}.$$

Then

$$\begin{aligned}
\mathbf{x}^{(m-1)^T}\mathbf{y}^{(m)} &= \mathbf{x}^{(m-1)^T}A\mathbf{x}^{(m-1)} \\
&= \frac{\lambda_1\left[\alpha_1^2\mathbf{v}_1^T\mathbf{v}_1 + \sum_{i=2}^n \alpha_i^2(\lambda_i/\lambda_1)^{2m-1}\mathbf{v}_i^T\mathbf{v}_i\right]}{\left[\alpha_1^2\mathbf{v}_1^T\mathbf{v}_1 + \sum_{i=2}^n \alpha_i^2(\lambda_i/\lambda_1)^{2m-2}\mathbf{v}_i^T\mathbf{v}_i\right]} \\
&= \lambda_1\left\{1 + O\left((\lambda_2/\lambda_1)^{2(m-1)}\right)\right\},
\end{aligned}$$

and $\mathbf{x}^{(m-1)^T}\mathbf{y}^{(m)} \to \lambda_1$ linearly with asymptotic rate constant $|\lambda_2/\lambda_1|^2$.

The power method

Approximated eigenvalue

An approximation for the dominant eigenvalue of A can be obtained from the sequence $\{\mathbf{x}^{(m)}\}$ as follows. Let i be an index for which $x_i^{(m-1)} \neq 0$, and consider the ratio of the ith element from the vector $\mathbf{x}^{(m)}$ to the ith element from $\mathbf{x}^{(m-1)}$. By equation (1),

$$\frac{x_i^{(m)}}{x_i^{(m-1)}} = \frac{\lambda_1^m \alpha_1 v_{1,i}\left[1 + O\left((\lambda_2/\lambda_1)^m\right)\right]}{\lambda_1^{m-1} \alpha_1 v_{1,i}\left[1 + O\left((\lambda_2/\lambda_1)^{m-1}\right)\right]} = \lambda_1 \left[1 + O\left((\lambda_2/\lambda_1)^{m-1}\right)\right],$$

provided $v_{1,i} \neq 0$, where $v_{1,i}$ denotes the ith element from the vector $\mathbf{v}_1$. Hence, the ratio $x_i^{(m)}/x_i^{(m-1)}$ converges toward the dominant eigenvalue, and the convergence is linear with asymptotic rate constant $|\lambda_2/\lambda_1|$.

Recall that

http://www.dianadepasquale.com/ThinkingMonkey.jpg

Any Questions?

The power method for symmetric matrices

Why to

Require the matrix to be symmetric?

Any Questions?

4.1 The power method

An application of eigenvalue

Application Problem 2: Eigenvalues and Undirected Graphs

Geometrically, an undirected graph consists of a set of marked points, called *vertices*, together with a set of lines which connect pairs of vertices, called *edges*. For example, Figure 4.2 displays an undirected graph with seven vertices (labeled 1, 2, 3, ..., 7) and nine edges.

