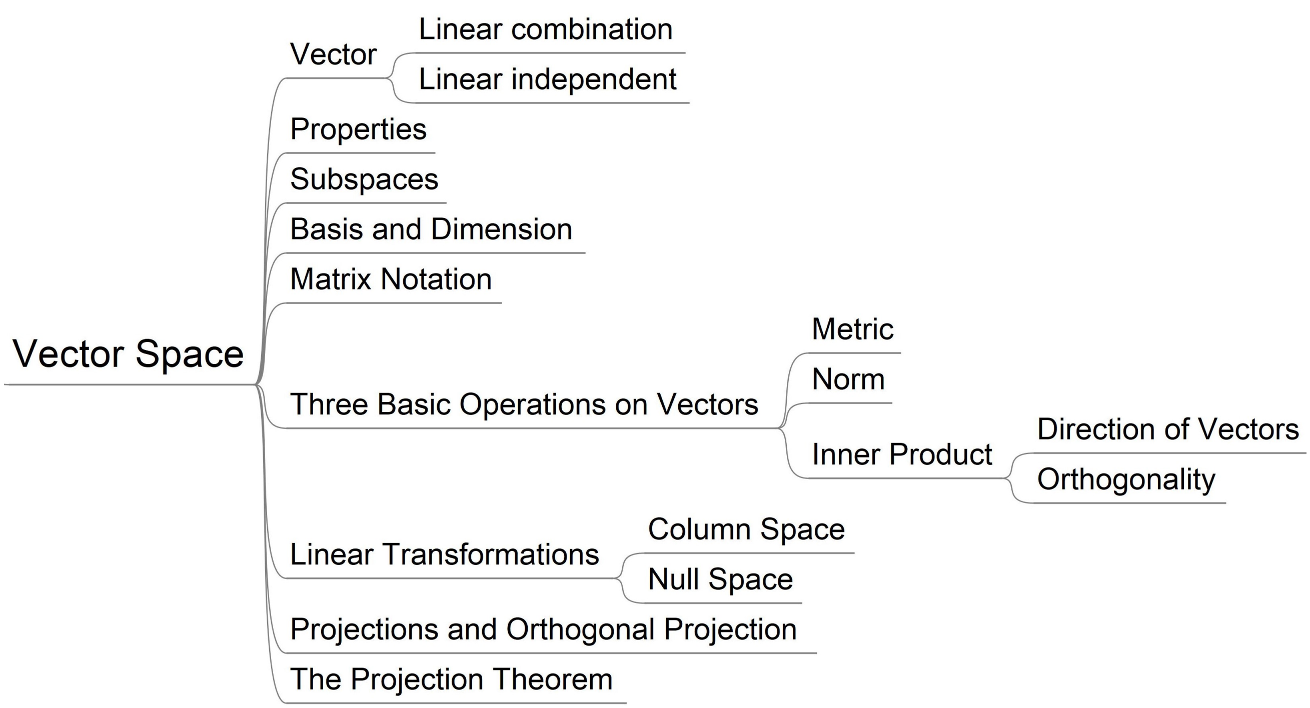
Linear Algebra

Zhenrong
February 9, 2021

1 Signal Spaces

The concept of vector space is a collection of vectors that can be represented by a smaller group of vectors using linear combination. A relationship between vectors and signal is that signals can be regarded as vectors. Treating signals as vectors, the theory of vector and vector space provide us a useful tool for signal analysis. In this section, the basic theory of vectors and vector spaces is introduced. We will give a brief overview of the knowledge using info-graphic. For some important concepts, it will followed by a detailed description.



1.1 Linear Transformation and Matrix Notation

Definition 1.1 (Linear Transformation)**.** A transformation $L : X \to Y$ from a vector space X to a vector space Y (where X and Y have the same scalar field R) is a linear transformation if for all vectors $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2} \in X$:

1. $L(\alpha \boldsymbol{x}) = \alpha L(\boldsymbol{x})$ for all $\mathbf{x} \in X$ and all scalars $\alpha \in \mathbb{R}$, and

2. $L(\boldsymbol{x}_1 + \boldsymbol{x}_2) = L(\boldsymbol{x}_1) + L(\boldsymbol{x}_2)$.

