线性方程组

本资料是微信公众号<机器人学家>编者的个人笔记，综合了各种线性代数课程、资料和自己的思考总结，仅供<机器人学家>公众号读者内部交流学习使用。

个人笔记难免有疏漏之处，关于笔记内容的任何疑问，可以在微信公众号<机器人学家>上与我直接交流，或者发送邮件给*gorobotics@126.com*，我尽量一一回复。

扫码关注微信公众号

机器人学家

扫码向我打赏

金额随意

目录

I. 解的存在性、唯一性

结合线性变换和线性空间理解。

1.1 AX=0

解即为A的null space。
有非零解 意味着A的null space不为空。

解性质：

- X的基础解系即Null（A）的一组基。包括n-r组解。
- AX = 0 -> BAX = 0，所以BA对应的解包括A对应的解

求基础解系：把自由变量取成自然基

1.2 AX=b

X是A的列向量表出b的系数。
所以有解等价于b在A的row space里。

有唯一解等价于说A的null space is empty

解性质：齐次解+非齐次特解

这些都要结合空间理解。

II. 矩阵分解factorization

矩阵分解是为了让求解线性方程组更方便。
计算机中实际实现的线性方程组求解方法都是基于矩阵分解。

Some decomposition came from this idea(LDU, Jordan), while some others (SVD) have clear geometry meaning, when view A as a linear transformation.

2.1 LDU分解

A is $m \times n, Rank(A) = n, m >= n$,
We could write $PA = LDU$, where:

P is a mxm permutation matrix. for example, $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

L is a mxm square low-triangle matrix with 1s on the diagonal: $\begin{bmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & * & 1 \end{bmatrix}$

D is a mxm square diagonal matrix.

U is a mxn upper-triangle matrix with 1s on the diagnal: $\begin{bmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{bmatrix}$

We call the diagnal elements of D “pivot”

examples:

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

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

2.1.1 Notes

P

The permutation matrix is used to change the ordering of base. We need to do this in case that a pivot is zero. PA changes row.
If needs to change coloum as well, we could write $P_1AP_2 = LDU$

U

Sometimes we treat D and U together as U, that is LU factorization: $PA = LU$

$det(A) = \pm det(D)$, sign depends on P.

if $A = A^T$ and $P = I$, then $U = L^T$

if $A = A^T$ and A is positive definite, Then we have:
$P = I, L = U^T, det(A) = det(D)$, diagonal entries of D are positive.

The LDU decomposition is **not unique**.
It is unique if:

1. A is square and invertable
2. P=I

2.1.2 Gaussian Elimination

Basic Idea

Perform row operations on A to obtain an Upper-triangular matrix.

Record the process

Two ways.

First one
While doing row operations on A, do the inverse, **row** operations on I
(Note: 初等矩阵$E_{i,j}(a)$左乘是i行的a倍加到j行，右乘是j列的a倍加到i列）
To see what happend:$L_3L_2L_1A = U$, then $A = L_1^{-1}L_2^{-1}L_3^{-1}U$
when A -> U'(upper triangular), I->L, L should be low-triangular.
then we have $A = LU'$.
When permutation is needed, say $L'A' = A$, A' needs to become P_1A'.
$LP_1(P_1^{-1}A') = A$ the problem is that how could ensure LP_1 is still a lower-triangular matrix.
The solution is $(P_1LP_1)(P_1^{-1}A') = P_1A$ it could be verified that when doing Gaussian Elimination, (P_1LP_1) is still a lower-triangular matrix.

Second one
Doing row operations on $[I\ A]$, obtain $[M\ U']$, then we have $MA = U', A = M^{-1}U'$

This is already LU decomposition. Next, simply get DU from U':

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

2.1.3 解线性方程组

$AX = B => LDUX = B$, then we could obtain X by solving two simpler linear systems:
$Ly = B$, then
$DUx = y$

2.2 对角化

见笔记“矩阵运算”

2.3 若当标准型

The following two methods are based on $A^T A$

2.4 QR分解

Suppose A is a $m \times n$ matrix with independent columns, we can factor A as:

$A = QR$,
where $Q \in m \times n, R \in n \times n$

The columns of Q forms the orthonormal basis of A's column space.
Which means $Q^T Q = I$.
(Note QQ^T may not be I, since A may not be square)

R is invertible and upper-triangular.

2.4.1 One Algorithm

First, form $A^T A$. This is a positive definite symmetric matrix.
(it is positive definite because the col of A are independent.)

Second, compute the LDU factorization of $A^T A$. you will get $A^T A = LDL^T$.

Finally, let $R = D^{1/2} L^T$, and $Q = AR^{-1}$

2.4.2 Notes on QR algorithms

Unfortunately, both of these straightforward algorithms have poor numerial stability. In practice one use a different algorithm.

