EE526 A review on Matrices

Darian Muresan

In these few pages, I would like to give a brief review of linear algebra and some of the matrix properties that we have seen quite often so far. If you have any questions or suggestions, please feel free to ask them. The presentation here tries to avoid very detailed mathematical proofs, and rather, I'm trying to give you a more intuitive view of matrices. In this short description, we will also think of all matrices as being of size $m \times n$ (with $m \geq n$), unless otherwise mentioned.

1 What's a Matrix

There are two general ways to visualize a matrix, both of which have their own advantages. First, think of a matrix as a collection of column vectors.

$$[A]_{m\times n} = \begin{bmatrix} | & \cdots & | \\ \phi_1 & \cdots & \phi_n \\ | & \cdots & | \end{bmatrix}$$

If you think of A this way, then

1. the *range* of A is the subspace spanned by all its columns; this is a subspace of R^m, since the columns have m components.
2. the *rank* of A is the dimension of the *range*. In other words, it's the number of linearly independent column vectors. As it turns out, the number of linearly independent column vectors is equivalent to the number of linearly independent row vectors.

A second way to think of $A_{m\times n}$ is as a linear transformation (linear function) from R^n to R^m

$$[A]_{m\times n}\,[x]_{n\times 1} = [y]_{m\times 1}$$

Viewing A this way we can talk about

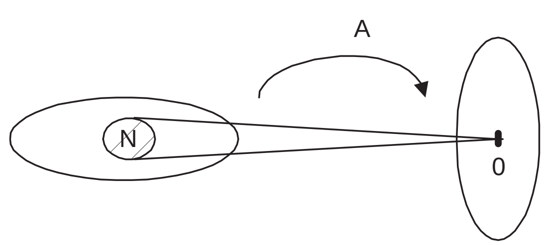

Figure 1: Null space of A gets mapped to $\{0\}$

1. the *null space* of A, is everything in the domain of function A, that gets mapped to zero (N in Fig. (1)).
2. the *eigenvectors* of A are those vectors in the domain that get stretched by transformation A (Fig. 2). The amount of stretching is equal to the corresponding *eigenvalue*. The definition of the eigenvector is those vectors x for which $Ax = \lambda x$.

3. if A is an $m \times m$ matrix, then we can also talk about its *determinant*. For a square matrix, the determinant tells how much (by volume) does a unitary cube get enlarged by transformation A ([4], pp. 156). If the unitary cube from the domain, gets mapped into a parallelogram of dimensions smaller than m, the determinant is zero.

Here are some general matrix properties:

1. For any matrix $A_{m\times n}$ we have
$$rank(A) + dim(null(A)) = n$$

2. For any matrix A the following are equivalent
   - All columns are linearly independent
   - Matrix A is full rank
   - Matrix A has non zero singular values
   - The null space of A is $\{0\}$

3. The determinant of a matrix A is the product of the eigenvalues. Intuitively this makes sense: the determinant measures how much a unit cube is increased in volume by transformation A. Eigenvalues, tell how much the eigenvectors are stretched by transformation A. Thus, considering the stretch done by A in all directions, is equivalent to knowing how much A stretched the unit cube.

4. For any square matrix A, the following are equivalent
   - Matrix A is invertible
   - Matrix A is full rank
   - Matrix A has non zero eigenvalues
   - Matrix A has non zero singular values
   - The null space of A is $\{0\}$
   - The determinant of A is non-zero

2 Eigenvalues

As mentioned above, the eigenvectors of $A_{m\times m}$ are the vectors x_i for which $Ax_i = \lambda_i x$ $(i = 1, \ldots, m)$. The eigenvectors are the only vectors that are stretched but not rotated by the linear transformation (Fig. 2). This means that

$$\begin{aligned} A\,[x_1, \ldots, x_m] &= [Ax_1, \ldots, Ax_m] \\ &= [\lambda_1 x_1, \ldots, \lambda_m x_m] \\ &= [x_1, \ldots, x_2] \underbrace{\begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ & \vdots & \vdots & \\ 0 & 0 & \ldots & \lambda_m \end{bmatrix}}_{\Lambda} \end{aligned} \tag{1}$$

Now let $[x_1, \ldots, x_m] = X$. Then equation (1) becomes

$$A = X\Lambda X^{-1} \tag{2}$$

Equation (2) is known as the diagonalization of A. A natural question arises: is every matrix A diagonalizable? The answer is No!, but almost all of them are. A non diagonalizable matrix is sometimes called *defective*. For completeness, when would a matrix be non-diagonalizable? Such a matrix would be one for which X is not invertible, one for which the eigenvectors are not linearly independent. One such matrix would be

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

As an exercise, you figure it out why.

