

Singular Value Decomposition (SVD)

Shusen Wang

Orthonormal Basis

Definition (Orthonormal Basis).

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} \subset \mathbb{R}^n$ form an orthonormal basis if:

- unit ℓ_2 -norm: $\|\mathbf{v}_i\|_2 = 1$ for all i ,
- orthogonal to each other: $\mathbf{v}_i^T \mathbf{v}_j = 0$ for all $i \neq j$.

SVD and Truncated SVD

Singular Value Decomposition (SVD)

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be any matrix.
- Rank: $r = \text{rank}(\mathbf{A})$. ($r \leq m, n$)
- SVD: $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$



A rank-1 matrix

Singular Value Decomposition (SVD)

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be any matrix.
- Rank: $r = \text{rank}(\mathbf{A})$. ($r \leq m, n$)
- SVD: $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
 - Singular values: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$.
 - Left singular vectors: $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\} \subset \mathbb{R}^m$ forms an orthonormal basis.
 - Right singular vectors: $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\} \subset \mathbb{R}^n$ forms an orthonormal basis.

Truncated SVD

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be any matrix.
- Rank: $r = \text{rank}(\mathbf{A})$. ($r \leq m, n$)
- SVD: $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- **Truncated SVD's definition:** for $0 < k < r$, $\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

Truncated SVD

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be any matrix.
- Rank: $r = \text{rank}(\mathbf{A})$. ($r \leq m, n$)
- SVD: $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- **Truncated SVD's definition:** for $0 < k < r$, $\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- **Truncated SVD's properties:**
 - \mathbf{A}_k is the best rank- k approximation to \mathbf{A} .
 - $\mathbf{A}_k = \underset{\mathbf{B}}{\text{argmin}} \|\mathbf{A} - \mathbf{B}\|_F^2$; s.t. $\text{rank}(\mathbf{B}) \leq k$.

Matrix Frobenius Norm

- Let $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m \times n}$ be any matrix.
- Frobenius norm: $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$.
 - A generalization of the ℓ_2 -vector norm to matrix.
- Property: $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$.
 - σ_i is the i -th singular value of \mathbf{A} .

Matrix Rank & Singular Values

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be any matrix. (Suppose $m \leq n$).
- $r = \text{rank}(\mathbf{A})$. ($r \leq m$)
- Singular values:

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_m.$
- $\text{rank}(\mathbf{A})$ equals # of positive singular values.

Error of Truncated SVD

- SVD: $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- Truncated SVD: for $0 < k < r$, $\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$.
- Error of the truncated SVD:

$$\| \mathbf{A} - \mathbf{A}_k \|_F^2 = \underbrace{\sum_{i=k+1}^r \sigma_i^2}_{\text{Bottom (the smallest) singular values}}.$$

