Linear algebra notes

Justin

1. Determinant

A determinant is a number:

$$D = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \text{a certain number}$$

Note a determinant always has the same number of rows and columns, in other words, it is a square.

a. Cofactor expansion

The remaining part after deleting the i-th row and j-th column is called a determinant D's minor, denoted by M_{ij}:

$$M_{ij} = \begin{vmatrix} a_{11} & \cdots & a_{1\,j-1} & a_{1\,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-1\,1} & \cdots & a_{i-1\,j-1} & a_{i-1\,j+1} & \cdots & a_{i-1\,n} \\ a_{i+1\,1} & \cdots & a_{i+1\,j-1} & a_{i+1\,j+1} & \cdots & a_{i+1\,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n\,j-1} & a_{n\,j+1} & \cdots & a_{nn} \end{vmatrix}$$

The associated cofactor is then defined as $A_{ij} = (-1)^{i+j} M_{ij}$. It is found that D can be expanded according to a certain column or row. For instance, the expansion according to the 1st column is the following:

$$D = \begin{cases} a_{11} & n = 1 \\ \sum_{i=1}^{n} a_{i1} A_{i1} & n > 1 \end{cases}$$

b. Triangle determinant (determinant of a triangle matrix)

If D is the determinant of a triangle matrix, then:

$$D = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = \prod_{i=1}^{n} a_{ii}$$

c. Properties

- $D = D^T$
- A determinant multiplied by a number k is equivalent to the scenario in which the elements in a certain row or column are all multiplied by k, i.e.,

$$kD = k \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ ka_{i1} & ka_{i2} & \cdots & ka_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

  Note this is very different from matrix-number multiplication.

- Interchanging two columns or rows changes the determinant's sign
- If all elements in one row (column) are k times that of the other, the determinant is 0

$$D = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ ka_{i1} & ka_{i2} & \cdots & ka_{in} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = k \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = 0 \text{ (Swapping two rows)}$$

- If all elements in one row (column) can be broken to the sum to two numbers, i.e., $a_{ij} = b_{ij} + c_{ij}$, then:

$$D = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ b_{i1} + c_{i1} & b_{i2} + c_{i2} & \cdots & b_{in} + c_{n1} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ b_{i1} & b_{i2} & \cdots & b_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ c_{i1} & c_{i2} & \cdots & c_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

- Multiplying one row (column) by k and adding it to another row (column) does not change the determinant's value

d. Cramer's rule

It's better to explain Cramer's rule via one example – assume the following system of linear equations:

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Then the solution is:

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, \qquad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}$$

Cramer's rule is quite computationally expensive and hence is not adopted as the major algorithm for solving the systems of linear equations.

e. Relation to the area of a parallelogram or the volume of a parallelepiped

I. Relation to the area of a parallelogram

The area of a parallelogram spanned by tow vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is the magnitude of $\boldsymbol{a} \times \boldsymbol{b}$:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Assume both $\boldsymbol{a}$ and $\boldsymbol{b}$ lie in the same plane so that $a_3 = b_3 = 0$, hence:

$$\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{k} \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

Therefore, the area of the parallelogram is given:

$$\text{Area} = |\boldsymbol{a} \times \boldsymbol{b}| = \left| \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \right|$$

II. Relation to volume of a parallelepiped

The volume of a parallelepiped spanned by three vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ is the magnitude of $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}$:

$$(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Hence, the volume of a parallelepiped can be expressed as:

$$\text{Volumn} = |(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}| = \left| \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \right|$$

2. Matrix

Unlike the determinant which is essentially a number, the matrix is a table of numbers. The identity matrix (equivalent to 1 in numbers) is:

$$\boldsymbol{I} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

a. Addition and number multiplication

I. Properties of matrix addition and number multiplication

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$$
$$(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$$
$$\boldsymbol{A} + \boldsymbol{0} = \boldsymbol{A}$$
$$k\boldsymbol{A} = k \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ \vdots & \vdots & & \vdots \\ ka_{i1} & ka_{i2} & \cdots & ka_{in} \\ \vdots & \vdots & & \vdots \\ ka_{n1} & ka_{n2} & \cdots & ka_{nn} \end{bmatrix}$$

Note multiplying a matrix by a number is equivalent to multiplying each of its elements by this number. This is different from determinant-number multiplication.

