

Discussion 3: SVD, PCA, and Singular Values

1. Review: Key Concepts

Before diving into the problems, we review the essential definitions of the Singular Value Decomposition.

1.1 Compact SVD

Definition 1 (Compact SVD). For $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$, the **compact SVD** is:

$$A = \underset{m \times n}{U_r} \underset{m \times r}{\Sigma_r} \underset{r \times r}{V_r^\top} \underset{r \times n}{V_r^\top}$$

where $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$ with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, the columns of U_r are orthonormal left singular vectors spanning $\mathcal{R}(A)$, and the columns of V_r are orthonormal right singular vectors spanning $\mathcal{R}(A^\top)$.

1.2 Full SVD

Definition 2 (Full SVD). The **full SVD** of $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$ is:

$$A = \underset{m \times n}{U} \underset{m \times m}{\Sigma} \underset{m \times n}{V}^\top \underset{n \times n}{V}^\top$$

where U and V are square orthonormal matrices ($U^\top U = U U^\top = I_m$, $V^\top V = V V^\top = I_n$), and Σ has Σ_r in its top-left block with zeros elsewhere:

$$\Sigma = \begin{bmatrix} \Sigma_r & 0_{r \times (n-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{bmatrix}$$

The extra columns of U (beyond U_r) span $\mathcal{N}(A^\top)$, and the extra columns of V (beyond V_r) span $\mathcal{N}(A)$.

1.3 Dyadic SVD

The matrix A can also be written as a sum of rank-1 matrices (dyads):

$$A = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top + \dots + \sigma_r u_r v_r^\top = \sum_{i=1}^r \sigma_i u_i v_i^\top$$

1.4 Principal Component Analysis (PCA)

Definition 3 (PCA via SVD). Given centered data matrix $X \in \mathbb{R}^{n \times d}$ (rows are data points), the **principal components** are the right singular vectors of X . The first principal component v_1 maximizes the variance $\|Xw\|_2^2$ over unit vectors w , equivalently solving $\arg \max_{\|w\|_2=1} w^\top X^\top X w$.

Key relationships: $X^\top X = V_d \Sigma_d^2 V_d^\top$ is an eigendecomposition. The eigenvalues of $X^\top X$ are σ_i^2 , and the eigenvectors are the right singular vectors v_i of X , which are the principal components.

2. Problem 1: SVD

Let $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$. Assume $m > n > r$.

2.1 Part (a): True/False on SVD Products

For each product, determine whether it equals the identity.

Mental model. In the full SVD, U and V are *square* orthonormal matrices, so both $MM^\top = I$ and $M^\top M = I$ hold for $M \in \{U, V\}$. In the compact SVD, U_r and V_r are *tall* matrices with orthonormal columns. For a tall matrix, only the “thin-side product” $M^\top M = I$ holds; the “fat-side product” MM^\top is a projector, not the identity.

- (a) $UU^\top = I$: **True.** U is $m \times m$ orthonormal (square), so $UU^\top = I_m$.
- (b) $U^\top U = I$: **True.** Same reason: U is square orthonormal, so $U^\top U = I_m$.
- (c) $V^\top V = I$: **True.** V is $n \times n$ orthonormal (square), so $V^\top V = I_n$.
- (d) $VV^\top = I$: **True.** Same reason: V is square orthonormal, so $VV^\top = I_n$.
- (e) $U_r^\top U_r = I$: **True.** U_r is $m \times r$ with orthonormal columns; the product $U_r^\top U_r$ is $r \times r$ and equals I_r .
- (f) $U_r U_r^\top = I$: **False.** $U_r U_r^\top$ is $m \times m$ but has rank $= r < m$. It is the orthogonal projector onto $\mathcal{R}(A)$, not the identity.
- (g) $V_r V_r^\top = I$: **False.** $V_r V_r^\top$ is $n \times n$ but has rank $= r < n$. It is the orthogonal projector onto $\mathcal{R}(A^\top)$, not the identity.
- (h) $V_r^\top V_r = I$: **True.** V_r is $n \times r$ with orthonormal columns; the product $V_r^\top V_r$ is $r \times r$ and equals I_r .