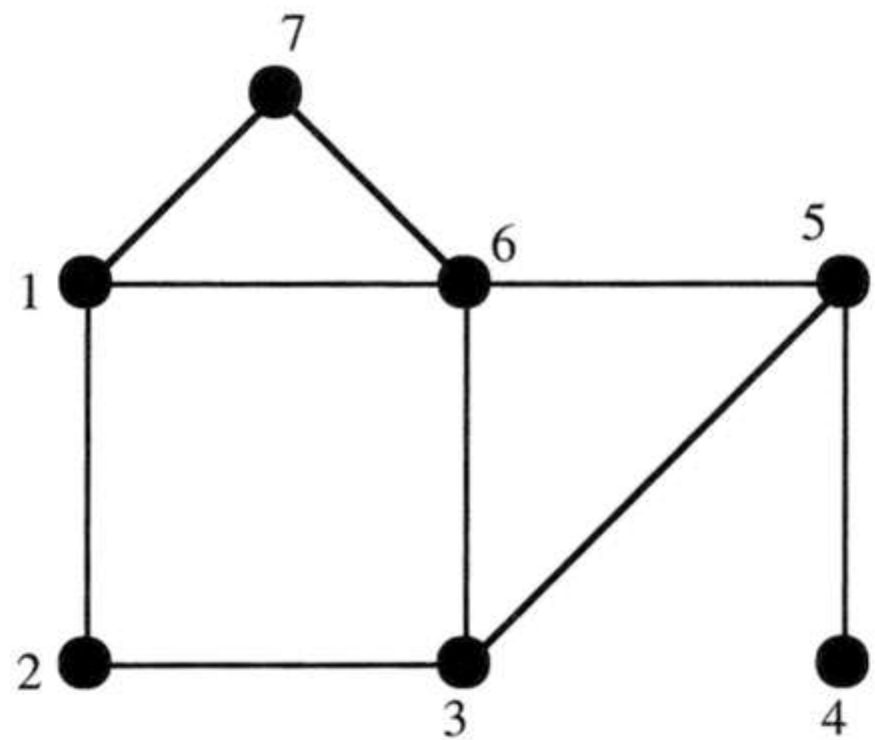


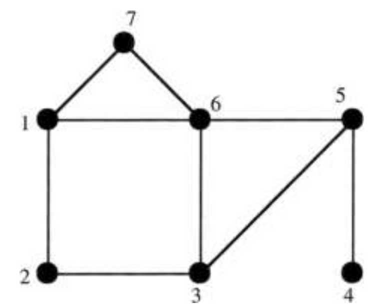
Figure 4.2 An undirected graph.

Two vertices that are connected by an edge are said to be *adjacent*. For instance, vertices 1 and 2 are adjacent in Figure 4.2, but vertices 1 and 3 are not. The overall adjacency structure of an undirected graph can be summarized in an *adjacency matrix*. For an undirected graph with n vertices, the adjacency matrix, which we shall denote by A, is the $n \times n$ matrix whose elements are defined by

$$a_{ij} = \begin{cases} 1, & \text{vertex } i \text{ is adjacent to vertex } j \\ 0, & \text{otherwise}. \end{cases}$$

The adjacency matrix for the undirected graph in Figure 4.2 is

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



Applying the power method for symmetric matrices to A produces the estimates

$$\lambda_1 \approx 2.86081$$

$$\mathbf{v}_1 \approx [\ 0.406691 \quad 0.290865 \quad 0.425420 \quad 0.134554 \quad 0.384933 \quad 0.541244 \quad 0.331352\ ]^T.$$

Relation to eigenvalue

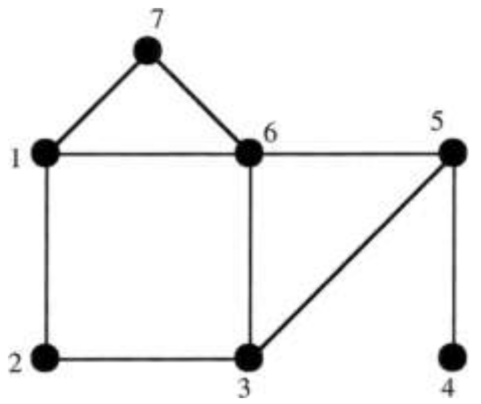
- Proper coloring
  - how to color the geographic regions on a map regions that share a common border receive different colors
- Chromatic number
  - the minimum number of colors that can be used in a proper coloring of a graph

The dominant eigenvalue

What significance does the dominant eigenvalue of an adjacency matrix have? Suppose we wish to assign a color to each vertex in an undirected graph in such a way that adjacent vertices are assigned different colors. Such an assignment is called a *proper coloring* of a graph and may be used by, say, a mapmaker to determine how to color the geographic regions on a map so that regions that share a common border receive different colors. The minimum number of colors that can be used in a proper coloring of a graph is called the *chromatic number* and is denoted by χ. It can be shown that $\frac{n}{n-\lambda_1} \leq \chi \leq 1 + \lambda_1$, where n is the number of vertices in the graph and λ_1 is the dominant eigenvalue of the adjacency matrix. See Cvetkovic, Doob, and Sachs [10] for a proof of the lower bound and van Lint and Wilson [11] for a proof of the upper bound. Hence, for the graph in Figure 4.2, $1.69 \leq \chi \leq 3.86$, or since the chromatic number must be an integer, $2 \leq \chi \leq 3$.

