Unit 9 : Eigenvectors and Eigenvalues

Motivation:

To understand the essence of a linear transformation, we need to find the eigenvectors and eigenvalues, the equation that represent that is:

$$A\vec{v} = \lambda \vec{v}$$

< We need to find $\vec{v}$ s.t when apply matrix A gives the same result as scalling that vector $\vec{v}$, $\vec{v}$ stay on it own span after the transformation >

Also, another way of looking at this. If we view a linear trasformation as:

$$y = Ax$$

then a change of basis will look like:

$$XX^{-1}y = \cancel{XA}\, AXX^{-1}x$$

$$\Leftrightarrow X\hat{y} = AX\hat{x}$$

$$\Leftrightarrow \hat{y} = X^{-1}AX\hat{x}$$

It'd be nice if $(X^{-1}AX)$ is diagnal:

$$\Rightarrow \hat{y} = D\hat{x}$$

$$\Rightarrow X^{-1}AX = D$$

$$\Leftrightarrow AX = XD$$

$$\Leftrightarrow (Ax_0 \dots x_{m-1}) = (x_0 \dots x_{m-1})\begin{pmatrix} \lambda_0 & & 0 \\ & \ddots & \\ 0 & & \lambda_{m-1} \end{pmatrix}$$

$$\Leftrightarrow (Ax_0 \dots Ax_{m-1}) = (\lambda_0 x_0 \dots \lambda_{m-1} \lambda_{m-1})$$

$$\Rightarrow Ax_i = \lambda_i x_i$$

< Which is the definition of eigenvectors and eigenvalues! >

All this to say that, the diagonalization of matrix correspond with computing eigenvectors and eigenvalues

9.2 Basics

Remember from introductory course, these statements are true about matrix $A \in \mathbb{C}^{m \times m}$:

- A is nonsingular
- A has linearly independent columns
- There does "not" exist a nonzero vector s.t $Ax = 0$
- $N(A) = \{0\}$ < Null space of A is trivial >
- $\dim(N(A)) = 0$
- $\det(A) \neq 0$

Since we are trying to find λ and x s.t $\boxed{(\lambda I - A)x = 0}$, these can said about matrix $(\lambda I - A)$:

- $(\lambda I - A)$ is singular
- $(\lambda I - A)$ has linearly dependent columns
- There exist vector $x \neq 0$ st $(\lambda I - A)x = 0$

- $\mathcal{N}(\lambda I - A) = \{0, \dots\}$ <Null space of $(\lambda I - A)$ is nontrivial>
- $\dim(\mathcal{N}(\lambda I - A)) > 0$
- $\det(\lambda I - A) = 0$

Definition 9.2.1.2 Spectrum of matrix

The set of all eigenvalues of A, denoted as $\Lambda(A)$

Example: $\{\lambda_0, \lambda_1, \dots, \lambda_{m-1}\} = \Lambda(A)$

Definition 9.2.1.3 Spectral radius

The eigenvalue with largest magnitude, denoted as $\rho(A)$

$$\rho(A) = \max_{\lambda \in \Lambda(A)} |\lambda|$$

Theorem 9.2.1.4 Gershgorin Disk Theorem

Let $A \in \mathbb{C}^{m \times m}$

$$\rho_i = \rho_i(A) = \sum_{j \neq i} |\alpha_{i,j}|$$ <radius of the disk>

and

$$R_i(A) = \{x \text{ s.t } |x - \alpha_{i,i}| \leq \rho_i\}$$

In other words: $\rho_i(A)$ is sum of all the off-diagonal elements of row i

$R_i(A)$ is all points in distance $\rho_i(A)$ from $\alpha_{i,i}$

Then: $\Lambda(A) \subset \cup_i R_i(A)$

In other words: the eigenvalues lie in the union of these disks

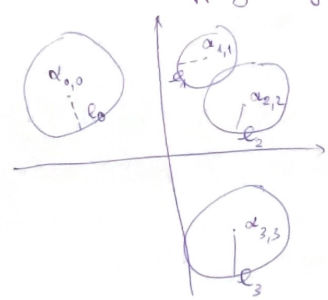
If we think of matrix A as:

$$A = \underbrace{D}_{\text{diagonal}} + \underbrace{(A - D)}_{\text{off-diagonal}}$$

and parameterize it: $A_\omega = D + \omega(A - D)$, with $0 \leq \omega \leq 1$

then:
- $\omega = 0 \Rightarrow$ only the diagonal part left, $\lambda_i = \alpha_{i,i}$
- $\omega = 1 \Rightarrow$ the conclusion from Gershgorin Disk Theorem

As ω ~~gets~~ approach 1, ρ_i gets larger and increase the chance of overlapping eigenvalues



Corallary 9.2.1.5

Let K and K^c be disjoint subsets of $\{0, \dots, m-1\}$ s.t $K \cup K^c = \{0, \dots, m-1\}$. If:

$$\left(\cup_{k \in K} R_k(A)\right) \cap \left(\cup_{j \in K^c} R_j(A)\right) = \emptyset$$

then $\cup_{k \in K} R_k(A)$ contains exactly $|K|$ eigenvalues of A.

