Some <u>special</u> symmetric matrices: Suppose A is an $m \times n$ matrix (we don't need $m=n$).

Then $A^T A$ and AA^T are both symmetric (not usually the same)

$\uparrow$ $n \times n$ $\qquad$ $\uparrow$ $m \times m$

What's special about $A^T A$? For one thing, its eigenvalues are not just <u>real</u> numbers. They are also <u>positive</u> (or 0).

Why? Suppose $A^T A \vec{x} = \lambda \vec{x}$ with $\vec{x} \neq \vec{0}$.

Then
$$\underbrace{\vec{x}^T A^T A \vec{x}}_{= (A\vec{x})^T Ax} = \underbrace{\lambda(\vec{x}^T \vec{x})}_{= \lambda(\vec{x}^T\vec{x})} \rightarrow \lambda = \frac{(A\vec{x})\cdot(A\vec{x})}{\vec{x}\cdot\vec{x}} = \frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} \geq 0$$

"<u>Spectral Theorem</u>"

Let's arrange the eigenvalues in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > \lambda_{r+1} = 0, \ldots, \lambda_n = 0$$

$$\vec{v}_1, \vec{v}_2, \ldots \vec{v}_r, \underbrace{\vec{v}_{r+1}, \ldots, \vec{v}_n}_{\text{Basis for } N(A^TA)} \leftarrow \boxed{\text{Orthonormal basis of eigenvectors}}$$

<u>Spectral Theorem</u>

Since each $\lambda_i \geq 0$, we can take their square roots: $\sigma_i = +\sqrt{\lambda_i}$

called the <u>singular values</u> of A

Also a basis for $N(A)$, because $N(A) = N(A^TA)$.

Why? If $\vec{x}$ in $N(A)$, the $A\vec{x} = \vec{0}$, so

$A^T A \vec{x} = A^T \vec{0} = \vec{0} \rightarrow \vec{x}$ is <u>also</u> in $N(A^TA)$.

On the other hand, if $A^T A \vec{x} = \vec{0}$, then

$$\underbrace{\vec{x}^T A^T A \vec{x}}_{= (A\vec{x})^T A\vec{x} = \|A\vec{x}\|^2} = \vec{x}^T \vec{0} = 0 \rightarrow A\vec{x} = \vec{0}.$$

So $\vec{x}$ is in $N(A)$ as well.

Note that $\dim N(A^TA) = \dim N(A) = n - r$, where $r = \text{rank}(A)$

same as rank of A^TA.

So the non-zero singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ go along with rank(A)-many orthonormal eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$: $A^TA\vec{v}_i = \sigma_i^2 \vec{v}_i$

For $i = 1, 2, \ldots, r$, let's define $\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i$ (vectors in $\mathbb{R}^m$ since A is $m \times n$)

What's special about $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r$?

① They are an orthonormal set in $\mathbb{R}^m$:

$$u_i^T \vec{u}_j = \frac{1}{\sigma_i \sigma_j}(Av_i)^T(A\vec{v}_j) = \frac{1}{\sigma_i\sigma_j} \vec{v}_i^T \underbrace{A^TA\vec{v}_j}_{\sigma_j^2 \vec{v}_j}$$

$$= \frac{\sigma_j}{\sigma_i} v_i^T \vec{v}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Because $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is orthonormal.

② They form a orthonormal basis of the column space $C(A)$:

- They are linearly independent because they are orthonormal.
- They are in the column space, because $C(A)$ = set of all $A\vec{x}$ for $\vec{x}$ in $\mathbb{R}^n$.
- They are enough for a basis since $\dim C(A) = \text{rank } r$.