For example, a linear transformation L from the vector space $\mathbb{R}^n$ to $\mathbb{R}^m$ can be expressed using the notation of an $m \times n$ matrix L. Let $L : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$L(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_2 + 4x_3).$$

This linear transformation can be placed in matrix notation. By writing an element in $\mathbb{R}^3$ in vector form as $[x_1, x_2, x_3]^T \in \mathbb{R}^3$, we can write

$$L = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 4 \end{bmatrix},$$

Then,

$$L\mathbf{x} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_2 + 4x_3 \end{bmatrix}.$$

Definition 1.2. Let $L : X \to Y$ be an operator (linear or otherwise). The column space $\mathcal{R}(L)$ is

$$\mathcal{R}(L) = \{\boldsymbol{y} = L\boldsymbol{x} : \boldsymbol{x} \in X\},$$

that is, it is the set of values in Y that are reached from X by application of L. The nullspace $\mathcal{N}(L)$ is

$$\mathcal{N}(L) = \{\boldsymbol{x} \in X : L\boldsymbol{x} = 0\},$$

that is, it is the set of values in x that are transformed to 0 in Y by L.

1.2 Projection and Orthogonal Projections

1.2.1 Projection

The projection decomposed the input vector into disjoint components. If V and W are disjoint subspaces of a linear space S and $S = V + W$, then any vector $\boldsymbol{x} \in S$ can be uniquely written as

$$\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{w},$$

where $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$.

The projection operator returns that component of $\boldsymbol{x}$ which lies in V, as $P\boldsymbol{x} = \boldsymbol{v}$.

1.2.2 Subspace Orthogonality and Orthogonal Complements

Definition 1.3 (Orthogonal Subspaces)**.** Two Hilbert subspaces are said to be orthogonal subspaces, $V \perp W$ if and only if every vector in V is orthogonal to every vector in W.

Definition 1.4 (Orthogonal Complement)**.** Given a subspace V of X, one defines the orthogonal complement $V^\perp$ of V to be the set $V^\perp$ of all vectors in X which are perpendicular to V.

Noted for subspaces $V^\perp$ and V:

1. $X = V \oplus V^\perp$.
2. $\dim X = \dim V + \dim V^\perp$.

1.2.3 Orthogonal Projections

Definition 1.5 (Orthogonal Projections)**.** In a Hilbert space the projection onto a subspace V along its orthogonal complement $V^{\perp}$ is an orthogonal projection operator.

Definition 1.6 (Orthogonal Projections)**.** Let P be a projection operator on an inner product space $S.P$ is said to be an orthogonal projection if its column and nullspace are orthogonal, $\mathcal{R}(P) \perp \mathcal{N}(P)$.

1.3 Projection Matrix and Geometrically Notion

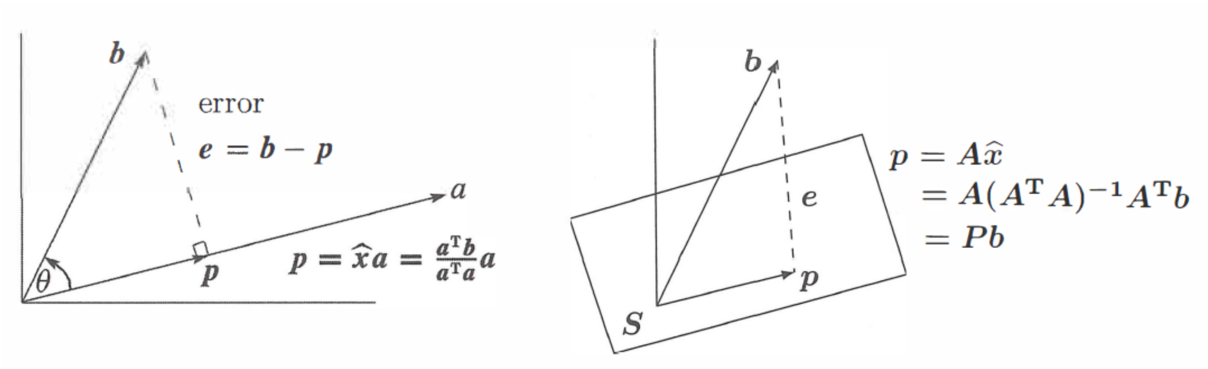


Figure 1: The projection $\boldsymbol{p}$ of $\boldsymbol{b}$ onto a line and onto S = column space of A.

The projection finds the nearest the point $\boldsymbol{p}$ in subspace P to $\boldsymbol{b}$ by minimize the approximation error vector $\boldsymbol{e}$. The approximation error $\boldsymbol{e}$ is orthogonal to the $\boldsymbol{a}$ or the column space of A. The approximation error $\boldsymbol{e} = (I - P)\boldsymbol{x}$ is in the subspaces $S^{\perp}$.

