Scientific Computing: An Introductory Survey

Chapter 4 – Eigenvalue Problems

Prof. Michael T. Heath

Department of Computer Science
University of Illinois at Urbana-Champaign

Outline

Eigenvalue Problems

- Eigenvalue problems occur in many areas of science and engineering, such as structural analysis
- Eigenvalues are also important in analyzing numerical methods
- Theory and algorithms apply to complex matrices as well as real matrices
- With complex matrices, we use conjugate transpose, $\boldsymbol{A}^H$, instead of usual transpose, $\boldsymbol{A}^T$

Eigenvalues and Eigenvectors

- Standard *eigenvalue problem*: Given $n \times n$ matrix $\boldsymbol{A}$, find scalar λ and nonzero vector $\boldsymbol{x}$ such that

$$\boldsymbol{A}\,\boldsymbol{x} = \lambda\,\boldsymbol{x}$$

- λ is *eigenvalue*, and $\boldsymbol{x}$ is corresponding *eigenvector*
- λ may be complex even if $\boldsymbol{A}$ is real
- *Spectrum* $= \lambda(\boldsymbol{A}) =$ set of eigenvalues of $\boldsymbol{A}$
- *Spectral radius* $= \rho(\boldsymbol{A}) = \max\{|\lambda| \,:\, \lambda \in \lambda(\boldsymbol{A})\}$

Geometric Interpretation

symmetric

$x = a + b + c \qquad Ax = \lambda_1 a + \lambda_2 b + \lambda_3 c + \dots$

- Matrix expands or shrinks any vector lying in direction of eigenvector by scalar factor
- Expansion or contraction factor is given by corresponding eigenvalue λ
- Eigenvalues and eigenvectors decompose complicated behavior of general linear transformation into simpler actions

Examples: Eigenvalues and Eigenvectors

- $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$: $\lambda_1 = 1$, $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\lambda_2 = 2$, $\boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
- $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$: $\lambda_1 = 1$, $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\lambda_2 = 2$, $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- $\boldsymbol{A} = \boxed{\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}}$: $\lambda_1 = 2$, $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\lambda_2 = 4$, $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
- $\boldsymbol{A} = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}$: $\lambda_1 = 2$, $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\lambda_2 = 1$, $\boldsymbol{x}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$
- $\boldsymbol{A} = \boxed{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}$: $\lambda_1 = i$, $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}$, $\lambda_2 = -i$, $\boldsymbol{x}_2 = \begin{bmatrix} i \\ 1 \end{bmatrix}$

  where $i = \sqrt{-1}$

Characteristic Polynomial

- Equation $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ is equivalent to
$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$$
which has nonzero solution $\boldsymbol{x}$ if, and only if, its matrix is singular
- Eigenvalues of $\boldsymbol{A}$ are roots λ_i of *characteristic polynomial*
$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$$
in λ of degree n
- *Fundamental Theorem of Algebra* implies that $n \times n$ matrix $\boldsymbol{A}$ always has n eigenvalues, but they may not be real nor distinct
- Complex eigenvalues of real matrix occur in complex conjugate pairs: if $\alpha + i\beta$ is eigenvalue of real matrix, then so is $\alpha - i\beta$, where $i = \sqrt{-1}$

Example: Characteristic Polynomial

- Characteristic polynomial of previous example matrix is

$$\det\left(\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) =$$

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$

$$\det\left(\begin{bmatrix} 3-\lambda & -1 \\ -1 & 3-\lambda \end{bmatrix}\right) =$$

$$(3-\lambda)(3-\lambda) - (-1)(-1) = \lambda^2 - 6\lambda + 8 = 0$$

so eigenvalues are given by

$$\lambda = \frac{6 \pm \sqrt{36-32}}{2}, \quad \text{or} \quad \lambda_1 = 2, \quad \lambda_2 = 4$$

Companion Matrix

- Monic polynomial

$$p(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n$$

is characteristic polynomial of *companion matrix*

Matrices $\xrightarrow{\det}$ polynomial

$$C_n = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix}$$

- Roots of polynomial of degree > 4 cannot always computed in finite number of steps (Abel 1824)
- So in general, computation of eigenvalues of matrices of order > 4 requires (theoretically infinite) iterative process