In other words, if $\cup_{k \in K} R_k(A)$ doesn't intersect with other disks, then it contains as many eigenvalues of A as there are elements of K.

9.2.1.1

TRUE: $0 \in \Lambda(A)$ i.o.i A is singular

9.2.1.2

Let $A \in \mathbb{C}^{m\times m}$ be Hermitian

ALWAYS: All eigenvalues of A are real-vallued

9.2.1.3

Let $A \in \mathbb{C}^{m\times m}$ be Hermitian positive definite (HPD)

ALWAYS: All eigenvallues of A are positive

The converse is also true.

9.2.1.4

Let $A \in \mathbb{C}^{m\times m}$ be Hermitian, (λ, x) and (μ, y) be the eigenpairs associated with A, and $\lambda \neq \mu$.

ALWAYS: $x^H y = 0$

9.2.1.5

Let $A \in \mathbb{C}^{m\times m}$, (λ, x) and (μ, y) be eigenpairs, and $\lambda \neq \mu$.

x and y are linearly independent

9.2.1.6

Let $A \in \mathbb{C}^{m\times m}$, $k \leq m$, and (λ_i, x_i) for $1 \leq i \leq k$ be eigenpairs of this matrix. Prove that if $\lambda_i \neq \lambda_j$ when $i \neq j$ then the eigenvectors x_i are linearly independent.

In other words, given a set of distinct eigenvalues, a set of vectors created by taking one eigenvector per eigenvalue is linearly independent.

9.2.1.7 Consider matrices $A = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}$

and $\begin{pmatrix} 4 & -1 & & -1 & & \\ -1 & 4 & \ddots & & -1 & \\ & \ddots & \ddots & \ddots & & \ddots \\ -1 & & \ddots & \ddots & \ddots & \\ & -1 & & \ddots & 4 & \ddots \\ & & \ddots & & \ddots & \ddots \end{pmatrix}$

ALWAYS: All eigenvalues of these matrices are nonnegative

ALWAYS: All eigenvalues of first matrix are positive

The characteristic polynomial

Thm 9.2.2.1

$$A = \begin{pmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{pmatrix}$$

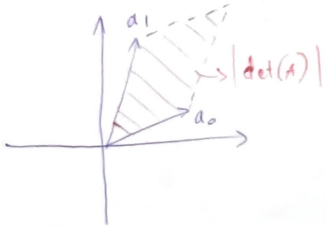
Then $$A^{-1} = \frac{1}{\underbrace{\alpha_{00}\alpha_{11} - \alpha_{10}\alpha_{01}}_{\det(A)}} \begin{pmatrix} \alpha_{11} & -\alpha_{01} \\ -\alpha_{10} & \alpha_{00} \end{pmatrix}$$

Definition 9.2.2.1 Determinant 2x2 matrix

$$A = \begin{pmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{pmatrix}$$

The determinant of A is $\det(A) = \alpha_{00}\alpha_{11} - \alpha_{10}\alpha_{01}$

The absolute value of $\det(A)$ is the area between $a_0, \dots a_{m-1}$ where $A = (a_0, \dots a_{m-1})$



Definition 9.2.2.2 Characteristic polynomial

The characteristic polynomial of $m \times m$ matrix A given by:

$$p_A(\lambda) = \det(\lambda I - A)$$

Theorem 9.2.2.3

If $A \in \mathbb{C}^{m\times m}$ then $p_A(\lambda) = \det(\lambda I - A)$ is a polynomial degree m

Theorem 9.2.2.4

Let $A \in \mathbb{C}^{m\times m}$ then $\lambda \in \Lambda(A)$ i.o.i $p_A(\lambda) = 0$

Recall that: $x^4 - 5x^3 + 18x^2 - 4x - 8$ can be refactored into: $(x-2)^3(x+1)^2$

where: 2 and -1 are the distinct roots

3 and 2 are the multiplicity of the root

Definition 9.2.2.5 Algebraic multiplicity of an eigenvalue

Let $A \in \mathbb{C}^{m\times m}$ and $p_m(\lambda)$ be its characteristic polynomial

Then the (algebraic) multiplicity of eigenvalue λ_i equals multiplicity of the roots of the polynomial

Lemma 9.2.2.6

If $A \in \mathbb{C}^{m\times m}$ then A has m eigenvalues (multiplicity counted ~~couted~~)

* Multiplicity counted means counting all the eigenvalues even if some are equals each other

Note: There is no general formula for the eigenvalues of an arbitrary $m \times m$ matrix with $m > 4$. This is because there is no general formula for the roots of a polynomial with degree > 4.