$$(kl)\boldsymbol{A} = k(l\boldsymbol{A})$$
$$k(\boldsymbol{A} + \boldsymbol{B}) = k\boldsymbol{A} + k\boldsymbol{B}$$
$$1\boldsymbol{A} = \boldsymbol{A}$$
$$0\boldsymbol{A} = \boldsymbol{0}$$
$$|\boldsymbol{A} + \boldsymbol{B}| \neq |\boldsymbol{A}| + |\boldsymbol{B}|$$

b. Matrix multiplication

Let $\boldsymbol{AB} = \boldsymbol{C}$, then each element in $\boldsymbol{C}$ is the inner product of the row vectors in $\boldsymbol{A}$ and column vectors in $\boldsymbol{B}$:

$$\boldsymbol{AB} = \begin{bmatrix} \vdots \\ \boldsymbol{a_i} \\ \vdots \end{bmatrix} [\cdots \quad \boldsymbol{b_j} \quad \cdots] = \begin{bmatrix} & \vdots & \\ \cdots & \boldsymbol{a_i b_j} & \cdots \\ & \vdots & \end{bmatrix}$$

Here, $\boldsymbol{a_i}$ and $\boldsymbol{b_j}$ are the row and column vectors of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

I. Properties of matrix multiplication

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$
$$(\boldsymbol{AB})\boldsymbol{c} = \boldsymbol{A}(\boldsymbol{Bc}), \boldsymbol{c}\text{: vector}$$
$$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}$$
$$k\boldsymbol{AB} = (k\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(k\boldsymbol{B})$$
$$\boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A}$$
$$(\boldsymbol{AB})^k \neq \boldsymbol{A}^k\boldsymbol{B}^k$$
$$|k\boldsymbol{A}| = k^n|\boldsymbol{A}|$$
$$|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$$

Please note, even though $\boldsymbol{A} \neq \boldsymbol{0}$ and $\boldsymbol{B} \neq \boldsymbol{0}$, it's still possible that $\boldsymbol{AB} = \boldsymbol{0}$.

I. Time complexity

Matrix multiplication is very computationally expensive. A $m \times n$ matrix multiplied by another $n \times p$ matrix has a time complexity of $O(mnp)$. One trick is that if the matrix multiplication is followed by vector multiplication, perform the vector multiplication first as it reduces the time complexity substantially.

c. Matrix transpose

I. Properties of matrix transpose

$$(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}$$
$$(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}^{\mathrm{T}}$$
$$(k\boldsymbol{A})^{\mathrm{T}} = k\boldsymbol{A}^{\mathrm{T}}$$
$$|\boldsymbol{A}^{\mathrm{T}}| = |\boldsymbol{A}|$$
$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$$

Given that $\boldsymbol{AB} \neq \boldsymbol{BA}$, the last property is exceptionally interesting. It can be generalized to finite number of matrices:

$$(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_k)^{\mathrm{T}} = \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_{k-1}^{\mathrm{T}}\cdots\boldsymbol{A}_1^{\mathrm{T}}$$

Both $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ are symmetric matrices.

Any square matrix can be decomposed to the sum of a symmetric and an anti-symmetric matrix:

$$\boldsymbol{A} = \underbrace{\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2}}_{\text{Symmetric}} + \underbrace{\frac{\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}}}{2}}_{\text{Anti-symmetric}}$$

d. Matrix inverse

Definition: if $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$, then $\boldsymbol{A}$ is invertible and the inverse of $\boldsymbol{A}$ is $\boldsymbol{B}$.

$\boldsymbol{A}$ to be invertible iff $|\boldsymbol{A}| \neq 0$. Under such a circumstance:

$$\boldsymbol{A}^{-1} = \frac{\boldsymbol{A}^*}{|\boldsymbol{A}|}$$

Here $\boldsymbol{A}^*$ is $\boldsymbol{A}$'s adjugate matrix. If $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$, then $\boldsymbol{A}^*$ has the following form:

$$\boldsymbol{A}^* = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}$$

In other words, $\boldsymbol{A}^*$ is the transpose of the matrix constructed by the cofactors of each element of $\boldsymbol{A}$.

