Eigenvalue Decomposition

Theory of Eigenvalue Decomposition

Eigenvalue and Eigenvector

Definition: Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. A nonzero vector x is an eigenvector of A with $\lambda \in \mathbb{C}$ being the corresponding eigenvalue if

$$Ax = \lambda x$$

Remarks:

- Even A is a real matrix, its eigenvalue and eigenvector can be complex.
- The set of eigenvalues of A is called the spectrum of A. The spectral radius $\rho(A)$ is the maximum value $|\lambda|$ over all eigenvalues of A.
- If (λ, x) is an eigenpair of A, then
  (λ^2, x) is an eigenpair of A^2,
  $(\lambda - \sigma, x)$ is an eigenpair of $A - \sigma I$
  $(\frac{1}{\lambda - \sigma}, x)$ is an eigenpair of $(A - \sigma I)^{-1}$.

Proof. Since (λ, x) is an eigenpair of A, $\quad Ax = \lambda x$.

Multiply both sides by A from the left,

$$A \cdot Ax = \lambda Ax \Rightarrow A^2 x = \lambda Ax = \lambda \cdot \lambda x = \lambda^2 x.$$

Also, $Ax - \sigma x = \lambda x - \sigma x \Rightarrow (A - \sigma I)x = (\lambda - \sigma)x$

$$\Rightarrow x = (\lambda - \sigma)(A - \sigma I)^{-1} x \Rightarrow \frac{1}{\lambda - \sigma} x = (A - \sigma I)^{-1} x \quad \blacksquare$$

Definition: Two matrices A and B are similar with each other if there exists a nonsingular matrix T such that

$$B = TAT^{-1}.$$

Theorem: If A and B are similar, then A and B have the same eigenvalues.

proof. Since A, B are similar, $B = TAT^{-1}$, which implies $A = T^{-1}BT$

If (λ, x) is an eigenpair of A, then $Ax = \lambda x$, so that

$$T^{-1}BTx = \lambda x \Rightarrow B(Tx) = \lambda(Tx).$$

Thus, (λ, Tx) is an eigenpair of B.

i.e., any eigenvalue of A is an eigenvalue of B.

The reverse is shown similarly ☒

Eigenvalue Decomposition:

An eigenvalue decomposition of a square matrix $A \in \mathbb{R}^{n\times n}$ is a factorization

$$A = X \Lambda X^{-1},$$

where $X \in \mathbb{C}^{n\times n}$ is non-singular and $\Lambda \in \mathbb{C}^{n\times n}$ is diagonal.

- If $A \in \mathbb{R}^{n\times n}$ admits an eigenvalue decomposition, then

$$AX = X\Lambda$$

If we rewrite $X = [x_1 \ x_2 \ \cdots \ x_n]$ with $x_i \in \mathbb{C}^n$ the i-th column of X,

and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2 \cdots, \lambda_n) \equiv \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$ with $\lambda_i \in \mathbb{C}$ being the i-th

diagonal of Λ, then

$$A[x_1 \ x_2 \ \cdots \ x_n] = [x_1 \ x_2 \ \cdots \ x_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

$$\Rightarrow \quad [Ax_1 \ Ax_2 \ \cdots \ Ax_n] = [\lambda_1 x_1 \ \lambda_2 x_2 \ \cdots \ \lambda_n x_n]$$

$$\Rightarrow \quad Ax_i = \lambda_i x_i, \quad i = 1, 2, \cdots, n.$$

In other words, (λ_i, x_i), $i = 1, \cdots, n$ are eigenpairs of A.

- Since X is nonsingular, x_i, $i = 1, \cdots, n$ are linearly independent. So, x_i, $i = 1, \cdots, n$, are n independent eigenvectors, which spans $\mathbb{C}^n$.

- Eigenvalue decomposition implies $X^{-1}AX = \Lambda$, so that we also says A is diagonalizable.
- Eigenvalue decomposition does not always exist, as a square matrix $A \in \mathbb{R}^{n\times n}$ does not always have n independent eigenvectors.
- Though $A \in \mathbb{R}^{n\times n}$ is real, the eigenvalue decomposition may be complex.

