The Singular Value Decomposition

This is the capstone of the class.
It's a fundamental application of linear algebra to:

- Statistics (PCA)
- Data Science
- Engineering
- etc.

Thm (SVD, outer product form): back to rectangular matrices

Let A be an $m \times n$ matrix of rank r. Then

$$\boxed{A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T}$$

where

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$
- $\{u_1, \ldots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$
- $\{v_1, \ldots, v_r\}$ is an orthonormal set in $\mathbb{R}^n$.

What does this mean?

Idea: columns of A are data points

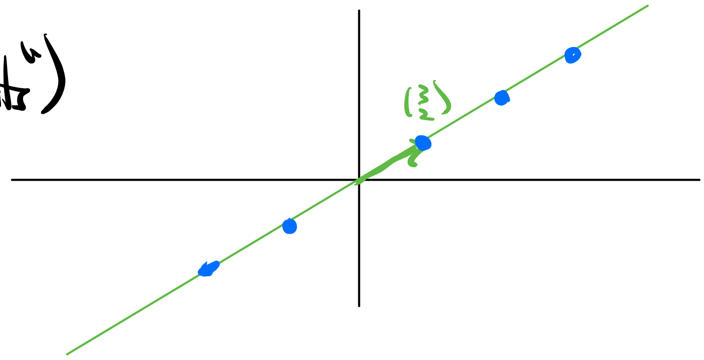
$r=1$: let $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$ be nonzero vectors.

$$uv^T = \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}}_{\text{vector}} \underbrace{(v_1 \cdots v_n)}_{\text{coefficients}} = \underbrace{\begin{pmatrix} | & & | \\ v_1 u & \cdots & v_n u \\ | & & | \end{pmatrix}}_{\text{multiples of } u} = \begin{pmatrix} u_1 v_1 & \cdots & u_1 v_n \\ \vdots & \ddots & \vdots \\ u_m v_1 & \cdots & u_m v_n \end{pmatrix}$$

This is an $m \times n$ matrix of rank 1: $\mathrm{Col}(uv^T) = \mathrm{Span}\{u\}$

Let's plot the columns ("data points")

$$\begin{pmatrix}3\\2\end{pmatrix}\begin{pmatrix}-1 & 2 & 1 & 3 & -2\end{pmatrix}$$



Upshot: A rank-1 matrix encodes data points (columns) that lie on a line.

$r=2$: $$A = u_1 v_1^T + u_2 v_2^T = \begin{pmatrix} | & & | \\ v_{11}u_1 & \cdots & v_{1n}u_1 \\ | & & | \end{pmatrix} + \begin{pmatrix} | & & | \\ v_{21}u_2 & \cdots & v_{2n}u_2 \\ | & & | \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ v_{11}u_1 + v_{21}u_2 & \cdots & v_{21}u_1 + v_{22}u_2 \\ | & & | \end{pmatrix}$$

The columns are linear combinations of u_1 & u_2.

Let's plot the columns ("data points"):

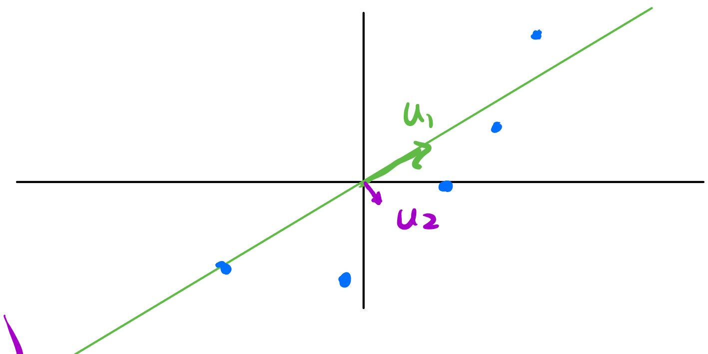
u_1 $\qquad$ $v_1 =$ coeffs of $\begin{pmatrix}3\\2\end{pmatrix}$

$$\begin{pmatrix}3\\2\end{pmatrix}\begin{pmatrix}-1 & 2 & 1 & 3 & -2\end{pmatrix}$$

