

MATH3322 - Matrix Computation

Taught by Jianfeng Cai

Notes by Aaron Wang

Contents

1	March 22nd, 2019	2
1.1	Eigenvalue Decomposition	2
1.2	Characteristic Polynomial	3
1.3	Special Case: Symmetric Matrix and SPD Matrix	5
2	March 27th, 2019	5
2.1	Computation of Eigenvalue Decomposition	5
2.2	Power Iteration	6
2.3	Analysis of Power Iteration	7
3	March 29th, 2019	9
3.1	Inverse Power Iteration	9
3.2	Shifted Inverse Power Iteration	10
	Index	11

1 March 22nd, 2019

1.1 Eigenvalue Decomposition

Definition 1.1. Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. A non-zero vector x is an **eigenvector** of A with $\lambda \in \mathbb{C}$ being the corresponding **eigenvalue** if:

$$Ax = \lambda x.$$

- Even if A is a real matrix, its eigenvalue and eigenvectors 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:

$$\begin{aligned} (\lambda^2, x) &\text{ is a eigenpair of } A^2 \\ (\lambda - \sigma, x) &\text{ is a eigenpair of } A - \sigma I \\ \left(\frac{1}{\lambda - \sigma}, x \right) &\text{ is a eigenpair of } (A - \sigma I)^{-1}. \end{aligned}$$

Proof. Since (λ, x) is an eigenpair of A , $Ax = \lambda x$. Multiplying both sides by A from the left:

$$\begin{aligned} A \cdot Ax &= \lambda Ax \implies A^2x = \lambda Ax = \lambda \cdot \lambda x = \lambda^2 x. \\ Ax - \sigma x &= \lambda x - \sigma x \implies (A - \sigma I)x = (\lambda - \sigma)x \\ \implies x &= (\lambda - \sigma)(A - \sigma I)^{-1}x \implies (A - \sigma I)^{-1}x. \end{aligned}$$

□

Definition 1.2. Two matrices A and B are **similar** with each other if there exists a nonsingular matrix T such that

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

Theorem 1.3

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

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

Thus, (λ, Tx) is an eigenpair of B . i.e. any eigenvalue of A is an eigenvalue of B . The reverse is similar. □

Definition 1.4. 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 $X = [x_1 x_2 \dots x_n]$ with $x_i \in \mathbb{C}^n$ the i -th column of x , and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ with $\lambda_i \in \mathbb{C}$ being the i -th diagonal of Λ , then

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

$$\implies [Ax_1 Ax_2 \dots Ax_n] = [\lambda_1 x_1 \lambda_2 x_2 \dots \lambda_n x_n].$$

$$\implies Ax_i = \lambda_i x_i, \quad i = 1, 2, \dots, n.$$

In other words $(\lambda_i, x_i), i = 1, 2, \dots, n$ are eigenpairs of A .

- Since X is nonsingular, x_i are linearly independent. So, x_i are n independent eigenvectors, which span \mathbb{C}^n .
- Eigenvalue decomposition implies $X^{-1}\Lambda X = A$, so that we also say 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.

1.2 Characteristic Polynomial

Definition 1.5. The **characteristic polynomial** of $A \in \mathbb{R}^{n \times n}$ denoted P_A is a degree n polynomial defined by

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

Let (λ_1, 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 iff $P_A(\lambda) = 0$, and the corresponding eigenvector x are non-zero solutions of $(\lambda I - A)x = 0$.

Example 1.6

$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. 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, $P_A(\lambda) = \lambda^2 = 0 \implies \lambda_1 = \lambda_2 = 0$ are the eigenvalues of A .

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

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

As there is only one independent eigenvector, A is not diagonalizable (i.e. no eigenvalue decomposition).

Example 1.7

$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. The characteristic polynomial is:

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

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

For eigenvector of $\lambda_1 = i$, solve $(iI - A)x = 0$, i.e.

$$\begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} x = 0 \implies x = \alpha \begin{bmatrix} i \\ 1 \end{bmatrix} \quad \alpha \in \mathbb{C}.$$

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

For eigenvector of $\lambda_2 = -i$:

$$\begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} x = 0 \implies x = \beta \begin{bmatrix} i \\ -1 \end{bmatrix} \quad \beta \in \mathbb{C}.$$

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

Define $X = [x_1 \ x_2] = \begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix}$, $\Lambda = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} = \begin{bmatrix} i & \\ & -i \end{bmatrix}$, $X^{-1} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2}i & -\frac{1}{2}i \end{bmatrix}$,

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

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

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

Remark 1.8 — However, we don't usually solve the characteristic equation, as polynomial root-finding is not numerically stable in general.