Characteristic Polynomial, continued

- Computing eigenvalues using characteristic polynomial is *not* recommended because of
  - work in computing coefficients of characteristic polynomial
  - sensitivity of coefficients of characteristic polynomial
  - work in solving for roots of characteristic polynomial
- Characteristic polynomial is powerful theoretical tool but usually not useful computationally

Example: Characteristic Polynomial

- Consider

$$\boldsymbol{A} = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix}$$

where ϵ is positive number slightly smaller than $\sqrt{\epsilon_{\text{mach}}}$

- Exact eigenvalues of $\boldsymbol{A}$ are $1 + \epsilon$ and $1 - \epsilon$

- Computing characteristic polynomial in floating-point arithmetic, we obtain

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \lambda^2 - 2\lambda + \overbrace{(1 - \epsilon^2)}^{fl(1-\varepsilon^2) = 1} = \lambda^2 - 2\lambda + 1$$

which has 1 as double root

- Thus, eigenvalues cannot be resolved by this method even though they are distinct in working precision

Multiplicity and Diagonalizability

algebraic

- *Multiplicity* is number of times root appears when polynomial is written as product of linear factors
- Eigenvalue of multiplicity 1 is *simple*
- *Defective* matrix has eigenvalue of multiplicity $k > 1$ with fewer than k linearly independent corresponding eigenvectors
- Nondefective matrix $\boldsymbol{A}$ has n linearly independent eigenvectors, so it is *diagonalizable*

$$\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \boldsymbol{D}$$

where $\boldsymbol{X}$ is nonsingular matrix of eigenvectors

Consider $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$

$y = 0$

$3x + y = \lambda x = 3x$

$3y = \lambda y$

Eigenspaces and Invariant Subspaces

- Eigenvectors can be scaled arbitrarily: if $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, then $\boldsymbol{A}(\gamma\boldsymbol{x}) = \lambda(\gamma\boldsymbol{x})$ for any scalar γ, so $\gamma\boldsymbol{x}$ is also eigenvector corresponding to λ
- Eigenvectors are usually *normalized* by requiring some norm of eigenvector to be 1
- *Eigenspace* $= \mathcal{S}_\lambda = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}\}$

$$\begin{bmatrix} 3I & 0 \\ 0 & A \end{bmatrix}$$

- Subspace $\mathcal{S}$ of $\mathbb{R}^n$ (or $\mathbb{C}^n$) is *invariant* if $\boldsymbol{A}\mathcal{S} \subseteq \mathcal{S}$
- For eigenvectors $\boldsymbol{x}_1 \cdots \boldsymbol{x}_p$, span($[\boldsymbol{x}_1 \cdots \boldsymbol{x}_p]$) is invariant subspace

Relevant Properties of Matrices

- Properties of matrix A relevant to eigenvalue problems

upper Hessenberg

Property	Definition
diagonal	$a_{ij} = 0$ for $i \neq j$
tridiagonal	$a_{ij} = 0$ for $\vert i - j \vert > 1$
triangular	$a_{ij} = 0$ for $i > j$ (upper)
	$a_{ij} = 0$ for $i < j$ (lower)
Hessenberg	$a_{ij} = 0$ for $i > j + 1$ (upper)
	$a_{ij} = 0$ for $i < j - 1$ (lower)
orthogonal	$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}$
unitary	$\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{I}$
symmetric	$\boldsymbol{A} = \boldsymbol{A}^T$
Hermitian	$\boldsymbol{A} = \boldsymbol{A}^H$
normal	$\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^H$

$$\begin{bmatrix} 2 & i \\ i & 3 \end{bmatrix}$$

Examples: Matrix Properties

- Transpose: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$
- Conjugate transpose: $\begin{bmatrix} 1+i & 1+2i \\ 2-i & 2-2i \end{bmatrix}^H = \begin{bmatrix} 1-i & 2+i \\ 1-2i & 2+2i \end{bmatrix}$
- Symmetric: $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$
- Nonsymmetric: $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$
- Hermitian: $\begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$
- NonHermitian: $\begin{bmatrix} 1 & 1+i \\ 1+i & 2 \end{bmatrix}$ symmetric