Characteristic Polynomial

The characteristic polynomial of $A \in \mathbb{R}^{n\times n}$, denoted P_A is a degree n polynomial defined by

$$P_A(z) = \det(zI - A) \quad , \quad \text{where } z \in \mathbb{C}$$

Let (λ, x) be an eigenpair of A. Then $Ax = \lambda x$, which is equivalent to

$$(\lambda I - A)x = 0.$$

Since x is non-zero, $\lambda I - A$ has a non-zero solution. Therefore, $\lambda I - A$ is singular. That is, $\det(\lambda I - A) = P_A(\lambda) = 0$. Thus, λ is an eigenvalue of A if and only if

$$P_A(\lambda) = 0,$$

and the corresponding eigenvector x are non-zero solutions of

$$(\lambda I - A)x = 0.$$

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

The characteristic polynomial is

$$P_A(z) = \det\left(zI - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \det\left(\begin{bmatrix} z & -1 \\ 0 & z \end{bmatrix}\right) = z^2$$

Therefore, $P_A(\lambda) = \lambda^2 = 0 \Rightarrow \lambda_1 = \lambda_2 = 0$ are eigenvalues of A.

For eigenvectors, solve $(0I - A)x = 0$, i.e.,

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

$\Rightarrow x = \begin{pmatrix} a \\ 0 \end{pmatrix} \ \forall a \in \mathbb{C}$. (Only one independent eigenvector)

So, A is not diagonalizable (i.e., no eigenvalue decomposition).

Example 2: $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

The characteristic polynomial is

$$p_A(z) = \det(zI - A) = \det\left(\begin{bmatrix} z & 1 \\ -1 & z \end{bmatrix}\right) = z^2+1$$

Therefore, $p_A(\lambda) = \lambda^2+1 = 0 \Rightarrow \lambda_1 = i,\ \lambda_2 = -i$ are eigenvalues.

For eigenvector of $\lambda_1 = i$; solve

$$(iI - A)x = 0, \text{ i.e., } \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} x = 0 \Rightarrow x = \alpha \begin{bmatrix} i \\ 1 \end{bmatrix} \quad \forall \alpha \in \mathbb{C}$$

Therefore, a corresponding eigenvector is $x_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}$.

For eigenvector of $\lambda_2 = -i$, solve

$$(-iI - A)x = 0, \text{ i.e., } \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} x = 0 \Rightarrow x = \beta \begin{bmatrix} i \\ -1 \end{bmatrix} \quad \forall \beta \in \mathbb{C}.$$

Therefore, a corresponding eigenvector is $x_2 = \begin{bmatrix} i \\ -1 \end{bmatrix}$.

Therefore, define $X = [x_1, x_2] = \begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix}$, $\Lambda = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} = \begin{bmatrix} i & \\ & -i \end{bmatrix}$

then $X^{-1} = \begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{1}{2}i & \frac{1}{2} \\ -\frac{1}{2}i & -\frac{1}{2} \end{bmatrix}$

Therefore $A = X \Lambda X^{-1}$

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix} \begin{bmatrix} i & \\ & -i \end{bmatrix} \begin{bmatrix} -\frac{1}{2}i & \frac{1}{2} \\ -\frac{1}{2}i & -\frac{1}{2} \end{bmatrix}$$

is an eigenvalue decomposition.

This shows that a real matrix may have a complex eigenvalue decomposition.

We will not use the method solve $p_A(\lambda) = 0$ and $(\lambda I - A)x = 0$ to compute the eigenvalue decomposition, because polynomial root-finding is not numerically stable in general.

Special Case: Symmetric matrix and SPD matrix.

Assume $A \in \mathbb{R}^{n\times n}$ is symmetric. Then

① The eigenvalues of A are real.

Proof. Let (λ, x) be an eigenpair of A. Then, $Ax = \lambda x$.

Multiply both sides by $x^* \equiv \overline{x^T}$ (conjugate transpose) from the left,

$$x^* A x = \lambda x^* x \Rightarrow \lambda = \frac{x^* A x}{x^* x}.$$

$x^* A x$ is real, because $\overline{x^* A x} = \overline{(x^* A x)^T} = \overline{x^T A^T \overline{x}} = \overline{x^T}\, \overline{A^T}\, \overline{\overline{x}}$

$$= x^* A x$$

$x^* x$ is also real, because $\overline{x^* x} = \overline{(x^* x)^T} = \overline{x^T \overline{x}} = x^* x$

So, $\lambda = \frac{x^* A x}{x^* x}$ is real. ☒

② The eigenvectors corresponding to distinct eigenvalues of A are orthogonal to each other.

proof. Let (λ_1, x_1) and (λ_2, x_2) are eigenpairs of A with $\lambda_1 \neq \lambda_2$.