1.4 Projection Theorem and Least-Square Solution

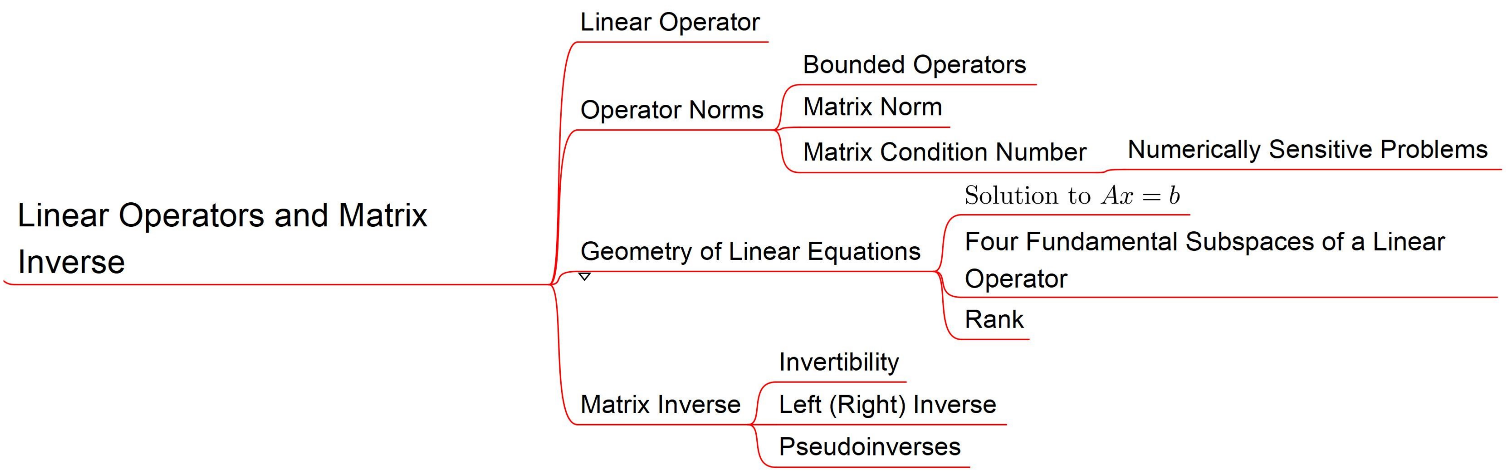
If $A\boldsymbol{x} = \boldsymbol{b}$ has no solution ($\boldsymbol{b}$ is not in the column space of A), we can not get the error $\boldsymbol{b} - A\boldsymbol{x}$ down to zero. The least-square solution $\hat{\boldsymbol{x}}$ makes $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ as small as possible. If we calculate the gradient of $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ and equating this result to zero we obtain $\hat{\boldsymbol{x}} = \left(A^T A\right)^{-1} A^T \boldsymbol{b}$. The approximation $\hat{\boldsymbol{b}} = A\left(A^T A\right)^{-1} A^T \boldsymbol{b}$ The matrix $P = A\left(A^T A\right)^{-1} A^T$ is a projection matrix, which projects $\boldsymbol{b}$ onto the the column space of A.

2 Linear Operator and Matrix Inverses

One of the problem in linear algebra is to solve a system of equations. In this section, we will discuss the solution of linear equation of the form $A\boldsymbol{x} = \boldsymbol{b}$, to understand when solutions are exist and unique, and to discuss factors that can affect the solutions.

2.1 Four Fundamental Subspaces of a Linear Operator

The first two subspaces is introduced section 1.1. The column space of A, $\mathcal{R}(A)$ and the nullspace of A, $\mathcal{N}(A)$. The other two subspaces are the column space of A^T, $\mathcal{R}(A^T)$ and the nullspace of



$A^T, \mathcal{N}(A^T)$.

The subspaces of the operator $A : X \to Y$ are summarized as follows:

$$\begin{gathered} \mathcal{R}(A) \subset Y, \\ \mathcal{N}(A) \subset X, \\ \mathcal{R}\left(A^T\right) \subset X, \\ \mathcal{N}\left(A^T\right) \subset Y. \end{gathered}$$

And these four fundamental subspaces have the following orthogonality properties.

1. The range is the orthogonal complement of the left nullspace:

$$[\mathcal{R}(A)]^{\perp} = \mathcal{N}\left(A^T\right).$$

2. The orthogonal complement of the row space is the nullspace:

$$\mathcal{R}\left(A^T\right)^{\perp} = \mathcal{N}(A).$$

Definition 2.1 (Rank)**.** The dimension of the column space (or the row space) of the matrix A is the rank of the matrix.

The rank of a matrix A is the number of linearly independent columns or rows of A. For an $m \times n$ matrix of rank r, the following size relationships hold:

1. Column space: $\dim(\mathcal{R}(A)) = r$.
2. Row space: $\dim\left(\mathcal{R}\left(A^T\right)\right) = r$.
3. Nullspace: $\dim(\mathcal{N}(A)) = n - r$.
4. Left nullspace: $\dim\left(\mathcal{N}\left(A^T\right)\right) = m - r$.