Examples, continued

- Orthogonal: $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$, $\begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$
- Unitary: $\begin{bmatrix} i\sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & -i\sqrt{2}/2 \end{bmatrix}$
- Nonorthogonal: $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$
- Normal: $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix}$

$A^T A = A A^T$

- Nonnormal: $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

Properties of Eigenvalue Problems

Properties of eigenvalue problem affecting choice of algorithm and software

- Are all eigenvalues needed, or only a few?
- Are only eigenvalues needed, or are corresponding eigenvectors also needed?
- Is matrix real or complex?
- Is matrix relatively small and dense, or large and sparse?
- Does matrix have any special properties, such as symmetry, or is it general matrix?

Conditioning of Eigenvalue Problems

- Condition of eigenvalue problem is sensitivity of eigenvalues and eigenvectors to changes in matrix
- Conditioning of eigenvalue problem is *not* same as conditioning of solution to linear system for same matrix
- Different eigenvalues and eigenvectors are not necessarily equally sensitive to perturbations in matrix

Conditioning of Eigenvalues

- If μ is eigenvalue of perturbation $\boldsymbol{A}+\boldsymbol{E}$ of nondefective matrix $\boldsymbol{A}$, then

$$|\mu - \lambda_k| \leq \text{cond}_2(\boldsymbol{X})\, \|\boldsymbol{E}\|_2$$

  where λ_k is closest eigenvalue of $\boldsymbol{A}$ to μ and $\boldsymbol{X}$ is nonsingular matrix of eigenvectors of $\boldsymbol{A}$
- Absolute condition number of eigenvalues is condition number of matrix of eigenvectors with respect to solving linear equations
- Eigenvalues may be sensitive if eigenvectors are nearly linearly dependent (i.e., matrix is nearly defective)
- For *normal* matrix ($\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^H$), eigenvectors are orthogonal, so eigenvalues are well-conditioned

Conditioning of Eigenvalues

- If $(\boldsymbol{A}+\boldsymbol{E})(\boldsymbol{x}+\Delta\boldsymbol{x}) = (\lambda + \Delta\lambda)(\boldsymbol{x}+\Delta\boldsymbol{x})$, where λ is simple eigenvalue of $\boldsymbol{A}$, then

$\lambda x = \lambda A$

$$|\Delta\lambda| \lessapprox \frac{\|\boldsymbol{y}\|_2 \cdot \|\boldsymbol{x}\|_2}{|\boldsymbol{y}^H\boldsymbol{x}|}\|\boldsymbol{E}\|_2 = \frac{1}{\cos(\theta)}\|\boldsymbol{E}\|_2$$

  where $\boldsymbol{x}$ and $\boldsymbol{y}$ are corresponding right and left eigenvectors and θ is angle between them
- For symmetric or Hermitian matrix, right and left eigenvectors are same, so $\cos(\theta) = 1$ and eigenvalues are inherently well-conditioned
- Eigenvalues of nonnormal matrices may be sensitive
- For multiple or closely clustered eigenvalues, corresponding eigenvectors may be sensitive

Problem Transformations

- *Shift*: If $\boldsymbol{Ax} = \lambda \boldsymbol{x}$ and σ is any scalar, then $(\boldsymbol{A} - \sigma \boldsymbol{I})\boldsymbol{x} = (\lambda - \sigma)\boldsymbol{x}$, so eigenvalues of shifted matrix are shifted eigenvalues of original matrix
- *Inversion*: If $\boldsymbol{A}$ is nonsingular and $\boldsymbol{Ax} = \lambda \boldsymbol{x}$ with $\boldsymbol{x} \neq \boldsymbol{0}$, then $\lambda \neq 0$ and $\boldsymbol{A}^{-1}\boldsymbol{x} = (1/\lambda)\boldsymbol{x}$, so eigenvalues of inverse are reciprocals of eigenvalues of original matrix
- *Powers*: If $\boldsymbol{Ax} = \lambda \boldsymbol{x}$, then $\boldsymbol{A}^k\boldsymbol{x} = \lambda^k \boldsymbol{x}$, so eigenvalues of power of matrix are same power of eigenvalues of original matrix
- *Polynomial*: If $\boldsymbol{Ax} = \lambda \boldsymbol{x}$ and $p(t)$ is polynomial, then $p(\boldsymbol{A})\boldsymbol{x} = p(\lambda)\boldsymbol{x}$, so eigenvalues of polynomial in matrix are values of polynomial evaluated at eigenvalues of original matrix