2.5 SVD（重要）

SVD is a generalization of diagonalization.

2.5.1 idea

Diagonalization could not work nicely when:

1. you could not find n eigenvector as a basis for both input and output space (for example, A is not square, or Algebraic multiplicity > geometric multiplicity for some eigen value)
2. some eigen values are complex.

In geometry, this means we could not choose a basis under which a dilation transformation could be found to be equivalence to A.

In this case, we would need to change basis for both input space and outspace, so as to form a dilation transformation.

The transformation matrix using B1 for input space, B2 for output space is denoted as ${}^{B1}f_{B2}$.
By definition, ${}^{B1}f_{B2} = A$ when $B_1 = B_2 = E$.

For example,

2.5.2 Summary

Matrix

Preparation
For Any $m \times n$ matrix A, we could form the following two decomposition:

$$A^TA = VDV^T$$
$$AA^T = UD'U^T$$

U is an $m \times m$ orthogonal matrix whose columns are the eigenvectors of AA^T.
The columns of U forms an orthognal basis for **the whole output space**.
The **first k column**s of U forms an orthognal basis for **column space** of A.
The rest n-k columns are null space of A^T(so that they form an orthogonal basis).
k

V is an $n \times n$ orthogonal matrix whose columns are the eigenvectors of A^TA
The columns of V forms an orthognal basis for **the whole input space**.
The **last n-k columns** of V forms an orthognal basis for **nullspace** of A.
The first k colums of V forms an orthognal basis for rowspace of A.

D is a $n \times n$ diagonal matrix, with the non-negative eigenvalues of A^TA.
D' is a $m \times m$ diagonal, with the non-negative eigenvalues of AA^T.

D and D' have the same non-zero diagonal entries(except that the order might be different)

Conclusion

Any $m \times n$ matrix A can be factored as $A = U\Sigma V^T$, where:

$$Rank(A) = Rank(\Sigma) = k$$

Σ is a diagonal $m \times n$ matrix of the form:

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_k & & & \\ & 0 & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$