1.3 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 $x^* \equiv \overline{x^*}$ (conjugate transpose) from the left:

$$x^* Ax = \lambda x^* x \implies \lambda = \frac{x^* Ax}{x^* x}.$$

- $x^* Ax$ is real because $\overline{x^* Ax} = \overline{(x^* Ax)^T} = \overline{x^T A^T x} = x^* Ax$
- $x^* x$ is also real, because $\overline{x^* x} = \overline{(x^* x)^T} = \overline{x^T x} = x^* x$.
- As such, $\lambda = \frac{x^* Ax}{x^* x}$ is real.

□

2. Eigenvectors corresponding to distinct eigenvalues are orthogonal.
3. A is always diagonalizable, and the eigenvalue decomposition has a special form

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

where $Q \in \mathbb{R}^{n \times n}$ is orthonormal and $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal.

4. If A is SPD, then all eigenvalues are positive.
5. If A is SPSD, then all eigenvalues are non-negative.

Proof. Let (λ, x) be an eigenpair of A . then $Ax = \lambda x$, and λ, x are real. So

$$x^T Ax = \lambda x^T x \implies \lambda = \frac{x^T Ax}{x^T x} > 0.$$

if A is SPD. If A is SPSD, then $\lambda = \frac{x^T Ax}{x^T x} \geq 0$, since $x^T Ax \geq 0$.

□

2 March 27th, 2019

2.1 Computation of Eigenvalue Decomposition

For simplicity, we assume that $A \in \mathbb{R}^{n \times n}$ is symmetric, so that all eigenvalues/eigenvectors are real. Let λ_i $i = 1, 2, \dots, n$ be the eigenvalues of A , which are sorted in magnitude, i.e.

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

The corresponding eigenvectors are denoted by q_i . We have

$$Q = [q_1 \quad q_2 \quad \dots \quad q_n] \in \mathbb{R}^{n \times n}$$

satisfying $Q^T Q = Q Q^T = I$.

Definition 2.1. Let $A \in \mathbb{R}^{n \times n}$ be symmetric. For a given vector $x \in \mathbb{R}^n$, the **Rayleigh Quotient** is defined by

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

If x is an eigenvector,

$$r(x) = \frac{x^T A x}{x^T x} = \frac{\lambda x^T x}{x^T x} = \lambda,$$

i.e. $r(x)$ is an eigenvalue.

The eigenvalues are critical points of $r(x)$, with $\nabla r(x) = 0$. It can be proven that

$$\min_i \lambda_i = \min_{x \neq 0} r(x).$$

Remark 2.2 — This can be extended to non-symmetric matrices/ matrices or eigenvalues that are complex.

2.2 Power Iteration

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

Algorithm 2.3 1. Choose $y^{(0)} \in \mathbb{R}^n$ s.t. $\|y^{(0)}\|_2 = 1$.

2. for $k = 1, 2, \dots, n$

$$z^{(k)} = A y^{(k-1)}$$

$$y^{(k)} = \frac{z^{(k)}}{\|z^{(k)}\|_2}$$

$$\mu^{(k)} = \frac{(y^{(k)})^T A y^{(k)}}{(y^{(k)})^T y^{(k)}} = (y^{(k)})^T A y^{(k)}.$$

Remark 2.4 — $y^{(k)}$ is an approximation to $\pm x_1$, $\mu^{(k)}$ is an approximation to λ_1 .

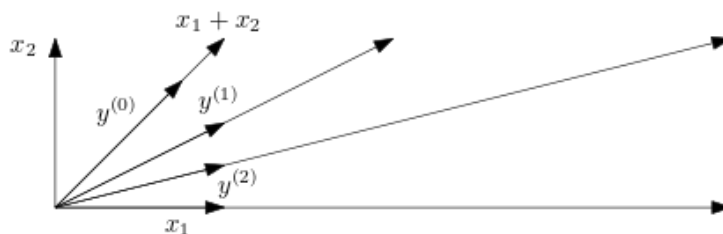


Figure 1

- Assume $(2, x_1), (1, x_2)$ are two eigenpairs of $A \in \mathbb{R}^{2 \times 2}$ (so that $x_1 \perp x_2$).
- Assume $y^{(0)} = \frac{1}{\sqrt{2}}(x_1 + x_2)$

- $k = 1$:

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

$$y^{(1)} = \frac{1}{\sqrt{5}}(2x_1 + x_2).$$

Note that $y^{(k)}$ approaches x_1 more than x_2 .