Undirected graph

The corresponding eigenvector



$$\mathbf{v}_1 \approx \begin{bmatrix} 0.406691 & 0.290865 & 0.425420 & 0.134554 \\ 0.384933 & 0.541244 & 0.331352 \end{bmatrix}^T.$$

Next, suppose the vertices in an undirected graph represent cities and an edge represents the existence of a direct traveling route between two cities. Geographers have shown that the entries in an eigenvector associated with the dominant eigenvalue of the adjacency matrix provide a measure of the accessibility of the cities (Straffin [12]). Thus, since the sixth entry in $\mathbf{v}_1$ is the largest, vertex 6 represents the most accessible city. Further, since the first, third, and fifth entries of $\mathbf{v}_1$ are roughly equal, the cities represented by vertices 1, 3, and 5 are nearly equal in terms of accessibility. Finally, as might have been expected, the city represented by vertex 4 is the least accessible.

Any Questions?

4.2

The inverse power method

The inverse power method

- An eigenvalue other than the dominant one
- To derive the inverse power method, we will need
  - the relationship between the eigenvalues of a matrix $\boldsymbol{A}$ to a class of matrices constructed from $\boldsymbol{A}$
- With that, we can
  - transform an eigenvalue of $\boldsymbol{A}$ the dominant eigenvalue of $\boldsymbol{B}$
  - $\boldsymbol{B} = (\boldsymbol{A} - q\boldsymbol{I})^{-1}$

later

B is a polynomial of A

Theorem. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ and associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$.

1. If $\underline{B = a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m = p(A)}$, where p is the polynomial $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$, then the eigenvalues of B are

$$p(\lambda_1), p(\lambda_2), p(\lambda_3), \ldots, p(\lambda_n)$$

with associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$.

2. If A is nonsingular, then $\underline{A^{-1}}$ has eigenvalues

$$\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \frac{1}{\lambda_3}, \ldots, \frac{1}{\lambda_n}$$

with associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$.

Proof. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ and associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$.

Part 1:
Note that for any positive integer k,

$$\begin{aligned} \underline{A^k \mathbf{v}_i} &= A^{k-1}(A\mathbf{v}_i) = \lambda_i A^{k-1}\mathbf{v}_i \\ &= \lambda_i A^{k-2}(A\mathbf{v}_i) = \lambda_i^2 A^{k-2}\mathbf{v}_i \\ &= \cdots \\ &= \lambda_i^{k-1}(A\mathbf{v}_i) = \underline{\lambda_i^k \mathbf{v}_i}. \end{aligned}$$

Now, let $B = a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m = p(A)$, where p is the polynomial $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$. Then, for each $i = 1, 2, 3, \ldots, n$,

$$\begin{aligned} \underline{B\mathbf{v}_i} &= (a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m)\mathbf{v}_i \\ &= a_0 \mathbf{v}_i + a_1 A\mathbf{v}_i + a_2 A^2 \mathbf{v}_i + \cdots + a_m A^m \mathbf{v}_i \\ &= a_0 \mathbf{v}_i + a_1 \lambda_i \mathbf{v}_i + a_2 \lambda_i^2 \mathbf{v}_i + \cdots + a_m \lambda_i^m \mathbf{v}_i \\ &= (a_0 + a_1 \lambda_i + a_2 \lambda_i^2 + \cdots + a_m \lambda_i^m)\mathbf{v}_i \\ &= \underline{p(\lambda_i)\mathbf{v}_i}. \end{aligned}$$

Hence, the eigenvalues of B are

$$p(\lambda_1), p(\lambda_2), p(\lambda_3), \ldots, p(\lambda_n)$$

with associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$.

