Thm: $A \in F^{m \times n}$ has SVD $A = W \Sigma V^*$ with
- $V^* \in O_n(F)$
- $\Sigma \in F^{m \times n}$ all $0$ except $\sigma_1, \ldots, \sigma_r$ on main diagonal
- $W \in O_m(F)$.

**Pf:** Complete bases in Schmidt decomposition to orthonormal bases of

ker $A$ (for $V^*$) and ker $A^*$ (for $W$).

**Cor:** $A \in F^{n \times n} \Rightarrow A = U |A|$ for some $U \in O_n(F)$.

**Pf:** $A = W \Sigma V^* = W \Sigma V^* U |A| |

\[ A^* A = \sum_{i=1}^r \sigma_i^2 v_i v_i^* = \Sigma^2 V^* = \Sigma W^* W \Sigma V^* \]

Note: SVD efficient numerically: fast + accurate

Q. How big can $\|Ax\|$ be, given that $\|x\| = 1$?

$B = \{ x \in F^n \mid \|x\| = 1 \}$ has image = ?

A. If $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$ with $\sigma_r = \ldots = \sigma_n = 0$ then

$y = \Sigma x$ for $x \in B$ $\Leftrightarrow y_i = \sigma_i x_i$ and $x \in B$

$\Leftrightarrow y_i = \sigma_i x_i$ and $\|x\|^2 + \ldots + \|x_n\|^2 \leq 1$

$\Leftrightarrow y_i = 0$ for $i > r$ and $\|y_1\|^2 + \ldots + \|y_r\|^2 \leq 1$

ellipsoid

- $F = R$: principal axes of length $2\sigma_1, \ldots, 2\sigma_r$ along $e_1, \ldots, e_r$
- $F = C$: $i^{th}$ real and imaginary principal axes of length $2\sigma_i$

A arbitrary $\Rightarrow [\lambda_A] V^* W = \Sigma$ unitary $\Rightarrow$ do not alter $\|$ normal basis

Thm: $A(\text{unit ball}) = \text{ellipsoid in im}(A)$ with principal half-axes along $w_1, \ldots, w_r$ of lengths $\sigma_1, \ldots, \sigma_r$.

**Cor:** $A$ has operator norm $\|A\| = \max_{x \in B} \|Ax\| = \sigma_1$.

**Pf:** $\|Ax\| \leq C \|x\| \forall x$, and $C = \|A\|$ is smallest such.

**Lemma:** $A \mapsto \|A\|$ is a norm on $F^{m \times n}$.

**Compare** Frobenius norm $\|A\|_2 = \sqrt{\text{tr}(A^*A)}$
(or Hilbert-Schmidt)

**Prop:** $\|A\|_2 \leq \|A\|_2$.

**Pf:** $\|A\| = \sigma_1 \leq \sigma_1^2 + \ldots + \sigma_n^2 = \|A\|_2$ since $\text{tr}(A^*A) = \Sigma$ eigenvalues$(A^*A)$. □
Eckart-Young Thm: Given $A \in \mathbb{R}^{m \times n}$, the $\hat{A}$ of rank $\leq k$ minimizing $\|A - \hat{A}\|_2$ is

$$
\hat{W} \hat{\Sigma} \hat{V}^* \\
\begin{bmatrix}
W \\
\Sigma \\
V^*
\end{bmatrix}
$$

Pf: omitted for time though we could totally do it

Principal Component Analysis (PCA)

$$
A = \begin{bmatrix}
A_1 \\
\vdots \\
A_m
\end{bmatrix} \leftrightarrow \text{m points in } \mathbb{R}^n
$$

PC 1 = direction $v_1 \in \mathbb{F}^n_{col}$ maximizing sample variance: $\|A_1 v_1\|^2 + \cdots + \|A_m v_1\|^2$

$\hat{A}_1 = \text{project rows of } A \text{ orthogonally to } v_1$

PC 2 = direction $v_2 \in v_1^\perp \subseteq \mathbb{F}^n_{col}$ maximizing $\hat{A}_1$ - sample variance

$\hat{A}_2 = \hat{A}_1 / v_2$

Def: The PC decomposition of $A$ is $T = AV$, where $V$ has columns $v_1, \ldots, v_n$.

ij entry is score of sample i along PC j.

Interpretation: cols($V$) $\leftrightarrow$ alternative features $\cdot$ linear combinations of original features

$\cdot$ explain variance in uncorrelated ($\perp$) way

Thm: $A = W \Sigma V^* \Rightarrow v_1, \ldots, v_n$ are the columns of $V$ and

$T = W \Sigma$ is polar decomposed

Pf: (sample variance in direction $v$) $= \|Av\|^2$ for $v \in \mathbb{F}^n_{col}$.

$v$ maximizes $\|Av\|^2 \iff v \rightarrow$ longest principal axis! (by Cor: $\|A\| = \sigma_1$)

$\Rightarrow V$ is SVD: cols($V$) $\perp$ normal basis of eigenvectors of $A^*A$ (by induction)

$\Rightarrow T = AV = W \Sigma V^* V = W \Sigma$. □

PCA $\iff$ low-rank projection of data: use only PC 1, ..., PC k

$\mathbb{F}^n \perp \mathbb{F}^k$

Note: PC 1, PC 2, ..., PC n $\rightarrow$ flag of best approximating subspaces.