Key Distinction

For a **tall** matrix M ($m \times r$, $m > r$) with orthonormal columns: $M^\top M = I_r$ always holds, but $MM^\top \neq I_m$. The product MM^\top is the orthogonal projector onto the column space of M . It equals I only when M is square.

Quick test: Is the matrix square? If yes, both MM^\top and $M^\top M$ give I . If no (tall or wide), only the “thin-side product” gives I .

Common Mistake

Students often confuse “orthonormal columns” with “orthogonal matrix.” A matrix with orthonormal columns satisfies $M^\top M = I$, but an **orthogonal matrix** must also be *square*, which additionally gives $MM^\top = I$. The compact SVD factors U_r, V_r have orthonormal columns but are *not* orthogonal matrices (they are not square).

2.2 Part (b): Compact SVD from Full SVD

Recipe: Compact SVD from Full SVD

Given the full SVD $A = U\Sigma V^\top$:

- Count the nonzero singular values in Σ to determine $\text{rank}(A) = r$.
- Keep the first r columns of $U \rightarrow U_r$.
- Keep the top-left $r \times r$ block of $\Sigma \rightarrow \Sigma_r$.
- Keep the first r rows of V^\top (equivalently, first r columns of V) $\rightarrow V_r^\top$.

The “first r ” columns correspond to the r largest singular values, which are already ordered $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ in Σ .

Goal: Find the compact SVD given the full SVD:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Step 1: Identify the rank.

The matrix Σ has only one non-zero singular value $\sigma_1 = 2$, so $r = 1$.

Step 2: Extract the compact components.

- $\Sigma_r = [2]$ (the 1×1 block of non-zero singular values).
- $U_r =$ first column of $U = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ (the left singular vector for σ_1).
- $V_r^\top =$ first row of $V^\top = [1 \ 0]$ (the right singular vector for σ_1).

Compact SVD:

$$A = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} [2] \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

Verification. Multiply out $U_r \Sigma_r V_r^\top$:

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} (2) \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

This matches the product from the full SVD: $U \Sigma V^\top = \begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$. ✓

2.3 Part (c): Full SVD from Compact SVD

Goal: Find the full SVD given the compact SVD:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Here $A \in \mathbb{R}^{3 \times 2}$ with $r = 2$ (both singular values are non-zero).

Step 1: Identify what we have and what we need.

We have $U_r \in \mathbb{R}^{3 \times 2}$, $\Sigma_r \in \mathbb{R}^{2 \times 2}$, and $V_r = I_2$.

For the full SVD, we need $U \in \mathbb{R}^{3 \times 3}$ (square orthonormal), $\Sigma \in \mathbb{R}^{3 \times 2}$, and $V \in \mathbb{R}^{2 \times 2}$. The key challenge is that U_r is 3×2 — it has orthonormal columns but is not square. We must find one additional column to make U a full 3×3 orthonormal matrix.

Step 2: Extend V .

Since $r = n = 2$, we have $V_r = V = I_2$. No additional columns needed.

Step 3: Extend U to a 3×3 orthonormal matrix.

The columns of U_r are $u_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. We need a third column u_3 orthogonal to both. Since we are in \mathbb{R}^3 , we can use the cross product:

$$u_3 = u_1 \times u_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \cdot 1 - 0 \cdot 0 \\ 0 \cdot 0 - \frac{1}{\sqrt{2}} \cdot 1 \\ \frac{1}{\sqrt{2}} \cdot 0 - \frac{1}{\sqrt{2}} \cdot 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

Step 4: Extend Σ .

Pad with a zero row: $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$.

Full SVD:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

What does the extra column represent?