Bottom (the smallest) singular values

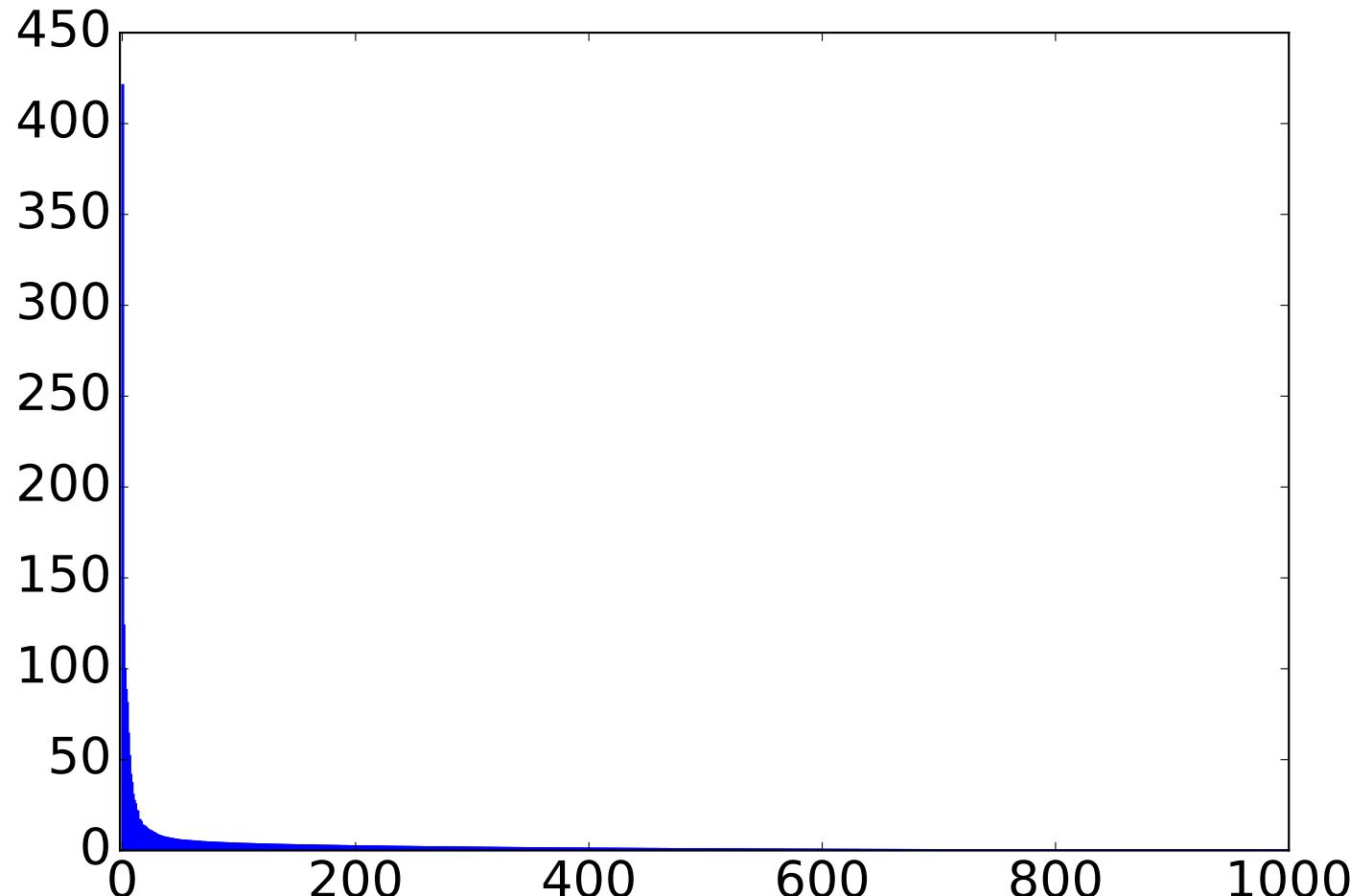
SVD: Example

Original image (1000×1500)