~~Collo~~ Corollary 9.2.2.7

If $A \in \mathbb{R}^{m\times m}$ is ~~real~~ real-valued then some or all of its eigenvalues maybe complex-valued. In that case, then its conjugate, $\bar{\lambda}$ is also an eigenvalue. Indeed, the complex eigenvalues of a real-valued matrix come in complex pairs

Corollary 9.2.2.8

If $A \in \mathbb{R}^{m\times m}$ is real-valued and m is odd, then at least one of the eigenvalues of A is real-valued

Note: As discussed, finding eigenvalues of $m > 4$ is a problem. Moreover, finding eigenvectors in the null space is also problematic in the presence of roundoff error.

For this reason, the strategy is to compute approximations of eigenvectors hand in hand with the eigenvalues.

More properties of eigenvalues and vectors

Hw 9.2.3.1

Let $\mathcal{E}_\lambda(A) = \{x \in \mathbb{C}^m \mid Ax = \lambda x\}$ be the set of all eigenvectors of A associated with λ plus the zero vector (which is not an eigenvector). Show that $\mathcal{E}_\lambda(A)$ is a subspace

Note: From hw 9.2.3.1, while there an infinite number of eigenvectors ~~for~~ associated with an eigenvalue, the fact that they form a subspace, means that they can be described by a finite number of vectors, which is the basis of that subspace

Hw 9.2.3.2

Let $D \in \mathbb{C}^{m\times m}$ be a diagonal matrix. Give all eigenvalues of D. For each eigenvalue, give a convenient eigenvector.

$$\Lambda(D) = \{\delta_0, \delta_1, \dots, \delta_{m-1}\}$$

And e_j is an eigenvector associated with δ_j.

Hw 9.2.3.3

Compute eigenvalues and associate eigenvectors of

$$A = \begin{pmatrix} -2 & 3 & -7 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

9.2.3.4

Let $U \in \mathbb{C}^{m \times m}$ be an upper triangular matrix. Give all eigenvalues of U. For each eigenvalue, give a convenient eigenvector

The Schur and Spectral Decompositions

Practical methods for computing eigenvectors and eigenvalues involve transforming the given matrix to a simpler matrix (diagonal or tridiagonal) known as similarity transformations

Definition 9.2.4.1 Similarity transformation

Given a nonsingular matrix Y, the transformation $Y^{-1}AY$ is called similarity transformation (applied to matrix A)

Definition 9.2.4.2 Similar matrices

Matrices A and B are considered similar if exist matrix Y such that $B = Y^{-1}AY$

9.2.4.1

Let $A, B, Y \in \mathbb{C}^{m \times m}$, Y is nonsingular, (λ, x) is an eigenpair of A, given $B = Y^{-1}AY$.

$(\lambda, Y^{-1}x)$ is an eigenpair of B

Theorem 9.2.4.3

Let $A, B, Y \in \mathbb{C}^{m \times m}$, assume Y is nonsingular, and let $B = Y^{-1}AY$. Then $\Lambda(A) = \Lambda(B)$

↳ expanding on this: If $\lambda \in \Lambda(A)$ has multiplicity of k then $\lambda \in \Lambda(B)$ has multiplicity of k.

Definition 9.2.4.4 Unitary similarity transformation

Given unitary matrix Q the transformation $Q^H A Q$ is called a unitary similarity transformation

Theorem 9.2.4.5 Schur Decomposition Theorem

Let $A \in \mathbb{C}^{m \times m}$. Then there exist a unitary matrix Q and upper triangular matrix U such that

$$A = QUQ^H$$

In other words, by applying unitary similarity to matrix A $(Q^H A Q)$ yields U, from previous hw we know that getting eigenvalues on U is easy, and $\Lambda(U) = \Lambda(A)$. Also, the eigenvectors of U can be computed and from those, the eigenvectors of A can be recovered

