Mathematical Tools for Neuroscience (NEU 314)
Princeton University, Fall 2021
Jonathan Pillow

Lecture 8-10 notes:
SVD and its applications

1 Singular Value Decomposition

The singular vector decomposition allows us to write *any* matrix A as

$$A = USV^\top,$$

where U and V are orthogonal matrices (square matrices whose columns form an orthonormal basis), and S is a diagonal matrix (a matrix whose only non-zero entries lie along the diagonal):

$$S = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_n \end{bmatrix}$$

The columns of U and V are called the *left singular vectors* and *right singular vectors*, respectively.

The diagonal entires $\{s_i\}$ are called *singular values*. The singular values are always ≥ 0.

The SVD tells us that we can think of the action of A upon any vector $\vec{x}$ in terms of three steps (Fig. 1):

1. rotation (multiplication by $V^\top$, which doesn't change vector length of $\vec{x}$).
2. stretching along the cardinal axes (where the $i'th$ component is stretched by s_i).
3. another rotation (multipication by U).

2 Inverses

The SVD makes it easy to compute (and understand) the inverse of a matrix. We exploit the fact that U and V are orthogonal, meaning their transposes are their inverses, i.e., $U^\top U = UU^\top = I$ and $V^\top V = VV^\top = I$.

$$\vec{y} = A\vec{x} = (USV^\top)\,\vec{x}$$

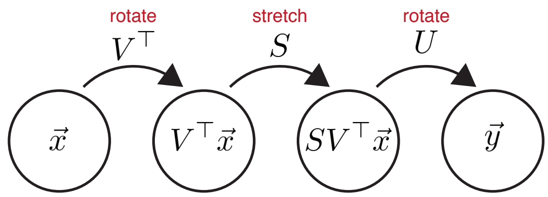


Figure 1: Schematic illustration of SVD in terms of three linear transformations.

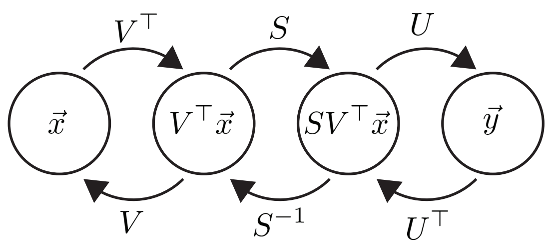
The inverse of A (if it exists) can be determined easily from the SVD, namely:

$$A^{-1} = VS^{-1}U^T, \tag{1}$$

where

$$S^{-1} = \begin{bmatrix} \frac{1}{s_1} & & & \\ & \frac{1}{s_2} & & \\ & & \ddots & \\ & & & \frac{1}{s_n} \end{bmatrix} \tag{2}$$

The logic is that we can find the inverse mapping by *undoing* each of the three operations we did when multiplying A: first, undo the last rotation by multiplying by $U^\top$; second, un-stretch by multiplying by $1/s_i$ along each axis, third, un-rotate by multiplying by V. (See Fig 2).



$$\vec{x} = A^{-1}\vec{y} = (VS^{-1}U^\top)\,\vec{y}$$

Figure 2: Illustrating the inverse of a matrix in terms of its SVD.

Another way to see that this definition of the inverse is correct is via:

$$\begin{aligned} A^{-1}A &= (VS^{-1}U^\top)(USV^\top) \\ &= VS^{-1}(U^\top U)SV^\top \\ &= V(S^{-1}S)V^\top \\ &= VV^\top \\ &= I \end{aligned}$$

We can do a similar analysis of AA^{-1}.

3 Pseudo-inverse

The SVD also makes it easy to see when the inverse of a matrix *doesn't* exist. Namely, if any of the singular values $s_i = 0$, then the S^{-1} doesn't exist, because the corresponding diagonal entry would be $1/s_i = 1/0$.

In other words, if a matrix A has any zero singular values (let's say $s_j = 0$), then multiplying by A effectively destroys information because it takes the component of the vector along the right singular vector $\vec{v}_j$ and multiplies it by zero. We can't recover this information, so there's no way to "invert" the mapping $A\vec{x}$ to recover the original $\vec{x}$ that came in. The best we can do is to recover the components of $\vec{x}$ that weren't destroyed via multiplication with zero.

The matrix that recovers all recoverable information is called the *pseudo-inverse*, and is often denoted $A^\dagger$. We can obtain the pseudoinverse from the SVD by inverting all singular values that are non-zero, and leaving all zero singular values at zero.

Suppose we have an $n \times n$ matrix A, which has only k non-zero singular values. Then the S matrix obtained from SVD will be

$$S = \begin{bmatrix} s_1 & & & & & \\ & \ddots & & & & \\ & & s_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}.$$

The pseudoinverse of A can then be written similarly to the inverse:

$$A^\dagger = V S^\dagger U^\top,$$

where

$$S^\dagger = \begin{bmatrix} \frac{1}{s_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{s_k} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}.$$

4 Condition number

In practical situations, a matrix may have singular values that are not exactly equal to zero, but are so close to zero that it is not possible to accurately compute them. In such cases, the matrix is what we call *ill-conditioned*, because dividing by the singular values $(1/s_i)$ for singular values s_i

that are arbitrarily close to zero will result in numerical errors. Such matrices are theoretically but not practically invertible. (If you try to invert such a matrix, you likely (hopefully) get a warning like: "Matrix is close to singular").