③ They are eigenvectors for AA^T!

same eigenvalue λ_i

$$AA^T\vec{u}_i = \frac{1}{\sigma_i} A\underbrace{A^TA\vec{v}_i}_{\sigma_i^2 \vec{v}_i} = \sigma_i A\vec{v}_i = \sigma_i^2\left(\frac{A\vec{v}_i}{\sigma_i}\right) = \sigma_i^2 \vec{u}_i$$

Now remember one of the big theorems:

$$C(A)^{\perp} = N(A^T) \text{ in } \mathbb{R}^m$$

We can get an orthonormal basis of $\mathbb{R}^m$ by combining

$$\underbrace{\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}}_{\substack{\text{orthonormal basis} \\ \text{of } C(A)\text{, also} \\ \text{eigenvectors for } AA^T}} \text{ with } \underbrace{\{\vec{u}_{r+1}, \ldots, \vec{u}_m\}}_{\substack{\text{orthonormal basis of } N(A^T)\text{, same as} \\ N(AA^T)\text{, so also eigenvectors for } AA^T \\ \text{(with eigenvalue 0).}}}$$

Conclusion: For any $m \times n$ matrix A, we've shown that we can find orthonormal bases of both $\mathbb{R}^m$ and $\mathbb{R}^n$ that are "good for A":

$$\mathbb{R}^m: \{\underbrace{\vec{u}_1, \ldots, \vec{u}_r}_{}, \underbrace{\vec{u}_{r+1}, \ldots, \vec{u}_m}_{\text{basis of } N(A^T)}\}$$ orthonormal basis of eigenvectors for AA^T

$$AA^T \vec{u}_i = \sigma_i^2 \vec{u}_i$$

$$\mathbb{R}^n: \{\underbrace{\vec{v}_1, \ldots, \vec{v}_r}_{}, \underbrace{\vec{v}_{r+1}, \ldots, \vec{v}_n}_{\text{basis of } N(A)}\}$$ orthonormal basis of eigenvectors for $A^T A$

$$A^T A \vec{v}_i = \sigma_i^2 \vec{v}_i$$

Moreover, $\boxed{A\vec{v}_i = \sigma_i \vec{u}_i}$

We can use this equation to derive the singular value decomposition (SVD) of A:

Write: $V = \underbrace{\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}}_{n\times n \text{ orthogonal matrix}}$, $U = \underbrace{\begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \end{bmatrix}}_{m \times m \text{ orthogonal matrix}}$, $\Sigma = \underbrace{\begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ 0 & & & & \ddots \; 0 \end{bmatrix}}_{m\times n \text{ "diagonal" matrix}}$

Now let's calculate:

$$AV = A\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \underbrace{A\vec{v}_1 \; A\vec{v}_2 \cdots A\vec{v}_r}_{A\vec{v}_i = \sigma_i \vec{u}_i \text{ if } i \le r} & \underbrace{\vec{0} \cdots \vec{0}}_{A\vec{v}_i = \vec{0} \text{ if } i > r.} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & \cdots & \sigma_r\vec{u}_r & \vec{0} & \cdots & \vec{0} \end{bmatrix}}_{}$$

Can write $\vec{0} = 0\vec{u}_{r+1}$, $\vec{0} = 0\vec{u}_{r+2}$ $\cdots$ $\vec{0} = 0\vec{u}_m$

$m \times n$ matrix; multiplying columns by σ_i = multiply on right with an $m \times n$ diagonal-like matrix

$$= \underbrace{\begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \end{bmatrix}}_{m \times m} \underbrace{\begin{bmatrix} \sigma_1 & & & & 0\text{'s} \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0 \ddots \\ 0\text{'s} & & & & & 0 \end{bmatrix}}_{m \times n} = U\Sigma$$

$m \times n$ (we will have columns of 0's on the right if $n > m$, and we will have rows of 0's at the bottom if $m > n$)

We calculated $AV = U\Sigma$. Also, V is orthogonal, $V^{-1} = V^T$.

So $\boxed{A = U\Sigma V^T}$ ← the singular value decomposition (SVD)

SVD shows that any $m \times n$ matrix A can be factored as:

($m \times m$ orthogonal)($m \times n$ diagonal-like)($n \times n$ orthogonal)

↑ columns are the left singular vectors (orthonormal basis of eigenvectors for AA^T

↑ diagonal entries are the singular values (square roots of the positive real eigenvalues of AA^T and A^TA

↑ ~~columns~~ rows are the right singular vectors (orthonormal basis of eigenvectors for A^TA.

Example (Problem 7.2.4)

$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$ Let's find the SVD. First find eigenvalues and eigenvectors of A^TA.