Then $A x_1 = \lambda_1 x_1$ and $A x_2 = \lambda_2 x_2$

Let us consider $x_2^T A x_1 \in \mathbb{R}$.

On the one hand, $x_2^T A x_1 = x_2^T (\lambda_1 x_1) = \lambda_1 x_2^T x_1 = \lambda_1 (x_1^T x_2)$.

On the other hand, $x_2^T A x_1 = (x_2^T A x_1)^T = x_1^T A^T x_2 = x_1^T A x_2$

$$= x_1^T (\lambda_2 x_2) = \lambda_2 (x_1^T x_2).$$

Therefore $\lambda_1 (x_1^T x_2) = \lambda_2 (x_1^T x_2)$. Since $\lambda_1 \neq \lambda_2$, we have

$$x_1^T x_2 = 0, \quad \text{i.e.,} \quad x_1 \perp x_2. \quad ☒.$$

③ A is always diagonalizable, and the eigenvalue decomposition has a special form (due to the orthogonality of eigenvectors)

$$A = Q \Lambda Q^T,$$

where $Q \in \mathbb{R}^{n \times n}$ is orthogonal and $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal. (Need some efforts to show this.)

④ If A is SPD, then all eigenvalues are positive.

If A is SPSD (symmetric positive semi-definite), then all eigenvalues are non-negative.

proof. Let (λ, x) be an eigenpair of A. Then $A x = \lambda x$.

So, $\quad x^TAx = \lambda x^Tx \Rightarrow \lambda = \dfrac{x^TAx}{x^Tx}$.

Since $x^Tx = \|x\|_2^2 > 0 \ (x \neq 0)$,

If A is SPD, $x^TAx > 0 \Rightarrow \lambda = \dfrac{x^TAx}{x^Tx} > 0$.

If A is SPSD, $x^TAx \geq 0 \Rightarrow \lambda = \dfrac{x^TAx}{x^Tx} \geq 0$. ⊠

Computation of Eigenvalue Decomposition

Computing a single eigenvalue / eigenvector

Problem Setup

For simplicity, we restrict our attention to real symmetric matrices whose eigenvalues/eigenvectors are all real. Many of the ideas can be similarly extended to non-symmetric matrices.

Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Let λ_i, $i = 1, 2, \cdots, n$ be its eigenvalues, which are sorted in magnitude, i.e.,

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

The corresponding eigenvectors are denoted by q_i, $i = 1, 2, \cdots, n$, which form an orthogonal matrix, i.e, $Q = [q_1\ q_2\ \cdots\ q_n] \in \mathbb{R}^{n\times n}$ with $Q^TQ = QQ^T = I$.

Rayleigh Quotient

Let $A \in \mathbb{R}^{n\times n}$ be symmetric. For a vector $x \in \mathbb{R}^n$, the Rayleigh quotient is defined by

$$r(x) = \frac{x^TAx}{x^Tx}.$$

- Typically, Rayleigh quotient is used to compute an estimate to an eigenvalue, given an estimate to an eigenvector. In particular, if x is an eigenvector of A with the corresponding eigenvalue λ, then it is easy to see that $r(x) = \frac{x^TAx}{x^Tx} = \frac{\lambda x^Tx}{x^Tx} = \lambda$.

- Actually, the eigenvalues of A are critical points of the optimizations of Rayleigh quotient, i.e.,

$$\max_{x\neq 0} r(x) \quad \text{and} \quad \min_{x\neq 0} r(x)$$

It can be proven that

$$\min_i \lambda_i = \min_{x\neq 0} r(x) \quad \text{and} \quad \max_i \lambda_i = \max_{x\neq 0} r(x)$$

and any eigenvalues that are not max or min are saddle points of $r(x)$.

Power Iteration

Purpose: Find λ_1 and its associated eigenvector x_1 with $\|x_1\|_2 = 1$.