I. Properties of matrix inverse

$$(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$$
$$(\boldsymbol{A}^{-1})^{\mathrm{T}} = \left(\boldsymbol{A}^{\mathrm{T}}\right)^{-1}$$
$$(k\boldsymbol{A})^{-1} = \frac{1}{k}\boldsymbol{A}^{-1}$$
$$|\boldsymbol{A}^{-1}| = \frac{1}{|\boldsymbol{A}|}$$
$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

The last property is very similar to matrix transpose. Its quick proof is in the following:

$$(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1})(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}(\boldsymbol{A}^{-1}\boldsymbol{A})\boldsymbol{B} = \boldsymbol{B}^{-1}(\boldsymbol{A}^{-1}\boldsymbol{A})\boldsymbol{B} = \boldsymbol{B}^{-1}\boldsymbol{I}\boldsymbol{B} = \boldsymbol{I}$$

Similar to matrix transpose, this property can be generalized to finite number of matrices:

$$(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_k)^{-1} = \boldsymbol{A}_k^{-1}\boldsymbol{A}_{k-1}^{-1}\cdots\boldsymbol{A}_1^{-1}$$

e. Block matrix

Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \cdots & \boldsymbol{A}_{1s} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \cdots & \boldsymbol{A}_{2s} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A}_{r1} & \boldsymbol{A}_{r2} & \cdots & \boldsymbol{A}_{rs} \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{1p} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \cdots & \boldsymbol{B}_{2p} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{B}_{s1} & \boldsymbol{B}_{s2} & \cdots & \boldsymbol{B}_{sp} \end{bmatrix}$, then the following holds:

- $\boldsymbol{A}^T = \begin{bmatrix} \boldsymbol{A}_{11}^T & \boldsymbol{A}_{21}^T & \cdots & \boldsymbol{A}_{r1}^T \\ \boldsymbol{A}_{12}^T & \boldsymbol{A}_{22}^T & \cdots & \boldsymbol{A}_{r2}^T \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A}_{1s}^T & \boldsymbol{A}_{2s}^T & \cdots & \boldsymbol{A}_{rs}^T \end{bmatrix}$
- $\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \cdots & \boldsymbol{C}_{1p} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & \cdots & \boldsymbol{C}_{2p} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{C}_{r1} & \boldsymbol{C}_{r2} & \cdots & \boldsymbol{C}_{rp} \end{bmatrix}$

Here $\boldsymbol{C}_{ij} = \sum_{i=1}^{s}\boldsymbol{A}_{it}\boldsymbol{B}_{tj}$. In other words, the rule is the same as ordinary matrix multiplication.

I. Some special cases

If $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_s \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & & & \\ & \boldsymbol{B}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{B}_s \end{bmatrix}$, then the following rules hold:

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_1\boldsymbol{B}_1 & & & \\ & \boldsymbol{A}_2\boldsymbol{B}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_s\boldsymbol{B}_s \end{bmatrix}$$
$$|\boldsymbol{A}| = |\boldsymbol{A}_1||\boldsymbol{A}_2|\cdots|\boldsymbol{A}_s|$$
$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_1^{-1} & & & \\ & \boldsymbol{A}_2^{-1} & & \\ & & \ddots & \\ & & & \boldsymbol{A}_s^{-1} \end{bmatrix}$$

Similar to multiplication, these rules are the same as the corresponding ordinary matrix operations.

f. Elementary matrix operations

There are three elementary matrix operations, each with its corresponding matrix form:

- Interchange two rows (or columns)

$$\boldsymbol{R}_{ij} = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & & 1 & & & \\ & & & & \ddots & & & & \\ & & & 1 & & 0 & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix}$$

- Multiply each element in a row (or column) by a non-zero number

$$\boldsymbol{R}_{i(k)} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & k & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

- Multiply a row (or column) by a non-zero number and add the result to another row (or column)

$$\boldsymbol{R}_{i+j(k)} = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 1 & & k & & & \\ & & & & \ddots & & & & \\ & & & & & 1 & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix}$$

Row-wise (column-wise) elementary operations on a matrix $\boldsymbol{A}$ are equivalent to $\boldsymbol{A}$ pre-multiply (post-multiply) by the corresponding elementary matrices.

Elementary matrices are invertible. Through elementary matrix operations, any matrix $\boldsymbol{A}$ can be converted to the standard form:

$$\boldsymbol{PAQ} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

Here, $\boldsymbol{P}$ and $\boldsymbol{Q}$ are product of a series of elementary matrices. It's apparent that both $\boldsymbol{P}$ and $\boldsymbol{Q}$ are invertible.

