Reconstruction and Principal Component Analysis (continued)

ECE 57000 — September 16, 2026

David I. Inouye

Monday reduced PCA to choosing one reconstruction subspace

- Choose an orthonormal decoder $D = W$ without losing reconstruction possibilities.
- For fixed W, the best encoder is $E = W^T$: reconstruct by perpendicular projection.
- Choose $W^TW = I_k$ to minimize $\|X - XWW^T\|_F^2$ for centered data X.
- Perpendicularity splits fixed total squared length into retained length plus reconstruction error.

Next: review the fixed-total decomposition, then use the SVD to find the best k directions.

Minimizing reconstruction error means keeping the most squared length

Let $\widehat{\boldsymbol{x}}_i = WW^T\boldsymbol{x}_i$. The reconstruction and residual are perpendicular.

$$
\begin{aligned}
\|X\|_F^2 &= \sum_i \|\boldsymbol{x}_i\|_2^2 && \text{(Sum squared row lengths)} \\
&= \sum_i \left(\|\widehat{\boldsymbol{x}}_i\|_2^2 + \|\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_i\|_2^2\right) && \text{(Pythagoras for each observation)} \\
&= \sum_i \left(\|W^T\boldsymbol{x}_i\|_2^2 + \|\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_i\|_2^2\right) && \text{(}W^TW = I_k\text{: decoding preserves length)} \\
&= \|XW\|_F^2 + \|X - XWW^T\|_F^2 && \text{(Collect the rows into matrices)}
\end{aligned}
$$

$$
\underbrace{\|X - XWW^T\|_F^2}_{\text{error to minimize}} = \underbrace{\|X\|_F^2}_{\text{fixed total}} - \underbrace{\|XW\|_F^2}_{\text{retained amount to maximize}}.
$$

Only W changes. The same centered data X give the same total squared length for every choice.

Three steps connect the PCA problem to its solution

The total squared length is fixed. We only need to maximize what the coordinates retain.

1. Rewrite the objective — Established

Least reconstruction error means most retained squared length.

2. Find the optimal directions — Next

Use the SVD to prove that the leading right singular vectors are optimal.

3. Interpret the solution

Read reconstruction error and coordinate variance from the singular values.

The SVD expresses a data matrix as a sum of simple patterns

Now decompose the centered data itself: $X = U\Sigma V^T$.

$$X = \sum_{j=1}^{r} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^T, \qquad r = \operatorname{rank}(X).$$

Part	Meaning for the Data Matrix
$\boldsymbol{v}_j \in \mathbb{R}^d$	One pattern across measured features.
$\sigma_j \boldsymbol{u}_j \in \mathbb{R}^n$	How strongly that pattern appears in each observation.
$\sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^T$	A matrix of rank one: one shared pattern with observation-specific weights.

Truncation keeps some patterns and drops the rest

For $1 \leq k \leq r$, keeping the leading k terms gives:

$$X_k = \sum_{j=1}^{k} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^T = U_k \Sigma_k V_k^T.$$

- U_k: first k columns of U.
- Σ_k: the $k \times k$ diagonal matrix of retained singular values.
- V_k: first k columns of V.

Does this choice retain the most squared length among all k-dimensional subspaces?

The SVD measures how much of each direction we keep

Write $W = [\boldsymbol{w}_1 \ \cdots \ \boldsymbol{w}_k]$ and $X = U\Sigma V^T$, with $\sigma_j = 0$ for $j > \operatorname{rank}(X)$.

$$\begin{aligned}
\|XW\|_F^2 &= \sum_{\ell=1}^{k} \|X\boldsymbol{w}_\ell\|_2^2 && \text{(Sum squared column lengths)} \\
&= \sum_{\ell=1}^{k}\sum_{j=1}^{d} \sigma_j^2 (\boldsymbol{v}_j^T \boldsymbol{w}_\ell)^2 && \text{(V^T reads coordinates; Σ scales them; U preserves length)} \\
&= \sum_{j=1}^{d} \sigma_j^2 \left[\sum_{\ell=1}^{k} (\boldsymbol{v}_j^T \boldsymbol{w}_\ell)^2\right] && \text{(Swap sums; factor out σ_j^2)} \\
&= \sum_{j=1}^{d} \sigma_j^2 \alpha_j && \text{(The bracketed sum is α_j)}
\end{aligned}$$

- $\alpha_j = \|W^T \boldsymbol{v}_j\|_2^2$: amount of direction $\boldsymbol{v}_j$ kept; $0 \le \alpha_j \le 1$.
- $\sum_j \alpha_j = \sum_{\ell=1}^{k} \|\boldsymbol{w}_\ell\|_2^2 = k$: k **units in total**.