⋮

- $k + 1$:

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

If the component of x_1 is non-zero, then it will converge to x_1 , i.e. as long as $y^{(0)}$ is not a multiple of x_2 , it will converge to x_1 .

Claim 2.5. Power iteration may not be convergent:

Example 2.6

Assume $(1, x_1), (-1, x_2)$ are two eigenpairs of $A \in \mathbb{R}^{2 \times 2}$. Assume $y^{(0)} = \frac{1}{\sqrt{2}}(x_1 + x_2)$.

$$k = 1 : z^{(1)} = Ay^{(0)} = \frac{1}{\sqrt{2}}(x_1 - x_2).$$

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

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

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

which just repeats itself.

Remark 2.7 — Try with $(-2, x_1), (1, x_2)$. Does not converge, but we can get the direction of x_1 since both x_1 and $-x_1$ are eigenvectors.

Remark 2.8 — Power iteration may not converge to (λ_1, x_1) , e.g. $y^{(0)} = x_2$. This is because there is no x_1 component.

2.3 Analysis of Power Iteration

We will show $|\langle y^{(k)}, x \rangle| \rightarrow 1$. It is the same as $1 - \langle y^{(k)}, x_1 \rangle^2 \rightarrow 0, k \rightarrow \infty$

Theorem 2.9

Assume $A \in \mathbb{R}^{n \times n}$ is symmetric and $|\lambda_1| > |\lambda_2|$ (otherwise they might be amplified at the same rate).

If $\langle y^{(0)}, x_1 \rangle \neq 0$, then $\exists C_0 > 0$ depending on $y^{(0)}$ only such that

$$(1 - \langle y^{(k)}, x_1 \rangle^2)^{\frac{1}{2}} \leq C_0 \left| \frac{\lambda_2}{\lambda_1} \right|^k \rightarrow 0, \text{ as } k \rightarrow \infty.$$

Consequently,

- $\min\{\|y^{(k)} - x_1\|_2, \|y^{(k)} + x_1\|_2\} \leq \sqrt{2}C_0 \left| \frac{\lambda_2}{\lambda_1} \right|^k$, i.e. $y^{(k)} \rightarrow \pm x_1$
- $|\mu^{(k)} - \lambda_1| \leq 2\sqrt{2}C_0 \left| \frac{\lambda_2}{\lambda_1} \right|^k \rightarrow 0$

Proof. Note that

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

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

$$A^k = X\Lambda X^T X\Lambda X^T \dots X\Lambda X^T = X\Lambda^k X^T.$$

So

$$A^k y^{(0)} = X\Lambda^k X^T y^{(0)} = X\Lambda^k v$$

$$A^k y^{(0)} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1^k v_1 \\ \vdots \\ \lambda_n^k v_n \end{bmatrix} = \sum_{i=1}^n \lambda_i^k v_i x_i, \quad v_i \in \mathbb{R}, \quad x_i \in \mathbb{R}^n.$$

Because x_i are orthonormal,

$$\|A^k y^{(0)}\|_2^2 = \sum_{i=1}^n \lambda_i^{2k} v_i^2 = \sum_{i=1}^n |\lambda_i|^{2k} |v_i|^2 = |\lambda_1|^{2k} |v_1|^2 (1 + \dots) \geq (|\lambda_1|^k |v_1|)^2$$

and

$$\begin{aligned} \langle y^{(k)}, x_1 \rangle^2 &= \frac{1}{\|A^k y^{(0)}\|_2^2} \langle A^k y^{(0)}, x_1 \rangle^2 = \frac{1}{\|A^k y^{(0)}\|_2^2} \left\langle \sum_{i=1}^n \lambda_i^k v_i x_i, x_1 \right\rangle^2 = \frac{1}{\|A^k y^{(0)}\|_2^2} (\lambda_1^k v_1)^2 \\ &\leq \left| \frac{\lambda_2}{\lambda_1} \right|^{2k} \left(\left| \frac{v_2}{v_1} \right|^2 + \left| \frac{v_3}{v_1} \right|^2 + \left| \frac{v_4}{v_1} \right|^2 + \dots + \left| \frac{v_n}{v_1} \right|^2 \right) = C_0^2 \left| \frac{\lambda_2}{\lambda_1} \right|^{2k}. \end{aligned}$$

Thus

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

Because $C_0 < +\infty$, as $v_1 = \langle x_1, y^{(0)} \rangle \neq 0$ by assumption,

$$\sqrt{1 - \langle y^{(k)}, x_1 \rangle^2} \leq C_0 \left| \frac{\lambda_2}{\lambda_1} \right|^k \rightarrow 0, \text{ as } k \rightarrow \infty.$$