Algorithm:

$$\boxed{\begin{array}{l} \text{Choose } y^{(0)} \in \mathbb{R}^n \text{ s.t. } \|y^{(0)}\|_2 = 1 \\ \text{for } k = 1, 2, \cdots \\ \quad z^{(k)} = A y^{(k-1)} \\ \quad y^{(k)} = z^{(k)} / \|z^{(k)}\|_2 \\ \quad \mu^{(k)} = (y^{(k)})^T A y^{(k)} \\ \text{end} \end{array}}$$

$\mu^{(k)}$ is an estimation of eigenvalue, because $\mu^{(k)} = r(y^{(k)})$ as $\|y^{(k)}\|_2 = 1$.

Illustration by Examples:

- Assume $(2, x_1), (1, x_2)$ are two eigenpairs of $A \in \mathbb{R}^{2\times 2}$. (So $x_1 \perp x_2$)

Assume $y^{(0)} = \frac{1}{\sqrt{2}}(x_1 + x_2)$

$k=1$: $z^{(1)} = Ay^{(0)} = A\left(\frac{1}{\sqrt{2}}(x_1+x_2)\right) = \frac{1}{\sqrt{2}}(Ax_1 + Ax_2) = \frac{1}{\sqrt{2}}(2x_1 + x_2)$

$\|z^{(1)}\|_2 = \frac{1}{\sqrt{2}}\|2x_1 + x_2\|_2 = \frac{1}{\sqrt{2}}\sqrt{2^2+1^2} = \sqrt{\frac{5}{2}}$

$y^{(1)} = z^{(1)}/\|z^{(1)}\|_2 = \frac{1}{\sqrt{5}}(2x_1 + x_2)$

$k=1$: $z^{(2)} = Ay^{(1)} = A\left(\frac{1}{\sqrt{5}}(2x_1+x_2)\right) = \frac{1}{\sqrt{5}}(2Ax_1 + Ax_2) = \frac{1}{\sqrt{5}}(2^2x_1 + x_2)$

$$\|z^{(2)}\|_2 = \frac{1}{\sqrt{5}}\|2^2x_1 + x_2\|_2 = \frac{1}{\sqrt{5}}\cdot\sqrt{(2^2)^2+1} = \sqrt{\frac{17}{5}}$$

$$y^{(2)} = z^{(2)}/\|z^{(2)}\|_2 = \frac{1}{\sqrt{17}}(2^2x_1 + x_2)$$

$$\vdots$$

$$k+1: \quad z^{(k+1)} = A\,y^{(k)} = A\left(\frac{1}{\sqrt{2^{2k}+1}}(2^k x_1 + x_2)\right)$$

$$= \frac{1}{\sqrt{2^{2k}+1}}(2^k Ax_1 + Ax_2) = \frac{1}{\sqrt{2^{2k}+1}}(2^{k+1}x_1 + x_2)$$

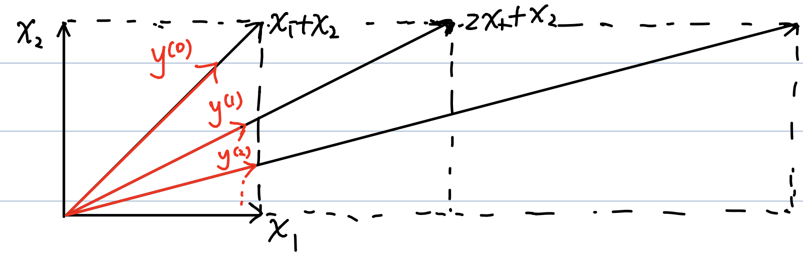
$$y^{(k+1)} = z^{(k+1)}/\|z^{(k+1)}\|_2 = (2^{k+1}x_1 + x_2)/\sqrt{2^{2k+2}+1}$$

Therefore, when k becomes larger and larger,

x_1 becomes more and more dominant in $y^{(k)}$, i.e.,

$\|y^{(k)} - x_1\|_2 \to 0$, as $k \to +\infty$

of



- Power iteration may not be convergent.