$A^TA = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ Eigenvalues: $\det(A^TA - \lambda I) =$

$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda-3)(\lambda-1) = 0$

So $\lambda_1 = 3 > \lambda_2 = 1 > 0 \rightsquigarrow \sigma_1 = \sqrt{3}, \sigma_2 = 1$ (square roots)

Eigenvectors for $\lambda = 3$: Solve $(A^TA - 3I)\vec{x} = \vec{0}$

$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightsquigarrow x_1 = x_2 \rightsquigarrow \vec{x} = x_2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. The orthonormal basis vector $\vec{v}_1$ should be a unit vector: $\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Eigenvectors for $\lambda = 1$: Solve $(A^TA - I)\vec{x} = \vec{0}$.

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightsquigarrow x_1 + x_2 = 0 \rightsquigarrow \vec{x} = x_2\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

$\vec{v}_2$ should be a unit vector: $\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Note that $\vec{v}_1 \perp \vec{v}_2$, ~~should~~ so $\{\vec{v}_1, \vec{v}_2\}$ is indeed an orthonormal basis of $\mathbb{R}^2$. Get orthogonal matrix $V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Left singular vectors $\vec{u}_1, \vec{u}_2, \vec{u}_3$

$\vec{u}_1 = \frac{1}{\sigma_1}A\vec{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

$\vec{u}_2 = \frac{1}{\sigma_2}A\vec{v}_2 = \frac{1}{1}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

What about $\vec{u}_3$? It needs to a unit basis vector of $N(AA^T) = N(A^T)$

Solve $A^T\vec{x}=\vec{0}$: $\begin{bmatrix}1&1&0\\0&1&1\end{bmatrix} \xrightarrow{\text{Row 1}-\text{Row 2}} \begin{bmatrix}1&0&-1\\0&1&1\end{bmatrix} \rightarrow \begin{matrix}x_1=x_3\\x_2=-x_3\end{matrix}$

So $\vec{x}=\vec{x}_3\begin{bmatrix}1\\-1\\1\end{bmatrix}$. $\vec{u}_3$ needs to be a <u>unit</u> vector: $\vec{u}_3=\frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}$

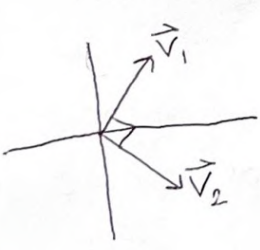
We can now write $U=\begin{bmatrix}1/\sqrt{6}&-1/\sqrt{2}&1/\sqrt{3}\\2/\sqrt{6}&0&-1/\sqrt{3}\\1/\sqrt{6}&1/\sqrt{2}&1/\sqrt{3}\end{bmatrix}$ ← Note that the columns are ⊥

Finally, the SVD is $A=U\Sigma V^T=\begin{bmatrix}1/\sqrt{6}&-1/\sqrt{2}&1/\sqrt{3}\\2/\sqrt{6}&0&-1/\sqrt{3}\\1/\sqrt{6}&1/\sqrt{2}&1/\sqrt{3}\end{bmatrix}\begin{bmatrix}\sqrt{3}&0\\0&1\\0&0\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&1/\sqrt{2}\\-1/\sqrt{2}&1/\sqrt{2}\end{bmatrix}$

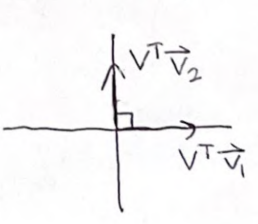
Check this is correct: $U\Sigma V^T=\begin{bmatrix}1/\sqrt{2}&-1/\sqrt{2}\\2/\sqrt{2}&0\\1/\sqrt{2}&1/\sqrt{2}\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&1/\sqrt{2}\\-1/\sqrt{2}&1/\sqrt{2}\end{bmatrix}=\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}=A$ ✓

What are some things we can do with SVD?