orthogonal

$$+\begin{pmatrix}.2\\-.3\end{pmatrix}\begin{pmatrix}3 & 1 & 2 & -1 & 0\end{pmatrix}$$

u_2 $\qquad$ $v_2 =$ coeffs of $\begin{pmatrix}.2\\-.3\end{pmatrix}$

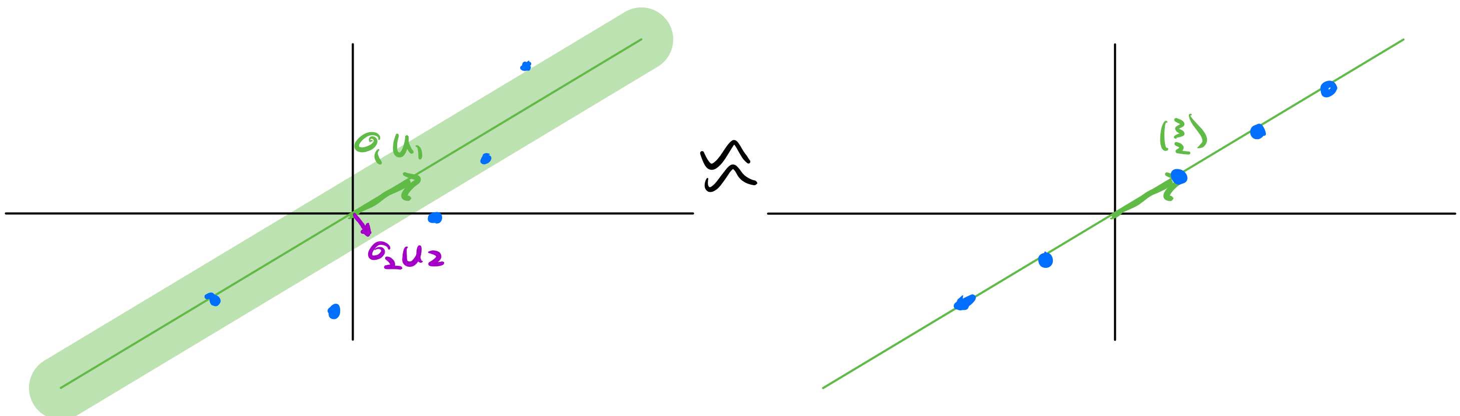


Upshot: A rank-2 matrix encodes data points that lie on the plane $\mathrm{Span}\{u_1, u_2\}$

But: $\left\|\begin{pmatrix}3\\2\end{pmatrix}\right\| \gg \left\|\begin{pmatrix}.2\\-.3\end{pmatrix}\right\|$ so the $\begin{pmatrix}.2\\-.3\end{pmatrix}$ direction is less important!

$$\begin{pmatrix}3\\2\end{pmatrix}\begin{pmatrix}-1 & 2 & 1 & 3 & -2\end{pmatrix} + \begin{pmatrix}.2\\-.3\end{pmatrix}\begin{pmatrix}3 & 1 & 2 & -1 & 0\end{pmatrix}$$

$$\approx \begin{pmatrix}3\\2\end{pmatrix}\begin{pmatrix}-1 & 2 & 1 & 3 & -2\end{pmatrix} \quad \text{(to one decimal place)}$$



We're extracted important information:

our data points almost lie on a line!

In general, the SVD will find the best-fit line, plane, 3-space, ..., r-space for our data, all at once. Not least-squares: the error will be in statistical language (variance).

Why might we care?

- Data compression: uv^T is 7 numbers instead of 10 for a 2×5 matrix.
- Data analysis: SVD will reveal all approximate linear relations among our data points.
- Statistics: SVD finds more & less important correlations

etc.

Back to the statement of the SVD:

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

where

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$
- $\{u_1, \ldots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$
- $\{v_1, \ldots, v_r\}$ is an orthonormal set in $\mathbb{R}^n$.