I. Use the elementary matrix operations to computer the matrix's inverse

A special case is when $\boldsymbol{A}$ is a square matrix and is invertible, $\boldsymbol{PAQ} = \boldsymbol{I}$, then $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{I}\boldsymbol{Q}^{-1}$, which means $\boldsymbol{A}$'s inverse $\boldsymbol{A}^{-1} = \boldsymbol{QP}$ is essentially the product of finite number of elementary matrices. This introduces a convenient approach to compute the inverse matrix.

If $\boldsymbol{A}$ is invertible, then its inverse can be expressed by $\boldsymbol{A}^{-1} = \boldsymbol{P}_1\boldsymbol{P}_2 \cdots \boldsymbol{P}_m$. Note if $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{P}_1\boldsymbol{P}_2 \cdots \boldsymbol{P}_m\boldsymbol{A} = \boldsymbol{I}$, then $\boldsymbol{P}_1\boldsymbol{P}_2 \cdots \boldsymbol{P}_m\boldsymbol{I} = \boldsymbol{A}^{-1}$**.** This indicates the same chain of row-wise operations that converts $\boldsymbol{A}$ to an identity matrix converts the identity matrix to its inverse $\boldsymbol{A}^{-1}$!

See the following example:

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

$$[\boldsymbol{A} \quad | \quad \boldsymbol{I}]=\left[\begin{matrix}0 & 2 & -1\\ 1 & 1 & 2\\ -1 & -1 & -1\end{matrix} \quad \middle| \quad \begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix}\right]$$

$$\xrightarrow{\text{Row−wise operations}}\left[\begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix} \quad \middle| \quad \begin{matrix}-\frac{1}{2} & -\frac{3}{2} & -\frac{5}{2}\\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\\ 0 & 1 & 1\end{matrix}\right]$$

Hence, the inverse of $\boldsymbol{A}$ is:

$$\boldsymbol{A}^{-1}=\begin{bmatrix}-\frac{1}{2} & -\frac{3}{2} & -\frac{5}{2}\\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\\ 0 & 1 & 1\end{bmatrix}$$

g. Rank of the matrix

Definition: for a given matrix $\boldsymbol{A}$, its maximum order (number of rows or columns) of sub-matrix with non-zero determinant is called $\boldsymbol{A}$'s rank.

I. Properties of matrix's rank

- $\mathrm{r}(\boldsymbol{A}) \le \min(n_{\mathrm{row}}, n_{\mathrm{col}})$
- $\mathrm{r}(\boldsymbol{A}^{\mathrm{T}}) = \mathrm{r}(\boldsymbol{A})$
- $\mathrm{r}(k\boldsymbol{A}) = \mathrm{r}(\boldsymbol{A})$
- If $\boldsymbol{A}$ is a $n \times n$ square matrix, $\mathrm{r}(\boldsymbol{A}) = n \Leftrightarrow |\boldsymbol{A}| \neq 0$ (full rank)
- Elementary matrix operations do not alter a matrix's rank
- Column-wise and row-wise ranks are equal

II. Use the elementary operations to determine a matrix's rank

Given the fact that the elementary matrix operations do not alter a matrix's rank, one can readily adopt this approach to compute a matrix's rank. See the following example:

$$\boldsymbol{A}=\begin{bmatrix}1 & 0 & 2 & 1 & 0\\ 7 & 1 & 14 & 7 & 1\\ 0 & 5 & 1 & 4 & 6\\ 2 & 1 & 1 & -10 & -2\end{bmatrix}\xrightarrow{\text{Elementary operations}}\begin{bmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

Hence, $\mathrm{r}(\boldsymbol{A}) = 3$.

3. Vector space

a. Maximal linearly independent subset

Given two vector sets $T_1\!:\{\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \cdots \quad \boldsymbol{\alpha}_r\}, T_2\!:\{\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2 \quad \cdots \quad \boldsymbol{\beta}_s\}$:

- If any $\boldsymbol{\alpha}$ in T_1 can be expressed as a linear combination of $\boldsymbol{\beta}$s in T_2, and vice versa, then the two sets are equivalent
- If any $\boldsymbol{\alpha}$ in T_1 can be expressed as a linear combination of $\boldsymbol{\beta}$s in T_2, and $r > s$, then $\boldsymbol{\alpha}$s in T_1 must be linearly dependent

- Assume $T_3 : \{\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \cdots \quad \boldsymbol{\alpha}_t\} \subseteq T_1$. If i) all $\boldsymbol{\alpha}$ in T_3 are linearly independent and ii) any $\boldsymbol{\alpha}$ in T_1can be written as a linear combination of T_3, then T_3 is one of T_1's maximal linearly independent subset

b. Rank of the vector set

The number of vectors in the maximal linearly independent subset is defined as the vector set's rank. If two vector sets are equivalent, they have the same rank.