The third column u_3 spans $\mathcal{N}(A^\top)$: it is a left singular vector corresponding to the zero singular value. This connects to the **four fundamental subspaces**: the columns of U_r span $\mathcal{R}(A)$, while the extra columns of U span $\mathcal{N}(A^\top)$. Similarly, extra columns of V (if any) would span $\mathcal{N}(A)$.

General method and non-uniqueness

General method. The cross product trick only works in \mathbb{R}^3 . In higher dimensions, find the extra columns by computing a basis for $\mathcal{N}(U_r^\top)$ (e.g., via Gram–Schmidt on any vectors that extend $\{u_1, \dots, u_r\}$ to a basis for \mathbb{R}^m).

Non-uniqueness. Any orthonormal basis for $\mathcal{N}(U_r^\top)$ works as the extra columns; sign flips (e.g., $-u_3$ instead of u_3) or rotations within the null space all give valid full SVDs. The compact SVD is essentially unique (up to sign flips of paired u_i, v_i), but the full SVD is not.

3. Problem 2: PCA and Regression

3.1 Part (a): Identifying Principal Components from a Scatter Plot

Goal: Given a scatter plot with candidate unit vectors v_1, v_2, v_3, v_4 , identify the first and second principal components.

Step 1: Recall the PCA criterion.

The first principal component is the direction of **maximum variance** in the data. Concretely, “variance along direction w ” means projecting every data point onto w and measuring the spread: $\text{Var} = \|Xw\|_2^2$ (for centered data X). The second principal component is orthogonal to the first and captures the next largest variance.

Step 2: Analyze the data cloud.

From the scatter plot, the data is elongated along a diagonal from lower-left to upper-right, indicating positive correlation between the two variables.

Step 3: Match directions to principal components.

- **First principal component:** v_3 — points along the direction of greatest spread (upper-right diagonal), capturing the maximum variance.
- **Second principal component:** v_2 — perpendicular to v_3 (upper-left diagonal), capturing the remaining variance.

Why not v_1 or v_4 ? These lie along the coordinate axes (x -axis and y -axis). Because the data has positive correlation, the direction of maximum spread is *diagonal*, not axis-aligned. Projecting onto a coordinate axis ignores the correlation structure and yields less variance than projecting onto v_3 .

Sign ambiguity: Both v_3 and $-v_3$ are valid first PCs — negating a direction does not change the variance of the projection.

PCA vs. Regression

The first principal component is *not* the same as the regression line.

- **Regression** (e.g., y on x) minimizes *vertical* distances (residuals in y). It treats x and y asymmetrically.
- **PCA** minimizes *perpendicular* distances to the line. It treats all variables symmetrically.

For positively correlated data, the PCA line is typically *steeper* than the regression line of y on x (but shallower than the regression of x on y). The two coincide only when the data lies exactly on a line.

3.2 Part (b): Top k Principal Components from SVD

Goal: Given centered data $X \in \mathbb{R}^{n \times d}$ with compact SVD $X = U_d \Sigma_d V_d^\top$, identify the top k principal components.

Given: Centered data matrix $X \in \mathbb{R}^{n \times d}$ (each row is one data point, columns are features), its compact SVD $X = U_d \Sigma_d V_d^\top$, and the optimization formulation for PCA: $\arg \max_{\|w\|_2=1} w^\top X^\top X w$.

Step 1: Relate PCA to the eigendecomposition of $X^\top X$.

The first principal component solves:

$$\arg \max_{\|w\|_2=1} w^\top X^\top X w$$

This is a **Rayleigh quotient** problem. The general theorem states: for a symmetric matrix M , the maximizer of $w^\top M w$ subject to $\|w\|_2 = 1$ is the eigenvector of M corresponding to its *largest* eigenvalue, and the maximum value is that eigenvalue. Applying this with $M = X^\top X$, the first PC is the top eigenvector of $X^\top X$.

Step 2: Connect $X^\top X$ to the SVD of X .