Suppose A is nonsingular. Since $\mathbf{v}_i$ is an eigenvector associated with the eigenvalue λ_i, it follows that

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

Premultiplying this equation by $(1/\lambda_i)A^{-1}$ yields

$$\frac{1}{\lambda_i} A^{-1} (A\mathbf{v}_i) = \frac{1}{\lambda_i} A^{-1} (\lambda_i \mathbf{v}_i),$$

or

$$\frac{1}{\lambda_i}\mathbf{v}_i = A^{-1}\mathbf{v}_i.$$

Therefore, for each $i = 1, 2, 3, \ldots, n$, $1/\lambda_i$ is an eigenvalue of A^{-1}, with associated eigenvector $\mathbf{v}_i$. □

The inverse power method

- An eigenvalue other than the dominant one
- To derive the inverse power method, we will need
  - the relationship between the eigenvalues of a matrix $\boldsymbol{A}$ to a class of matrices constructed from $\boldsymbol{A}$
- With that, we can
  - transform an eigenvalue of $\boldsymbol{A}$ the dominant eigenvalue of $\boldsymbol{B}$
  - $\boldsymbol{B} = (\boldsymbol{A} - q\boldsymbol{I})^{-1}$

Once again, let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ and associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$. Let q be any constant for which $A - qI$ is nonsingular (this will hold true for any q that is not an eigenvalue of A), and consider the matrix $B = (A - qI)^{-1}$. As a consequence of the theorem we just finished proving, the eigenvalues of B are

$$\mu_1 = \frac{1}{\lambda_1 - q}, \mu_2 = \frac{1}{\lambda_2 - q}, \mu_3 = \frac{1}{\lambda_3 - q}, \ldots, \mu_n = \frac{1}{\lambda_n - q}$$

with associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$.

If we apply the power method to the matrix B, the eigenvalue estimates $\lambda^{(m)}$ will converge to the dominant eigenvalue, say μ_k. Note, however, that μ_k will be the dominant eigenvalue of B if and only if λ_k is the eigenvalue of A that is closest to the number q. Hence, if by some means we determine that A has an eigenvalue in the vicinity of q, we can obtain an approximation to that eigenvalue by applying the power method to the matrix $B = (A - qI)^{-1}$. This procedure is known as the inverse power method.

Any Questions?

How to

Find the eigenvalue smallest in magnitude

Any Questions?

4.2 The inverse power method

4.3

Deflation

Deflation

- So far, we can approximate
  - the dominant eigenvalue of a matrix
  - the one smallest in magnitude
  - the one closest to a specific value
- What if we need several of the largest/smallest eigenvalues?
- Deflation
  - to remove an already determined solution, while leaving the remainder solutions unchanged

- Within the context of polynomial rootfinding
  - remove each root by dividing out the monomial
  - $x^3 - 6x^2 + 11x - 6 = (x-1)(x^2 - 5x + 6) = (x-1)(x-2)(x-3)$
  - $x^2 - 5x + 6 = (x-2)(x-3)$ is a deflation of $x^3 - 6x^2 + 11x - 6$
- For the matrix eigenvalue problem
  - shift the previously determined eigenvalue to zero (while leaving the remainder eigenvalues unchanged)
  - to do this, we need the relationship among the eigenvalues of a matrix $\boldsymbol{A}$ and $\boldsymbol{A}^T$

Eigenvalues and Eigenvectors of a Matrix and Its Transpose

To establish the key theorem behind Wielandt deflation, we must first establish the relationship among the eigenvalues and eigenvectors of a matrix A and its transpose. Recall that the eigenvalues of A^T satisfy the equation $\det(A^T - \lambda I) = 0$. Since

$$\det(A^T - \lambda I) = \det\left[(A - \lambda I)^T\right] = \det(A - \lambda I),$$

it follows that the eigenvalues of A^T are the same as those of A. The eigenvectors are generally not the same, but there is an important relationship among the eigenvectors associated with different eigenvalues.