Assume $(1, x_1), (-1, x_2)$ are two eigen pairs of $A \in \mathbb{R}^{2\times 2}$

Assume $y^{(0)} = \frac{1}{\sqrt{2}}(x_1 + x_2)$.

$$k=1: \quad z^{(1)} = A\,y^{(0)} = A\left(\frac{1}{\sqrt{2}}(x_1+x_2)\right) = \frac{1}{\sqrt{2}}(Ax_1 + Ax_2) = \frac{1}{\sqrt{2}}(x_1 - x_2)$$

$$\|z^{(1)}\|_2 = \frac{1}{\sqrt{2}}\|x_1 - x_2\|_2 = 1$$

$$y^{(1)} = z^{(1)}/\|z^{(1)}\|_2 = \frac{1}{\sqrt{2}}(x_1 - x_2)$$

$$k=1: \quad z^{(2)} = Ay^{(1)} = A\left(\frac{1}{\sqrt{2}}(x_1 - x_2)\right) = \frac{1}{\sqrt{2}}(Ax_1 - Ax_2) = \frac{1}{\sqrt{2}}(x_1 + x_2)$$

$$\|z^{(2)}\|_2 = \frac{1}{\sqrt{2}}\|x_1 + x_2\|_2 = 1$$

$$y^{(2)} = z^{(2)}/\|z^{(2)}\|_2 = \frac{1}{\sqrt{2}}(x_1 + x_2)$$

$$\vdots$$

$$k+1: \quad z^{(k+1)} = A\,y^{(k)} = \frac{1}{\sqrt{2}}(x_1 + (-1)^k x_2)$$

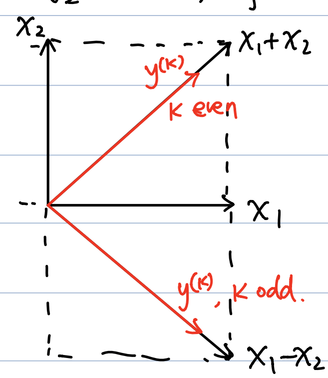
$$\|z^{(k+1)}\|_2 = 1$$

$$y^{(k+1)} = z^{(k+1)}/\|z^{(k+1)}\|_2 = \frac{1}{\sqrt{2}}(x_1 + (-1)^k x_2)$$

Therefore $y^{(k)}$ is $\frac{1}{\sqrt{2}}(x_1+x_2)$ if k even
$\frac{1}{\sqrt{2}}(x_1-x_2)$ if k odd

(Neither $\frac{1}{\sqrt{2}}(x_1+x_2)$ nor $\frac{1}{\sqrt{2}}(x_1-x_2)$ is an eigenvector)

∴



- Power iteration may not be convergent to (λ_1, x_1)

i) Assume $(-2, x_1)$, $(1, x_2)$ are two eigenpairs of $A \in \mathbb{R}^{2\times 2}$

$y^{(0)} = \frac{1}{\sqrt{2}}(x_1+x_2)$

$$\Rightarrow \quad y^{(k)} = \frac{1}{\sqrt{2^{2k}+1}}\left((-2)^k x_1 + x_2\right).$$

So that $\|y^{(k)} - x_1\|_2$ is small when k is even,
$\|y^{(k)} + x_1\|_2$ is small when k is odd,

$$\Rightarrow \quad y^{(k)} \to \pm x_1 \quad \text{(only the direction is correct)}$$

ii) Assume $(2, x_1)$ and $(1, x_2)$ are two eigenvectors of $A \in \mathbb{R}^{2\times 2}$

Assume $y^{(0)} = x_2$

$$\Rightarrow \quad z^{(1)} = Ax_2 = x_2 \Rightarrow z^{(1)} = x_2$$

$$\cdots \quad \Rightarrow \quad y^{(k)} = x_2 \quad \forall k.$$

i.e., $y^{(k)} \to x_2$.

Analysis of Power Iteration:

Since $y^{(k)}$ may converge to x_1 or $-x_1$, or alternatively, we can use $\min\{\|y^{(k)} - x_1\|_2^2, \|y^{(k)} + x_1\|_2^2\}$ to measure the convergence.

Note that $\|y^{(k)} - x_1\|_2^2 = \|y^{(k)}\|_2^2 + \|x_1\|_2^2 - 2\langle y^{(k)}, x_1\rangle = 2 - 2\langle y^{(k)}, x_1\rangle$

$$\|y^{(k)}+x_1\|_2^2=\|y^{(k)}\|_2^2+\|x_1\|_2^2+2\langle y^{(k)},x_1\rangle=2+2\langle y^{(k)},x_1\rangle$$

Therefore, we use

$1-\langle y^{(k)},x_1\rangle^2$ to measure the convergence.

we will show $1-\langle y^{(k)},x_1\rangle^2\to 0$ as $k\to+\infty$

Theorem: Assume $A\in\mathbb{R}^{n\times n}$ is symmetric and $|\lambda_1|>|\lambda_2|$.