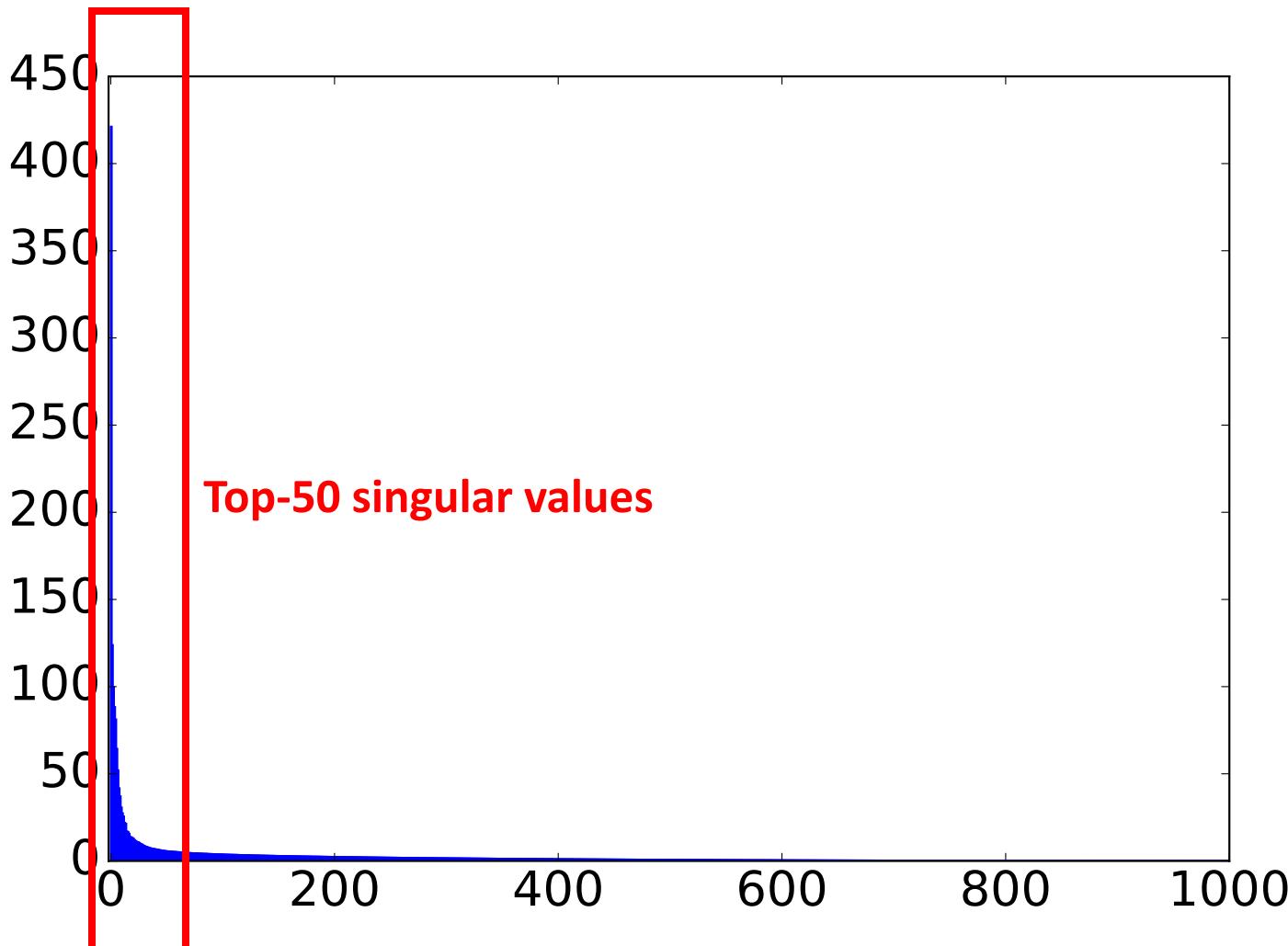
SVD: Example

Singular values



Indices

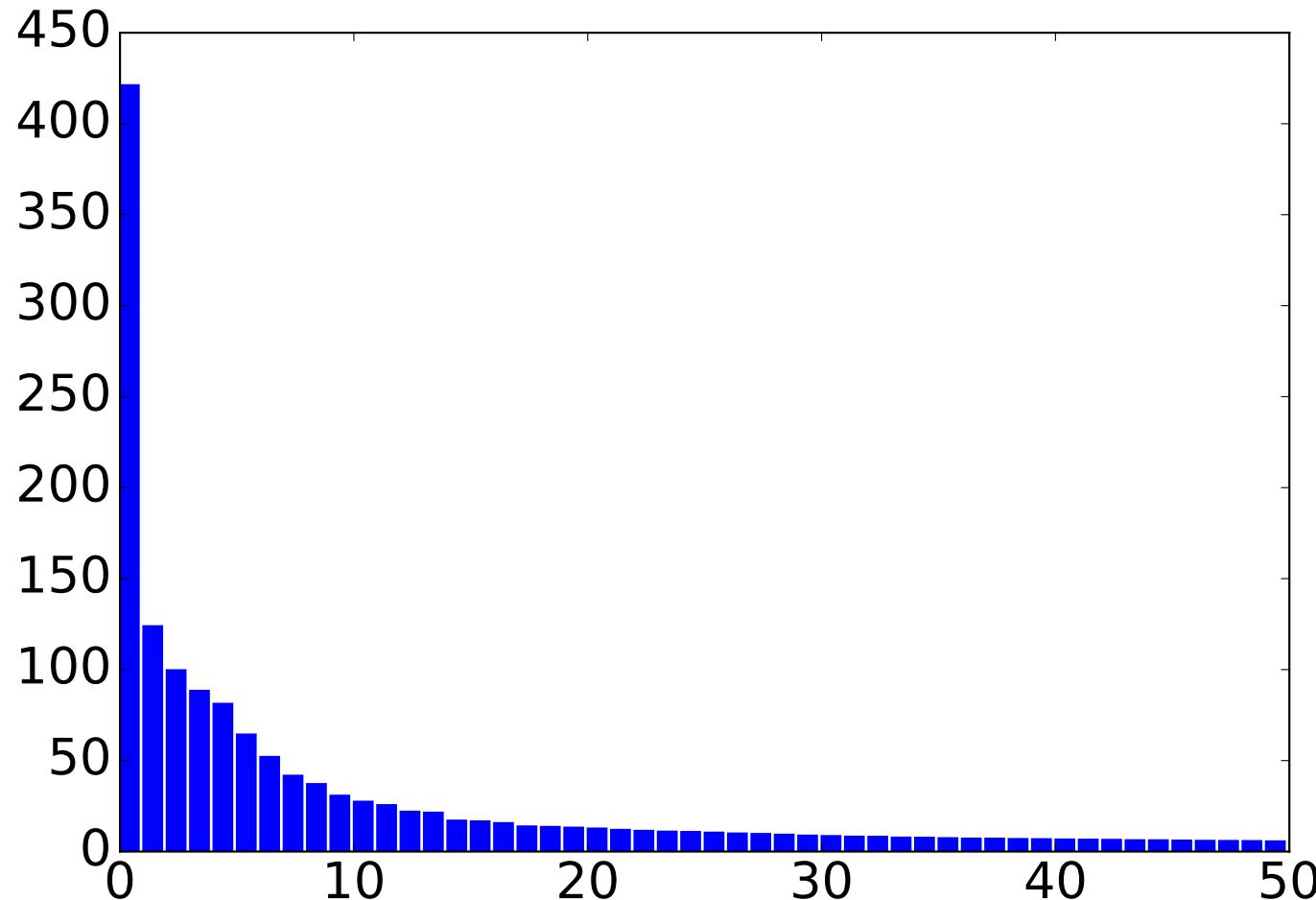