The degree to which ill-conditioning prevents a matrix from being inverted accurately depends on the ratio of its largest to smallest singular value, a quantity known as the **condition number**:

$$\text{condition number} = \frac{s_1}{s_n}.$$

The larger the condition number, the more practically non-invertible it is. When using double floating point precision, matrices with condition numbers greater than $\approx 10^{14}$ cannot be stably inverted.

You can compute the condition number yourself from the SVD, or using the built-in Matlab command `cond`, or the numpy command `numpy.linalg.cond`.

5 SVD of non-square matrix

If $A_{m\times n}$ is a non-square matrix, then U is $m \times m$ and V is $n \times n$, and $S_{m\times n}$ is non-square (and therefore has only $\min(\mathrm{m}, \mathrm{n})$ non-zero singular values. Such matrices are (obviously) non-invertible, though we can compute their pseudo-inverses using the formula above.

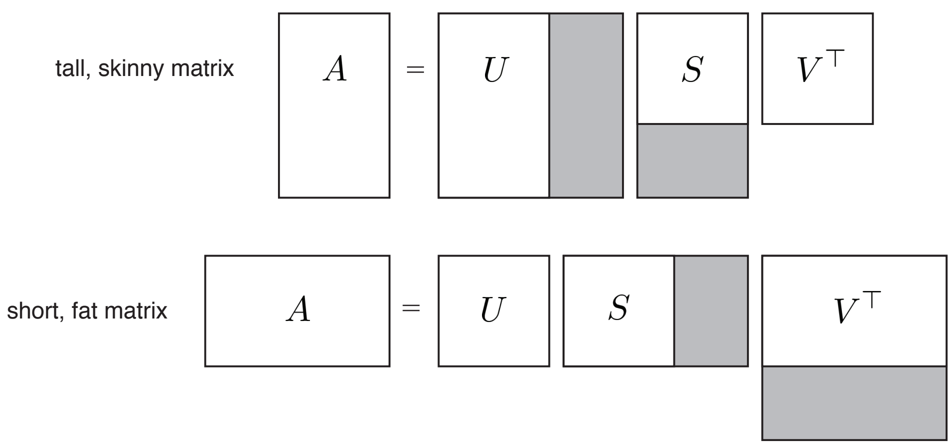


Figure 3: SVD of non-square matrices. The gray regions of the matrices are not needed, since they consist of all zeros in the S matrix, and are only hit by zeros in those portions of the U or $V^\top$ matrices. Dropping them results in the more compact "reduced" SVD for tall, skinny (above) or short, fat (below) matrices.

6 Matrix rank and bases for column / row / null space

Recall that the *rank* of a matrix is equal to: (i) its number of linearly independent columns; (ii) its number of linearly independent rows. (Remarkably, these are always the same!). The rank can be thought of as the *dimensionality* of the vector space spanned by its rows or its columns. Here we are ready to add another definition: the rank of a matrix is equal to the number of non-zero singular values it has!

Consider the SVD of a matrix A that has rank k:

$$A = USV^\top$$

Column space: Since A is rank k, the first k left singular vectors, $\{\vec{u}_1, \ldots \vec{u}_k\}$ (the columns of U), provide an orthonormal basis for the column space of A.

Row space: Similarly, the first k right singular vectors, $\{\vec{v}_1, \ldots \vec{v}_k\}$ (the columns of V, or the rows of $V^\top$), provide an orthonormal basis for the row space of A.

Null space: The last right singular vectors, $\{\vec{v}_{k+1}, \ldots \vec{v}_n\}$ (the last columns of V, or the last rows of $V^\top$), provide an orthonormal basis for the null space of A.

Proof: Let's prove this last one, just to see what such a proof looks like. First, consider a vector $\vec{x}$ that can be expressed as a linear combination of the last $n - k$ columns of V:

$$\vec{x} = \sum_{i=k+1}^{n} w_i \vec{v}_i,$$

for some real-valued weights $\{w_i\}$. To show that $\vec{x}$ lives in the null space of A, we need to show that $A\vec{x} = 0$. Let's go ahead and do this now. (It isn't that hard, and this gives the flavor of what a lot of proofs in linear algebra look like)

$$A\vec{x} = A\left(\sum_{i=k+1}^{n} w_i \vec{v}_i\right) \qquad \text{(by definition of } \vec{x}\text{)} \tag{3}$$

$$= \sum_{i=k+1}^{n} w_i \left(A\vec{v}_i\right). \qquad \text{(by definition of linearity)} \tag{4}$$

Now let's look at any one of the terms in this sum:

$$A\vec{v}_i = (USV^\top)\vec{v}_i \;=\; US(V^\top \vec{v}_i) = US\,\vec{e}_i, \tag{5}$$

where $\vec{e}_i$ is the "identity" basis vector consisting of all 0's except for a single 1 in the i'th row. This follows from the fact that $\vec{v}_i$ is orthogonal to every row of $V^\top$ except the i'th row, which gives $\vec{v}_i \cdot \vec{v}_i = 1$ because $\vec{v}_i$ is a unit vector.

