

# ECE 586: Vector Space Methods

## Lecture 20: Frequently Asked Questions

Henry D. Pfister  
Duke University

November 15, 2022

### 1 Singular Value Decomposition

**Are there cases when eigenvalues are negative?** Yes, this occurs when a matrix reverses the direction of an eigenvector. For example, the every vector is an eigenvector of the identity  $I$  with eigenvalue 1. Likewise, every vector is an eigenvector of the negative identity  $-I$  with eigenvalue -1.

**Why is it useful to have k-truncated SVD expansion?** For data matrices where the columns are data vectors, it often assumed that the larger singular values can be associated with signal while the smaller singular values are associated with noise. If that is true, we can increase the SNR by truncating the SVD and keeping only the top- $k$ .

**Why do we need to diagonalize matrices?** If a matrix is diagonal some basis, then it's typically quite easy to solve problems associated with that matrix in that basis. One example is taking powers of a matrix  $A$  given its diagonalization  $A = V\Lambda V^{-1}$ , which is easy because

$$A^m = (V\Lambda V^{-1}) \cdots (V\Lambda V^{-1}) = V\Lambda^m V^{-1}.$$

Similarly, solving a system of linear ordinary differential equations (e.g.,  $n$  equations in  $n$  variables) is easy via diagonalization because the answer separates into  $n$  univariate problems.

**What is the most important property of singular value decomposition? I am wondering why we need to do the SVD?** SVD gives the optimal approximation (in a Frobenius norm sense) of a rank- $r$  linear transform by a rank- $k$  (with  $k < r$ ). This can be used for noise reduction and/or complexity reduction for questions related to the matrix.

**It looks like the compact SVD is a necessary generalization for matrices whose rank is smaller than their dimensions. Is this a correct statement? Because in practice, I don't think I've seen the need for a compact SVD vs the SVD.** I would say no, it is not necessary for rank deficient matrices. Any problem that be solved using the compact SVD can also be solved using the full SVD. But, many problems have simpler (both computationally and descriptively) answers using the compact SVD.

**What is the relation between the SVD and the fundamental subspaces? I am having some trouble conceptualizing them together.** The SVD  $A = U\Sigma V^H$  computes orthonormal bases for all 4 subspaces and they can be extracted from the  $U$  and  $V$  matrices. In particular, columns of  $U$  associated with non-zero singular values form a basis for  $\mathcal{R}(A)$  and the remaining columns for a basis for  $\mathcal{N}(A^H)$ . Similarly, columns of  $V$  associated with non-zero singular values form a basis for  $\mathcal{R}(A^H)$  and the remaining columns for a basis for  $\mathcal{N}(A)$ .

**Do all matrices have eigenvalues and eigenvectors?** Over  $\mathbb{C}$  (i.e., an algebraically closed field), every finite-dimensional matrix has at least one eigenvector. But, the number of eigenvectors may not equal the size of the matrix. In general, this is a rather complicated topic and a thorough discussion can be found in Chapters 5 and 8 of LADR.

**How efficiently can we determine the SVD for a matrix currently? Are current techniques for finding it more or less "good enough", or does determining this still represent a not insignificant bottleneck?** For dense unstructured matrices, the standard computational complexity of computing the full SVD is  $O(mn \min(m, n))$ . This may be significantly reduced in a number of scenarios including combinations of:

- the matrix is sparse or has some specific structure,
- only the top- $k$  singular values / vectors are needed, or
- only rough approximations are required (e.g., randomized algorithms can be used).

**Is a positive definite matrix required to be symmetric (or Hermitian)? Or does it only need to satisfy  $x^T Ax > 0$  for all  $x$ ?** In general, the term positive definite also implies symmetry but in very rare cases some authors use it more generally to only imply  $x^T Ax > 0$  for all  $x$ .

**Does  $A$  have to be invertible for the compact SVD of  $A$  to equal the full SVD of  $A$ ?** Yes, the compact SVD of  $A \in F^{m \times n}$  only equals the full SVD if  $m = n = \text{rank}(A)$  which is equivalent to  $A$  being square and invertible.

**What is the process for finding the columns of  $V$ ? Each  $v$  must live in nullspace of  $A^H A - \lambda I$  but that requires knowing  $\lambda$ , right?**

**On slide 3, in the equation  $v_j^H A^H A v_i = \delta_{i,j}$ , what does the  $\delta_{i,j}$  mean?** In this context,  $\delta_{i,j}$  is the Kronecker delta function and it equals 0 unless  $i = j$  and, if  $i = j$ , then it equals 1.

## 2 Principal Component Analysis

**Is principal component analysis (PCA) equivalent to singular vector decomposition (SVD)?** Strictly speaking the SVD is a matrix decomposition used to solve the PCA problem.

**For PCA, I don't fully understand what is being minimized? What is  $P_W(x_i)$  doing to  $x_i$  conceptually?** When vectors in  $\mathbb{R}^n$  are approximated by a subspace  $W$  of dimension  $p < n$ , this will typically introduce some error. PCA allows one to choose the subspace  $W$  in a way that minimizes this error.  $P_W(\underline{x})$  is projecting  $\underline{x}$  onto the subspace  $W$ . In other words, it is finding the best approximation of  $\underline{x}$  by vectors in  $W$ .

**In class, you briefly analogized PCA with aspects of signal detection theory, could you elaborate on that again please?** If a signal lives a low-dimensional subspace of high-dimensional space, then projecting onto the subspace reduces the noise while keeping the signal. PCA allows the same approach in an approximate way – if most of the signal lives in a low-dimensional subspace, then we can project to reduce the noise while keeping most of the signal.

**What do you mean by an affine subspace description of PCA?** An affine subspace is a shifted subspace. For a subspace  $W \subseteq V$  and a non-zero vector  $\underline{v}$ , we say that  $W + \underline{v} = \{\underline{u} \in V \mid \underline{u} - \underline{v} \in W\}$ .