Univ. of Washington

Lecture : Matrix Decompositions

Winter 2021

Lecturer: Dan Calderone

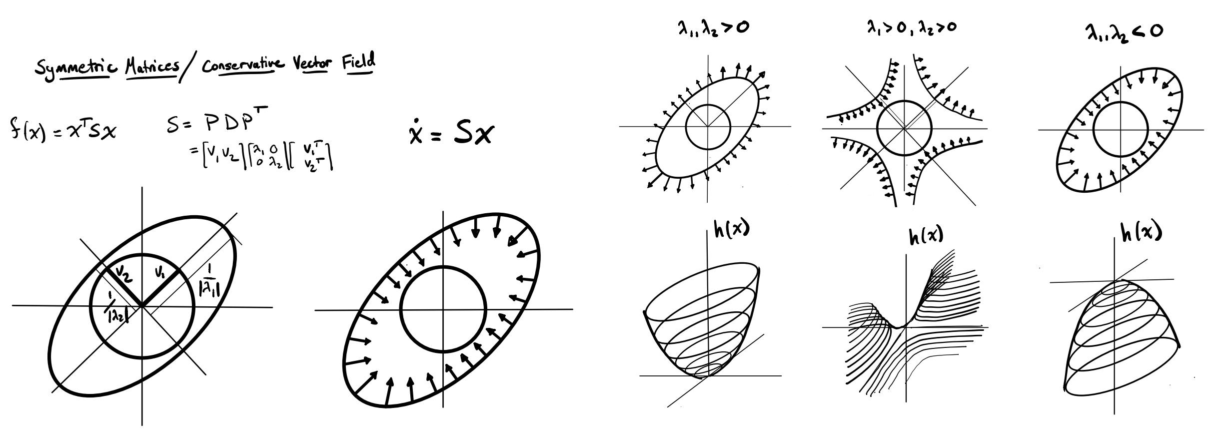
1 Symmetric Matrices

1.1 Symmetric and Hermitian Matrices

A *symmetric* matrix is a real matrix $S \in \mathbb{R}^{n\times n}$ such that $S = S^T$. A Hermitian matrix $H \in \mathbb{C}^{n\times n}$ is a matrix such that $H = H^*$. Symmetric matrices have all real eigenvalues and can be diagonalized by rotation matrices, ie. for every symmetric matrix S, there exists a rotation matrix R and a diagonal matrix of real eigenvalues D such that

$$S = RDR^T = \begin{bmatrix} | & & | \\ r_1 & \cdots & r_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} - & r_1^T & - \\ & \vdots & \\ - & r_n^T & - \end{bmatrix} \tag{1}$$

This means that multiplying by an $n\times n$ symmetric matrix corresponds to stretching in n orthogonal coordinate directions. Symmetric matrices can be thought of as defining level sets of the quadratic form $h(x) = \frac{1}{2}x^T Sx$. The vector field $\dot{x} = Sx$ is then a *gradient field*, $\dot{x} = \frac{\partial h}{\partial x}^T$. Intuitively, $\dot{x} = Sx$ can be thought of as flowing up a surface defined by $h(x)$. This is also called a *conservative* vector field in physics and the function $h(x)$ is typically related to the energy of a system (or some analog). The general condition for a nonlinear vector field $\dot{x} = f(x)$ to be conservative, ie. that $f(x) = \frac{\partial h}{\partial x}^T$ for some $h(x)$ is given by $\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}$ for all i, j. This condition is derived from the fact that if such and $h(x)$ exists, then $\frac{\partial^2 h}{\partial x_i \partial x_j} = \frac{\partial^2 h}{\partial x_j \partial x_i}$.



1.2 Positive Definiteness

We say a symmetric matrix S is *positive definite* if

$$x^T S x > 0, \qquad \text{for all } x \in \mathbb{R}^n \tag{2}$$

Since we could pick x to be any r_i, it follows that for a positive definite matrix, $\lambda_i > 0$ for all i. If the "greater than" signs above are replaced with "greater than or equal" signs, ie. "$>$" is replaced with "$\geq$" then we say the matrix is *positive semi-definite*. Similarly, if we replace the "greater than" signs with "less than" signs we say that the matrix is *negative definite* or *negative semi-definite*.