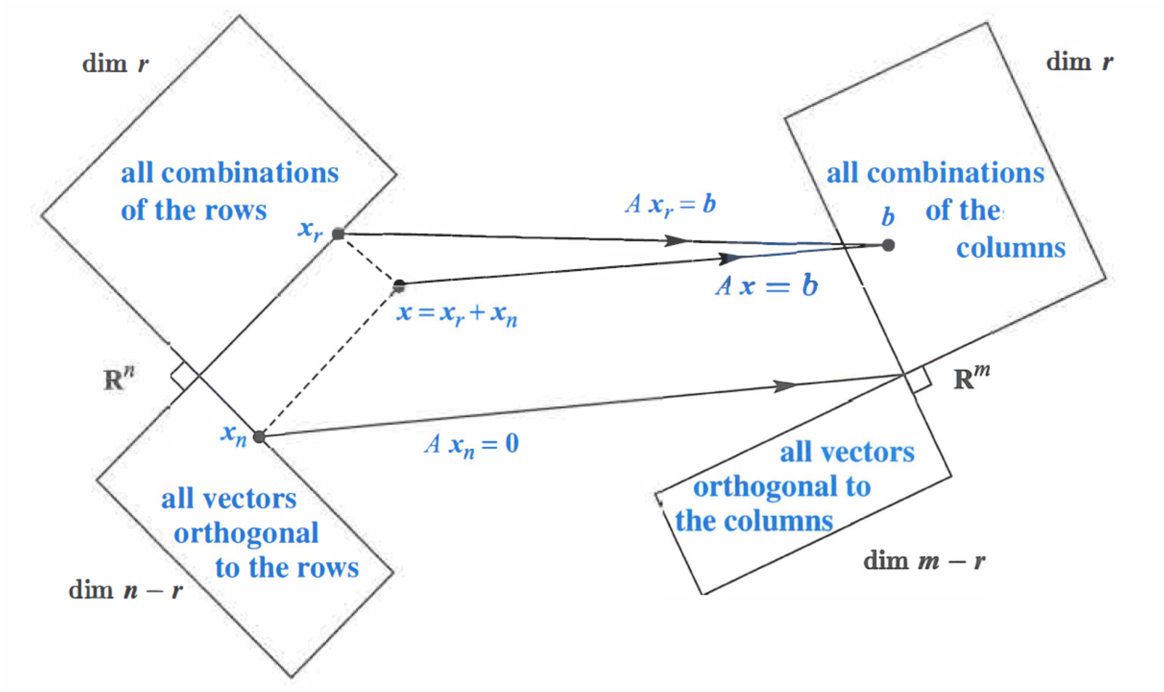


Figure 2: The four fundamental subspaces of a matrix operator

2.2 Inverse

In linear equation, finding an inverse equivalent to finding a solution to $A\boldsymbol{x} = \boldsymbol{b}$. For a full rank square matrix A, A^{-1} exist and $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. If matrix A is not a square matrix, regarding the existence of the solution $A\boldsymbol{x} = \boldsymbol{b}$, left inverse and right inverse can be defined.

Definition 2.2. A matrix A is said to have a left inverse if there is a matrix B such that $BA = I$, and a right inverse if there is a matrix C such that $AC = I$.

If matrix A is a $m \times n$ matrix with full row rank ($rank = m \leq n$), there is an $n \times m$ right inverse C and a solution $\boldsymbol{x} = C\boldsymbol{b}$. Since the column of A must contain a basis of $\mathbb{R}^m$ and $N(A)$ is not empty, it has more than one solutions.

If matrix A is a $m \times n$ matrix with full column rank ($rank = n \leq m$), The equation $A\boldsymbol{x} = \boldsymbol{b}$ has at most one solution for any $\boldsymbol{b}$. The column of A did not contain a basis of $\mathbb{R}^m$. If $\boldsymbol{b}$ is not well chosen, then there is no solution.

2.2.1 Pseudoinverse

The idea of pseudoinverse is shown in Figure 3. The pseudoinverse first projects $\boldsymbol{b}$ onto $\mathcal{R}(A)$, then inverse back to vector $\hat{\boldsymbol{x}}$ in $\mathcal{R}(A^T)$. The projection P insures that the length of error $\boldsymbol{e} = \boldsymbol{b} - A\hat{\boldsymbol{x}}$ is minimized. The inverse $\hat{\boldsymbol{x}} = A^{+}\boldsymbol{p}$ that mapping the $\hat{\boldsymbol{x}}$ in $\mathcal{R}(A^T)$ insures that $\hat{\boldsymbol{x}}$ has the minimum norm.