I. Relation to the rank of matrix

Think of the matrix as a set to row (column) vectors, the rank of the matrix equals to that of the row (column) vector set.

c. Vector space and the base

The base of a vector space is essentially the maximal linearly independent subset of all vectors in the vector space. The number of the vectors in the base is called the vector space's dimension.

The determinant of the matrix formed by the base vectors is not 0. This is because of the equality between matrix's rank and vector set's rank.

d. Inner product

Let $\langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle$ denote the inner product of two vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

I. Properties of inner product

- $\langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\beta} = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\alpha} = \langle \boldsymbol{\beta}, \boldsymbol{\alpha} \rangle$
- $\langle \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2, \boldsymbol{\beta} \rangle$=$\langle \boldsymbol{\alpha}_1, \boldsymbol{\beta} \rangle + \langle \boldsymbol{\alpha}_2, \boldsymbol{\beta} \rangle$
- $\langle k\boldsymbol{\alpha}, \boldsymbol{\beta} \rangle = k\langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle$
- $\langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \rangle \geq 0$
- Cauchy-Schwarz inequality: $\langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle^2 \leq \langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \rangle \langle \boldsymbol{\beta}, \boldsymbol{\beta} \rangle$. This is because $\frac{\langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle}{|\boldsymbol{\alpha}|} \leq |\boldsymbol{\beta}|$ (projection of $\boldsymbol{\beta}$ along $\boldsymbol{\alpha}$ is shorter than the magnitude of $\boldsymbol{\beta}$)

4. System of linear equations

a. Possible cases of solutions

For a system of linear equations with m equations and n variables: $\boldsymbol{Ax} = \boldsymbol{b}$, there are three possible cases:

- If $\mathrm{r}(\boldsymbol{A}) < \mathrm{r}([\boldsymbol{A}|\boldsymbol{b}])$, this system is inconsistent and there are no solutions
- If $\mathrm{r}(\boldsymbol{A}) = \mathrm{r}([\boldsymbol{A}|\boldsymbol{b}]) = n$, there is only one solution
- If $\mathrm{r}(\boldsymbol{A}) = \mathrm{r}([\boldsymbol{A}|\boldsymbol{b}]) < n$, there are infinite solutions

Here $[\boldsymbol{A}|\boldsymbol{b}]$ is the augmented matrix of the coefficient matrix $\boldsymbol{A}$.

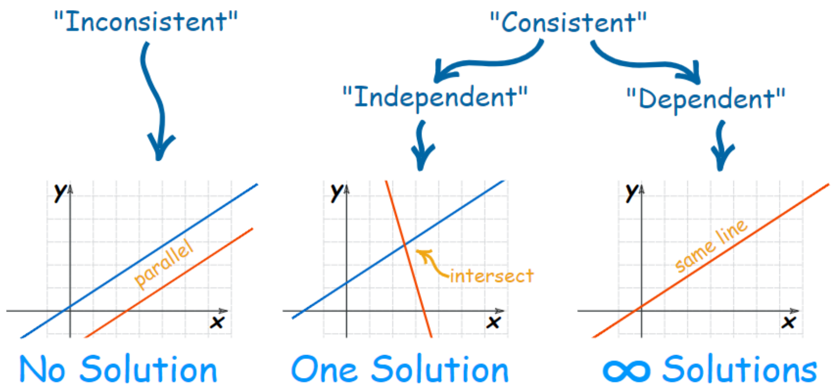

Figure 1. Three possible cases of solutions: i) no solution, ii) one solution, iii) infinity solutions

b. Homogeneous systems

For a homogenous system with m equations and n variables: $\boldsymbol{Ax}=\mathbf{0}$, the dimension of the solution set is $\dim\mathcal{N}(\boldsymbol{A})=n-r(\boldsymbol{A})$. One special case is when $n=\mathrm{r}(\boldsymbol{A})$, there is only one solution ($\boldsymbol{x}=\mathbf{0}$) and $\dim\mathcal{N}(\boldsymbol{A})=0$. In order to have non-zero solutions, $n>\mathrm{r}(\boldsymbol{A})$.