If $\langle y^{(0)},x_1\rangle\neq 0$, then $\exists$ C_0 depending on $y^{(0)}$ s.t.

$$(1-\langle y^{(k)},x_1\rangle^2)^{1/2}\leq C_0\left|\frac{\lambda_2}{\lambda_1}\right|^k\to 0$$

Cosequently, ① $\min\{\|y^{(k)}-x_1\|_2,\|y^{(k)}+x_1\|_2\}\leq\sqrt{2}C_0\left|\frac{\lambda_2}{\lambda_1}\right|^k\to 0$

i.e., the limit of $y^{(k)}$ is $\pm x_1$

and ② $|\mu^{(k)}-\lambda_1|\leq 2\sqrt{2}C_0\left|\frac{\lambda_2}{\lambda_1}\right|^k\to 0$.

proof. By induction,

$$y^{(k)}=A^ky^{(0)}/\|A^ky^{(0)}\|_2.$$

Let $A=X\Lambda X^T$ be an eigenvalue decomposition of A.

Then $A^k=X\Lambda X^TX\Lambda X^T\cdots X\Lambda X^T=X\Lambda^kX^T$ (Because X is orthogonal)

Thus, $A^ky^{(0)}=X\Lambda^kX^Ty^{(0)}$

Let $v=X^Ty^{(0)}$

$$A^ky^{(0)}=X\Lambda^kv=\sum_{i=1}^n\lambda_i^kv_ix_i\text{ , where }v_i\in\mathbb{R},\ x_i\in\mathbb{R}^n.$$

$$\|A^ky^{(0)}\|_2^2=\left(\sum_{i=1}^n|\lambda_i|^{2k}|v_i|^2\right)$$

$$=|\lambda_1|^{2k}|v_1|^2\left(1+\left(\frac{|\lambda_2|}{|\lambda_1|}\right)^{2k}\left(\frac{|v_2|}{|v_1|}\right)^2+\cdots+\left(\frac{|\lambda_n|}{|\lambda_1|}\right)^{2k}\left(\frac{|v_n|}{|v_1|}\right)^2\right)$$

$$\geq(|\lambda_1|^k|v_1|)^2$$

and

$$\langle y^{(k)},x_1\rangle^2=\frac{1}{\|A^ky^{(0)}\|_2^2}\left\langle\sum_{i=1}^n\lambda_i^kv_ix_i,x_1\right\rangle^2=\frac{1}{\|A^ky^{(0)}\|_2^2}|\lambda_1|^{2k}|v_1|^2$$

(Because $\langle x_i,x_1\rangle=\begin{cases}1 & \text{if } i=1\\ 0 & \text{otherwise}\end{cases}$)

Therefore,

$$1-\langle y^{(k)},x_1\rangle^2=\frac{1}{\|A^ky^{(0)}\|_2^2}\left(\|A^ky^{(0)}\|_2^2-\left\langle\sum_{i=1}^n\lambda_i^kv_ix_i,x_1\right\rangle^2\right)$$

$$= \frac{|\lambda_1|^{2k}|v_1|^2}{\|A^{(k)}y^{(0)}\|_2^2}\left(\left(\frac{|\lambda_2|}{|\lambda_1|}\right)^{2k}\left(\frac{|v_2|}{|v_1|}\right)^2 + \cdots + \left(\frac{|\lambda_n|}{|\lambda_1|}\right)^{2k}\left|\frac{v_n}{v_1}\right|\right)$$

$$\leq \left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\left|\frac{v_2}{v_1}\right|^2 + \left|\frac{\lambda_3}{\lambda_1}\right|^{2k}\left|\frac{v_3}{v_1}\right| + \cdots + \left|\frac{\lambda_n}{\lambda_1}\right|^{2k}\left|\frac{v_n}{v_1}\right|$$

$$\leq \left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\cdot\left(\sum_{i=2}^{n}\left|\frac{v_i}{v_1}\right|^2\right) \equiv C_0 < +\infty \ (\text{because } v_1 \neq 0)$$

(Because $|\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$)

Thus,

$$\sqrt{1-\langle y^{(k)}, x_1\rangle^2} \leq C_0\left|\frac{\lambda_2}{\lambda_1}\right|^k \to 0 \quad \left(\text{since } \frac{|\lambda_2|}{|\lambda_1|} < 1\right)$$

Consequently,

① $\langle y^{(k)}, x_1\rangle^2 \geq 1 - C_0^2\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}$.