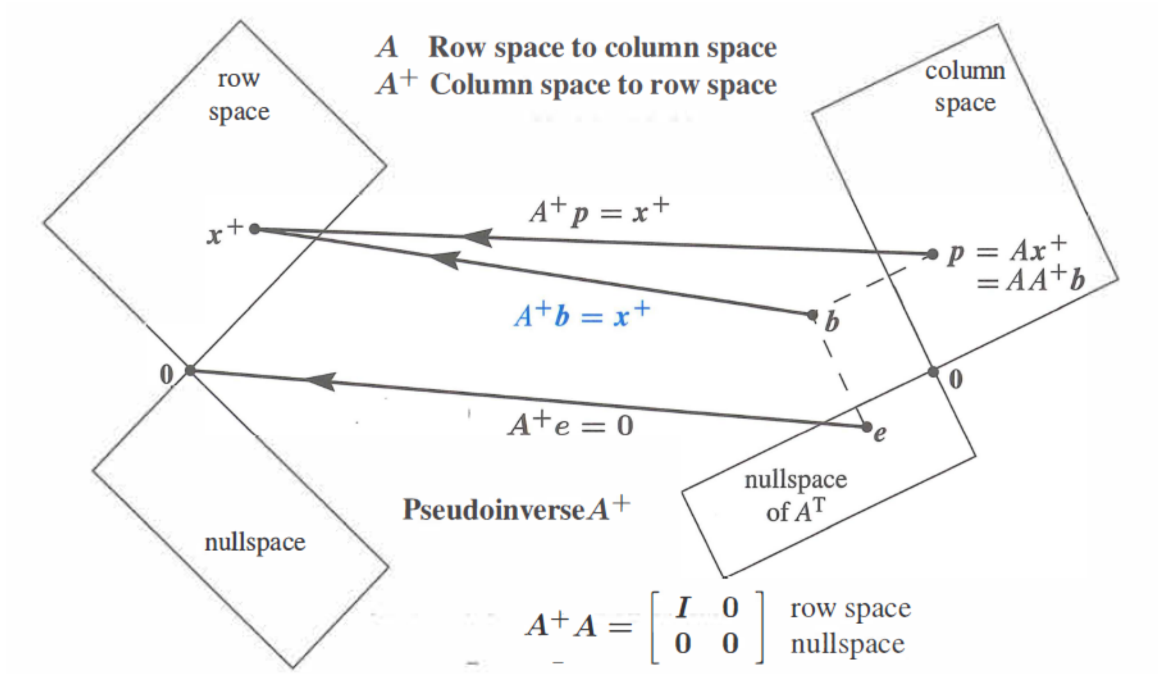


Figure 3: Operation of the pseudoinverse

2.2.2 Pseudoinverse and Least Squares Solution

In section 1.4, we obtain the least square solution $\hat{\boldsymbol{x}} = \left(A^T A\right)^{-1} A^T \boldsymbol{b}$ and it is an example of pseudoinverse. We first project $\boldsymbol{b}$ onto $\mathcal{R}(A)$ and obtain $\boldsymbol{p} = A(A^T A)^{-1} A^T \boldsymbol{b}$. $\boldsymbol{p}$ is in the column space of A, so $(A^T A)^{-1} A^T \boldsymbol{b}$ is in the row space of A.

3 Eigenvalues, Eigenvectors and Matrix Factorization

3.1 Eigenvectors and Eigenvalues

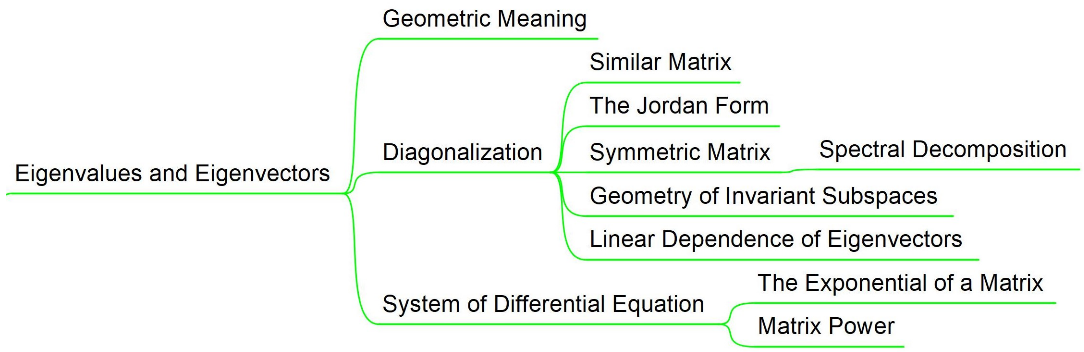
The eigenvectors of A are those vectors that are not changed in direction by the operation of A, in the mean time, the eigenvalue scaled the vector by λ. For example, the projection matrix P have $\lambda = 1$ or $\lambda = 0$, since only the vectors that lie in the nullspace and column space multiply A does not change their direction.