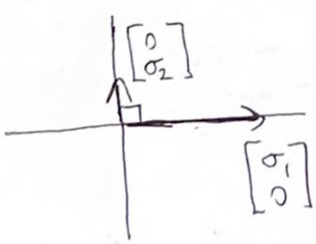
<u>Geometry</u> $A=U\Sigma V^T$



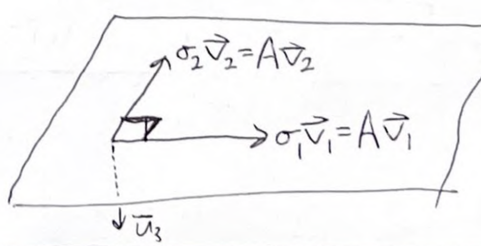
V^T, orthogonal, doesn't change length or angle



Σ, just scales the x and y-axis vectors



U, orthogonal, doesn't change length or angle



So SVD breaks A up into three pieces, and only Σ changes the lengths of vectors.

This gives us a way to measure the "size" of A, i.e., what is the maximum possible amount that A can stretch a vector.

Remember: length of a vector $\vec{x}$: $\|\vec{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$

The size of A, or norm of A, $\|A\|$, is the maximum possible ratio $\|A\vec{x}\| / \|\vec{x}\|$ (where $\vec{x} \neq \vec{0}$)

This is the factor that A stretches $\vec{x}$.

Theorem $\|A\| =$ largest singular value σ_1.

Proof: First let's show $\|A\| \geq \sigma_1$ by showing $\|A\vec{v}_1\| / \|\vec{v}_1\| = \sigma_1$:

$$\|A\vec{v}_1\| / \underbrace{\|\vec{v}_1\|}_{1} = \|U\Sigma V^T \vec{v}_1\| = \left\|U\Sigma \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right\| = \left\|U \begin{bmatrix} \sigma_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right\| = \|\sigma_1 \vec{u}_1\|$$

$$= \sigma_1 \|\vec{u}_1\| = \sigma_1.$$

Now let's show $\|A\| \leq \sigma_1$ by showing $\|A\vec{x}\| / \|\vec{x}\| \leq \sigma_1$ for all $\vec{x} \neq \vec{0}$:

$$\frac{\|A\vec{x}\|}{\|\vec{x}\|} = \frac{\|U\Sigma V^T \vec{x}\|}{\|\vec{x}\|} = \frac{\|\Sigma V^T \vec{x}\|}{\|\vec{x}\|} = \frac{\sqrt{(\sigma_1 (V^T\vec{x})_1)^2 + \dots + (\sigma_r (V^T\vec{x})_r)^2}}{\|\vec{x}\|}$$

Because U is orthogonal, doesn't change length.

Because σ_1 is the biggest singular value

Because V^T is orthogonal, doesn't change lengths.

$$\leq \frac{\sqrt{\sigma_1^2 \left((V^T\vec{x})_1^2 + \dots + (V^T\vec{x})_n^2 \right)}}{\|\vec{x}\|} = \frac{\sigma_1 \|V^T \vec{x}\|}{\|\vec{x}\|} = \frac{\sigma_1 \|\vec{x}\|}{\|\vec{x}\|} = \sigma_1$$

So $\|A\| \geq \sigma_1$ and $\|A\| \leq \sigma_1$ $\rightsquigarrow$ $\|A\| = \sigma_1$ ✓

This shows that σ_1 is the maximum amount that A stretches vectors, and that the vectors that get stretched the most are in span$(\vec{v}_1)$.

Example ~~[illegible]~~ $A=\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}$. We saw that the biggest singular vector is $\sigma_1=\sqrt{3}$, and $\vec{v}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$.

So for example $\begin{bmatrix}1\\1\end{bmatrix}$ gets stretched by a factor of $\sqrt{3}$:

$$\frac{\|A\vec{x}\|}{\|\vec{x}\|}=\frac{\left\|\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\right\|}{\left\|\begin{bmatrix}1\\1\end{bmatrix}\right\|}=\frac{\left\|\begin{bmatrix}1\\2\\1\end{bmatrix}\right\|}{\left\|\begin{bmatrix}1\\1\end{bmatrix}\right\|}=\frac{\sqrt{1+4+1}}{\sqrt{1+1}}=\frac{\sqrt{6}}{\sqrt{2}}=\sqrt{3}$$

Another application: SVD gives a good way writing a rank r matrix A as a sum of r rank-1 matrices:

Theorem $A=\sigma_1\vec{u}_1\vec{v}_1^T+\sigma_2\vec{u}_2\vec{v}_2^T+\dots+\sigma_r\vec{u}_r\vec{v}_r^T$

"outer product," column vector x row vector, has rank = 1 because every row is a multiple of $\vec{v}_1^T$.