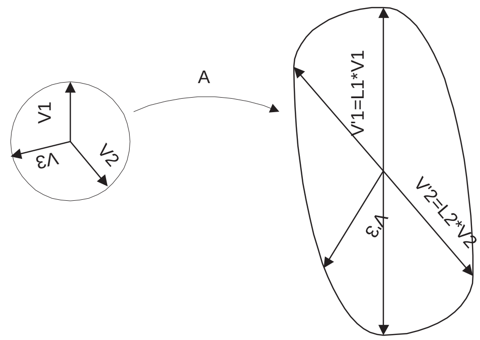

Figure 2: Eigenvectors V_1 and V_2 are stretched and not rotated by transformation A, while vector V_3 is stretched **and** rotated. Also note that if x is an eigenvector, so is $-x$, which means that under a linear transformation, a circle will always be stretched in an ellipsoid looking figure.

3 Singular Value Decomposition (SVD)

The idea behind SVD is similar to diagonalization, except that it can be applied to any matrix A, regardless of it's dimensions. As we already know, $A_{m\times n}$ is a linear transformation from R^n to R^m. The idea behind SVD is that there exists n orthonormal vectors $\{v_1, \ldots, v_n\}$ in the domain of A that get mapped to n orthogonal vectors $\{t_1, \ldots, t_n\}$ in the range of A (Fig. 3)

$$\begin{aligned} A\underbrace{[v_1, \ldots, v_n]}_{V} &= [t_1, \ldots, t_n] \\ &= \underbrace{[u_1, \ldots, u_n]}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ & \vdots & \vdots & \\ 0 & 0 & \ldots & \sigma_n \end{bmatrix}}_{\Sigma} \end{aligned} \tag{3}$$

where u_i is the normalized version of t_i ($\|t_i\| = \sigma_i$). So what are the dimensions of all these matrices? Matrix V contains n column vectors that are in the domain of matrix A, hence $V_{n\times n}$. And U contains n column vectors that are in the range of A, hence $U_{m\times n}$. Here are some important things to notice from equation (3)

- matrix V is orthonormal of dimensions $n \times n$ (hence $V^{-1} = V^T$)
- the columns of matrix V form an orthonormal basis for the domain of matrix A.
- the columns of matrix V corresponding to zero singular values form a basis for the null space of A.
- the columns of matrix U, corresponding to non-zero singular values form an orthonormal basis for the range of A.

The common form of equation (3) is

$$\begin{aligned} A &= U\Sigma V^{-1} \\ &= U\Sigma V^T \end{aligned} \tag{4}$$

SVD decomposition can be thought of as the skeleton of matrix A, it's the most "bare bones" decomposition of a matrix that you can have. Once a matrix is in the SVD form, you can tell almost everything you want to know about the matrix. For the interested reader, I would strongly recommend reading chapters 4 and 5 of [1].

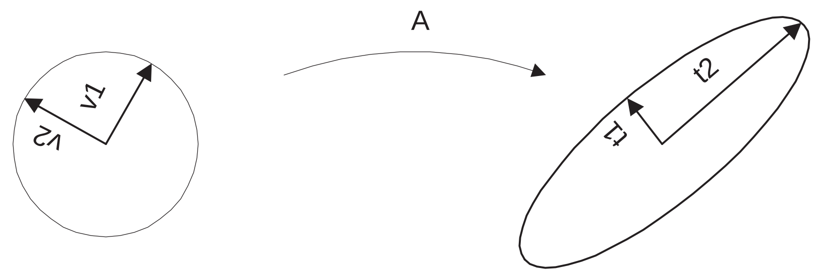

Figure 3: Transformation A maps orthonormal vectors v_1 and v_2 to the orthogonal vectors t_1 and t_2

4 QR Decomposition

Given a matrix $A_{m\times n}$, the idea behind QR factorization is to use A to generate a second matrix $Q_{m\times n}$ which has orthonormal column vectors that span the same subspace as the column vectors of A.

$$Range(A) = Range(Q) \text{ and } Q^TQ = I_{n\times n}$$

Since the range of A and Q is the same, linear combinations of the column vectors of Q can generate all the column vectors of A. This linear combination can be described by a matrix $R_{n\times n}$. In other words:

$$A_{m\times n} = Q_{m\times n}R_{n\times n}$$

Additionally, if we require that the first column of A must have the same direction as the first column of Q, the second column of A to be a linear combination of the first two columns of Q, the third column of A to be a linear combination of the first three columns of Q and so on, matrix R will be an upper triangular matrix. When we say *QR Decomposition*, we mean R to be upper triangular. In fact, QR factorization is nothing else but Graham Schmidt applied to the columns of A.

5 General Matrix Properties

Here are some general matrix properties that can be easily proved with the material discussed so far

- For any matrix A, the number of linearly independent column vectors is equal with the number of linearly independent row vectors (i.e. row rank is equal to column rank)

  Proof. This is almost trivial from the SVD decomposition. The column rank of A is equal with the number of non-zero singular values of A. If $A = U\Sigma V^T$ then $A^T = V\Sigma^T U^T$. Since $\Sigma^T = \Sigma$, it follows that the column rank of A is equivalent to the column rank of A^T. In other words, the number of linearly independent columns of A is equal with the number of linearly independent rows of A. Q.E.D.