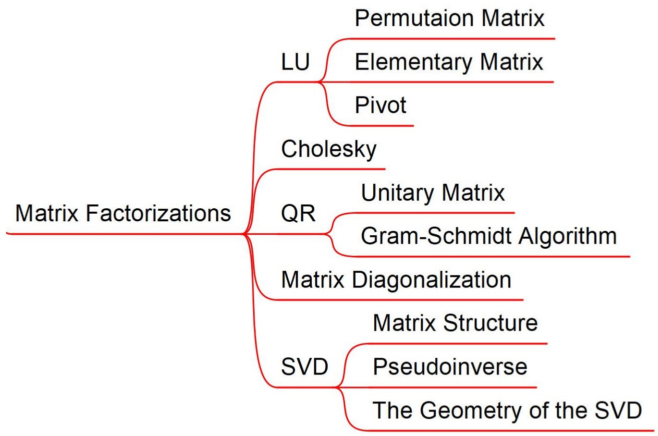
3.2 Determinant and Trace

The product of the n eigenvalues equals the determinant. The sum of the n eigenvalues equals the sum of the n diagonal entries (trace).

3.3 Diagonalization of a Matrix

Suppose the n by n matrix A has n linearly independent eigenvectors $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$. Put them into the columns of an eigenvector matrix X. Then $X^{-1}AX$ is the eigenvalue matrix Λ :





$$\begin{array}{l}\text{Eigenvector matrix } X \\ \text{Eigenvalue matrix } \Lambda\end{array} \qquad X^{-1}AX = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

If the eigenvalues of an n by n matrix A are all distinct, then the eigenvectors of A are linearly independent.

3.4 Solution to Difference and Differential Equation

For first order difference equation $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$, $\boldsymbol{u}_k$ can be obtained by:

$$\boldsymbol{u}_k = A^k \boldsymbol{u}_0 = c_1 \left(\lambda_1\right)^k \boldsymbol{x}_1 + \cdots + c_n \left(\lambda_n\right)^k \boldsymbol{x}_n.$$

The coefficient vector can be calculated as $\boldsymbol{c} = X^{-1}\boldsymbol{u}_0$.

For first order differential equation $\frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u}$, $\boldsymbol{u}_t$ can by obtained by:

$$\boldsymbol{u}(t) = c_1 e^{\lambda_1 t}\boldsymbol{x}_1 + \cdots + c_n e^{\lambda_n t}\boldsymbol{x}_n.$$

The coefficient vector can be calculated as $\boldsymbol{c} = X^{-1}\boldsymbol{u}(0)$.

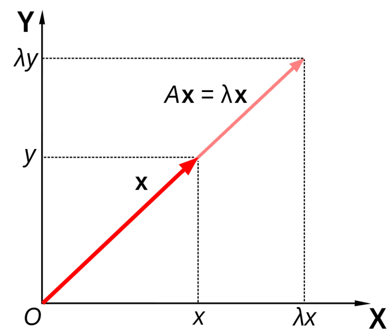


Figure 4: The direction of eigenvectors

For n-th order difference/differential equation, we can still apply these equation by substituting the variable and describe it by a first order equation.

3.5 Symmetric Matrices

- A symmetric matrix S has n real eigenvalues λ_i and n orthonormal eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$.
- Antisymmetric matrices $A = -A^T$ have imaginary λ 's and orthonormal (complex) $\boldsymbol{q}$'s.
- The number of positive eigenvalues of S equals the number of positive pivots.

  Every Hermitian matrix S can be by diagonalized by a unitary matrix

$$S = Q\Lambda Q^{-1} = Q\Lambda Q^T.$$

3.6 Positive Definite Matrices

- Symmetric S : all eigenvalues $> 0 \Leftrightarrow$ all pivots > 0.
- The matrix S is then positive definite. The equation $\boldsymbol{x}^T S\boldsymbol{x} > 0$ for all vectors $\boldsymbol{x} \neq 0$.
- $S = A^T A$ with independent columns in A.
- Positive semidefinite S allows $\lambda = 0$, pivot $= 0$, determinant $= 0$, equation $\boldsymbol{x}^T S\boldsymbol{x} = 0$.