Let $\mathbf{v}_i$ be an eigenvector for the matrix A associated with the eigenvalue λ_i, and let $\mathbf{w}_j$ be an eigenvector for the matrix A^T associated with the eigenvalue λ_j, where $\lambda_i \neq \lambda_j$. Taking the transpose of the eigenvalue equation $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ and postmultiplying the result by $\mathbf{w}_j$ yields $\mathbf{v}_i^T A^T \mathbf{w}_j = \lambda_i \mathbf{v}_i^T \mathbf{w}_j$. Using the eigenvalue equation $A^T \mathbf{w}_j = \lambda_j \mathbf{w}_j$ then leads to $\lambda_j \mathbf{v}_i^T \mathbf{w}_j = \lambda_i \mathbf{v}_i^T \mathbf{w}_j$ or $(\lambda_j - \lambda_i)\mathbf{v}_i^T \mathbf{w}_j = 0$. Since we have assumed $\lambda_i \neq \lambda_j$, it follows that $\mathbf{v}_i^T \mathbf{w}_j = 0$. In other words, eigenvectors from A and A^T that are associated with different eigenvalues are orthogonal with respect to the standard inner product on $\mathbf{R}^n$.

Theorem. Let A be an $n \times n$ matrix.

1. If A has a row or column consisting only of zero entries, then $\det(A) = 0$;
2. If A has two rows the same or two columns the same, then $\det(A) = 0$;
3. $\det(A^T) = \det(A)$;
4. If A is nonsingular, then $\det(A^{-1}) = (\det(A))^{-1}$;
5. If B is an $n \times n$ matrix, then $\det(AB) = \det(A)\det(B)$.

Recall that

http://www.dianadepasquale.com/ThinkingMonkey.jpg

Deflation

Shift an eigenvalue to zero

An Important Matrix Transformation

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ and associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$. Suppose that by some means (*e.g.*, the power method or the inverse power method) we have obtained approximations for λ_1 and $\mathbf{v}_1$, and we now wish to deflate the spectrum of the matrix A.

Consider the matrix

$$B = A - \lambda_1 \mathbf{v}_1 \mathbf{x}^T, \tag{1}$$

where $\mathbf{x}$ is an arbitrary n-vector. If we postmultiply equation (1) by the vector $\mathbf{v}_1$, we find

$$\begin{aligned} B\mathbf{v}_1 &= A\mathbf{v}_1 - \lambda_1 \mathbf{v}_1 \mathbf{x}^T \mathbf{v}_1 \\ &= \lambda_1 \mathbf{v}_1 - \lambda_1 \mathbf{v}_1 \mathbf{x}^T \mathbf{v}_1 \\ &= \lambda_1 \mathbf{v}_1 \left(1 - \mathbf{x}^T \mathbf{v}_1\right). \end{aligned}$$

Hence, provided $\mathbf{x}$ is chosen so that $\mathbf{x}^T \mathbf{v}_1 = 1$, zero is an eigenvalue of the matrix B with associated eigenvector $\mathbf{v}_1$.

What about the other eigenvalues of A? Have they been changed by the transformation performed in equation (1)? Taking the transpose of equation (1) and postmultiplying the result by $\mathbf{w}_i$, where $i = 2, 3, 4, \ldots, n$ and $\mathbf{w}_i$ is an eigenvector of A^T associated with the eigenvalue λ_i, gives

$$\begin{aligned} B^T\mathbf{w}_i &= A^T\mathbf{w}_i - \lambda_1 \mathbf{x}\mathbf{v}_1^T\mathbf{w}_i \\ &= \lambda_i\mathbf{w}_i - 0 \\ &= \lambda_i\mathbf{w}_i. \end{aligned}$$

Note that in going from the first line to the second, we have used the orthogonality of the eigenvectors of A and A^T that are associated with different eigenvalues. This establishes that $\lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_n$ are eigenvalues of B^T, which implies that they are also eigenvalues of B.