Def:
- $\sigma_1, \ldots, \sigma_r$ are the singular values of A
- $u_1, \ldots, u_r$ are the left singular vectors
- $v_1, \ldots, v_r$ are the right singular vectors

Note 1: For any vector $x \in \mathbb{R}^n$,

$$\begin{aligned} Ax &= (\sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T)x \\ &= \sigma_1 u_1 v_1^T x + \cdots + \sigma_r u_r v_r^T x \\ &= \sigma_1 (v_1 \cdot x) u_1 + \cdots + \sigma_r (v_r \cdot x) u_r \end{aligned}$$

Note 2: Taking $x = v_i$, we have

$$Av_i = \sigma_1 \cancelto{0}{(v_1 \cdot v_i)} u_1 + \cdots + \sigma_i \cancelto{1}{(v_i \cdot v_i)} u_i + \cdots + \sigma_r \cancelto{0}{(v_r \cdot v_i)} u_r$$

So the singular vectors are related by

$$\boxed{Av_i = \sigma_i u_i} \quad \text{and thus} \quad \boxed{\|Av_i\| = \sigma_i}$$

Note 3: Take transposes:

$$A^T = (\sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T)^T$$
$$= \sigma_1 v_1 u_1^T + \cdots + \sigma_r v_r u_r^T$$

Therefore,

$$\boxed{A^T = \sigma_1 v_1 u_1^T + \cdots + \sigma_r v_r u_r^T}$$

is the SVD of A^T!

So A & A^T have the same

- singular values and
- singular vectors (switch right & left).

Note 4: Note 2 + Note 3 $\Rightarrow A^T u_i = \sigma_i v_i$, so

$$A^T A v_i = A^T(\sigma_i u_i) = \sigma_i A^T u_i = \sigma_i^2 v_i$$
$$A A^T u_i = A(\sigma_i v_i) = \sigma_i A v_i = \sigma_i^2 u_i$$

In particular,

> $\{v_1, \ldots, v_r\}$ are orthonormal eigenvectors of A^TA with eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$.
>
> $\{u_1, \ldots, u_r\}$ are orthonormal eigenvectors of AA^T with eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$.

This tells us how to prove/compute the SVD:

orthogonally diagonalize A^TA

Recall: Let v be an eigenvector of A^TA with eigenvalue λ and let $u=Av$.

(1) $\|u\|^2 = \lambda\|v\|^2$

(2) If $\lambda>0$ then u is an eigenvector of AA^T with eigenvalue λ

(3) Let v' be another eigenvector of A^TA with eigenvalue λ', and let $u'=Av'$. If $v\cdot v'=0$ then $u\cdot u'=0$.

Proof of SVD:

Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A^TA

(the λ_i's show up multiple times if $AM \geq 1$)

Note $\lambda_n \geq 0$ because A^TA is positive-semidefinite.

Step 1: $\lambda_{r+1} = \cdots = \lambda_n = 0$

- $Nul(A^TA) = Nul(A)$ has dimension $n-r$.
- $Nul(A^TA) =$ the 0-eigenspace of A^TA.
- $AM(0) = GM(0)$ in A^TA

  because A^TA is symmetric $\Rightarrow$ diagonalizable

So $n-r$ of the λ_i's are $=0$

$$\Rightarrow \lambda_{r+1} = \cdots = \lambda_n = 0$$

Now: $\lambda_1 \geq \cdots \geq \lambda_r > 0$ are the nonzero eigenvalues of A^TA.

Set:

- $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_r = \sqrt{\lambda_r}$
- Let $v_1, \ldots, v_r$ be orthonormal eigenvectors with $A^TAv_i = \lambda_i v_i$.
- Let $u_1 = \frac{1}{\sigma_1} Av_1, \ldots, u_r = \frac{1}{\sigma_r} Av_r$

Check:

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$
- $\{v_1, \ldots, v_r\}$ is orthonormal
- $\|u_i\| = \frac{1}{\sigma_i^2} \|Av_i\| = \frac{1}{\lambda_i} \cdot \lambda_i \|v_i\| = 1$ by fact (1)

  $u_i \cdot u_j = 0$ for $i \neq j$ by fact (3)

  $\Rightarrow \{u_1, \ldots, u_r\}$ is orthonormal

It remains to show

$$A \stackrel{?}{=} \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T.$$

Let $A' = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$, so we want $A' = A$.