Hw 9.2.4.2

Let $A \in \mathbb{C}^{m\times m}$, $A = QUQ^H$ be its Schur decomposition and $X^{-1}UX = \Lambda$, where Λ is a diagonal matrix and X is nonsingular

- The diagonal elements of U equal the diagonal elements of Λ
- How are the columns of X related to the eigenvector of A?

$$A = QUQ^H = \cancel{Q X^{-1} U X}$$
$$= Q X \Lambda X^{-1} Q^H$$
$$= (QX)\, \Lambda\, (QX)^{-1}$$

$\Rightarrow$ Columns of QX equal eigenvectors of A

Spectral Decomposition Theorem

Theorem 9.2.4.6

Let $A \in \mathbb{C}^{m\times m}$ be Hermitian. Then exist a unitary matrix Q, diagonal matrix $\underline{D \in \mathbb{R}^{m\times m}}$ such that $A = QDQ^H$.

Theorem 9.2.4.7 Breaking to smaller problems

Let $A \in \mathbb{C}^{m\times m}$ be of form $A = \begin{pmatrix} A_{TL} & A_{TR} \\ 0 & A_{BR} \end{pmatrix}$, where A_{TL} and A_{BR} are square sub matrices. Then:

$$\Lambda(A) = \Lambda(A_{TL}) \cup \Lambda(A_{BR})$$

Hw 9.2.4.3

Let $A = \begin{pmatrix} A_{TL} & A_{TR} \\ 0 & A_{BR} \end{pmatrix}$ and $\begin{cases} A_{TL} = Q_{TL} U_{TL} Q_{TL}^H \\ A_{BR} = Q_{BR} U_{BR} Q_{BR}^H \end{cases}$

$$\Rightarrow A = \underbrace{\begin{pmatrix} Q_{TL} & 0 \\ 0 & Q_{BR} \end{pmatrix}}_{Q} \underbrace{\begin{pmatrix} U_{TL} & Q_{TL}^H A_{TR} Q_{BR} \\ 0 & U_{BR} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} Q_{TL} & 0 \\ 0 & Q_{BR} \end{pmatrix}^H}_{Q^H}$$

Hw 9.2.4.4

$$A = \begin{pmatrix} Q_0 & 0 & \cdots \\ 0 & Q_1 & \\ \vdots & & \ddots \end{pmatrix} \begin{pmatrix} U_0 & Q_0^H A_{0,1} Q_1 & \cdots & Q_0^H A_{0,N-1} Q_{N-1} \\ 0 & U_1 & & Q_0^H A_{1,N-1} Q_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & U_{N-1} \end{pmatrix} \begin{pmatrix} Q_0 & & 0 \\ & Q_1 & \\ & & \ddots \\ 0 & & & Q_{N-1} \end{pmatrix}^H$$

left $\xleftarrow{\text{read}}$ right

Why diagonalize is important?

It make thing easier to deal with. For example:

Consider: $\omega = A v$ (1)

$$\Leftrightarrow \quad \forall \quad X\underbrace{(X^{-1}\omega)}_{\tilde{\omega}} = AX\underbrace{(X^{-1}v)}_{\tilde{v}} \qquad (2)$$

$\langle X^{-1}\omega$ is the vector of coefficients when ω change to basis that consists of columns of X, similar can be said for $X^{-1}v$ ~~and~~ $\rangle$

Now with (1) and (2), we have:

$$X^{-1}\omega = D(X^{-1}v)$$

Remark 9.2.5.2

If we choose the right basis, then the transformation $\omega = Av$ becomes $\tilde{\omega} = D\tilde{v}$, which is much simplier

Theorem 9.2.5.3 Diagonizable

Matrix $A \in \mathbb{C}^{m \times m}$ is diagonizable i.o.i it has m linearly independent eigenvectors

Diagonalizing a matrix

A matrix A is diagonalizable if $X^{-1}AX = D$

If you look close enough, this is actually a discussion about eigenvectors and eigenvalues.

$$X^{-1}AX = D$$

$$\Leftrightarrow \quad AX = XD$$

$$\Leftrightarrow \quad A(x_0 | \dots | x_{m-1}) = (x_0 | \dots | x_{m-1})\begin{pmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_{m-1} \end{pmatrix}$$

$$\Leftrightarrow \quad (Ax_0 | \dots | Ax_{m-1}) = (\lambda_0 x_0 | \dots | \lambda_{m-1} x_{m-1})$$

$$\Leftrightarrow \quad \boxed{A \underbrace{x_i}_{\substack{\text{eigenvector} \\ \text{(columns of } X)}} = \underbrace{\lambda_i}_{\substack{\text{eigenvalue} \\ \text{(diagonal elements of } D)}} x_i}$$