1.3 Polar Decomposition

For any matrix $A \in \mathbb{R}^{m\times n}$, there are two positive semi-definite matrices $P = (A^T A)^{\frac{1}{2}}$ and $P' = (AA^T)^{\frac{1}{2}}$ (where the positive square root of each eigenvalue is taken) that are closely related to the "shape" of the matrix A. Similarly to how the magnitude of a complex number is defined by $|z| = \sqrt{z^* z}$, we can say that the "magnitude and shape" of A is defined either by $(A^T A)^{\frac{1}{2}}$ or $(AA^T)^{\frac{1}{2}}$. Expanding out $A^T A$, we get

$$\begin{aligned} A^T A &= \begin{bmatrix} - & A_1^T & - \\ & \vdots & \\ - & A_n^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ A_1 & \cdots & A_n \\ | & & | \end{bmatrix} \\ &= \begin{bmatrix} A_1^T A_1 & \cdots & A_1^T A_n \\ \vdots & & \vdots \\ A_n^T A_1 & \cdots & A_n^T A_n \end{bmatrix} = \begin{bmatrix} |A_1||A_1|\cos(\theta_{11}) & \cdots & |A_1||A_n|\cos(\theta_{1n}) \\ \vdots & & \vdots \\ |A_n||A_1|\cos(\theta_{n1}) & \cdots & |A_n||A_n|\cos(\theta_{nn}) \end{bmatrix} \end{aligned}$$

We note that this matrix is fully determined by the size and relative orientation of the columns of A. Another way to say this is that applying an orthonormal transformation to all the columns of A does not change $A^T A$. Indeed $(RA)^T(RA) = A^T R^T R A = A^T A$. Similarly the size and relative orientation of the rows of A full determines AA^T. We can make precise the sense in which A has the same shape as $P = (A^T A)^{\frac{1}{2}}$ by noting that P and A differ by a orthonormal transformation. Indeed,

$$A = \underbrace{A(A^T A)^{-\frac{1}{2}}}_{R} \underbrace{(A^T A)^{\frac{1}{2}}}_{P}$$

We note that we can check that $R^T R = I$

$$\begin{aligned} R^T R &= (A^T A)^{-\frac{1}{2}} A^T A (A^T A)^{-\frac{1}{2}} \\ &= \underbrace{(A^T A)^{-\frac{1}{2}} (A^T A)^{\frac{1}{2}}}_{I} \underbrace{(A^T A)^{\frac{1}{2}} (A^T A)^{-\frac{1}{2}}}_{I} = I \end{aligned}$$

and thus we have that rotating (and possibly reflecting) all the columns of the positive semidefinite matrix $P \succeq 0$ by one rotation gives A. Similarly a complex number $z = |z|e^{i\theta}$ can be created by starting with it's norm $|z| \geq 0$ and rotating it in the complex plane by $e^{i\theta}$. Similarly A can be created from $(AA^T)^{\frac{1}{2}}$ by applying the orthonormal transformation $R' = (AA^T)^{-\frac{1}{2}}A$.

$$A = \underbrace{(AA^T)^{\frac{1}{2}}}_{P'} \underbrace{(AA^T)^{-\frac{1}{2}}A}_{R'}$$

One can check that in fact for a square matrix, $R = R'$.

This leads us to the polar decomposition. A square, invertible $A \in \mathbb{R}^{n \times n}$ can be written in a polar form similar to the polar decomposition of a complex number $z = \sqrt{z^* z} e^{i\theta}$. The

$$A = RP = P'R$$

where

$$\begin{aligned} P &= (A^*A)^{\frac{1}{2}} = V\Sigma V^* \\ P &= (AA^*)^{\frac{1}{2}} = U\Sigma U^* \\ R &= (AA^*)^{\frac{-1}{2}}A = A(A^*A)^{\frac{-1}{2}} = UV^* \end{aligned}$$

Note that $P \succeq 0$ and $P' \succeq 0$. Note also the There are two separate versions of the polar decomposition one with P and one with P'. Note also the connections between the polar decomposition and the singular value decomposition (see below). These relationships can be checked directly.

2 Singular Value Decomposition

The *singular value decomposition* (SVD) provides even more insight beyond the polar decomposition. The SVD is very general and can apply to any matrix $A \in \mathbb{C}^{m \times n}$ even if the matrix is not-invertible, not diagonalizable, or even not square or full-rank. As a result, it is an often used, powerful analysis tool. We will perform the derivation below for $A \in \mathbb{R}^{m \times n}$ but we note that the same derivation works for complex matrices using conjugate transposes.

To construct the SVD, we will assume that A is fat or square and analyze $A^T A$. (A similar construction can be done if A is tall using AA^T.). We first diagonalize $A^T A$ as

$$A^T A = V D V^T$$

where $V \in \mathbb{R}^{n \times n}$ is orthonormal, ie. $V^T V = I$, and $D \succeq 0$ is diagonal, real and positive semi-definite. We know this is possible since $A^T A$ is symmetric and thus has all real eigenvalues and orthogonal eigenvectors. We will also assume that the diagonal of D is arranged in descending order from the largest eigenvalue to the smallest. Since A is fat or square, some portion of the diagonal of D will be 0's. We can take the positive square root of D to get