6 Positive Definite and Symmetric Matrices

So far in the course, on numerous occasions, we have encountered positive definite matrices. I would like to give here a brief review of the properties we have made so much use of. By definition, a square matrix A is positive definite if

$$x^TAx > 0 \text{ for all } x \neq 0$$

Here are some of the properties of a positive definite matrix A

1. All the eigen values of a positive definite matrix A are positive.

   Proof. Let x be an eigenvector of A and λ the corresponding eigenvalue. Then we have

   $$0 < x^TAx = x^T\lambda x = \lambda\|x\|^2$$

So $\lambda > 0$ Q.E.D.

2. In general, eigenvectors are not orthogonal, but if A is symmetric ($A^T = A$), eigenvectors corresponding to distinct eigenvalues are orthogonal.

   Proof. This proof is taken from [1]. Suppose $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ with $\lambda_1 \neq \lambda_2$. Then

$$\lambda_2 x_1^T x_2 = x_1^T A x_2 = (x_2^T A x_1)^T = (\lambda_1 x_2^T x_1)^T = \lambda_1 x_1^T x_2,$$

   so $(\lambda_1 - \lambda_2) x_1^T x_2 = 0$. Since $\lambda_1 \neq \lambda_2$, we have $x_1^T x_2 = 0$ (i.e. orthogonal). Q.E.D.

3. An even stronger statement can be easily proven. Any symmetrical, positive definite matrix A has orthonormal eigenvectors.

   Proof. So we just have to show that if two eigenvectors have the same eigenvalue, then we can find two eigenvectors with that same eigenvalue, which are orthogonal. But if two linearly independent eigenvectors have the same eigenvalue, then any linear combination of the two eigenvectors will also be an eigenvector with the same eigenvalue.

$$\left.\begin{array}{c} Ax = \lambda x \\ Ay = \lambda y \end{array}\right\} \rightarrow A(c_1 x + c_2 y) = \lambda(c_1 x + c_2 y)$$

   So apply Gram Schmidt to all the vectors corresponding to the same eigenvalue, to get a set of orthogonal eigenvectors. Done!

   Q.E.D.

4. If matrix S is full rank, then $A = S^T S$ is symmetric, positive definite.

   Proof. The fact that A is symmetric is easy to see. If S is full rank, then any vector y in the range of S can be written as a linear combination of the column vectors of S ($y = Sx$, for some x). Next, for $y \neq 0$ we have $x \neq 0$ and

$$\begin{aligned} 0 &< \|y\|^2 \\ &= y^T y \\ &= x^T S^T S x \end{aligned}$$

   Hence $S^T S$ is positive definite Q.E.D.

5. Any symmetric, positive definite matrix A can be written as $A = G^T G$ for some G

   Proof. From problem (3) we have $A = U \Sigma U^T$. Since Σ is a diagonal matrix, its square root is obtained by taking the square root of each of the diagonal entries. Now let $G^T = U \sqrt{\Sigma}^T$. Then

$$G^T G = U \sqrt{\Sigma}^T \sqrt{\Sigma} U^T = U \Sigma U^T = A$$

   And this proves it. Q.E.D.

7 Matrix Calculations

In this section, we'll present different methods of taking inverses and obtaining determinants of blocked matrices.

1.

$$\begin{aligned} \begin{bmatrix} I & -CB^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & C \\ D & B \end{bmatrix} \begin{bmatrix} I & 0 \\ -B^{-1}D & I \end{bmatrix} &= \begin{bmatrix} A - CB^{-1}D & 0 \\ D & B \end{bmatrix} \begin{bmatrix} I & 0 \\ -B^{-1}D & I \end{bmatrix} \\ &= \begin{bmatrix} A - CB^{-1}D & 0 \\ 0 & B \end{bmatrix} \end{aligned} \quad (5)$$

2. Let A, B, C, and D be square matrices. Here is how we compute the inverse of the block matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

Proof.

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \underbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}_{M} = \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix}$$

$$\begin{bmatrix} I & -B(D - CA^{-1}B)^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} M = \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} A^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & -B(D - CA^{-1}B)^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}}_{\text{This product is } M^{-1}} M = I$$

Hence,

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

Q.E.D.

References

[1] Lloyd N. Trefethen and David Bau, *Numerical Linear Algebra,* SIAM 1997.

[2] Gilbert Strang, *Linear Algebra and its Applications,* Third Editions, Sounders SBJ 1986.

[3] Golub and Van Loan *Matrix Computations,* Johns Hopkins 1996 (3rd ed.).

[4] Hans Schneider & George P. Barker *Matrices and Linear Algebra,* Holt, Rinehart and Winston 1968