From the compact SVD $X = U_d \Sigma_d V_d^\top$:

$$X^\top X = (U_d \Sigma_d V_d^\top)^\top (U_d \Sigma_d V_d^\top) = V_d \Sigma_d^\top \underbrace{U_d^\top U_d}_{=I_d} \Sigma_d V_d^\top = V_d \Sigma_d^2 V_d^\top$$

The key simplification is $U_d^\top U_d = I_d$, which holds because U_d has orthonormal columns (as discussed in Problem 1a). This is precisely the eigendecomposition of $X^\top X$: the eigenvalues are $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_d^2$, and the eigenvectors are the columns v_1, \dots, v_d of V_d .

Step 3: Identify the top k principal components.

Since $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d > 0$, the eigenvector corresponding to the largest eigenvalue σ_1^2 is v_1 , the next is v_2 , and so on.

The top k principal components are the first k right singular vectors of X :

$$v_1, v_2, \dots, v_k$$

These are the columns of V_d corresponding to the k largest singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k$.

Practical Interpretation: Variance Explained

The i -th principal component v_i explains a fraction of the total variance:

$$\text{Fraction explained by } v_i = \frac{\sigma_i^2}{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_d^2}$$

The top k PCs together explain $\sum_{i=1}^k \sigma_i^2 / \sum_{j=1}^d \sigma_j^2$ of the total variance. Projecting data onto the top k components gives the best rank- k approximation to the data in the least-squares sense (Eckart–Young theorem).

4. Problem 3: Singular Values

4.1 Part (a): Singular Values from Eigendecomposition of $A^\top A$

Goal: Find the singular values of $A \in \mathbb{R}^{3 \times 2}$ given:

$$A^\top A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 1: Recognize the eigendecomposition.

How to recognize PDP^\top : We are given a product of three matrices where the outer matrices are transposes of each other and the middle matrix is diagonal. Check that P is orthogonal (its columns are orthonormal). If so, this is an eigendecomposition: the diagonal entries of D are the eigenvalues, and the columns of P are the corresponding eigenvectors.

Here $P = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$ is orthogonal and $D = \text{diag}(5, 3)$, so the eigenvalues of $A^\top A$ are 5 and 3.

Step 2: Recall the relationship between singular values and eigenvalues.

The singular values of A are the square roots of the eigenvalues of $A^\top A$:

$$\sigma_i = \sqrt{\lambda_i(A^\top A)}$$

Why is taking square roots always valid? The matrix $A^\top A$ is always positive semidefinite (PSD), since for any x :

$$x^\top (A^\top A)x = \|Ax\|_2^2 \geq 0$$

Therefore all eigenvalues of $A^\top A$ are ≥ 0 , and taking square roots is well-defined.

Step 3: Compute.

The singular values of A are:

$$\sigma_1 = \sqrt{5}, \quad \sigma_2 = \sqrt{3}$$

Bonus connection. The columns of P in the decomposition $A^\top A = PDP^\top$ are actually the right singular vectors V of A . This is because $A^\top A = V\Sigma^2V^\top$ (from the SVD). So the eigendecomposition of $A^\top A$ simultaneously gives us both the singular values (via D) and the right singular vectors (via P).

4.2 Part (b): Singular Values of $C = \begin{bmatrix} B & -B & 3I_3 \end{bmatrix}$

Goal: Given $B \in \mathbb{R}^{3 \times 2}$ with singular values 0, $\sqrt{2}$, $\sqrt{7}$, find the singular values of $C = \begin{bmatrix} B & -B & 3I_3 \end{bmatrix} \in \mathbb{R}^{3 \times 7}$.

Step 1: Choose CC^\top over $C^\top C$.

Since $C \in \mathbb{R}^{3 \times 7}$, we have two options: $C^\top C \in \mathbb{R}^{7 \times 7}$ or $CC^\top \in \mathbb{R}^{3 \times 3}$. Both share the same nonzero eigenvalues (recall Discussion 2, Problem 2c), so always work with the **smaller** one. Here CC^\top is 3×3 — much easier.

Step 2: Compute CC^\top using the block product rule.