SVD: Example



Indices

SVD: Example

Singular values



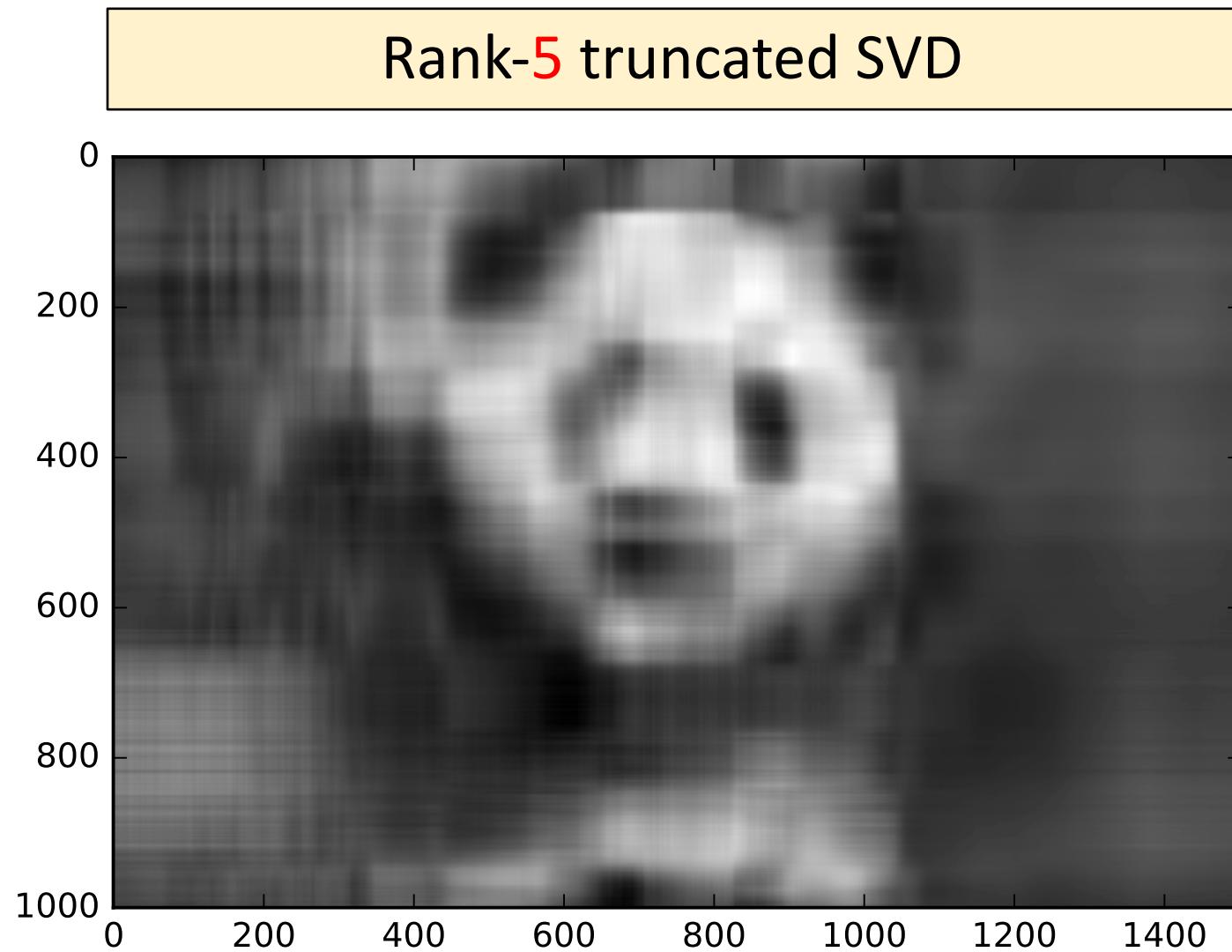
