C: Fundamental matrix factorizations and the pseudo-inverse

C.1 Fundamental matrix factorizations

In this section we discuss the fundamental matrix factorizations known as the singular value decomposition and the eigenvalue decomposition of square symmetric matrices, and end by describing the so-called pseudo-inverse of a matrix. Mathematical proofs showing the existence of these factorizations can be found in any linear algebra textbook.

C.1.1 The singular value decomposition

The singular value decomposition (SVD) is a fundamental factorization of matrices that arises in a variety of contexts: from calculating the inverse of a matrix and the solution to the Least Squares problem, to a natural encoding of matrix rank. In this section we review the SVD, focusing especially on the motivation for its existence. This motivation for the SVD is to understand, in the simplest possible terms, how a given $M \times N$ matrix $\mathbf{A}$ acts on N-dimensional vectors $\mathbf{w}$ via the multiplication $\mathbf{A}\mathbf{w}$. We refer to this as *parsimonious representation* or, in other words, the drive to represent $\mathbf{A}\mathbf{w}$ in the simplest way possible. For ease of exposition we will assume that the matrix $\mathbf{A}$ has at least as many rows as it has columns, i.e., $N \leq M$, but what follows generalizes easily to the case when $N > M$.

Through the product $\mathbf{A}\mathbf{w} = \mathbf{y}$ the matrix $\mathbf{A}$ sends the vector $\mathbf{w} \in \mathbb{R}^N$ to $\mathbf{y} \in \mathbb{R}^M$. Using any two sets of linearly independent vectors which span $\mathbb{R}^N$ and $\mathbb{R}^M$, denoted as $\mathbb{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ and $\mathbb{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M\}$ respectively, we can decompose an arbitrary $\mathbf{w}$ over $\mathbb{V}$ as

$$\mathbf{w} = \sum_{n=1}^{N} \alpha_n \mathbf{v}_n, \tag{C.1}$$

for some coefficients α_n for $n = 1 \ldots N$. Further, since for each n the product $\mathbf{A}\mathbf{v}_n$ is some vector in $\mathbb{R}^M$, each product itself can be decomposed over $\mathbb{U}$ as

$$\mathbf{A}\mathbf{v}_n = \sum_{m=1}^{M} \beta_{n,m} \mathbf{u}_m, \tag{C.2}$$

for some coefficients $\beta_{n,m}$ for $m = 1 \ldots M$. Together these two facts allow us to decompose the action of $\mathbf{A}$ on an arbitrary vector $\mathbf{w}$ in terms of how $\mathbf{A}$ acts on the individual $\mathbf{v}_n$s as

$$\mathbf{A}\mathbf{w} = \mathbf{A}\left(\sum_{n=1}^{N}\alpha_n\mathbf{v}_n\right) = \sum_{n=1}^{N}\alpha_n\mathbf{A}\mathbf{v}_n = \sum_{n=1}^{N}\sum_{m=1}^{M}\alpha_n\beta_{n,m}\mathbf{u}_m. \tag{C.3}$$

This representation would be much simpler if $\mathbb{U}$ and $\mathbb{V}$ were such that $\mathbf{A}$ acted on each $\mathbf{v}_n$ via direct proportion, sending it to a weighted version of one of the $\mathbf{u}_m$s. In other words, if $\mathbb{U}$ and $\mathbb{V}$ existed such that

$$\mathbf{A}\mathbf{v}_n = s_n\mathbf{u}_n \quad \text{for all } n, \tag{C.4}$$

this would considerably simplify the expression for $\mathbf{A}\mathbf{w}$ in (C.3), giving instead

$$\mathbf{A}\mathbf{w} = \sum_{n=1}^{N}\alpha_n s_n\mathbf{u}_n. \tag{C.5}$$

If such a pair of bases for $\mathbf{A}$ indeed exists, (C.4) can be written equivalently in matrix form as

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{S}, \tag{C.6}$$

where $\mathbf{V}$ and $\mathbf{U}$ are $N \times N$ and $M \times M$ matrices formed by concatenating the respective basis vectors column-wise, and $\mathbf{S}$ is an $M \times N$ matrix with the s_i values on its diagonal (and zero elsewhere). That is,

$$\mathbf{A}\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_M \\ | & | & & | \end{bmatrix}\begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_N \\ & & & \\ & & & \end{bmatrix}. \tag{C.7}$$

If, in addition, the basis matrices were *orthogonal*,[1] we can rearrange (C.6) for $\mathbf{A}$ alone giving the factorization

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T. \tag{C.8}$$

This ideal factorization can in fact be shown to hold rigorously (see e.g., [81]) and is referred to as the singular value decomposition of $\mathbf{A}$. The matrices $\mathbf{U}$ and $\mathbf{V}$ each have orthonormal columns (meaning the columns of $\mathbf{U}$ all have unit length and are orthogonal to each other, and likewise for $\mathbf{V}$) and are typically referred to as *left* and *right singular matrices* of $\mathbf{A}$, with the *real nonnegative values* along the diagonal of $\mathbf{S}$ referred to as *singular values*.