Similarity Transformation

- B is *similar* to A if there is nonsingular matrix T such that

$$B = T^{-1} A \, T$$

T can be X

- Then

$TBT^{-1} = A$

$$By = \lambda y \;\Rightarrow\; T^{-1}ATy = \lambda y \;\Rightarrow\; A(Ty) = \lambda(Ty)$$

so A and B have same eigenvalues, and if y is eigenvector of B, then $x = Ty$ is eigenvector of A

- Similarity transformations preserve eigenvalues and eigenvectors are easily recovered

Example: Similarity Transformation

- From eigenvalues and eigenvectors for previous example,

$$\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

-1

and hence

$$\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

- So original matrix is similar to diagonal matrix, and eigenvectors form columns of similarity transformation matrix

Diagonal Form

- Eigenvalues of diagonal matrix are diagonal entries, and eigenvectors are columns of identity matrix
- Diagonal form is desirable in simplifying eigenvalue problems for general matrices by similarity transformations
- But not all matrices are diagonalizable by similarity transformation
- Closest one can get, in general, is *Jordan form*, which is nearly diagonal but may have some nonzero entries on first superdiagonal, corresponding to one or more multiple eigenvalues

Triangular Form

- Any matrix can be transformed into triangular (*Schur*) form by similarity, and eigenvalues of triangular matrix are diagonal entries
- Eigenvectors of triangular matrix less obvious, but still straightforward to compute
- If

$$\boldsymbol{A} - \lambda \boldsymbol{I} = \begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{u} & \boldsymbol{U}_{13} \\ \boldsymbol{0} & 0 & \boldsymbol{v}^T \\ \boldsymbol{O} & \boldsymbol{0} & \boldsymbol{U}_{33} \end{bmatrix}$$

is triangular, then $\boldsymbol{U}_{11}\boldsymbol{y} = \boldsymbol{u}$ can be solved for $\boldsymbol{y}$, so that

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{y} \\ -1 \\ \boldsymbol{0} \end{bmatrix}$$

is corresponding eigenvector

Block Triangular Form

- If
$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ & A_{22} & \cdots & A_{2p} \\ & & \ddots & \vdots \\ & & & A_{pp} \end{bmatrix}$$
with *square* diagonal blocks, then
$$\lambda(A) = \bigcup_{j=1}^{p} \lambda(A_{jj})$$
so eigenvalue problem breaks into p smaller eigenvalue problems
- *Real* Schur form has 1×1 diagonal blocks corresponding to real eigenvalues and 2×2 diagonal blocks corresponding to pairs of complex conjugate eigenvalues

Forms Attainable by Similarity

A	T	B
distinct eigenvalues	nonsingular	diagonal
real symmetric	orthogonal	real diagonal
complex Hermitian	unitary	real diagonal
normal	unitary	diagonal
arbitrary real	orthogonal	real block triangular (real Schur)
arbitrary	unitary	upper triangular (Schur)
arbitrary	nonsingular	almost diagonal (Jordan)

- Given matrix A with indicated property, matrices B and T exist with indicated properties such that $B = T^{-1}AT$
- If B is diagonal or triangular, eigenvalues are its diagonal entries
- If B is diagonal, eigenvectors are columns of T

Power Iteration

$x = a + b + \dots \quad A^k x = \lambda_1^k a + \lambda_2^k b + \dots$

- Simplest method for computing one eigenvalue-eigenvector pair is *power iteration*, which repeatedly multiplies matrix times initial starting vector
- Assume $\boldsymbol{A}$ has unique eigenvalue of maximum modulus, say λ_1, with corresponding eigenvector $\boldsymbol{v}_1$
- Then, starting from nonzero vector $\boldsymbol{x}_0$, iteration scheme