I. Examples

- Problem: solve the following system of linear equations:

$$\begin{cases} x_1+x_2+x_3+x_4 &= 0\\ 2x_1+2x_2+x_3+3x_4 &= 0\\ x_1+x_2+2x_3 &= 0\end{cases}$$

Solution: the coefficient matrix $\boldsymbol{A}=\begin{bmatrix}1&1&1&1\\2&2&1&3\\1&1&2&0\end{bmatrix}\xRightarrow{\text{Elementary operations}}\begin{bmatrix}1&1&0&2\\0&0&1&-1\\0&0&0&0\end{bmatrix}$.

Hence, $r(\boldsymbol{A})=2$ and the dimension of the solution set is $\dim\mathcal{N}(\boldsymbol{A})=n-\mathrm{r}(\boldsymbol{A})=2$. Let x_2 and x_4 be the free variables, the solutions can be expressed as:

$$\begin{cases} x_1 &= -x_2-2x_4\\ x_2 &= x_2\\ x_3 &= x_4\\ x_4 &= x_4\end{cases}$$

c. Non-homogeneous systems

Let $\boldsymbol{\eta}$ be one special solution of the non-homogenous system with m equations and n variables: $\boldsymbol{Ax}=\boldsymbol{b}$, and $\boldsymbol{x}_\mathrm{c}$ be the solution set of its homogenous counterpart $\boldsymbol{Ax}=\mathbf{0}$, then the solution of $\boldsymbol{Ax}=\boldsymbol{b}$ is $\boldsymbol{x}=\boldsymbol{x}_\mathrm{c}+\boldsymbol{\eta}$.

5. Matrix similarity

a. Eigenvalues and eigenvectors

Definition: let $\boldsymbol{A}$ be a $n\times n$ square matrix, if there exist a number λ and a non-zero vector $\boldsymbol{x}$ that satisfies $\boldsymbol{Ax}=\lambda\boldsymbol{x}$, then λ and $\boldsymbol{x}$ are called $\boldsymbol{A}$'s eigenvalues and eigenvectors, respectively.

Reshaping the equation $\boldsymbol{Ax}=\lambda\boldsymbol{x}$ leads to $(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\mathbf{0}$. In order to have non-zero solutions $\boldsymbol{x}\neq\mathbf{0}$, one must have $|\lambda\boldsymbol{I}-\boldsymbol{A}|=0$:

$$f(\lambda)=|\lambda\boldsymbol{I}-\boldsymbol{A}|=\begin{vmatrix}\lambda-a_{11} & -a_{12} & \cdots & -a_{1n}\\ -a_{21} & \lambda-a_{22} & \cdots & -a_{2n}\\ \vdots & \vdots & & \vdots\\ -a_{n1} & -a_{n2} & \cdots & \lambda-a_{nn}\end{vmatrix}=\lambda^n+b_1\lambda^{n-1}+\cdots=0$$

Since $|\lambda\boldsymbol{I}-\boldsymbol{A}|$ is an n-order polynomial of λ, an n-order squared matrix always has n complex eigenvalues.

I. Properties of eigenvalues

Assume the n eigenvalues of an n-order square matrix $\boldsymbol{A}$ are: $\lambda_1, \lambda_2, \ldots, \lambda_n$, then:

- $\lambda_1\lambda_2\cdots\lambda_n=|\boldsymbol{A}|$
  - Proof: $|\lambda\boldsymbol{I}-\boldsymbol{A}|=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)$. When $\lambda=0$, $|-\boldsymbol{A}|=(-\lambda_1)(-\lambda_2)\cdots(-\lambda_n)=(-1)^n\lambda_1\lambda_2\cdots\lambda_n$. Hence $\lambda_1\lambda_2\cdots\lambda_n=|\boldsymbol{A}|$.
- $\mathrm{tr}(\boldsymbol{A})=a_{11}+a_{22}+\cdots+a_{nn}$
- The eigenvalues of $k\boldsymbol{A}$ are $k\lambda_1, k\lambda_2, \ldots, k\lambda_n$
- The eigenvalues of $\boldsymbol{A}^l$ are $\lambda_1^l, \lambda_2^l, \ldots, \lambda_n^l$
- If $\boldsymbol{A}$ is invertible, the eigenvalues of $\boldsymbol{A}^{-1}$ are $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1}$