  Rewrite in sum of squares:

$$\boldsymbol{x}^T S\boldsymbol{x} = \boldsymbol{x}^T A^T A\boldsymbol{x},$$

  with $A = L\sqrt{D}$ or $A = Q\sqrt{\Lambda}Q^T$.

3.7 Bases and Matrices in the SVD

As shown in Figure 4, the $\boldsymbol{u}$'s and $\boldsymbol{v}$'s give bases for the four fundamental subspaces:

$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ is an orthonormal basis for the column space.

$\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$ is an orthonormal basis for the left nullspace $\mathcal{N}\left(A^T\right)$.

$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ is an orthonormal basis for the row space.

$\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$ is an orthonormal basis for the nullspace $\mathcal{N}(A)$.

The singular values σ_1 to σ_r will be positive numbers: σ_i is the length of $A\boldsymbol{v}_i$. The σ 's go into a diagonal matrix that is otherwise zero.

$$A\left[\boldsymbol{v}_1 \cdot \cdot \boldsymbol{v}_n\right] = \left[\boldsymbol{u}_1 \cdot \boldsymbol{u}_m\right] \begin{bmatrix} \sigma_1 & & \\ & \sigma_r & \\ & & 0 \end{bmatrix}.$$

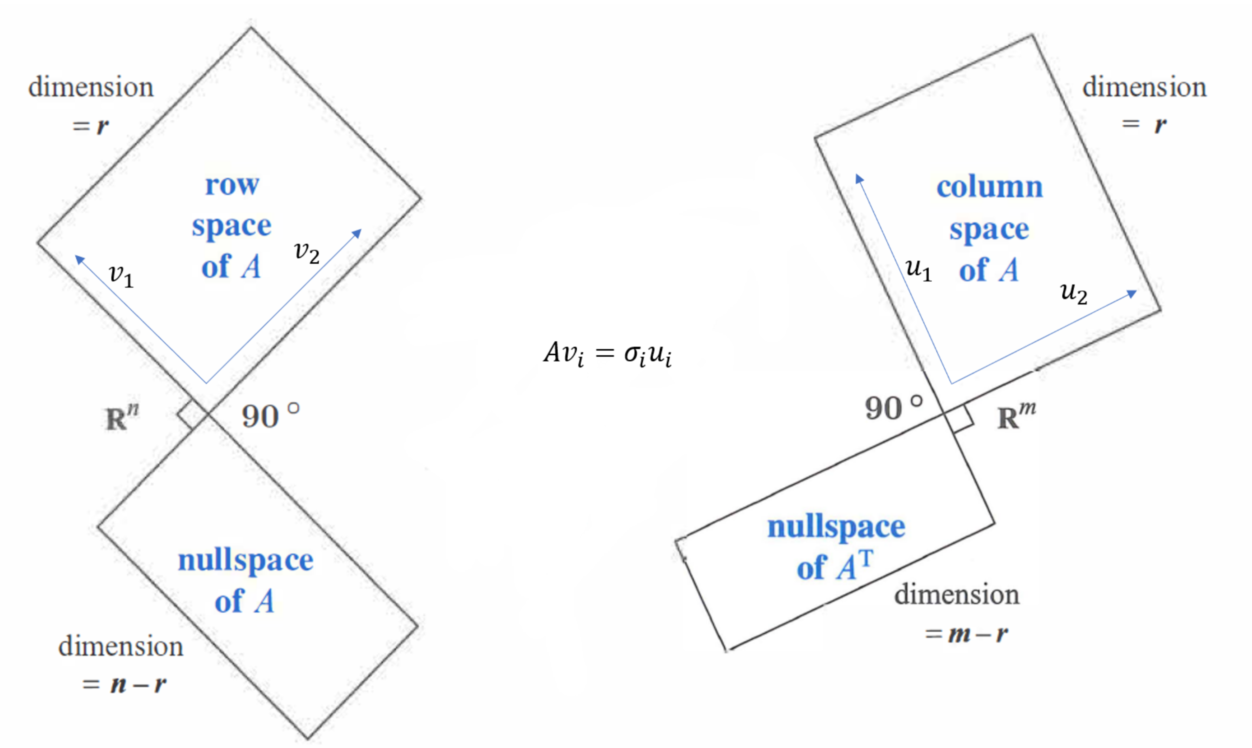


Figure 5: Bases and Matrices in the SVD

$$A = U\Sigma V^T = \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^T + \cdots + \boldsymbol{u}_r\sigma_r\boldsymbol{v}_r^T.$$