For any block matrix $C = \begin{bmatrix} A_1 & A_2 & \cdots & A_k \end{bmatrix}$, we have:

$$CC^\top = A_1A_1^\top + A_2A_2^\top + \cdots + A_kA_k^\top$$

Applying this with $A_1 = B$, $A_2 = -B$, $A_3 = 3I_3$:

$$\begin{aligned} CC^\top &= BB^\top + (-B)(-B)^\top + (3I_3)(3I_3)^\top \\ &= BB^\top + BB^\top + 9I_3 \\ &= 2BB^\top + 9I_3 \end{aligned}$$

Step 3: Find the eigenvalues of BB^\top .

The singular values of $B \in \mathbb{R}^{3 \times 2}$ are given as 0, $\sqrt{2}$, $\sqrt{7}$. Note: B has $\min(3, 2) = 2$ nonzero singular values at most. The “third singular value” 0 arises because BB^\top is 3×3 but has rank ≤ 2 , giving a zero eigenvalue. The eigenvalues of BB^\top (squares of singular values) are:

$$0, \quad 2, \quad 7$$

Step 4: Find the eigenvalues of $CC^\top = 2BB^\top + 9I$.

We use two eigenvalue properties:

- **Scaling:** $\lambda(cM) = c \cdot \lambda(M)$. So $\lambda(2BB^\top) = 2 \cdot \lambda(BB^\top) = 0, 4, 14$.
- **Shifting:** $\lambda(M + cI) = \lambda(M) + c$. So $\lambda(2BB^\top + 9I) = 0 + 9, 4 + 9, 14 + 9$.

Eigenvalues of CC^\top : 9, 13, 23

Step 5: Take square roots.

The singular values of C are the square roots of the eigenvalues of CC^\top :

The singular values of C are:

$$\sigma_1 = \sqrt{23}, \quad \sigma_2 = \sqrt{13}, \quad \sigma_3 = 3$$

Strategy

When C is wide (3×7), computing CC^\top (3×3) is much easier than $C^\top C$ (7×7). Both share the same *nonzero* eigenvalues, so always work with the smaller product. The singular values of C are the same either way.

Key Takeaways

1. **Compact vs. full SVD:** The compact SVD keeps only the r non-zero singular values and their singular vectors. The full SVD extends U and V to square orthonormal matrices by adding vectors spanning $\mathcal{N}(A^\top)$ and $\mathcal{N}(A)$.
2. **Orthonormality of SVD factors:** For the compact SVD, $U_r^\top U_r = I_r$ and $V_r^\top V_r = I_r$ (columns are orthonormal), but $U_r U_r^\top \neq I$ and $V_r V_r^\top \neq I$ when r is less than the ambient dimension. These products are projectors, not identities.
3. **PCA from SVD:** The top k principal components of centered data X are the first k right singular vectors of X (columns of V), corresponding to the k largest singular values.
4. **Singular values from $A^\top A$:** The singular values of A are $\sigma_i = \sqrt{\lambda_i(A^\top A)}$. When given an eigendecomposition of $A^\top A$, simply read off the eigenvalues and take square roots.
5. **Block matrix trick:** For $C = [B \ -B \ 3I]$, compute $CC^\top = 2BB^\top + 9I$ and use eigenvalue shifting. Always work with the smaller of CC^\top and $C^\top C$.
6. **Recognizing eigendecompositions:** When you see PDP^\top with P orthogonal and D diagonal, the diagonal entries of D are eigenvalues and the columns of P are eigenvectors. For $A^\top A = PDP^\top$, the columns of P are the right singular vectors of A .
7. **Non-uniqueness in SVD:** The compact SVD is essentially unique up to sign flips of paired singular vectors ($u_i, v_i \rightarrow -u_i, -v_i$). The full SVD is *not* unique: any orthonormal basis for $\mathcal{N}(A^\top)$ or $\mathcal{N}(A)$ can serve as the extra columns of U or V .