II. Properties of eigenvectors

- Eigenvectors associated with different eigenvalues are linearly independent
- Let λ_0 be an n-order square matrix $\boldsymbol{A}$'s t-degenerate eigenvalue, the dimension of the vector space span by the eigenvectors associated with λ_0 is $\leq t$. Hence, an n-order square matrix $\boldsymbol{A}$ has at most n eigenvectors

III. Condition number

Definition: the condition number of a matrix $\boldsymbol{A}$ is the ratio of its maximal and minimal eigenvalues.

$$\kappa(\boldsymbol{A})=\frac{\lambda_{\max}}{\lambda_{\min}}$$

The eigenvalues can be thought of as the matrix's ability to "magnify" the eigenvectors. Hence, the condition number reflects the matrix's maximum and minimum magnifications.

The matrix is said to be ill-conditioned if the condition number is large. If the coefficient matrix of the system of linear equations is ill-conditioned, the solutions are instable.

In reality, when dealing with numeric matrices, the condition number is a better gauge to test the singularity of a matrix than its determinant. For instance, $10^{-10}\boldsymbol{I}$ has a tiny determinant, but it's actually well-conditioned with $\kappa=1$. Check this link for details.

IV. Examples

- Problem: find the eigenvalues and eigenvectors of the following matrix:

$$\boldsymbol{A}=\begin{bmatrix}1 & -3 & 3\\ 3 & -5 & 3\\ 6 & -6 & 4\end{bmatrix}$$

Solution: first of all let's compute $\boldsymbol{A}$'s eigenvalues:

$$|\lambda\boldsymbol{I}-\boldsymbol{A}|=(\lambda+2)^2(\lambda-4)\Longrightarrow\lambda_1=\lambda_2=-2, \lambda_3=4.$$

When $\lambda_1=\lambda_2=-2$:

$$(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=(-2\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$$

$$\Longrightarrow\boldsymbol{x}=k_1[1\quad 1\quad 0]^{\mathrm{T}}+k_2[-1\quad 0\quad 1]^{\mathrm{T}}$$

When $\lambda_3=4$:

$$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = (4\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$$
$$\Rightarrow \boldsymbol{x} = k[1 \quad 1 \quad 2]^{\mathrm{T}}$$

- Problem: find the steady-state of the Markov chain transition matrix:

$$\boldsymbol{A} = \begin{bmatrix} \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

Solution: the steady-state of the Markov chain transition matrix satisfies $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}$:

$$|\lambda \boldsymbol{I} - \boldsymbol{A}| = (\lambda - 1)\left[\lambda - \frac{1}{8}(1 + \sqrt{3})\right]\left[\lambda - \frac{1}{8}(1 - \sqrt{3})\right]$$

Hence, $\lambda = 1$ is $\boldsymbol{A}$'s eigenvalue. Using $(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ lets us find the associated eigenvector:

$$\boldsymbol{x} = k\begin{bmatrix} \frac{14}{23} & \frac{5}{23} & \frac{4}{23} \end{bmatrix}^{\mathrm{T}}$$

Note, one can solve the equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}$ directly and obtain the same solution.

b. Similar matrices

Definition: let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $n \times n$ square matrices, if there exists an invertible n-order square matrix $\boldsymbol{P}$ so that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{B}$, then $\boldsymbol{A}$ and $\boldsymbol{B}$ are called similar matrices, denoted as $\boldsymbol{A} \sim \boldsymbol{B}$.

I. Properties of similar matrices

- $\boldsymbol{A} \sim \boldsymbol{A}$
- $\boldsymbol{A} \sim \boldsymbol{B} \Rightarrow \boldsymbol{B} \sim \boldsymbol{A}$
- $\boldsymbol{A} \sim \boldsymbol{B}$ and $\boldsymbol{B} \sim \boldsymbol{C}$, then $\boldsymbol{A} \sim \boldsymbol{C}$
- If $\boldsymbol{A} \sim \boldsymbol{B}$, then i) $\mathrm{r}(\boldsymbol{A}) = \mathrm{r}(\boldsymbol{B})$, ii) $|\boldsymbol{A}| = |\boldsymbol{B}|$, and iii) $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same eigenvectors

c. Matrix diagonalization

The major problem of similar matrices is to find an invertible square matrix $\boldsymbol{P}$, so that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix $\boldsymbol{\Lambda}$.