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

So

$$1 - C_0^2 \left| \frac{\lambda_2^{2k}}{\lambda_1} \right| \leq 1.$$

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

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

I give up will do this later

□

Remark 2.10 — 1. $\langle y^{(k)}, x_1 \rangle = \cos \theta$. Generally,

$$\cos \angle(x, y) = \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2}.$$

2. The convergence rate depends on $\left| \frac{\lambda_2}{\lambda_1} \right| < 1$, the smaller $\left| \frac{\lambda_2}{\lambda_1} \right|$, the faster the convergence. When $|\lambda_1| = |\lambda_2|$, the power iteration may not converge.
3. When $\langle y^{(0)}, x_1 \rangle = 0$, then $C_0 = +\infty$, so y will not converge to λ_1 .
4. In power iteration, only one matrix-product and several vector operations are used, the cost per step is $O(n^2)$. If we want an approximate eigenvalue/eigenvector of error ϵ , we need to choose k , s.t.

$$C \left| \frac{\lambda_2}{\lambda_1} \right|^{\frac{k}{2}} \leq \epsilon \implies \left| \frac{\lambda_1}{\lambda_2} \right|^{\frac{k}{2}} \geq \frac{C}{\epsilon}.$$

$$\frac{k}{2} \log \left| \frac{\lambda_1}{\lambda_2} \right| \geq \log \left(\frac{C}{\epsilon} \right) \implies k \geq \frac{\log \left(\frac{C}{\epsilon} \right)}{\log \left(\left| \frac{\lambda_1}{\lambda_2} \right| \right)} \sim O \left(\log \left(\frac{1}{\epsilon} \right) \right).$$

Then the total computational cost is

$$O \left(\log \left(\frac{1}{\epsilon} \right) \cdot n^2 \right).$$

5. Only the matrix-vector product involving A is needed. This means that A is not needed explicitly, only the subroutine to compute Ax is sufficient.

3 March 29th, 2019

3.1 Inverse Power Iteration

If λ_i , $i \in 1, \dots, n$ with $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ are eigenvalues of A , then $\frac{1}{\lambda_i}$ are eigenvalues of A^{-1} and

$$\frac{1}{|\lambda_1|} \leq \frac{1}{|\lambda_2|} \leq \dots \leq \frac{1}{|\lambda_n|}.$$

Therefore, we can apply power iteration to A^{-1} to get λ_n and hence x_n . This is called the inverse power iteration.

Algorithm 3.1 1. Choose $y^{(0)} \in \mathbb{R}^n$ s.t. $\|y^{(0)}\|_2 = 1$
 2. for $k = 1, 2, \dots$

$$z^{(k)} = A^{-1}y^{(k-1)}$$

$$y^{(k)} = \frac{z^{(k)}}{\|z^{(k)}\|_2}$$

$$\mu^{(k)} = (y^{(k)})^T A y^{(k)}.$$

Remark 3.2 — 1. From the convergence of power iteration, if:

- $\langle y^{(0)}, x_n \rangle \neq 0$
- $\frac{1}{|\lambda_n|} > \frac{1}{|\lambda_{n-1}|}$ (i.e. $|\lambda_n| < |\lambda_{n-1}|$)
- A^{-1} is symmetric (always true because A is symmetric).

then the limit of the iteration is:

$$y^{(k)} \rightarrow \pm x_n, \quad \mu^{(k)} \rightarrow \lambda_n,$$

with a rate $\left(\frac{|\lambda_n|}{|\lambda_{n-1}|}\right)^{\frac{k}{2}}$

2. We need to solve $Az^{(k)} = y^{(k-1)}$ in each iteration, which can be done by Gaussian Elimination. But we only need to compute $A = LU$ before the iteration and then, in each iteration, we obtain:

$$z^{(k)} = U^{-1}L^{-1}y^{(k-1)},$$

which is just a forward and backward substitution.

- Thus the total computational cost is:

$$O(n^3) + O\left(n^2 \cdot \log\left(\frac{1}{\epsilon}\right)\right)$$

for an ϵ -solution, ($O(n^3)$ for the LU decomposition)

3. If $|\lambda_n|$ is very close to 0, then, A is very close to singular, meaning that the solution of $Az^{(k)} = y^{(k-1)}$ may have a large error. However, we can still get a very accurate solution.

3.2 Shifted Inverse Power Iteration

Index

characteristic polynomial, 3

eigenvalue decomposition, 2

power iteration, 6

Rayleigh quotient, 6

similar matrices, 2