$$D^{\frac{1}{2}} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

where

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

and the block zeros have the appropriate sizes. $\{\sigma_i\}_{i=1}^k$ are called *singular values* and are the positive square roots of the nonzero eigenvalues of A^TA. We can enumerate V as

$$V = \begin{bmatrix} V^1 & V^2 \end{bmatrix}$$

where

$$V^1 = \begin{bmatrix} | & & | \\ V_1 & \cdots & V_k \\ | & & | \end{bmatrix}, \qquad V^2 = \begin{bmatrix} | & & | \\ V_{k+1} & \cdots & V_n \\ | & & | \end{bmatrix}$$

where the columns of V^1 correspond to positive, non-zero singular values and the columns of V^2 are the eigenvectors for the zero eigenvalues of A^TA. We note that the columns of V^2 can be chosen somewhat arbitrarily as long as they are orthonormal and span $\mathcal{N}(A^TA)$. Since $\mathcal{N}(A^TA) = \mathcal{N}(A)$, $\mathcal{R}(V^2) \in \mathcal{N}(A)$ as well.

We note that any columns corresponding to repeated eigenvalues of A^TA (including repeated zeros), may be arbitrarily chosen (as long as they are orthonormal). We can now define a matrix $U \in \mathbb{R}^{m \times m}$ as

$$U = \begin{bmatrix} U^1 & U^2 \end{bmatrix}$$

where

$$U^1 = \begin{bmatrix} | & & | \\ U_1 & \cdots & U_k \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \frac{AV_1}{\sigma_1} & \cdots & \frac{AV_k}{\sigma_k} \\ | & & | \end{bmatrix} \qquad U^2 = \begin{bmatrix} | & & | \\ U_{k+1} & \cdots & U_m \\ | & & | \end{bmatrix}$$

Note that the columns of U^1 are orthonormal.

$$\frac{V_i^TA^T}{\sigma_i}\frac{AV_i}{\sigma_i} = \frac{\sigma_i^2}{\sigma_i^2}V_i^TV_i = 1, \qquad \frac{V_i^TA^T}{\sigma_i}\frac{AV_j}{\sigma_j} = \frac{\sigma_i^2}{\sigma_j^2}V_i^TV_j = 0$$

The columns of U_2 can be chosen to complete an orthonormal basis for $\mathbb{R}^m$ and thus $U^TU = I$. We can then write

$$U^1\Sigma = AV^1$$

By adding $(0)U^2$ to the left-hand side, and since $V^2 \in \mathcal{N}(A)$, we can write

$$\begin{bmatrix} U^1\Sigma + (0)U^2 & 0 \end{bmatrix} = A \begin{bmatrix} V^1 V^2 \end{bmatrix}$$
$$\begin{bmatrix} U^1 & U^2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} = A \begin{bmatrix} V^1 V^2 \end{bmatrix}$$

right multiplying by V^T gives

$$\begin{aligned} A &= U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T \\ &= \begin{bmatrix} U^1 & U^2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^{1T} \\ V^{2T} \end{bmatrix} \end{aligned}$$

This is the singular value decomposition. Note the following relationships

$$\mathcal{R}(U^1) = \mathcal{R}(A), \qquad \mathcal{R}(V^1) = \mathcal{R}(A^T)$$
$$\mathcal{R}(U^2) = \mathcal{N}(A^T), \qquad \mathcal{R}(V^2) = \mathcal{N}(A)$$

Note also that the columns of U are orthonormal eigenvectors of AA^T with the singular values squared as eigenvalues. Indeed,

$$AA^T U_i = \tfrac{AA^T AV_i}{\sigma_i} = \tfrac{\sigma_i^2}{\sigma_i} AV_i = (\sigma_i)^2 U_i$$

Thus this construction would have worked using AA^T instead of $A^T A$ if desired.

2.1 Symmetric-Skew Symmetric/Helmholtz Decomposition

A square matrix $A \in \mathbb{R}^{n\times n}$ can be decomposed as follows

$$A = \underbrace{\tfrac{1}{2}\Big(A + A^T\Big)}_{S} + \underbrace{\tfrac{1}{2}\Big(A - A^T\Big)}_{K}$$

Note that $S = S^T$ is symmetric and $K = -K^T$ is skew-symmetric. This decomposition says that the space of real matrices is actually the direct sum of the space of symmetric matrices and the space of skew-symmetric matrices. Under the vectorized matrix inner product $\langle \cdot, \cdot \rangle = \mathrm{Tr}\big((\cdot)^T(\cdot)\big)$, we have that

$$\begin{aligned} \langle S, K \rangle = \mathrm{Tr}(S^T K) &= \sum_{i,j} S_{ij} K_{ij} \\ &= \sum_{i<j} S_{ij} K_{ij} + \sum_{i>j} S_{ij} K_{ij} + \sum_{i} S_{ii} K_{ii} \\ &= \sum_{i<j} S_{ij} K_{ij} + \sum_{j>i} S_{ji} K_{ji} + \sum_{i} S_{ii} K_{ii} \\ &= \sum_{i<j} S_{ij}\big(K_{ij} + K_{ji}\big) + \sum_{i} S_{ii} K_{ii} = 0 \end{aligned}$$