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1}$$

converges to multiple of eigenvector $\boldsymbol{v}_1$ corresponding to *dominant* eigenvalue λ_1

Convergence of Power Iteration

- To see why power iteration converges to dominant eigenvector, express starting vector $\boldsymbol{x}_0$ as linear combination

$$\boldsymbol{x}_0 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

where $\boldsymbol{v}_i$ are eigenvectors of $\boldsymbol{A}$

- Then

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} = \boldsymbol{A}^2\boldsymbol{x}_{k-2} = \cdots = \boldsymbol{A}^k\boldsymbol{x}_0 =$$

$$\sum_{i=1}^{n} \lambda_i^k \alpha_i \boldsymbol{v}_i = \lambda_1^k \left(\alpha_1 \boldsymbol{v}_1 + \sum_{i=2}^{n} (\lambda_i/\lambda_1)^k \alpha_i \boldsymbol{v}_i \right)$$

- Since $|\lambda_i/\lambda_1| < 1$ for $i > 1$, successively higher powers go to zero, leaving only component corresponding to $\boldsymbol{v}_1$

Example: Power Iteration

- Ratio of values of given component of $\boldsymbol{x}_k$ from one iteration to next converges to dominant eigenvalue λ_1
- For example, if $\boldsymbol{A} = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}$ and $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we obtain

k	$\boldsymbol{x}_k^T$		ratio
0	0.0	1.0	
1	0.5	1.5	1.500
2	1.5	2.5	1.667
3	3.5	4.5	1.800
4	7.5	8.5	1.889
5	15.5	16.5	1.941
6	31.5	32.5	1.970
7	63.5	64.5	1.985
8	127.5	128.5	1.992

- Ratio is converging to dominant eigenvalue, which is 2

Limitations of Power Iteration

Power iteration can fail for various reasons

- Starting vector may have *no* component in dominant eigenvector $\boldsymbol{v}_1$ (i.e., $\alpha_1 = 0$) — not problem in practice because rounding error usually introduces such component in any case
- There may be more than one eigenvalue having same (maximum) modulus, in which case iteration may converge to linear combination of corresponding eigenvectors
- For real matrix and starting vector, iteration can never converge to complex vector

Normalized Power Iteration

- Geometric growth of components at each iteration risks eventual overflow (or underflow if $\lambda_1 < 1$)
- Approximate eigenvector should be normalized at each iteration, say, by requiring its largest component to be 1 in modulus, giving iteration scheme

$$\begin{aligned} \boldsymbol{y}_k &= \boldsymbol{A}\boldsymbol{x}_{k-1} \\ \boldsymbol{x}_k &= \boldsymbol{y}_k / \|\boldsymbol{y}_k\|_\infty \end{aligned}$$

- With normalization, $\|\boldsymbol{y}_k\|_\infty \to |\lambda_1|$, and $\boldsymbol{x}_k \to \boldsymbol{v}_1 / \|\boldsymbol{v}_1\|_\infty$

Example: Normalized Power Iteration

- Repeating previous example with normalized scheme,

k	$\boldsymbol{x}_k^T$		$\lVert \boldsymbol{y}_k \rVert_\infty$
0	0.000	1.0	
1	0.333	1.0	1.500
2	0.600	1.0	1.667
3	0.778	1.0	1.800
4	0.882	1.0	1.889
5	0.939	1.0	1.941
6	0.969	1.0	1.970
7	0.984	1.0	1.985
8	0.992	1.0	1.992

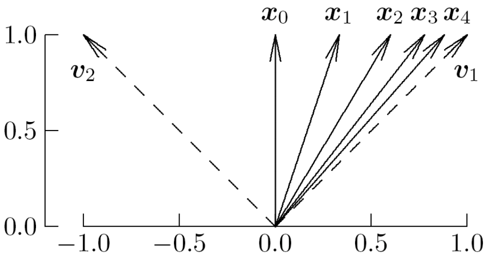
normalized

$y_k = A x_k$

< interactive example >

Geometric Interpretation

- Behavior of power iteration depicted geometrically



- Initial vector $x_0 = v_1 + v_2$ contains equal components in eigenvectors v_1 and v_2 (dashed arrows)
- Repeated multiplication by A causes component in v_1 (corresponding to larger eigenvalue, 2) to dominate, so sequence of vectors x_k converges to v_1

