Recap: spectral norm of matrix

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A)$$

Proof: Use SVD

Recap: spectral norm of matrix

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A)$$

Proof: Use SVD

$$\begin{aligned}
\|Ax\|_2 &= \|U\Sigma V^T x\|_2 \\
&= \|\Sigma V^T x\|_2 \quad \text{by unitary invariance} \\
&= \|\Sigma y\|_2 \quad \text{with } \|y\|_2 = 1 \\
&= \sqrt{\sum_{i=1}^n \sigma_i^2 y_i^2} \\
&\leq \sqrt{\sum_{i=1}^n \sigma_1^2 y_i^2} = \sigma_1 \|y\|_2^2 = \sigma_1.
\end{aligned}$$

Recap: spectral norm of matrix

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A)$$

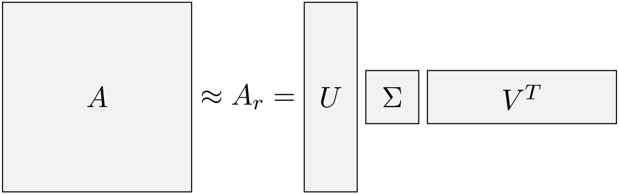
Proof: Use SVD

$$\begin{aligned}
\|Ax\|_2 &= \|U\Sigma V^T x\|_2 \\
&= \|\Sigma V^T x\|_2 \quad \text{by unitary invariance} \\
&= \|\Sigma y\|_2 \quad \text{with } \|y\|_2 = 1 \\
&= \sqrt{\sum_{i=1}^n \sigma_i^2 y_i^2} \\
&\le \sqrt{\sum_{i=1}^n \sigma_1^2 y_i^2} = \sigma_1 \|y\|_2^2 = \sigma_1.
\end{aligned}$$

Frobenius norm: $\|A\|_F = \sqrt{\sum_i \sum_j |A_{ij}|^2} = \sqrt{\sum_{i=1}^n (\sigma_i(A))^2}$ (exercise)

Low-rank approximation of a matrix

Given $A \in \mathbb{R}^{m \times n}$, find A_r such that



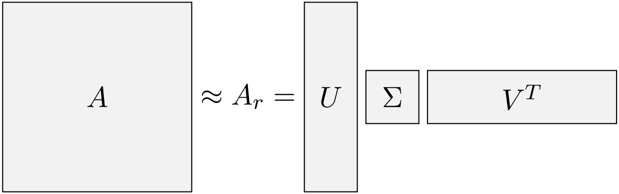
- Storage savings (data compression)

Optimal low-rank approximation by SVD

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\mathrm{rank}(B)=r} \|A - B\|_2$$

- Good approximation if $\sigma_{r+1} \ll \sigma_1$:



- Optimality holds for any unitarily invariant norm
- Prominent application: PCA
- Many matrices have explicit or hidden low-rank structure (nonexaminable)

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\text{rank}(B)=r} \|A - B\|_2$$

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\text{rank}(B)=r} \|A - B\|_2$$

- Since $\operatorname{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\text{rank}(B)=r} \|A - B\|_2$$

- Since $\text{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.
- There exists orthonormal $W \in \mathbb{C}^{n\times(n-r)}$ s.t. $BW = 0$. Then $\|A - B\|_2 \ge \|(A - B)W\|_2 = \|AW\|_2 = \|U\Sigma(V^*W)\|_2$.

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\mathsf{rank}(B)=r} \|A - B\|_2$$

- Since $\mathrm{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.
- There exists orthonormal $W \in \mathbb{C}^{n \times (n-r)}$ s.t. $BW = 0$. Then $\|A - B\|_2 \ge \|(A - B)W\|_2 = \|AW\|_2 = \|U\Sigma(V^*W)\|_2$.
- Now since W is $(n-r)$-dimensional, there is is an intersection between W and $[v_1, \ldots, v_{r+1}]$, the $(r+1)$-dimensional subspace spanned by the leading $r+1$ left singular vectors ($[W, v_1, \ldots, v_{r+1}]\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = 0$ has a solution; then Wx_1 is such a vector).

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$

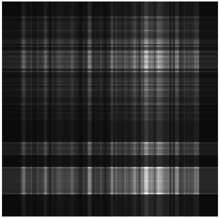
$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\mathsf{rank}(B)=r} \|A - B\|_2$$

- Since $\operatorname{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.
- There exists orthonormal $W \in \mathbb{C}^{n\times(n-r)}$ s.t. $BW = 0$. Then $\|A - B\|_2 \ge \|(A - B)W\|_2 = \|AW\|_2 = \|U\Sigma(V^*W)\|_2$.
- Now since W is $(n-r)$-dimensional, there is is an intersection between W and $[v_1, \ldots, v_{r+1}]$, the $(r+1)$-dimensional subspace spanned by the leading $r+1$ left singular vectors ($[W, v_1, \ldots, v_{r+1}]\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = 0$ has a solution; then Wx_1 is such a vector).
- Then scale x_1 to have unit norm, and $\|U\Sigma V^* W x_1\|_2 = \|U\Sigma_{r+1} y_1\|_2$, where $\|y_1\|_2 = 1$ and Σ_{r+1} is the leading $r+1$ part of Σ. Then $\|U\Sigma_{r+1} y_1\|_2 \ge \sigma_{r+1}$ can be verified directly.

Low-rank approximation: image compression

grayscale image=matrix

original



rank 1

rank 5

rank 10

rank 20

rank 50