Definition 9.2.5.1 Diagonalizable matrix

Matrix $A \in \mathbb{C}^{m \times m}$ is diagonalizable i.o.i exist a nonsingular matrix X ~~that~~ and diagonal matrix D s.t

$$X^{-1}AX = D$$

Some classes of matrices that are diagonalizable:

- Diagonal matrices

  Choose $X = I$ and $A = D$ yields $X^{-1}AX = D$

- Hermitian matrices

  Choose $X = Q$ yields $X^{-1}AX = D$

  $\Rightarrow A = QDQ^H$ <Spectral decomposition theorem>

- Triangular matrices with distinct diagonal elements

  thr. 9.2.3.4

Thr 9.2.5.2

Let $A \in \mathbb{C}^{m\times m}$ have distinct eigenvalues.

ALWAYS A is diagonalizable

One scenario you might find diagonalizing helpful is when you want to apply A to vector x many times:

$$A \dots A x$$
$$= A^k x$$

With diagonalizing A, we can prove that:

$$A^k x = (X D X^{-1})^k x = X D^k X^{-1} x$$

which is much easier to compute!

Un-diagonalizable matrices: Jordan Canonical Form

Consider matrix

$$\underset{(k\times k)}{J_k(\mu)} = \begin{pmatrix} \mu & 1 & & 0 \\ & \mu & \ddots & \\ & & \ddots & 1 \\ 0 & & & \mu \end{pmatrix}$$

We can ~~see~~ prove that the matrix has a eigenvalue (μ) of multiplicity k, then it does not necessarily have k linearly independent eigenvectors, meaning there are matrices that do not have a full set of eigenvectors, ~~and~~ we conclude that these matrices are not diagonalizable

Definition 9.2.6.1 Defective matrix

Matrix $A \in \mathbb{C}^{m\times m}$ that does not have m linearly independent eigenvectors is defective

Corollary 9.2.6.2

Matrix $A \in \mathbb{C}^{m\times m}$ is diagonalizable i.o.i it is not defective

Definition 9.2.6.3 Geometric multiplicity

Let $\lambda \in \Lambda(A)$. Then the geometric multiplicity of λ is defined to be the dimensions of $\mathcal{E}_\lambda(A)$ defined by:

$$\mathcal{E}_\lambda(A) = \{ x \in \mathbb{C}^m \mid Ax = \lambda x \}$$

In other words, the geometric multiplicity of λ equals the number of linearly independent eigenvectors that are associated with λ.

thm 9.2.6.2

Let $A = \begin{pmatrix} A_{00} & 0 \\ 0 & A_{11} \end{pmatrix}$ where A_{00}, A_{11} are square

- If (λ, x) is an eigenpair of A_{00} then $\left(\lambda, \begin{pmatrix} x \\ 0 \end{pmatrix}\right)$ is an eigenpair of A
- If (μ, y) is an eigenpair of A_{11} then $\left(\mu, \begin{pmatrix} 0 \\ y \end{pmatrix}\right)$ is an eigenpair of A
- If $\left(\lambda, \begin{pmatrix} x \\ y \end{pmatrix}\right)$ is an eigenpair of A, then:
  $\begin{cases} (\lambda, x) \text{ is an eigenpair of } A_{00} \\ (\lambda, y) \text{ is an eigenpair of } A_{11} \end{cases}$
- $\Lambda(A) = \Lambda(A_{00}) \cup \Lambda(A_{11})$

Geometric multiplicity vs Algebraic multiplicity

Let λ be an eigenvalue of A

- Algebraic multiplicity of λ: number of times λ appears as root of $p(A)$

  For example: $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $\quad p(A) = (\lambda-2)^2(\lambda-1)$

  λ of value 2 has algebraic multiplicity of 2.

- Geometric multiplicity of λ: is the ~~dim (N(~~ $\dim(\underbrace{\mathcal{E}_\lambda(A)}_{\text{eigenspace for } \lambda})$

For example: ~~Given~~ Sine $\mathcal{E}_\lambda(A)$ is $N(\lambda I - A)$

Looking at λ with value 2, then:

$2I - A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ has $\text{rank}(A) = 2$

By ~~null~~ rank-nullity theorem:

$\Rightarrow \dim(N(2I - A)) = 1$

That means $\lambda = 2$ has geometric multiplicity 1.