While leaving the remaining eigenvalues unchanged

The last issue to address is the eigenvectors of the matrix B. Let $\mathbf{u}_i$ denote an eigenvector of B associated with the eigenvalue λ_i. We have already established that $\lambda_1 = 0$ and $\mathbf{u}_1 = \mathbf{v}_1$. Given the construction of the matrix B, we can assume that for $i = 2, 3, 4, \ldots, n$,

$$\mathbf{v}_i = \alpha \mathbf{u}_i + \beta \mathbf{v}_1, \tag{2}$$

where α and β are constants whose value is to be determined. If we postmultiply equation (1) by $\mathbf{v}_i$,

$$\begin{aligned} B\mathbf{v}_i &= A\mathbf{v}_i - \lambda_1 \mathbf{v}_1 \mathbf{x}^T \mathbf{v}_i \\ &= \lambda_i \mathbf{v}_i - \lambda_1 \mathbf{v}_1 \mathbf{x}^T \mathbf{v}_i \end{aligned}$$

and then substitute for $\mathbf{v}_i$ from equation (2), we find

$$B(\alpha \mathbf{u}_i + \beta \mathbf{v}_1) = \lambda_i(\alpha \mathbf{u}_i + \beta \mathbf{v}_1) - \lambda_1 \mathbf{v}_1 \mathbf{x}^T (\alpha \mathbf{u}_i + \beta \mathbf{v}_1).$$

Clearing parentheses and using the relations $\mathbf{x}^T \mathbf{v}_1 = 1$, $B\mathbf{v}_1 = 0$, and $B\mathbf{u}_i = \lambda_i \mathbf{u}_i$ leads to

$$\alpha \lambda_i \mathbf{u}_i = \lambda_i \alpha \mathbf{u}_i + \lambda_i \beta \mathbf{v}_1 - \alpha \lambda_1 (\mathbf{x}^T \mathbf{u}_i) \mathbf{v}_1 - \beta \lambda_1 \mathbf{v}_1,$$

or

$$0 = \left[\beta(\lambda_i - \lambda_1) - \alpha \lambda_1 (\mathbf{x}^T \mathbf{u}_i)\right] \mathbf{v}_1. \tag{3}$$

One solution of this equation is

$$\alpha = (\lambda_i - \lambda_1) \quad \text{and} \quad \beta = \lambda_1 (\mathbf{x}^T \mathbf{u}_i),$$

which yields

$$\mathbf{v}_i = (\lambda_i - \lambda_1) \mathbf{u}_i + \lambda_1 (\mathbf{x}^T \mathbf{u}_i) \mathbf{v}_1. \tag{4}$$

Any other solution to equation (3) will produce a multiple of the eigenvector given by this last equation.

Deflation

Summary

Theorem. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ and associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$, and let $\mathbf{x}$ be any n-vector for which $\mathbf{x}^T\mathbf{v}_1 = 1$. Then the matrix

$$B = A - \lambda_1 \mathbf{v}_1 \mathbf{x}^T$$

has eigenvalues $0, \lambda_2, \lambda_3, \ldots, \lambda_n$ with associated eigenvectors $\mathbf{v}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n$ where for $i = 2, 3, 4, \ldots, n$,

$$\mathbf{v}_i = (\lambda_i - \lambda_1)\mathbf{u}_i + \lambda_1(\mathbf{x}^T\mathbf{u}_i)\mathbf{v}_1.$$

Any Questions?

Do we

Miss something?

An Important Matrix Transformation

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ and associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$. Suppose that by some means (*e.g.*, the power method or the inverse power method) we have obtained approximations for λ_1 and $\mathbf{v}_1$, and we now wish to deflate the spectrum of the matrix A.

Consider the matrix

$$B = A - \lambda_1 \mathbf{v}_1 \mathbf{x}^T, \tag{1}$$

where $\mathbf{x}$ is an arbitrary n-vector. If we postmultiply equation (1) by the vector $\mathbf{v}_1$, we find

$$\begin{aligned} B\mathbf{v}_1 &= A\mathbf{v}_1 - \lambda_1 \mathbf{v}_1 \mathbf{x}^T \mathbf{v}_1 \\ &= \lambda_1 \mathbf{v}_1 - \lambda_1 \mathbf{v}_1 \mathbf{x}^T \mathbf{v}_1 \\ &= \lambda_1 \mathbf{v}_1 \left(1 - \mathbf{x}^T \mathbf{v}_1\right). \end{aligned}$$

Hence, provided $\mathbf{x}$ is chosen so that $\mathbf{x}^T \mathbf{v}_1 = 1$, zero is an eigenvalue of the matrix B with associated eigenvector $\mathbf{v}_1$.