Indices (first 50)

SVD: Example

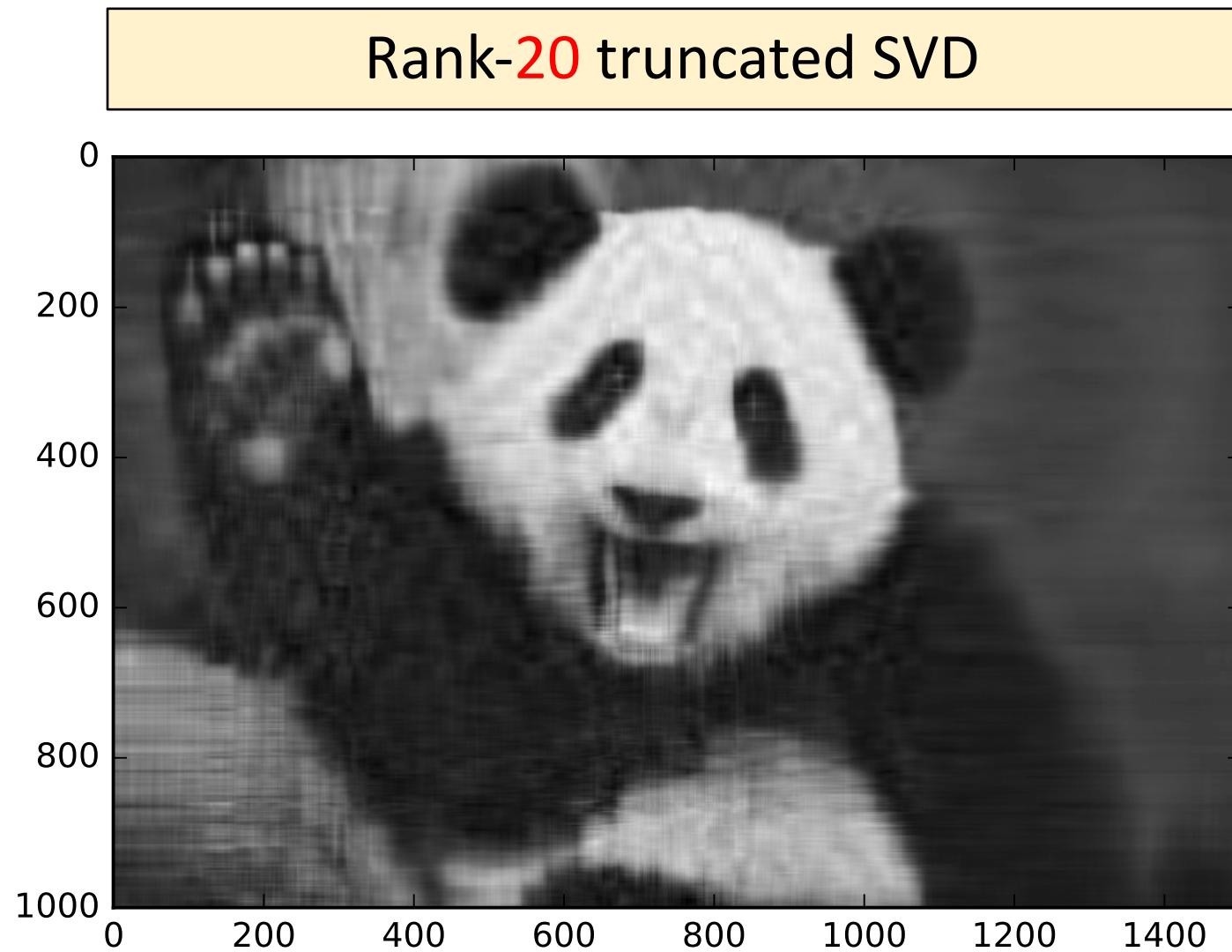
Original image (1000×1500)