Proof Two matrices A and B are equal if $A\vec{x}=B\vec{x}$ for every $\vec{x}$ in $\mathbb{R}^n$.

Here $\{\vec{v}_1,\dots,\vec{v}_n\}$ is a basis, so can write $\vec{x}=c_1\vec{v}_1+c_2\vec{v}_2+\dots+c_n\vec{v}_n$.

Then $(\sigma_1\vec{u}_1\vec{v}_1^T+\dots+\sigma_r\vec{u}_r\vec{v}_r^T)\underbrace{(c_1\vec{v}_1+\dots+c_n\vec{v}_n)}_{\vec{x}}=$

$\sigma_1c_1\vec{u}_1\vec{v}_1^T\vec{v}_1+\dots+\sigma_rc_r\vec{u}_r\vec{v}_r^T\vec{v}_r+$ a bunch of 0's (because $\vec{v}_i^Tv_j=0$ if $i\neq j$.)

$=c_1\underbrace{(\sigma_1\vec{u}_1)}_{A\vec{v}_1}+\dots+c_r\underbrace{(\sigma_r\vec{u}_r)}_{A\vec{v}_r}+c_{r+1}\underbrace{\vec{0}}_{Av_{r+1}}+\dots+c_n\underbrace{\vec{0}}_{A\vec{v}_n}$

$=A(c_1\vec{v}_1+\dots+c_r\vec{v}_r+c_{r+1}\vec{v}_{r+1}+\dots+c_n\vec{v}_n)=A\vec{x}$ ✓

Example If $A=\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}$, then $A=\sigma_1\vec{u}_1\vec{v}_1^T+\sigma_2\vec{u}_2\vec{v}_2^T$

$$=\underset{\sigma_1}{\sqrt{3}}\underset{\vec{u}_1}{\begin{bmatrix}1/\sqrt{6}\\2/\sqrt{6}\\1/\sqrt{6}\end{bmatrix}}\underset{\vec{v}_1^T}{\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2}\end{bmatrix}}+\underset{\sigma_2}{1}\underset{\vec{u}_2}{\begin{bmatrix}-1/\sqrt{2}\\0\\1/\sqrt{2}\end{bmatrix}}\underset{\vec{v}_2}{\begin{bmatrix}-1/\sqrt{2} & 1/\sqrt{2}\end{bmatrix}}$$

$$=\frac{1}{2}\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}1&1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}-1\\0\\1\end{bmatrix}\begin{bmatrix}-1&1\end{bmatrix}=\frac{1}{2}\left(\begin{bmatrix}1&1\\2&2\\1&1\end{bmatrix}+\begin{bmatrix}1&-1\\0&0\\-1&1\end{bmatrix}\right)=\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}=A\ \checkmark$$

Application in image processing: Represent a digital photograph as an $m\times n$ matrix:

$$A=\underbrace{\begin{bmatrix} & & \\ & \cdots\cdots & \\ & & \end{bmatrix}}_{n}\Big\}\, m$$

mn pixels in the image

$$\text{Or } A=\sigma_1\underset{\substack{\uparrow\\ m \text{ components}}}{\vec{u}_1}\overset{\substack{n\text{ components}\\ \downarrow}}{\vec{v}_1^T}+\cdots+\sigma_r\vec{u}_r\vec{v}_r^T$$

$r(m+n)$ vector components

If r is much smaller than m,n, then it's more efficient to transmit or store $r(m+n)$ vector components than mn matrix entries.

<u>However</u>, even if r is <u>not</u> much smaller than m or n, many of the singular values σ_i are often very small. So we may be able to write

$$A\approx\sigma_1\vec{u}_1\vec{v}_1^T+\cdots+\sigma_s\vec{u}_s\vec{v}_s^T$$

where s <u>is</u> much smaller than m and n.

Image compression. We lose a little information by throwing out some terms, but it might not make a difference.