Recall that

http://www.dianadepasquale.com/ThinkingMonkey.jpg

How to choose $\mathbf{x}$ for the formula $\boldsymbol{B} = \boldsymbol{A} - \lambda_1 \mathbf{v}_1 \mathbf{x}^T$?

Wielandt deflation

Wielandt Deflation

In Wielandt Deflation, the deflation vector $\mathbf{x}$ is chosen to be

$$\mathbf{x} = \frac{1}{\lambda_1 v_{1,k}} \begin{bmatrix} a_{k1} \\ a_{k2} \\ a_{k3} \\ \cdot \\ \cdot \\ \cdot \\ a_{kn} \end{bmatrix},$$

where $v_{1,k}$ denotes the kth element of the vector $\mathbf{v}_1$. The values $a_{k1}, a_{k2}, a_{k3}, \ldots, a_{kn}$ correspond to the kth row of the matrix A written as a column vector. The value of k can be any index for which $v_{1,k}$ is nonzero, but we will consistently choose the smallest index for which $|v_{1,k}|$ is equal to the infinity norm of the vector $\mathbf{v}_1$. With this choice for $\mathbf{x}$,

$$\begin{aligned} \mathbf{x}^T \mathbf{v}_1 &= \frac{1}{\lambda_1 v_{1,k}} [k\text{th row of } A]\, \mathbf{v}_1 \\ &= \frac{1}{\lambda_1 v_{1,k}} [k\text{th element of the product } A\mathbf{v}_1] \\ &= \frac{1}{\lambda_1 v_{1,k}} [k\text{th element of } \lambda_1 \mathbf{v}_1] \\ &= \frac{1}{\lambda_1 v_{1,k}} \lambda_1 v_{1,k} = 1. \end{aligned}$$

Therefore, the hypothesis of the deflation theorem is satisfied.

Wielandt deflation

Bonus

We get an extra bonus with the Wielandt deflation vector. Each row of the matrix $\lambda_1 \mathbf{v}_1 \mathbf{x}^T$ is a multiple of the kth row of A. In particular, the ith row of $\lambda_1 \mathbf{v}_1 \mathbf{x}^T$ is $v_{1,i}/v_{1,k}$ times the kth row of A. This implies that the kth row of $B = A - \lambda_1 \mathbf{v}_1 \mathbf{x}^T$ consists entirely of zeros. Suppose that $\mathbf{u}$ is an eigenvector of B associated with the eigenvalue $\lambda \neq 0$. Given that B has all zeros along the kth row, the kth element of the product $B\mathbf{u}$, which is just λu_k, must be zero, and therefore $u_k = 0$. This, in turn, implies that the kth column of B has no influence on the product $B\mathbf{u}$. Thus, before searching for the next eigenpair, we can reduce the size of B by deleting the kth row and the kth column.

When we take advantage of this reduction in size, an extra detail must be accounted for when equation (4) is applied to convert the eigenvectors of B into the eigenvectors of A. The vector $\mathbf{u}$ that appears on the right-hand side of the equation will be one element smaller than the other vectors. To compensate for the size difference, a zero must be placed between the $(k-1)$st and kth elements of $\mathbf{u}$ before equation (4) is used.

http://thomashawk.com/hello/209/1017/1024/Jackson%20Running.jpg

In action

EXAMPLE 4.6 Wielandt Deflation in Action

Consider the 4×4 matrix

$$A = \begin{bmatrix} 11 & -6 & 4 & -2 \\ 4 & 1 & 0 & 0 \\ -9 & 9 & -6 & 5 \\ -6 & 6 & -6 & 7 \end{bmatrix}.$$

Let's determine the two largest eigenvalues and associated eigenvectors of A.