3.8 Singular Vectors and Eigenvectors

Symmetric S:

$$S = Q\Lambda Q^T = \lambda_1\boldsymbol{q}_1\boldsymbol{q}_1^T + \lambda_2\boldsymbol{q}_2\boldsymbol{q}_2^T + \cdots + \lambda_r\boldsymbol{q}_r\boldsymbol{q}_r^T.$$

Any matrix A:

$$A = U\Sigma V^T = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^T + \cdots + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^T.$$

The singular vectors $\boldsymbol{v}_i$ are the eigenvectors $\boldsymbol{q}_i$ of $S = A^T A$. The eigenvalues λ_i of S are the same as σ_i^2 for A. The rank r of S equals the rank of A. This is true since

$$A^T A = \left(U\Sigma V^T\right)^T \left(U\Sigma V^T\right) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T.$$

Therefore the vectors $\boldsymbol{u}_i$ and singular values σ can be computed as the eigenvectors and the square root of the eigenvalues of AA^T.

The eigenvalue λ and singular value σ also solves the problem of

$\lambda_1 =$ maximum ratio $\frac{\boldsymbol{x}^T S\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}$. The winning vector is $\boldsymbol{x} = \boldsymbol{q}_1$ with $S\boldsymbol{q}_1 = \lambda_1\boldsymbol{q}_1$. Compare with the largest singular value σ_1 of A. It solves this problem: $\sigma_1 =$ maximum ratio $\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$.

3.9 The Geometry of the SVD

The SVD separates a linear transformation into three steps: (orthogonal) $\times$(diagonal) $\times$ (orthogonal). The geometry behind it can be expressed by: (rotation) $\times$(stretching) $\times$ (rotation). $U\Sigma V^T\boldsymbol{x}$ starts with the rotation to $V^T\boldsymbol{x}$. Then Σ stretches that vector to $\Sigma V^T\boldsymbol{x}$, and U rotates to $A\boldsymbol{x} = U\Sigma V^T\boldsymbol{x}$.

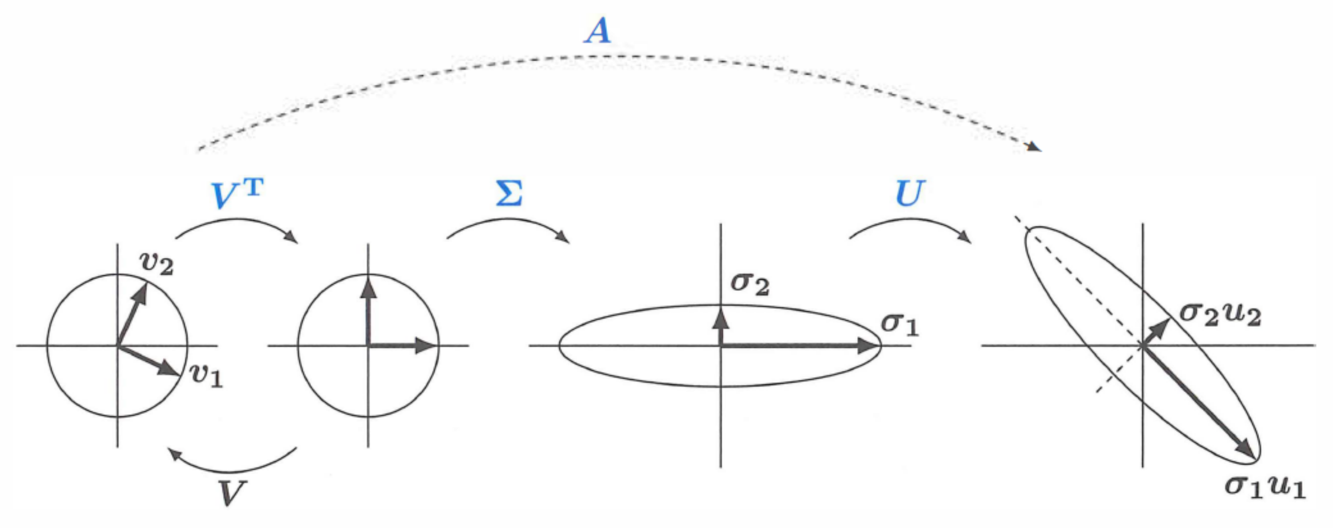


Figure 6: U and V are rotations. Σ : stretches circle to ellipse.

3.10 Polar Decomposition

The polar decomposition separates a $n \times n$ matrix by an orthogonal matrix and a positive definite matrix, $A = QS$.

The polar decomposition separates the rotation (in $Q = UV^T$) from the stretching (in S). The eigenvalues of S give the stretching factors as in Figure 5. The eigenvectors of S give the stretching directions (the principal axes of the ellipse). The orthogonal matrix Q includes both rotations U and V^T.