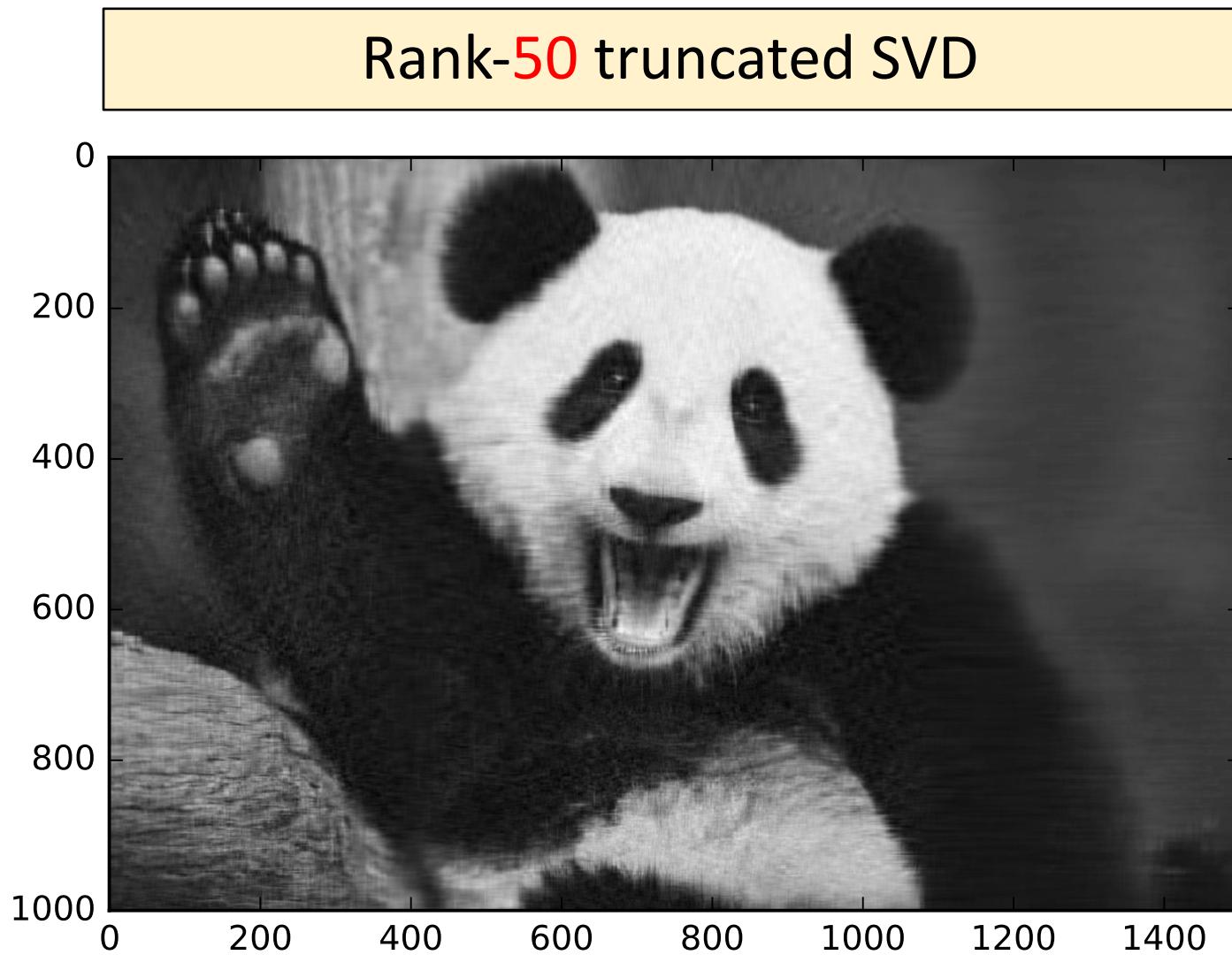
SVD: Example



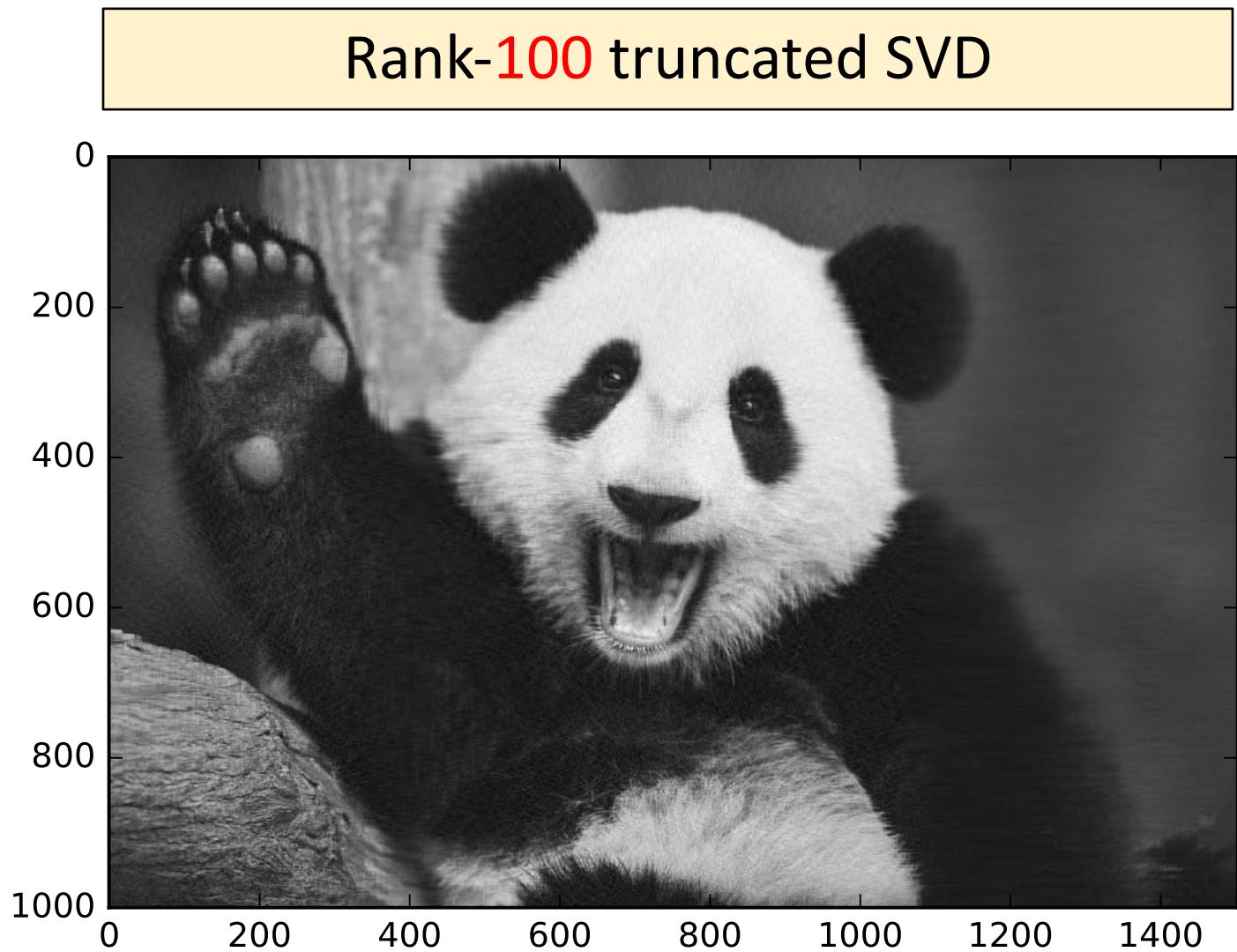
SVD: Example



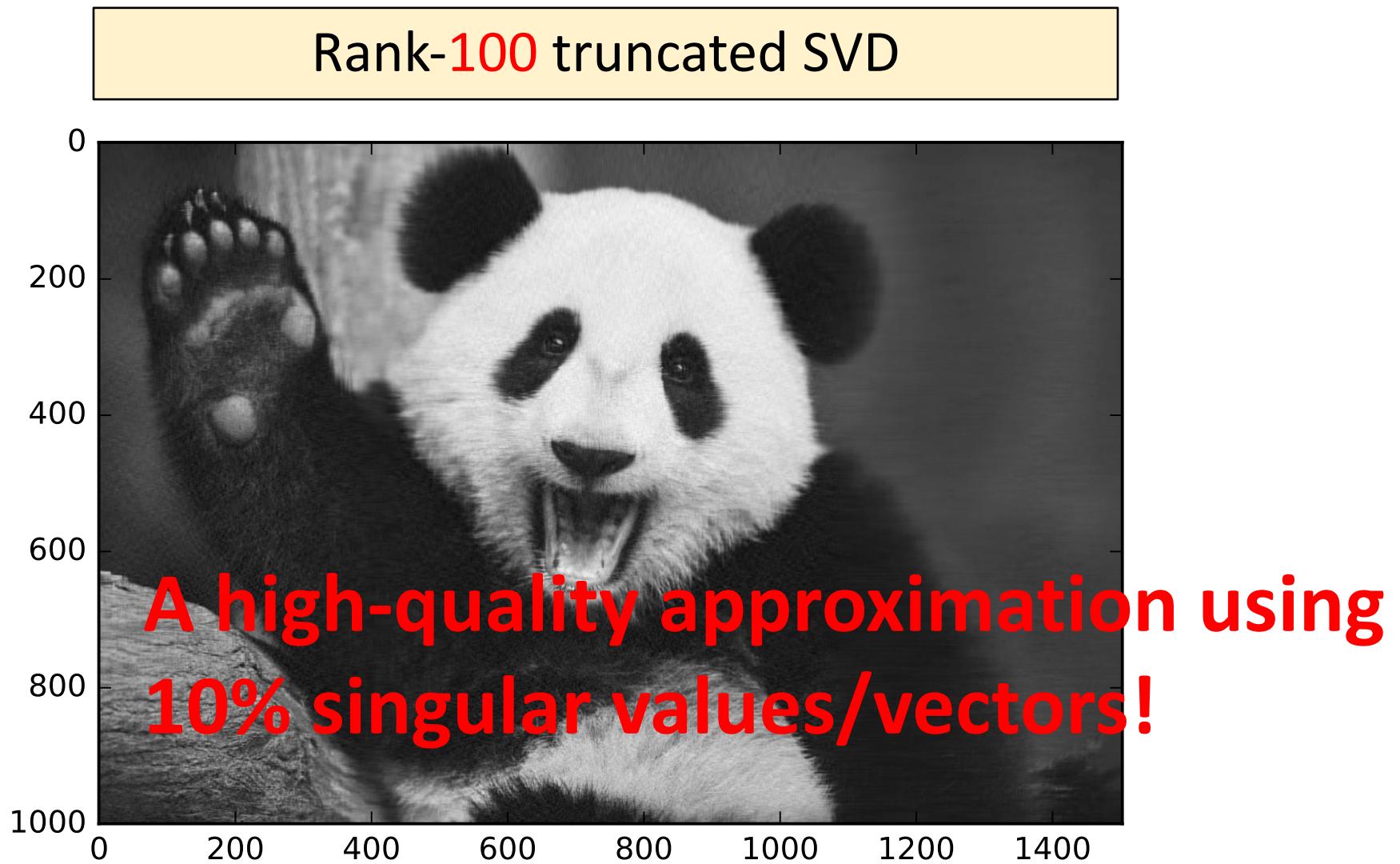
SVD: Example



SVD: Example



SVD: Example



SVD: Example

- The original matrix
 - Shape: 1000×1500
 - #Entries: 1.5M
- The rank-100 truncated SVD
 - Shape: 100×1 , 100×1500 , and 100×1000
 - #Entries: 0.25M
- Truncated SVD saves 83% storage

Power Iteration for Computing Truncated SVD

A Property

Theorem. If $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the SVD of \mathbf{A} , then $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Proof.

- $\mathbf{A}^T \mathbf{A} = (\sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T)^T (\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T) = (\sum_{i=1}^r \sigma_i \mathbf{v}_i \mathbf{u}_i^T) (\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T)$.

A Property

Theorem. If $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the SVD of \mathbf{A} , then $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Proof.

- $\mathbf{A}^T \mathbf{A} = (\sum_{i=1}^r \sigma_i \mathbf{v}_i \mathbf{u}_i^T)(\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T)$.
- $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{u}_i^T \mathbf{u}_i \mathbf{v}_i^T + \sum_{i \neq j} \sigma_i \sigma_j \mathbf{v}_i \mathbf{u}_i^T \mathbf{u}_j \mathbf{v}_j^T$.

Using $(\sum_i \mathbf{X}_i^T)(\sum_j \mathbf{X}_j) = \sum_i \mathbf{X}_i^T \mathbf{X}_i + \sum_{i \neq j} \mathbf{X}_i^T \mathbf{X}_j$.

A Property