Note: Eigenvectors and Eigenvalues

Let A be an $m \times m$ matrix

- An eigenvector ~~of A~~ of A is a nonzero vector v in $\mathbb{R}^m$ such that $Av = \lambda v$, ~~by definition, are~~

An eigenvalue of A is a scalar λ s.t the equation $Av = \lambda v$ has a nontrivial solution

$\mathcal{E}_\lambda(A) \Leftrightarrow N(\lambda I - A)$

made of eigenvectors associated with λ

Practical algorithms to find eigenvalue and eigenvector

The Power Method

We will assume that: $\begin{cases} A \text{ is diagonalizable, so } A = X \Lambda X^{-1} \\ |\lambda_0| > |\lambda_1| \geq \ldots \geq |\lambda_{m-1}|, \\ \quad \text{so } \lambda_0 \text{ is the eigenvalue with maximal absolute value} \end{cases}$

Algorithm:

> Pick $v^{(0)}$
> for $k = 0, \ldots$
> $\quad v^{(k+1)} = A v^{(k)}$
> end for

Explain:

We are trying to find $v^{(k)}$ that is eigenvector, meaning it is also a column in X inside ~~XAX~~ $X \Lambda X^{-1}$.

We can express the initial guess as:

$$v^{(0)} = X y$$
$$= \psi_0 x_0 + \psi_1 x_1 + \ldots + \psi_{m-1} x_{m-1}$$

Apply A: $$A v^{(0)} = \psi_0 A x_0 + \psi_1 A x_1 + \ldots + \psi_{m-1} A x_{m-1}$$
$$= \psi_0 \lambda_0 x_0 + \psi_1 \lambda_1 x_1 + \ldots + \psi_{m-1} \lambda_{m-1} x_{m-1}$$
$$\langle Ax = \lambda x \rangle$$

Apply A k times: $$A v^{(k-1)} = \psi_0 \lambda_0^k x_0 + \ldots + \psi_{m-1} \lambda_{m-1}^k x_{m-1}$$

Under our assumption, $\psi_0 \lambda_0^k x_0$ will dominate, hence found (kinda) the eigenvector ~~x_0~~ correspond to λ_0

The problem is: $\begin{cases} \text{If } |\lambda_0| > 1, \text{ "k times" it will make } \cancel{x_0} \text{ eigenvector infinitely long} \\ \text{If } |\lambda_0| < 1, \text{ "k times" it will make eigenvector infinitely short} \end{cases}$

Everything good if $|\lambda_0| = 1$

"We could" at each iteration divide by $|\lambda_0|$ to solve this issue; but that is cheating since we don't know λ_0 yet.

So instead, we recognize that we only interested in the "direction" of the eigenvector, therefore we "normalize" in each iteration

Practical Algorithm

> Pick $v^{(0)}$ of unit length
> for $k = 0, \ldots$
> $\quad v^{(k+1)} = A v^{(k)}$
> $\quad v^{(k+1)} = \dfrac{v^{(k+1)}}{\| v^{(k+1)} \|}$
> endfor

The Rayleigh quotient (find the λ after knowing x)

eigenvalue ↑ (λ), eigenvector ↑ (x)

A question is how to extract ~~appropriate~~ approx λ_0 given approx x_0

Definition 9.3.1.1 Rayleigh quotient

If $A \in \mathbb{C}^{m \times m}$ and $x \neq 0 \in \mathbb{C}^m$ then

$$\frac{x^H A x}{x^H x}$$

is the Rayleigh quotient

Thm 9.3.1.1

Let x be eigenvector of A

ALWAYS: $\lambda = \frac{x^H A x}{x^H x}$ is the associated eigenvalue of A

The convergence power method

Definition 9.3.2.1 Converge of sequence of scalars

Let $\alpha_0, \alpha_1, \alpha_2, \ldots \in \mathbb{C}^m$ be an infinite sequence of ~~vectors~~ scarlars. Then α_k converge to α if:

$$\lim_{k \to \infty} |\alpha_k - \alpha| = 0$$

Definition 9.3.2.2 Converge of sequence of vectors

Let $x_0, x_1, x_2, \ldots \in \mathbb{C}^m$ be an infinite sequence of vectors.