An n-order square matrix $\boldsymbol{A}$ to be diagonalizable iff $\boldsymbol{A}$ has n linearly independent eigenvectors. Then $\boldsymbol{A}$ is diagonalized by its eigenvalues and eigenvectors:

$$\mathrm{P}^{-1}\mathrm{AP} = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}, \mathrm{P} = [\mathrm{x}_1 \quad \mathrm{x}_2 \quad \cdots \quad \mathrm{x}_n]$$

Here $\lambda_1, \lambda_2, \ldots, \lambda_n$ are n eigenvalues, $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ are the associated n linearly independent eigenvectors.

Note, matrix inversion and matrix diagonalization are different concepts. They do not have any direct connections!

I. Examples

- Problem: assume annually 10% of the city population moves to the country yard. On the other hand, 20% of the country population moves to the cities. Will all the population be concentrated in the cities many years later?

Solution: let the initial city and country population be y_0 and z_0, respectively. The transition matrix is therefore:

$$\begin{bmatrix} y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix}$$

Let $\boldsymbol{A} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}$, the goal is to figure out $\begin{bmatrix} y_k \\ z_k \end{bmatrix} = \boldsymbol{A}^k \begin{bmatrix} y_0 \\ z_0 \end{bmatrix}$ $(k \to \infty)$. To solve this, it's more convenient to diagonalize $\boldsymbol{A}$ first.

It's not hard to find $\boldsymbol{A}$'s two eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = 0.7$. The associated eigenvectors are $\boldsymbol{x}_1 = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}^{\boldsymbol{T}}$ and $\boldsymbol{x}_2 = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \end{bmatrix}^{\boldsymbol{T}}$. Hence

$$\begin{aligned} \boldsymbol{A}^k &= \left\{ [\boldsymbol{x}_1 \quad \boldsymbol{x}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} [\boldsymbol{x}_1 \quad \boldsymbol{x}_2]^{-1} \right\}^k \\ &= [\boldsymbol{x}_1 \quad \boldsymbol{x}_2] \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix} [\boldsymbol{x}_1 \quad \boldsymbol{x}_2]^{-1} \\ &= [\boldsymbol{x}_1 \quad \boldsymbol{x}_2] \begin{bmatrix} 1 & 0 \\ 0 & 0.7^k \end{bmatrix} [\boldsymbol{x}_1 \quad \boldsymbol{x}_2]^{-1} \end{aligned}$$

As $k \to \infty$, $\boldsymbol{A}^k \to [\boldsymbol{x}_1 \quad \boldsymbol{x}_2] \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} [\boldsymbol{x}_1 \quad \boldsymbol{x}_2]^{-1} = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}$. Hence, $\begin{bmatrix} y_\infty \\ z_\infty \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix}$, i.e., the ratio between the city and country population reaches a steady $2:1$, which is fairly counter-intuitive.

- Problem: solve the following system of differential equations:

$$\begin{cases} \dot{x_1} &= 4x_1 + x_3 \\ \dot{x_2} &= 2x_1 + 3x_2 + 2x_3 \\ \dot{x_3} &= x_1 + 4x_3 \end{cases}$$

Solution: the given system of differential equations can be expressed in the matrix form:

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \dot{x_3} \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Let $\boldsymbol{A} = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix}$, $\boldsymbol{A}$ can expressed as:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 5 & & \\ & 3 & \\ & & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}^{-1}$$

Plug $\boldsymbol{A}$ back to the original system of differential equations, yielding:

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \dot{x_3} \end{bmatrix} = \begin{bmatrix} 5 & & \\ & 3 & \\ & & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Let $\begin{bmatrix} \dot{z_1} \\ \dot{z_2} \\ \dot{z_3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, then the original system is reduced:

$$\begin{bmatrix} \dot{z_1} \\ \dot{z_2} \\ \dot{z_3} \end{bmatrix} = \begin{bmatrix} 5 & & \\ & 3 & \\ & & 3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

The new system is much easier to solve!