Theorem. If $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the SVD of \mathbf{A} , then $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Proof.

- $\mathbf{A}^T \mathbf{A} = (\sum_{i=1}^r \sigma_i \mathbf{v}_i \mathbf{u}_i^T)(\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T)$.
- $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \boxed{\mathbf{u}_i^T \mathbf{u}_i} \mathbf{v}_i^T + \sum_{i \neq j} \sigma_i \sigma_j \mathbf{v}_i \boxed{\mathbf{u}_i^T \mathbf{u}_j} \mathbf{v}_j^T$.
- $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i 1 \mathbf{v}_i^T + \sum_{i \neq j} \sigma_i \sigma_j \mathbf{v}_i 0 \mathbf{v}_j^T$.

Using the properties of orthonormal basis: $\mathbf{u}_i^T \mathbf{u}_i = 1$ and $\mathbf{u}_i^T \mathbf{u}_j = 0$ for $i \neq j$.

A Property

Theorem. If $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the SVD of \mathbf{A} , then $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Proof.

- $\mathbf{A}^T \mathbf{A} = (\sum_{i=1}^r \sigma_i \mathbf{v}_i \mathbf{u}_i^T)(\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T)$.
- $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{u}_i^T \mathbf{u}_i \mathbf{v}_i^T + \sum_{i \neq j} \sigma_i \sigma_j \mathbf{v}_i \mathbf{u}_i^T \mathbf{u}_j \mathbf{v}_j^T$.
- $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i 1 \mathbf{v}_i^T + \sum_{i \neq j} \sigma_i \sigma_j \mathbf{v}_i 0 \mathbf{v}_j^T$.
- $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$

A Property

Theorem. If $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the SVD of \mathbf{A} , then $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.