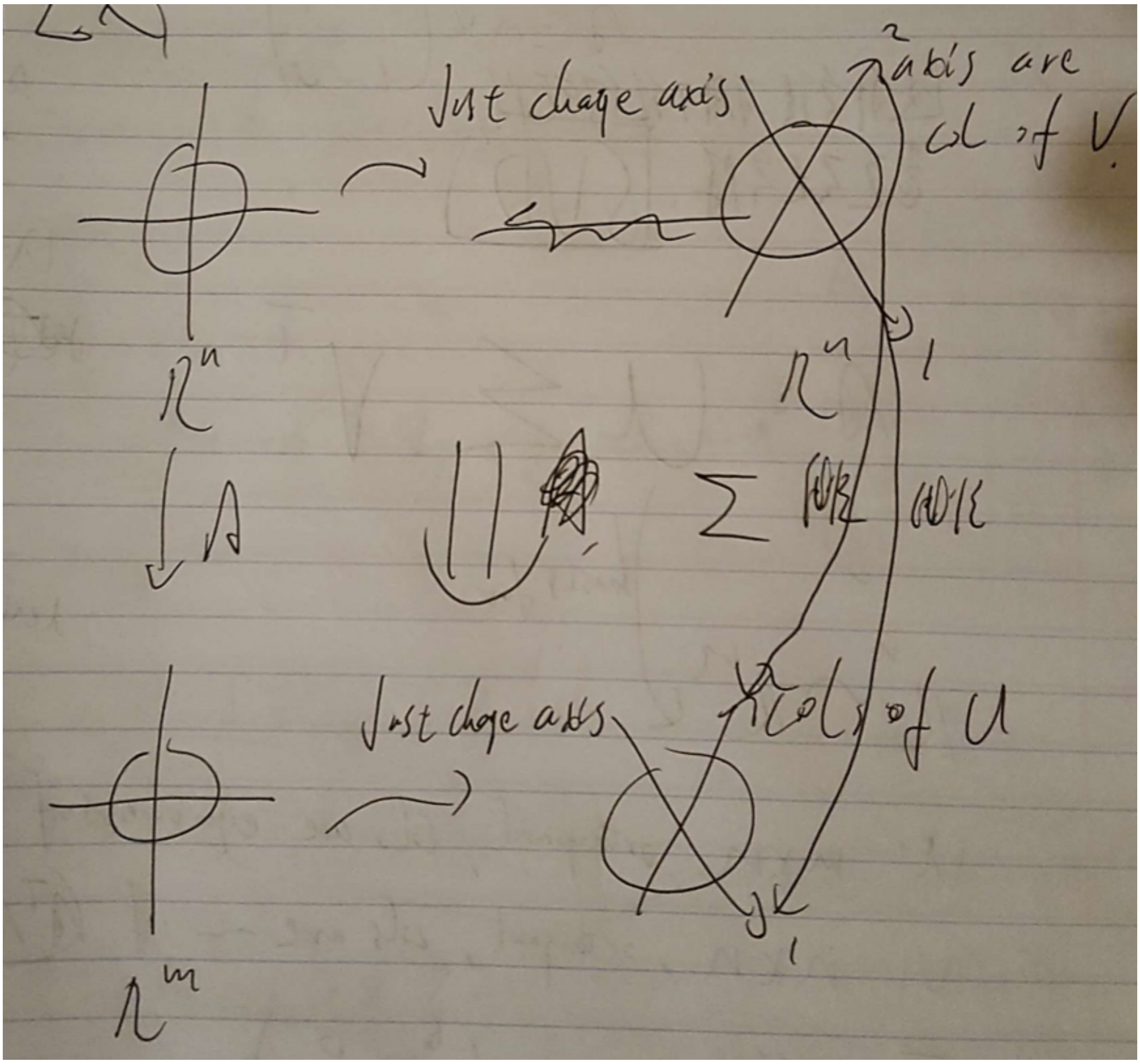
and $k = \text{rank}(A)$

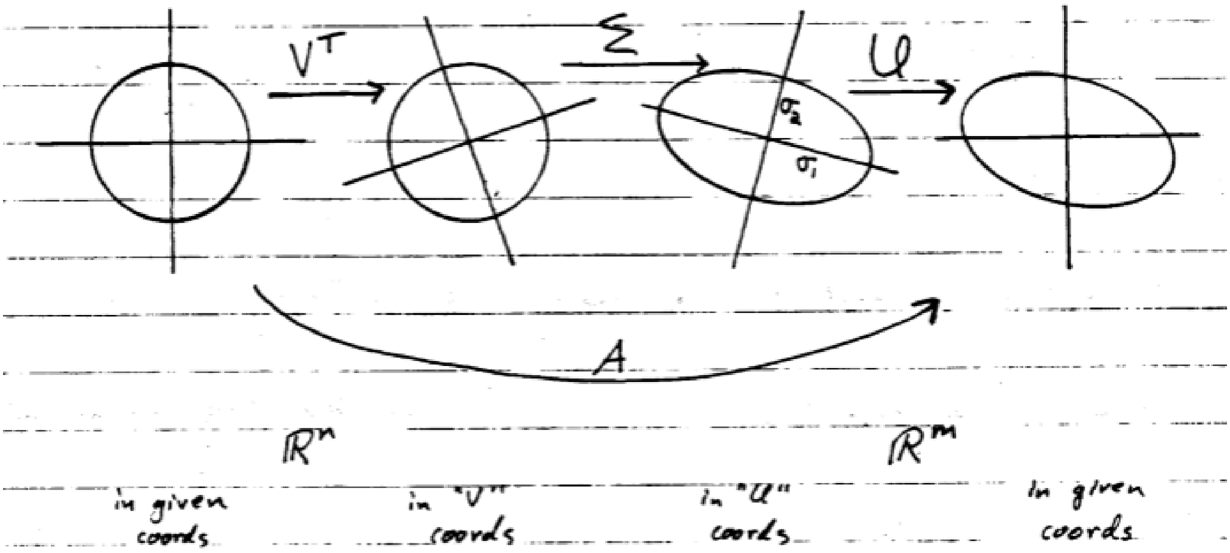
$\sigma_i = \sqrt{\lambda_i}$, where λ_i is the nonzero eigenvalue of $A^T A$.

In fact, the nonzero eigenvalues of $A^T A$ are the same as of AA^T

When A is symmetric positive-definite, then SVD reduce to diagonalization.