Choose an orthonormal basis $\{v_{r+1}, \dots, v_n\}$ for $\mathrm{Nul}(A) = \mathrm{Nul}(A^TA) = 0$-eigenspace of A^TA, so $\{v_1, \dots, v_r, v_{r+1}, \dots, v_n\}$ is an orthonormal eigenbasis of A^TA. Claim: $Av_i = A'v_i$ for all $i = 1, \dots, n$

$i \le r$: By definition $Av_i = \sigma_i u_i$

and we know $A'v_i = \sigma_i u_i$ by Note 2.

$i > r$: We have $Av_i = 0$ because $v_i \in \mathrm{Nul}(A)$

$$A'v_i = \sigma_1 (\cancelto{0}{v_1 \cdot v_i}) u_1 + \dots + \sigma_r (\cancelto{0}{v_r \cdot v_i}) u_r$$

All $v_1 \cdot v_i = 0$ because the 0-eigenspace of A^TA is $\perp$ the λ_i-eigenspace.

So $Av_i = A'v_i$ for all $i = 1, \dots, n$.

Any $x \in \mathbb{R}^n$ can be written $x = a_1 v_1 + \dots + a_n v_n$

$$\begin{aligned} Ax &= a_1 Av_1 + \dots + a_n Av_n \\ &= a_1 A'v_1 + \dots + a_n A'v_n = A'x \end{aligned}$$

Since $Ax = A'x$ for all $x \in \mathbb{R}^n$ we have $A = A'$

This also gives us a procedure to compute the SVD. It is not the algorithm used in practice!

$\rightarrow$ Efficient computation of the SVD is a difficult problem!

Naive Schoolbook Procedure to Compute the SVD:

Let A be an $m \times n$ matrix of rank r.

(1) Compute the nonzero eigenvalues of A^TA,
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$$
(where λ_i appears multiple times if $AM > 1$)

→ There are automatically r of them, and they're positive.

(2) Find an orthonormal eigenbasis for each eigenspace: get an orthonormal set $\{v_1, \ldots, v_r\}$ with $A^TAv_i = \lambda_i v_i$.

(3) Set $\sigma_i = \sqrt{\lambda_i}$ $\quad u_i = \frac{1}{\sigma_i} A v_i$.

Then $\{u_1, \ldots, u_r\}$ is orthonormal and
$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T.$$

Eg: $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$ NB: $r = 2$ (2 pivots)

(1) $A^TA = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix}$ $\quad p(\lambda) = \lambda^2 - 50\lambda + 225$
$$= (\lambda - 45)(\lambda - 5)$$

$$\lambda_1 = 45 \qquad \lambda_2 = 5$$

(2) Compute eigenspaces:

$$\begin{pmatrix} -b \\ a-\lambda \end{pmatrix}$$

$$A^TA - 45 I_2 = \begin{pmatrix} -20 & 20 \end{pmatrix} \overset{\text{trick}}{\rightsquigarrow} v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$A^TA - 5 I_2 = \begin{pmatrix} 20 & 20 \end{pmatrix} \rightsquigarrow v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

(3) $\sigma_1 = \sqrt{\lambda_1} = \sqrt{45} = 3\sqrt{5}$ $\qquad \sigma_2 = \sqrt{\lambda_2} = \sqrt{5}$

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{3\sqrt{5}}\begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{3\sqrt{10}}\begin{pmatrix} 3 \\ 9 \end{pmatrix} = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

SVD:

$$\begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} = 3\sqrt{5} \cdot \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 3 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix} + \sqrt{5} \cdot \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ -1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \end{pmatrix}$$

Check: $\|u_1\| = \frac{1}{\sqrt{10}}\sqrt{1^2+3^2} = 1$ $\qquad \|u_2\| = \frac{1}{\sqrt{10}}\sqrt{3^2+(-1)^2} = 1$

$u_1 \cdot u_2 = 0$ ✓

Summary:

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

$$A v_i = \sigma_i u_i \qquad \underbrace{A^TA v_i = \sigma_i^2 v_i \qquad AA^T u_i = \sigma_i^2 u_i}_{\text{eigenvectors}}$$

$$A^T = \sigma_1 v_1 u_1^T + \cdots + \sigma_r v_r u_r^T$$