> Any matrix $\mathbf{A}$ may be factorized as $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ where $\mathbf{U}$ and $\mathbf{V}$ have orthonormal columns and $\mathbf{S}$ is a diagonal matrix containing the (real and nonnegative) singular values of $\mathbf{A}$ along its diagonal.

[1] A square matrix $\mathbf{Q}$ is called orthogonal if $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$.

The SVD can also be written equivalently as a weighted sum of outer product matrices

$$\mathbf{A} = \sum_{i=1}^{\min(N,M)} \mathbf{u}_i s_i \mathbf{v}_i^T. \tag{C.9}$$

Note that we can (and do) assume that the singular values are placed in descending order along the diagonal of $\mathbf{S}$.

C.1.2 Eigenvalue decomposition

When $\mathbf{A}$ is square and symmetric, i.e., when $N = M$ and $\mathbf{A} = \mathbf{A}^T$, there is an additional factorization given by

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T, \tag{C.10}$$

where $\mathbf{E}$ is an $N \times N$ matrix with orthonormal columns referred to as *eigenvectors*, and $\mathbf{D}$ is a diagonal matrix whose diagonal elements *are always real numbers* and are referred to as *eigenvalues*.

> A square symmetric matrix $\mathbf{A}$ may be factorized as $\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T$ where $\mathbf{E}$ is an orthogonal matrix of eigenvectors and $\mathbf{D}$ a diagonal matrix of all real eigenvalues.

We may also write this spectral decomposition equivalently as a sum of N weighted outer product matrices:

$$\mathbf{A} = \sum_{i=1}^{N} d_i \mathbf{e}_i \mathbf{e}_i^T. \tag{C.11}$$

This factorization can be motivated analogously to the SVD in the case of square symmetric $\mathbf{A}$, and is therefore highly related to $\mathbf{A}$'s SVD (for a proof of this fact, commonly referred to as the spectral theorem of symmetric matrices, see e.g., [81]). Specifically, when $\mathbf{A}$ is additionally positive (semi) definite one can show that this factorization is precisely the SVD of $\mathbf{A}$.

Note also that a symmetric matrix is invertible if and only if it has all nonzero eigenvalues. In this case the inverse of $\mathbf{A}$, denoted as $\mathbf{A}^{-1}$, can be written as $\mathbf{A}^{-1} = \mathbf{E}^T\mathbf{D}^{-1}\mathbf{E}$ where $\mathbf{D}^{-1}$ is a diagonal matrix containing the reciprocal of the eigenvalues in $\mathbf{D}$ along its diagonal.

> A square symmetric matrix $\mathbf{A}$ is invertible if and only if it has all nonzero eigenvalues.

C.1.3 The pseudo-inverse

Here we describe the so-called pseudo-inverse solution to the linear system of equations

$$\mathbf{A}\mathbf{w} = \mathbf{b}, \tag{C.12}$$

where $\mathbf{w}$ is an $N \times 1$ vector, $\mathbf{A}$ is an $M \times N$, and $\mathbf{b}$ an $M \times 1$ vector, and where we assume the system has at least one solution. By taking the SVD of $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, and removing all columns of $\mathbf{U}$ and $\mathbf{V}$ associated to any zero singular values, we may then write a solution to this system using the fact that the columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal,

$$\mathbf{w} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T\mathbf{b}. \tag{C.13}$$

Note that since $\mathbf{S}$ is a diagonal matrix the matrix $\mathbf{S}^{-1}$ is also diagonal, containing the reciprocal of the nonzero singular values along its diagonal. The matrix $\mathbf{A}^\dagger = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T$ is referred to as the pseudo-inverse of the matrix $\mathbf{A}$, and we generally write the solution above as

$$\mathbf{w} = \mathbf{A}^\dagger\mathbf{b}. \tag{C.14}$$

When $\mathbf{A}$ is square and invertible the pseudo-inverse equals the matrix inverse itself, i.e., $\mathbf{A}^\dagger = \mathbf{A}^{-1}$. Otherwise, if there are infinitely many solutions to the system $\mathbf{A}\mathbf{w} = \mathbf{b}$ then the pseudo-inverse solution provides the *smallest* solution to this system.

> The smallest solution to the system $\mathbf{A}\mathbf{w} = \mathbf{b}$ is given by $\mathbf{w} = \mathbf{A}^\dagger\mathbf{b}$ where $\mathbf{A}^\dagger$ is the pseudo-inverse of $\mathbf{A}$.