Applying the power method, we find the dominant eigenvalue of A to be $\lambda_1 = 5$, with corresponding eigenvector $\mathbf{v}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^T$. In order to focus on the deflation process, we will ignore the effects of roundoff error in this example. With $k = 1$, the deflation vector is

$$\mathbf{x} = \frac{1}{\lambda_1 v_{1,1}} [\text{first row of } A]^T = \frac{1}{5} \begin{bmatrix} 11 & -6 & 4 & -2 \end{bmatrix}^T.$$

Forming the matrix $\lambda_1 \mathbf{v}_1 \mathbf{x}^T$ and subtracting the result from A gives the matrix B:

$$\lambda_1 \mathbf{v}_1 \mathbf{x}^T = \begin{bmatrix} 11 & -6 & 4 & -2 \\ 11 & -6 & 4 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -7 & 7 & -4 & 2 \\ -9 & 9 & -6 & 5 \\ -6 & 6 & -6 & 7 \end{bmatrix}.$$

Deleting the first row and first column from this matrix produces the matrix

$$B' = \begin{bmatrix} 7 & -4 & 2 \\ 9 & -6 & 5 \\ 6 & -6 & 7 \end{bmatrix}.$$

Applying the power method to B' generates $\lambda_2 = 4$ and $\mathbf{u}_2' = \begin{bmatrix} 0 & 1/2 & 1 \end{bmatrix}^T$.

To complete the deflation process, we must convert the eigenvector $\mathbf{u}_2'$ to correspond to the original matrix. Since $k = 1$, we first prepend a zero to $\mathbf{u}_2'$ to create the 4-vector $\mathbf{u}_2$ which appears on the right-hand side of equation (4). Using equation (4), we then obtain

$$\begin{aligned}
\mathbf{v}_2 &= (\lambda_2 - \lambda_1)\mathbf{u}_2 + \lambda_1 \left(\mathbf{x}^T\mathbf{u}_2\right)\mathbf{v}_1 \\
&= -\begin{bmatrix} 0 \\ 0 \\ 1/2 \\ 1 \end{bmatrix} + 5 \cdot \frac{1}{5}\left(\begin{bmatrix} 11 & -6 & 4 & -2 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1/2 \\ 1 \end{bmatrix}\right)\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 0 \\ -1/2 \\ -1 \end{bmatrix}.
\end{aligned}$$

We therefore have the eigenpairs

$$\left(5, \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^T\right) \quad \text{and} \quad \left(4, \begin{bmatrix} 0 & 0 & -1/2 & -1 \end{bmatrix}^T\right).$$

Hotelling deflation

- Recall that we choose $v_{1,k}$ based on infinity norm
- Like the power method, there is another deflation variation for symmetric matrices

for λ_1 and $\mathbf{v}_1$, and consider the matrix

$$B = A - \frac{\lambda_1}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1 \mathbf{v}_1^T. \tag{5}$$

To begin, note that

$$\begin{aligned} B^T &= A^T - \frac{\lambda_1}{\mathbf{v}_1^T \mathbf{v}_1} (\mathbf{v}_1 \mathbf{v}_1^T)^T \\ &= A - \frac{\lambda_1}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1 \mathbf{v}_1^T = B, \end{aligned}$$

so B is a symmetric matrix. Next, by direct calculation, we find

$$\begin{aligned} B\mathbf{v}_1 &= A\mathbf{v}_1 - \frac{\lambda_1}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1 \mathbf{v}_1^T \mathbf{v}_1 \\ &= \lambda_1 \mathbf{v}_1 - \lambda_1 \mathbf{v}_1 \\ &= 0, \end{aligned}$$

$i = 2, 3, 4, \ldots, n$

$$\begin{aligned} B\mathbf{v}_i &= A\mathbf{v}_i - \frac{\lambda_1}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1 \mathbf{v}_1^T \mathbf{v}_i \\ &= \lambda_i \mathbf{v}_i - 0 \\ &= \lambda_i \mathbf{v}_i, \end{aligned}$$

where, in going from the first line to the second, we have used the orthogonality of the eigenvectors of a symmetric matrix. Thus the transformation given by equation (5) shifts the eigenvalue λ_1 to zero, but preserves every other eigenvalue and every eigenvector of the matrix A.

Any Questions?

4.3 Deflation