Also, by Cauchy-Schwartz, $\langle y^{(k)}, x_1\rangle^2 \leq \|y^{(k)}\|_2^2\,\|x_1\|_2^2 = 1$

So $1 - C_0^2\left|\frac{\lambda_2}{\lambda_1}\right|^{2k} \leq \langle y^{(k)}, x_1\rangle^2 \leq 1$

If $\langle y^{(k)}, x_1\rangle \geq 0$, then

$$\|y^{(k)} - x_1\|_2 = \left(2 - 2\langle y^{(k)}, x_1\rangle\right)^{1/2} \leq \sqrt{2C_0^2}\left|\frac{\lambda_2}{\lambda_1}\right|^k$$

If $\langle y^{(k)}, x_1\rangle \leq 0$, then

$$\|y^{(k)} + x_1\|_2 = \left(2 + 2\langle y^{(k)}, x_1\rangle\right)^{1/2} \leq \sqrt{2C_0^2}\left|\frac{\lambda_2}{\lambda_1}\right|^k$$

So $y^{(k)}$ is close to either x_1 or $-x_1$ when k is sufficiently large.

② $\mu^{(k)} = (y^{(k)})^T A y^{(k)}$

If $\langle y^{(k)}, x_1\rangle \geq 0$, then,

$$|\mu^{(k)} - \lambda_1| = |(y^{(k)})^T A y^{(k)} - x_1^T A x_1|$$

$$= |(y^{(k)})^T A (y^{(k)} - x_1) - (x_1 - y^{(k)})^T A x_1|$$

$$\leq |(y^{(k)})^T A (y^{(k)} - x_1)| + |(x_1 - y^{(k)})^T A x_1|$$

$$= |\langle y^{(k)}, A(y^{(k)} - x_1)\rangle| + |\langle x_1, A^T(x_1 - y^{(k)})\rangle|$$

$$\leq \overset{=1}{\|y^{(k)}\|_2}\,\|A(y^{(k)} - x_1)\|_2 + \overset{=1}{\|x_1\|_2}\,\|A^T(x_1 - y^{(k)})\|_2$$

$$\leq \|A\|_2\,\|y^{(k)} - x_1\|_2 + \overset{=\|A\|_2}{\|A^T\|_2}\,\|x_1 - y^{(k)}\|_2$$

$$\leq 2\|A\|_2\,\|y^{(k)} - x_1\|_2 \leq 2\sqrt{2C_0^2}\left|\frac{\lambda_2}{\lambda_1}\right|^k$$

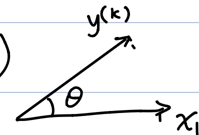
If $\langle y^{(k)}, x_1 \rangle \le 0$, similarly,

$$|\mu^{(k)} - \lambda_1| \le |(y^{(k)})^T A (y^{(k)} + x_1)| + |(x_1 + y^{(k)})^T A x_1|$$

$$\cdots \le 2\sqrt{2C_0^2}\left|\frac{\lambda_2}{\lambda_1}\right|^k.$$

In either case,

$$|\mu^{(k)} - \lambda_1| \le 2\sqrt{2C_0^2}\left|\frac{\lambda_2}{\lambda_1}\right|^k.$$



Remark : ① $\langle y^{(k)}, x_1 \rangle = \cos \angle (y^{(k)}, x_1)$ (since $\|y^{(k)}\|_2 = \|x_1\|_2 = 1$)

So $(1 - \langle y^{(k)}, x_1 \rangle^2)^{1/2} = \sin \angle (y^{(k)}, x_1)$

② The convergence rate depends on $\frac{|\lambda_2|}{|\lambda_1|} < 1$

The smaller $\left|\frac{\lambda_2}{\lambda_1}\right|$, the faster convergence.

When $|\lambda_2| = |\lambda_1|$, the power iteration may not converge to x_1.

③ When $\langle y^{(0)}, x_1 \rangle = 0$, C_0 may be infinity,

so that $|\langle y^{(k)}, x_1 \rangle| \not\to 0$. Consequently, $y^{(k)} \not\to \pm x_1$.