Power Iteration with Shift

- Convergence rate of power iteration depends on ratio $|\lambda_2/\lambda_1|$, where λ_2 is eigenvalue having second largest modulus
- May be possible to choose shift, $\boldsymbol{A} - \sigma \boldsymbol{I}$, such that

$$\left|\frac{\lambda_2 - \sigma}{\lambda_1 - \sigma}\right| < \left|\frac{\lambda_2}{\lambda_1}\right|$$

so convergence is accelerated

- Shift must then be added to result to obtain eigenvalue of original matrix

Example: Power Iteration with Shift

- In earlier example, for instance, if we pick shift of $\sigma = 1$, (which is equal to other eigenvalue) then ratio becomes zero and method converges in one iteration
- In general, we would not be able to make such fortuitous choice, but shifts can still be extremely useful in some contexts, as we will see later

Inverse Iteration

- If smallest eigenvalue of matrix required rather than largest, can make use of fact that eigenvalues of $\boldsymbol{A}^{-1}$ are reciprocals of those of $\boldsymbol{A}$, so smallest eigenvalue of $\boldsymbol{A}$ is reciprocal of largest eigenvalue of $\boldsymbol{A}^{-1}$

- This leads to *inverse iteration* scheme

$$\begin{aligned} \boldsymbol{A}\boldsymbol{y}_k &= \boldsymbol{x}_{k-1} \quad \text{solve} \\ \boldsymbol{x}_k &= \boldsymbol{y}_k / \|\boldsymbol{y}_k\|_\infty \end{aligned}$$

  which is equivalent to power iteration applied to $\boldsymbol{A}^{-1}$

- Inverse of $\boldsymbol{A}$ not computed explicitly, but factorization of $\boldsymbol{A}$ used to solve system of linear equations at each iteration

Inverse Iteration, continued

- Inverse iteration converges to eigenvector corresponding to *smallest* eigenvalue of $\boldsymbol{A}$
- Eigenvalue obtained is dominant eigenvalue of $\boldsymbol{A}^{-1}$, and hence its reciprocal is smallest eigenvalue of $\boldsymbol{A}$ in modulus

Example: Inverse Iteration

- Applying inverse iteration to previous example to compute smallest eigenvalue yields sequence

k	$\boldsymbol{x}_k^T$		$\\|\boldsymbol{y}_k\\|_\infty$
0	0.000	1.0	
1	−0.333	1.0	0.750
2	−0.600	1.0	0.833
3	−0.778	1.0	0.900
4	−0.882	1.0	0.944
5	−0.939	1.0	0.971
6	−0.969	1.0	0.985

which is indeed converging to 1 (which is its own reciprocal in this case)

< interactive example >

Inverse Iteration with Shift

- As before, shifting strategy, working with $\boldsymbol{A} - \sigma \boldsymbol{I}$ for some scalar σ, can greatly improve convergence
- Inverse iteration is particularly useful for computing eigenvector corresponding to approximate eigenvalue, since it converges rapidly when applied to shifted matrix $\boldsymbol{A} - \lambda \boldsymbol{I}$, where λ is approximate eigenvalue
- Inverse iteration is also useful for computing eigenvalue closest to given value β, since if β is used as shift, then desired eigenvalue corresponds to smallest eigenvalue of shifted matrix

Rayleigh Quotient

- Given approximate eigenvector $\boldsymbol{x}$ for real matrix $\boldsymbol{A}$, determining best estimate for corresponding eigenvalue λ can be considered as $n \times 1$ linear least squares approximation problem

$$\boldsymbol{x}\lambda \cong \boldsymbol{A}\boldsymbol{x}$$

- From normal equation $\boldsymbol{x}^T\boldsymbol{x}\lambda = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$, least squares solution is given by

$$\lambda = \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}$$

- This quantity, known as *Rayleigh quotient*, has many useful properties

$$\sigma = \frac{x_k^T A x_k}{x_k^T x_k}$$