Then x_k converge to x if:

$$\lim_{k \to \infty} \| x_k - x \| = 0$$

Note: If vectors converge in one norm, then it will converge in other norms

Definition 9.3.2.3 Rate of convergence

Let $\alpha_0, \alpha_1, \alpha_2, \ldots \in \mathbb{R}$ be an infinite sequence of scalars that converges to α. Then:

- α_k converge linearly to α if for a sufficiently large k:

$$|\alpha_{k+1} - \alpha| \leq C |\alpha_k - \alpha|$$

for some constant $C < 1$. In other words:

$$\lim_{k \to \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|} = C < 1$$

- α_k converge superlinearly to α if for a sufficiently large k:

$$|\alpha_{k+1} - \alpha| \leq C_k |\alpha_k - \alpha|$$

with $C_k \to 0$. In other words:

$$\lim_{k \to \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|} = 0$$

- α_k converge quadratically to α if for a sufficiently large k

$$|\alpha_{k+1} - \alpha| \leq C |\alpha_k - \alpha|^2$$

for some constant C. In other words:

$$\lim_{k \to \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^2} = C$$

- α_k is said to converge cubically to α if for large k:

$$|\alpha_{k+1} - \alpha| \leq C |\alpha_k - \alpha|^3$$

for some constant C. In other words:

$$\lim_{k \to \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|} = C$$

Linear coverge can be slow, especially if $C = 0.99$.

For n iterations:

$$|\alpha_{k+n} - \alpha| \leq C^n |\alpha_k - \alpha|$$

Quadratic covergence is fast. For n iterations:

$$|\alpha_{k+1} - \alpha| \leq C^{2n-1} |\alpha_k - \alpha|^{2n}$$

Analyzing number of correct digits, we can prove that:

$$\underbrace{-\log_{10} |\alpha_{k+1} - \alpha|}_{\text{number of correct digits in } \alpha_{k+1}} \geq 2\underbrace{\left(-\log_{10} |\alpha_k - \alpha|\right)}_{\text{number of correct digits in } \alpha_k}$$

HW 9.3.2.1

Let $X \in \mathbb{C}^{m \times m}$ be nonsingular. Define $\|.\|_{X^{-1}} : \mathbb{C}^m \to \mathbb{R}$ by $\|y\|_{X^{-1}} = \|X^{-1}y\|$ for some norm $\|.\| : \mathbb{C}^m \to \mathbb{R}$.

ALWAYS: $\|.\|_{X^{-1}}$ is a norm

Understand how convergence works:

We can express: $v^{(0)} = Xy$

After k iterations: $v^{(k)} = \psi_0 x_0$

So the convergence is given by: $\| v^{(k)} - \psi_0 x_0 \|$

And since $\| v^{(k)} - \psi_0 x_0 \| = \| \frac{1}{\lambda_0^k} A^k v^{(0)} - \psi_0 x_0 \|$

$$= \| \frac{1}{\lambda_0^k} A^k X y - \psi_0 x_0 \|$$

$$= \cancel{\| \frac{1}{\lambda_0^k} X A^k y}$$

$$= \| \frac{1}{\lambda_0^k} X \Lambda^k y - \psi_0 x_0 \|$$

$$= \| \psi_1 \left(\frac{\lambda_1}{\lambda_0}\right)^k x_1 + \dots + \psi_{m-1} \left(\frac{\lambda_{m-1}}{\lambda_0}\right)^k x_{m-1} \|$$

$$= \| (x_0 | x_1 | \dots | x_{m-1}) \begin{pmatrix} 0 & & & \\ & \left(\frac{\lambda_1}{\lambda_0}\right)^k & & \\ & & \ddots & \\ & & & \left(\frac{\lambda_{m-1}}{\lambda_0}\right)^k \end{pmatrix} \begin{pmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \psi_{m-1} \end{pmatrix} \|$$

$$\leq \|X\| \left(\frac{|\lambda_1|}{|\lambda_0|}\right)^k \|y\|$$

Doing similar analysis on $v^{(k+1)}$ we can prove that:

$$\| v^{(k+1)} - \psi_0 x_0 \| \leq \|X\| \left(\frac{|\lambda_1|}{|\lambda_0|}\right)^{k+1} \|y\|$$

$$= \left(\frac{|\lambda_1|}{|\lambda_0|}\right) \left(\|X\| \left(\frac{|\lambda_1|}{|\lambda_0|}\right)^k \|y\| \right)$$

$$= \frac{|\lambda_1|}{|\lambda_0|} \| v^{(k)} - \psi_0 x_0 \|$$

~~There~~ Since the difference between $v^{(k)}$ and $\psi_0 x_0$ is reduced by at least a constant factor $\frac{|\lambda_1|}{|\lambda_0|}$, the converge is linear.

The Inverse Power Method

Thm 9.3.3.1

Let $A \in \mathbb{C}^{m \times m}$ be nonsingular, and (λ, x) is an eigenpair of A, then $(\frac{1}{\lambda}, x)$ is an eigenpair of A^{-1}.