If the definition of positive definite is extended to non-symmetric matrices we have that

$$x^T Ax = x^T Sx + x^T Kx = x^T Sx$$

and thus A is positive definite if and only if $S = \frac{1}{2}(A + A^T)$ is positive definite. In the context of vector fields, this means that any linear vector field $\dot{x} = Ax$ can be decomposed into a conservative piece and a rotational piece.

$$\dot{x} = Ax = \underbrace{Sx}_{\text{conservative}} + \underbrace{Kx}_{\text{rotational}}$$

which is a special application of the Helmholtz decomposition to linear vector fields.

3 Matrix vs. Complex Number Analogies

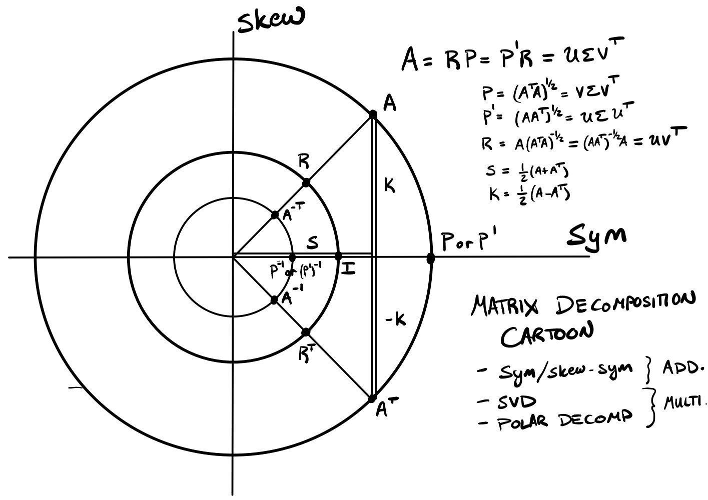
The polar decomposition and the symmetric/skew-symmetric decomposition provide several deep analogies between matrices and complex numbers. Symmetric matrices act a lot like real numbers and positive definite symmetric matrices act like positive real numbers. Among other things, they have real and positive real eigenvalues respectively. Skew-symmetric matrices act a lot like purely imaginary numbers, encoding rotational flow and having purely imaginary eigenvalues. We can even draw a "complex plane" of sorts for matrices with symmetric matrices as the real axis and skew-symmetric matrices as the vertical axis. As shown above, the fact that these two subspaces are orthogonal is accurate. The Cartesian description of complex numbers is analogous to the symmetric-skew symmetric decomposition and the polar description of complex numbers is analogous to the polar decomposition.

$$\begin{aligned} z = a + bi, \qquad &\Rightarrow \qquad A = S + K \\ z = |z|e^{i\theta} \qquad &\Rightarrow \qquad A = RP = P'R \end{aligned}$$

This analogy extends in the following ways detailed in the diagram. For complex numbers and square invertible matrices and

$$z = a + bi = |z|e^{i\theta}, \qquad\qquad A = U\Sigma V^T$$

- $z \Rightarrow A$, $z^* \Rightarrow A^T$, $z^{-1} \Rightarrow A^{-1}$, $z^{-*} \Rightarrow A^{-T}$
- $a \Rightarrow S = \frac{1}{2}(A + A^T)$, $bi \Rightarrow K = \frac{1}{2}(A - A^T)$.
- $|z| = (z^* z)^{\frac{1}{2}} \Rightarrow P = (A^T A)^{\frac{1}{2}} = V\Sigma V^T, P' = (AA^T)^{\frac{1}{2}} = U\Sigma U^T,$
- $e^{i\theta} \Rightarrow R = A(A^T A)^{-\frac{1}{2}} = (AA^T)^{-\frac{1}{2}} A = UV^T$, $e^{-i\theta} \Rightarrow R^T$,


Skew
Sym
A = RP = P'R = UΣV^T
P = (A^T A)^½ = VΣV^T
P' = (AA^T)^½ = UΣU^T
R = A(A^T A)^-½ = (AA^T)^-½ A = UV^T
S = ½(A+A^T)
K = ½(A−A^T)
A
R
A^-T
S
P or P'
P^-1 or (P')^-1
I
K
-K
A^-1
R^T
A^T
MATRIX DECOMPOSITION
CARTOON
- Sym/skew-sym } ADD.
- SVD
- POLAR DECOMP } MULTI.