Now, because i in the sum only ranged over $k+1$ to n, then when we multiply $\vec{e}_i$ by S (which has non-zeros along the diagonal only up to the k'th row / column), we get zero:

$$S\vec{e}_i = 0 \qquad \text{for } i > k.$$

Thus

$$US\vec{e}_i = 0$$

which means that the entire sum

$$\sum_{i=k+1}^{n} US\vec{e}_i = 0.$$

So this shows that $A\vec{x} = 0$ for *any* vector $\vec{x}$ that lives in the subspace spanned by the last $n-k$ columns of V, meaning it lies in the null space. This is of course equivalent to showing that the last $n-k$ columns of V provide an (orthonormal) basis for the null space!

7 Relationship between SVD and eigenvector decomposition

Definition: An *eigenvector* of a square matrix A is defined as a vector $\vec{x}$ satisfying the equation

$$A\vec{x} = \lambda\vec{x},$$

and λ is the corresponding *eigenvalue*. In other words, an eigenvector of A is any vector that, when multiplied by A, comes back as a scaled version of itself.

Definition: A *positive semi-definite (PSD)* matrix is a matrix that has all eignevalues ≥ 0, or equivalently, a matrix A for which $\vec{x}^\top A\vec{x} \geq 0$ for any vector $\vec{x}$.

To *generate* an $n \times n$ positive semi-definite matrix, we can take any matrix X that has n columns and let $A = X^\top X$.

Spectral theorem: If a matrix A is symmetric and positive semi-definite, then the SVD also an eigendecomposition, that is, a decomposition in terms of an orthonormal basis of eigenvectors:

$$A = USU^\top,$$

where the columns of U are eigenvectors and the diagonal entries $\{s_i\}$ of S are the eigenvalues. Note that for such matrices, $U = V$, meaning the left and right singular vectors are identical.

Exercise: prove to yourself that: $A\vec{u}_i = s_i\vec{u}_i$

SVD of matrix times its transpose. In class we showed that if $A = USV^\top$, then $A^\top A$ (which it turns out, is symmetric and PSD) has the singular value decomposition (which is also an eigendecomposition): $A^\top A = VS^2V^\top$. Test yourself by deriving the SVD of $AA^\top$.

8 Frobenius norm

The Frobenius norm of a matrix is the equivalent of the Euclidean norm for vectors. It is equal to the square root of the sum of the squared elements of the matrix. Thus, for an $m \times n$ matrix A, the Frobenius norm is:

$$||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}, \tag{6}$$

where a_{ij} is the i,j'th entry of A. This is the same value we would get if we re-shaped A into a vector and evaluted its vector norm!

An interesting property of the Frobenius norm is that it is also equal to the square root of the sum of the squared singular values:

$$||A||_F = \sqrt{\sum_{i=1}^{r} s_i^2}, \tag{7}$$

where we have assumed r is the rank of A and $s_1, \ldots, s_r$ are its singular values.

A simple proof of this fact follows from the fact that we also can write the Frobenious norm of a matrix A as $\sqrt{\mathrm{Tr}[A^\top A]}$, where $\mathrm{Tr}[\cdot]$ denotes the matrix *trace* function, which corresponds to summing up the diagonal elements of a matrix. The trace has the property of being invariant to circular shifts: $\mathrm{Tr}[ABC] = \mathrm{Tr}[CAB] = \mathrm{Tr}[BCA]$ for any matrices A, B, and C.

Combining these facts gives us:

$$||A||_F^2 = \mathrm{Tr}[A^\top A] = \mathrm{Tr}[VSU^\top USV^\top] = \mathrm{Tr}[S^2] = \sum s_i^2. \tag{8}$$

9 Low-rank approximations to a matrix using SVD

Note that an equivalent way to write the SVD is as a sum of rank-1 matrices, each given by the outer product of the left singular vector and correspnoding right singular vector, weighted by singular value:

$$A = USV^\top = s_1 u_1 v_1^\top + \cdots + s_n u_n v_n^\top \tag{9}$$

To get the best low-rank approximation to a matrix A (in a mean-squared error sense), we simply truncate this expansion after k terms:

$$A^{(k)} = s_1 u_1 v_1^\top + \cdots + s_k u_k v_k^\top. \tag{10}$$

Thus, for example, the best rank-1 approximation to A (also known as a *separable* approximation) is given by

$$A \approx s_1 u_1 v_1^\top \tag{11}$$

10 Fraction of variance accounted for

The variance accounted by a rank-k approximation to a rank-n matrix A, computed as above using SVD, is given by:

$$\text{Fraction of variance accounted for} = \frac{\sum_{i=1}^{k} s_i^2}{\sum_{j=1}^{n} s_n^2} \tag{12}$$

Note that the denominator in the right-hand expression is equal the squared Frobenius norm of A, and the numerator is the squared Frobenius norm of its rank-k approximation.