Eigenvalue decomposition of $\mathbf{A}^T \mathbf{A}$.

Implication: To compute the top singular values σ_i and right singular vectors \mathbf{v}_i , we can do **eigenvalue decomposition** instead of SVD.

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Power Iteration for Truncated SVD

Goal: Compute the top **1** eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \|\mathbf{x}_q\|_2$.



Merely 2 matrix-vector multiplications.
Very Cheap!

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$.
- Every vector can be written as a linear combination of the orthonormal basis.

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- Every vector can be written as a linear combination of the orthonormal basis.
- Because \mathbf{x}_0 is randomly initialized, $|\alpha_1| = \mathbf{x}_0^T \mathbf{v}_1 = \Omega(1/n)$ with high probability.

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto (\mathbf{A}^T \mathbf{A})^q \mathbf{x}_0$

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto (\mathbf{A}^T \mathbf{A})^q \mathbf{x}_0$

Power Iteration for Truncated SVD

Goal: Compute the top 1 eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \|\mathbf{x}_q\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto (\mathbf{A}^T \mathbf{A})^q \mathbf{x}_0 = \left(\sum_{i=1}^r \sigma_i^{2q} \mathbf{v}_i \mathbf{v}_i^T \right) \left(\sum_{j=1}^n \alpha_j \mathbf{v}_j \right)$.

- It can be proved that $(\mathbf{A}^T \mathbf{A})^q = \sum_{i=1}^r \sigma_i^{2q} \mathbf{v}_i \mathbf{v}_i^T$.

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto (\mathbf{A}^T \mathbf{A})^q \mathbf{x}_0 = (\sum_{i=1}^r \sigma_i^{2q} \mathbf{v}_i \mathbf{v}_i^T) (\sum_{j=1}^n \alpha_j \mathbf{v}_j) = \sum_{i=1}^r \alpha_i \sigma_i^{2q} \mathbf{v}_i$.

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto \sum_{i=1}^r \alpha_i \sigma_i^{2q} \mathbf{v}_i$.

Power Iteration for Truncated SVD

Goal: Compute the top $\textcolor{brown}{1}$ eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto \sum_{i=1}^r \alpha_i \sigma_i^{2q} \mathbf{v}_i$.
- $\mathbf{x}_q \propto \sum_{i=1}^r \alpha_i \left(\frac{\sigma_i}{\sigma_1} \right)^{2q} \mathbf{v}_i$

Power Iteration for Truncated SVD

Goal: Compute the top 1 eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \|\mathbf{x}_q\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto \sum_{i=1}^r \alpha_i \sigma_i^{2q} \mathbf{v}_i$.
- $\mathbf{x}_q \propto \sum_{i=1}^r \alpha_i \left(\frac{\sigma_i}{\sigma_1}\right)^{2q} \mathbf{v}_i = \alpha_1 \mathbf{v}_1 + \sum_{i=2}^r \alpha_i \left(\frac{\sigma_i}{\sigma_1}\right)^{2q} \mathbf{v}_i$.

Power Iteration for Truncated SVD

Goal: Compute the top 1 eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a vector \mathbf{x}_0 (with unit ℓ_2 -norm);
2. Repeat the power iteration: $\mathbf{x}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{x}_{q-1}$ and $\mathbf{x}_q \leftarrow \mathbf{x}_q / \|\mathbf{x}_q\|_2$.

Convergence Analysis (\mathbf{x}_q converges to \mathbf{v}_1):

- $\mathbf{x}_0 = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Here $|\alpha_1| = \Omega(1/n)$ with high probability.
- $\mathbf{x}_q \propto \sum_{i=1}^r \alpha_i \sigma_i^{2q} \mathbf{v}_i$.
- $\mathbf{x}_q \propto \sum_{i=1}^r \alpha_i \left(\frac{\sigma_i}{\sigma_1}\right)^{2q} \mathbf{v}_i = \underbrace{\alpha_1 \mathbf{v}_1 + \sum_{i=2}^r \alpha_i \left(\frac{\sigma_i}{\sigma_1}\right)^{2q} \mathbf{v}_i}_{\text{It converge to 0 because } \frac{\sigma_i}{\sigma_1} < 1}$.

It converge to 0 because $\frac{\sigma_i}{\sigma_1} < 1$.

Power Iteration for Truncated SVD

Goal: Compute the top k eigenvalue/eigenvector of $\mathbf{A}^T \mathbf{A} = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$.

Algorithm:

1. Randomly initialize a matrix $\mathbf{X}_0 \in \mathbb{R}^{n \times k}$ (entries are i.i.d. standard Gaussian);
2. Orthogonalize the columns: $\mathbf{X}_0 \leftarrow \text{orth}(\mathbf{X}_0)$;
3. Repeat the power iteration:
 - i. $\mathbf{X}_q \leftarrow \mathbf{A}^T \mathbf{A} \mathbf{X}_{q-1}$;
 - ii. $\mathbf{X}_q \leftarrow \text{orth}(\mathbf{X}_q)$.

Summary

- SVD: $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- Truncated SVD: abandon the bottom singular values/vectors.
- Power iteration (algorithm) for computing truncated SVD.