Geometry



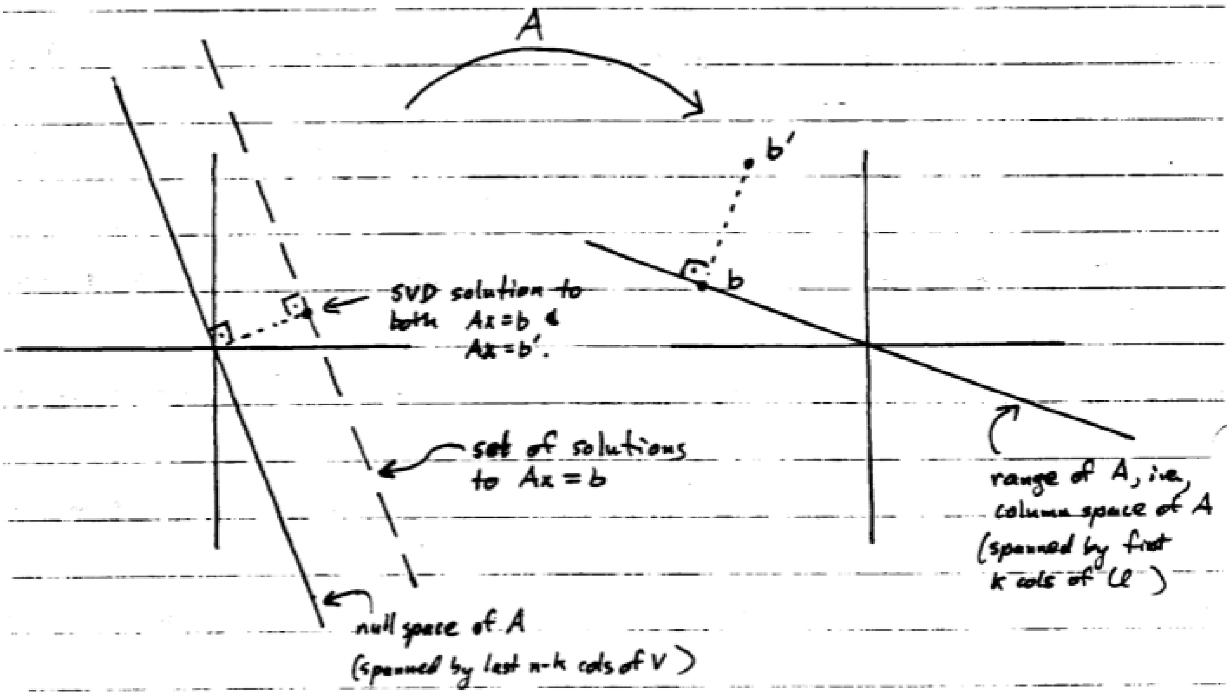


2.5.3 Solving linear systems

To solve Ax=b, we first calculate the SVD decomposition $A = U\Sigma V^T$, then compute $\bar{x} = V\,\frac{1}{\Sigma}\,U^T b$

what SVD done

The following sketch shows how SVD solve for $AX = b$ and $AX = b'$, where $b \in colspace(A), b' \notin colspace(A)$



SVD solve Ax=b by choosing x entirely from row space, which is closest to the origin among all possible solutions. $V \frac{1}{\Sigma} U^T$ is a pseudo-inverse.

SVD solve Ax=b' by projecting b' onto the colspace of A, obtaining b then solve Ax=b. In other words, SVD obtain the least-square solution.

Pseudo-inverse: $Ax = b$, the pseudo-inverse solution to it is $x_p = (A^T A)^{-1} A^T b$.

$$(A^T A)^{-1} A^T = V \frac{1}{\Sigma} U^T$$

How SVD do it

Before we do it, we define $\frac{1}{\Sigma}$ as the diagonal matrix whose diagonal entries are of the form:

$$\left(\frac{1}{\Sigma}\right)_{ii} = \begin{cases} \dfrac{1}{\Sigma_{ii}} & \text{if } \Sigma_{ii} \neq 0 \\ 0 & \text{if } \Sigma_{ii} = 0 \end{cases}$$

Observe:

a. if Σ is invertible, then $\frac{1}{\Sigma} = \Sigma^{-1}$

b. if $rank(\Sigma) = k$ then

$$\frac{1}{\Sigma} \cdot \Sigma = \Sigma \cdot \frac{1}{\Sigma} = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 0 \end{pmatrix}$$

Ok, now takes a look at SVD.
Firstly, $Ax = b$ ->

$$U\Sigma V^T x = b$$

$$\Sigma V^T x = U^T b$$

$U^T b$ is the coordinate of b under an orthogonal basis, and the first k vectors of the basis forms the colspace(A). In other word, $U^T b$ is the projection of b on Colspace(A).
Thus, if $b \in colspace(A)$, $U^T b = [x1, ..xk, 0, \ldots, 0]^T$
Next,

$$V^T x = \frac{1}{\Sigma} U^T b$$

$V^T x$ gives the coordinate of x under a basis whose last n-k vectors spans Null(A). In other words, x is projected on the row space of A,

Considering the structure of $\frac{1}{\Sigma}$, it only do dilation on first k elements, then ignore the other elements.

Now, lets take an inverse look. what is the meanning of SVD solution $\bar{x}$? What they do with b?
clearly, it project b on colspace(A), dilation, then project what remains onto rowspace(A).

Now, lets take an look at what b could we get using a SVD solution $\bar{x}$.

$$A\bar{x} = U\Sigma V^T V \frac{1}{\Sigma} U^T b$$

$$= U \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 0 \end{bmatrix} U^T b$$

If $b \in colspace(A)$, then the matrix

$$\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 0 \end{bmatrix}$$

actually do nothing, and $A\bar{x} = b$.

If $b \notin colspace(A)$, then we only got b projected on colspace, eliminated other elements, then project back to whole space. That is the projection of b onto colspace(A).