From this insight (~~the~~ an eigenvector associated with the smallest eigenvalue of A is an eigenvector associated with the largest eigenvalue of A^{-1}

Algorithm

for $k = 0, \ldots$

$v^{(k+1)} = A^{-1} v^{(k)}$

$v^{(k+1)} = \lambda_{m-1} v^{(k+1)}$

endfor

From the convergence analysis, we can prove that;

$$\| v^{(k+1)} - \psi_{m-1} x_{m-1} \|_{X^{-1}} \leq \left| \frac{\lambda_{m-1}}{\lambda_{m-2}} \right| \| v^{(k)} - \psi_{m-1} x_{m-1} \|_{X^{-1}}$$

Practical Algorithm

Pick $v^{(0)}$ of unit length

for $k = 0, \ldots$

$v^{(k+1)} = A^{-1} v^{(k)}$

$v^{(k+1)} = \frac{v^{(k+1)}}{\| v^{(k+1)} \|}$

end for

We probably should factor $PA = LU$ once and solve $L(Uv^{(k+1)}) = Pv^{(k)}$

The Rayleigh Quotient Iteration

Thm 9.3.4.1

Let $A \in \mathbb{C}^{m \times m}$, $\rho \in \mathbb{C}$ and (λ, x) an eigenpair of A.

~~$(\frac{1}{(\lambda - \rho)}, x)$~~ $((\lambda - \rho), x)$ is an eigenpair of the shifted matrix $A - \rho I$

From previous hw, if we choose appropriate ρ (closer to λ_{m-1}), then convergence happens faster.

Lemma 9.3.4.1

Let $A \in \mathbb{C}^{m \times m}$, $A = X \Lambda X^{-1}$ and $\rho \in \mathbb{C}$ Then

$A - \rho I = X(\Lambda - \rho I)X^{-1}$

Hw 9.3.4.2

Prove above lemma 9.3.4.1

Inverse algorithm with ρ shift

Pick $v^{(0)}$ of unit length

for $k = 0, \ldots$

$v^{(k+1)} = (A - \rho I)^{-1} v^{(k)}$

$v^{(k+1)} = (\lambda_{m-1} - \rho) v^{(k+1)}$

end for

should solve $(A - \rho I) v^{(k+1)} = v^{(k)}$ instead of computing $(A - \rho I)^{-1}$

If we index eigenvalues as:

$$|\lambda_{m-1} - \rho| < |\lambda_{m-2} - \rho| \leq \dots \leq |\lambda_0 - \rho|$$

then: $\| v^{(k+1)} - \varphi_{m-1} x_{m-1} \|_{X^{-1}} \leq \left| \frac{\lambda_{m-1} - \rho}{\lambda_{m-2} - \rho} \right| \| v^{(k)} - \varphi_{m-1} x_{m-1} \|_{X^{-1}}$

Practical inverse with ρ shift algorithm

Pick $v^{(0)}$ of unit length
for $k = 0, \dots$
$v^{(k+1)} = (A - \rho I)^{-1} v^{(k)}$
$v^{(k+1)} = \frac{v^{(k+1)}}{\| v^{(k+1)} \|}$
end for

How to choose a good value for ρ ?

Rayleigh quotient stated: $\lambda = \frac{x^T A x}{x^T x}$

So we can come up with this algorithm

Practical Inverse method with ρ and Rayleigh Quotient

$v^{(0)} = \frac{v^{(0)}}{\| v^{(0)} \|_2}$
for $k = 0, \dots$
$\rho_k = v^{(k)H} A v^{(k)}$
$v^{(k+1)} = (A - \rho I)^{-1} v^{(k)}$
$v^{(k+1)} = \frac{v^{(k+1)}}{\| v^{(k+1)} \|}$
end for

Summarize on Power Method:

- The Power Method finds the eigenvector associated with the largest eigenvalue, cost $2m^2$ flops per iteration, convergence is linear
- The Inverse Power Method finds the eigenvector associated with the smallest eigenvalue, cost an initial investment of m^3 flops by using partial LU factorization, and then $2m^2$ flops per iteration, convergence is linear.
- The Rayleigh Quotient Iteration finds an eigenvector, but with which eigenvalue is not clear from the start. It cost m^3 flops per iteration if solve via partial LU factorization.
  The convergence is quadratic if A is not Hermitian, cubic if it is Hermitian.

* The cost of these methods is greatly reduced if the matrix is sparse, which case each iteration may costs as little as $O(m)$ flops.