Directions beyond k contribute at most the cutoff weight

Order $\sigma_1^2 \geq \cdots \geq \sigma_d^2$. Recall $\alpha_j \geq 0$ and $\sum_j \alpha_j = k$.

$$\begin{aligned}
\sum_{j=1}^{d} \sigma_j^2 \alpha_j &= \sum_{j=1}^{k} \sigma_j^2 \alpha_j + \sum_{j>k} \sigma_j^2 \alpha_j && \text{(Separate the first } k \text{ directions)} \\
&\leq \sum_{j=1}^{k} \sigma_j^2 \alpha_j + \sigma_k^2 \sum_{j>k} \alpha_j && \text{(For } j > k\text{: } \sigma_j^2 \leq \sigma_k^2, \alpha_j \geq 0\text{)} \\
&= \sum_{j=1}^{k} \sigma_j^2 \alpha_j + \sigma_k^2 \left(k - \sum_{j=1}^{k} \alpha_j\right) && \text{(All } \alpha_j \text{ sum to } k\text{)}
\end{aligned}$$

Why the inequality? Replace each smaller weight by σ_k^2. Multiplying by a nonnegative α_j preserves the inequality.

Each leading direction contributes at most one unit

Continue from the previous bound. For $j \le k$, $\sigma_j^2 - \sigma_k^2 \ge 0$ and $\alpha_j \le 1$.

$$\begin{aligned}
\|XW\|_F^2 &\le \sum_{j=1}^{k} \sigma_j^2 \alpha_j + \sigma_k^2 \Big(k - \sum_{j=1}^{k} \alpha_j\Big) && \text{(Previous slide's bound)} \\
&= k\sigma_k^2 + \sum_{j=1}^{k} (\sigma_j^2 - \sigma_k^2)\alpha_j && \text{(Expand and regroup)} \\
&\le k\sigma_k^2 + \sum_{j=1}^{k} (\sigma_j^2 - \sigma_k^2) \cdot 1 && \text{(Nonnegative coefficient; } \alpha_j \le 1\text{)} \\
&= \sum_{j=1}^{k} \sigma_j^2 && \text{(The } k \text{ copies of } \sigma_k^2 \text{ cancel)}
\end{aligned}$$

Thus **every feasible** W retains at most $\sum_{j=1}^{k} \sigma_j^2$.

Choosing $W = V_k$ attains the bound

Choose $W = V_k = [\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_k]$. These columns are orthonormal, so $W^T W = I_k$.

$$\alpha_j = \sum_{\ell=1}^{k} (\boldsymbol{v}_j^T \boldsymbol{v}_\ell)^2$$

(Substitute $\boldsymbol{w}_\ell = \boldsymbol{v}_\ell$)

$$= \begin{cases} 1, & j \le k, \\ 0, & j > k. \end{cases}$$

(Inner products are 1 for $j = \ell$ and 0 otherwise)

For $j \le k$, exactly one term is 1. For $j > k$, every term is 0.

$$\|XV_k\|_F^2 = \sum_{j=1}^{d} \sigma_j^2 \alpha_j = \sum_{j=1}^{k} \sigma_j^2.$$

The upper bound is achieved: V_k maximizes retained squared length and therefore minimizes reconstruction error.

Three steps connect the PCA problem to its solution

The leading right singular vectors attain the optimum. Now read what this solution keeps and loses.

1. Rewrite the objective — Established

Least reconstruction error means most retained squared length.

2. Find the optimal directions — Established

Use the SVD to prove that the leading right singular vectors are optimal.

3. Interpret the solution — Next

Read reconstruction error and coordinate variance from the singular values.

Discarded singular values quantify what reconstruction loses

Subtract the optimal retained squared length from the fixed total:

$$\|X - X_k\|_F^2 = \underbrace{\sum_{j=1}^{r} \sigma_j^2}_{\text{total}} - \underbrace{\sum_{j=1}^{k} \sigma_j^2}_{\text{retained}} = \sum_{j>k} \sigma_j^2 .$$

For singular values $6, 3, 1$:

Retained Features	Squared Reconstruction Error
$k = 1$	$3^2 + 1^2 = 10$
$k = 2$	$1^2 = 1$
$k = 3$	0

Squared singular values give the variance of each PCA coordinate

The jth coordinate across observations is $Z_{:j} = X\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j$. Since X is centered, this coordinate also has mean zero.

$$
\begin{aligned}
\mathrm{Var}_{\mathrm{sample}}(Z_{:j}) &= \frac{1}{n-1}\sum_{i=1}^{n} Z_{ij}^2 && \text{(Variance of a centered coordinate; } n > 1\text{)} \\
&= \frac{\sigma_j^2}{n-1}\|\boldsymbol{u}_j\|_2^2 && \text{(Substitute } Z_{:j} = \sigma_j \boldsymbol{u}_j\text{)} \\
&= \frac{\sigma_j^2}{n-1} && (\boldsymbol{u}_j \text{ is a unit vector})
\end{aligned}
$$

The leading PCA coordinates therefore have the **largest variances**. This interprets the solution we already proved.