特征值分解（EVD）

为什么一个向量和一个数相乘的效果与一个矩阵和一个向量相乘的效果是一样的呢？

矩阵变换遵循：左乘是进行初等行变换，右乘是进行初等列变换

Singular Value Decomposition (SVD)

奇异值分解

注意到要进行特征分解，矩阵 必须为方阵。那么如果 不是方阵，即行和列不相同时，此时需要借助SVD对矩阵进行分解。

Shusen Wang

ATA的特征向量组成的就是我们SVD中的U矩阵

AAT的特征向量组成的矩阵就是我们SVD中的V矩阵

https://blog.csdn.net/qfikh/article/details/103994319

Orthonormal Basis

标准正交基

Definition (Orthonormal Basis).

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

- unit ℓ_2-norm: $\left\|\mathbf{v}_i\right\|_2 = 1$ 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: $\mathbf{A} = \sum_{i=1}^{r} \boxed{\sigma_i \, \mathbf{u}_i \, \mathbf{v}_i^T}$

A rank-1 matrix

Singular Value Decomposition (SVD)

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

Truncated SVD

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be any matrix.
- Rank: $r = \operatorname{rank}(\mathbf{A})$. $(r \leq m, n)$
- 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: $r = \mathrm{rank}(\mathbf{A})$. $(r \leq m, n)$
- 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 $\mathbf{A}$.
    - $\mathbf{A}_k = \mathrm{argmin}_{\mathbf{B}} \; \left\| \mathbf{A} - \mathbf{B} \right\|_F^2; \quad s.t. \quad \mathrm{rank}(\mathbf{B}) \leq k.$

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:

$$\boxed{\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0} = \sigma_{r+1} = \cdots = \sigma_m.$$

- $\text{rank}(\mathbf{A})$ equals # of positive singular values.

Error of Truncated SVD

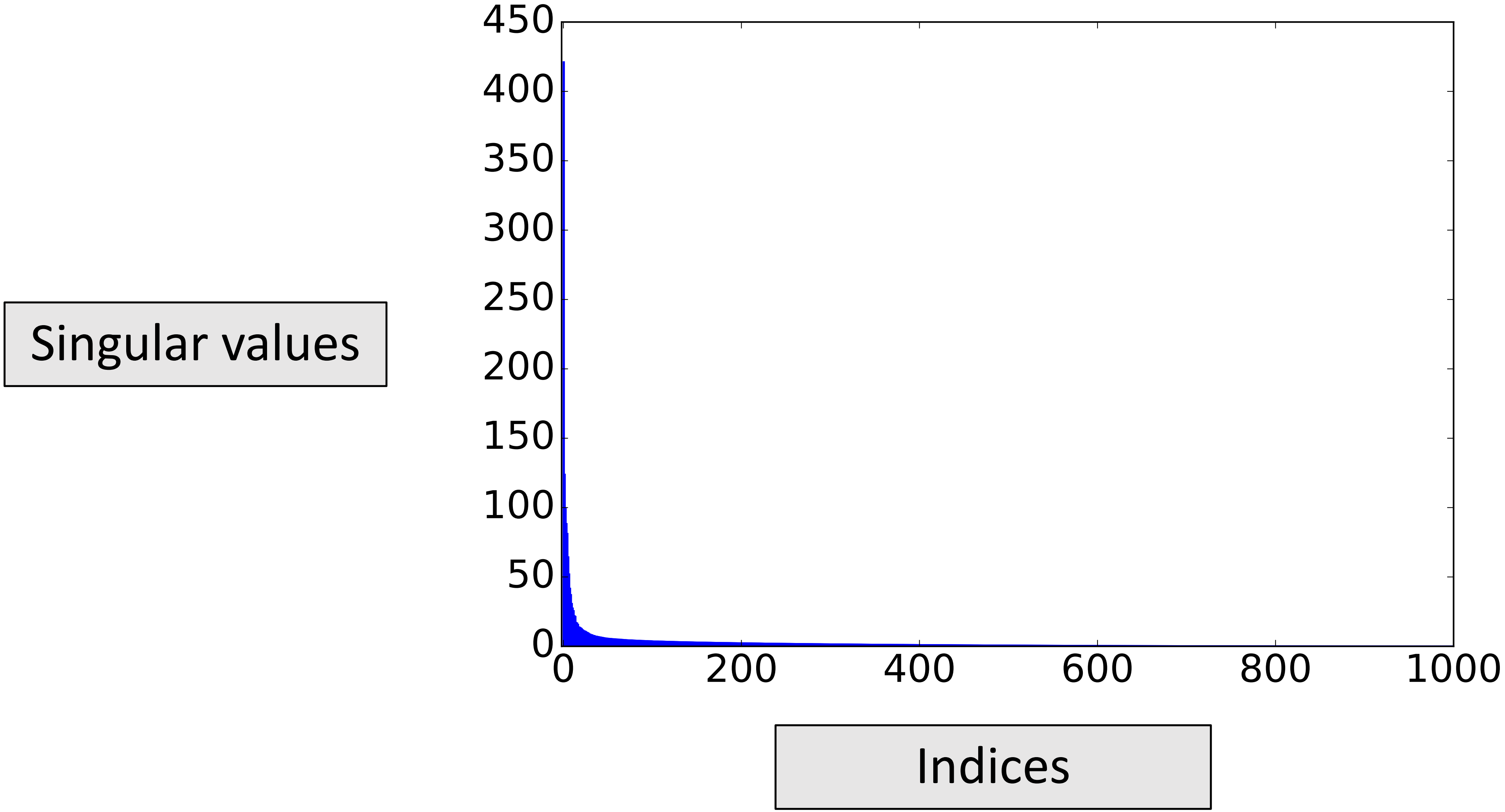
- SVD: $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \, \mathbf{u}_i \, \mathbf{v}_i^T$
- Truncated SVD: for $0 < k < r$, $\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \, \mathbf{u}_i \, \mathbf{v}_i^T$.
- Error of the truncated SVD:

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

SVD: Example

Original image (1000×1500)

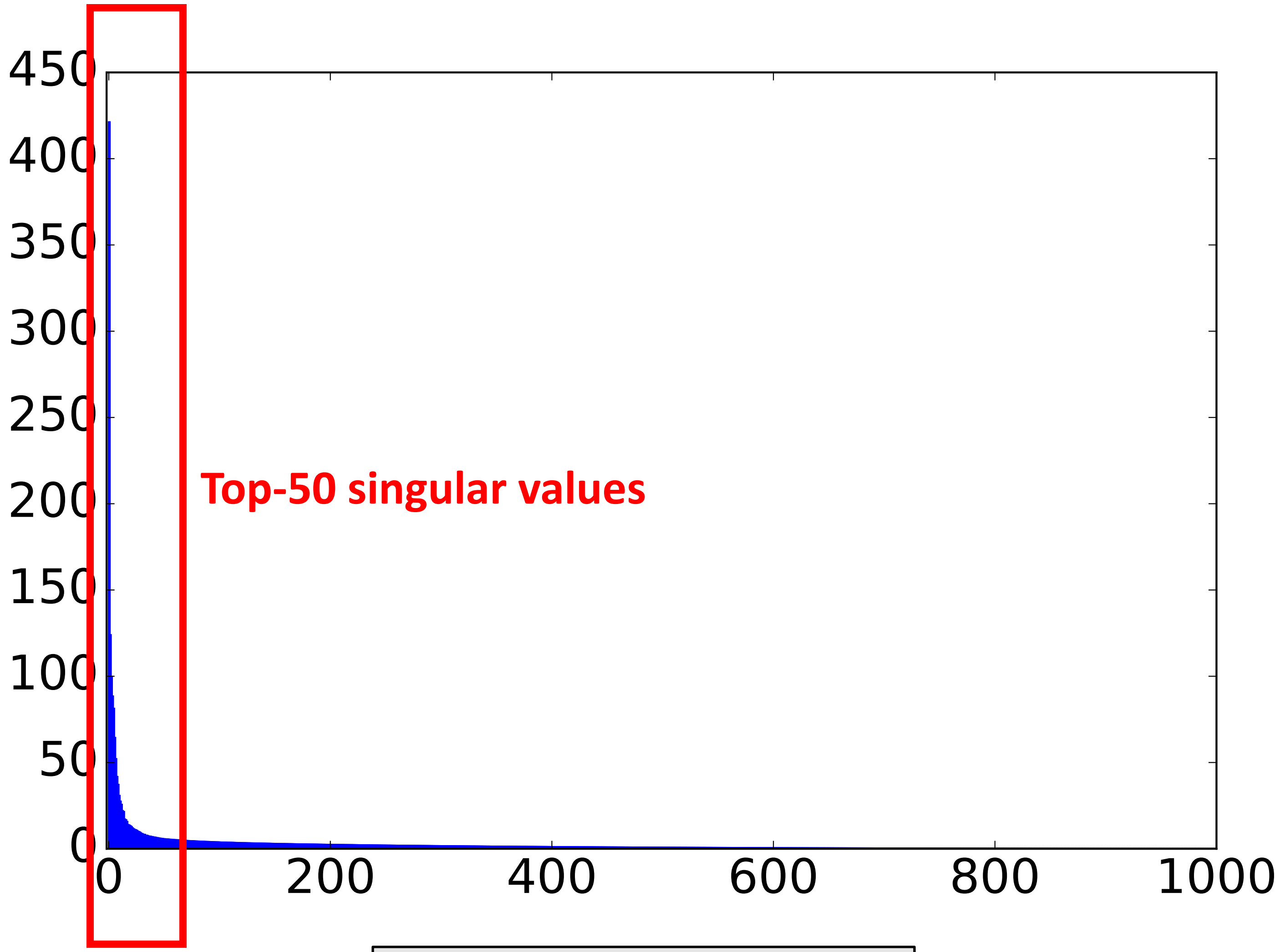
SVD: Example



Singular values
450
400
350
300
250
200
150
100
50
0
0
200
400
600
800
1000
Indices

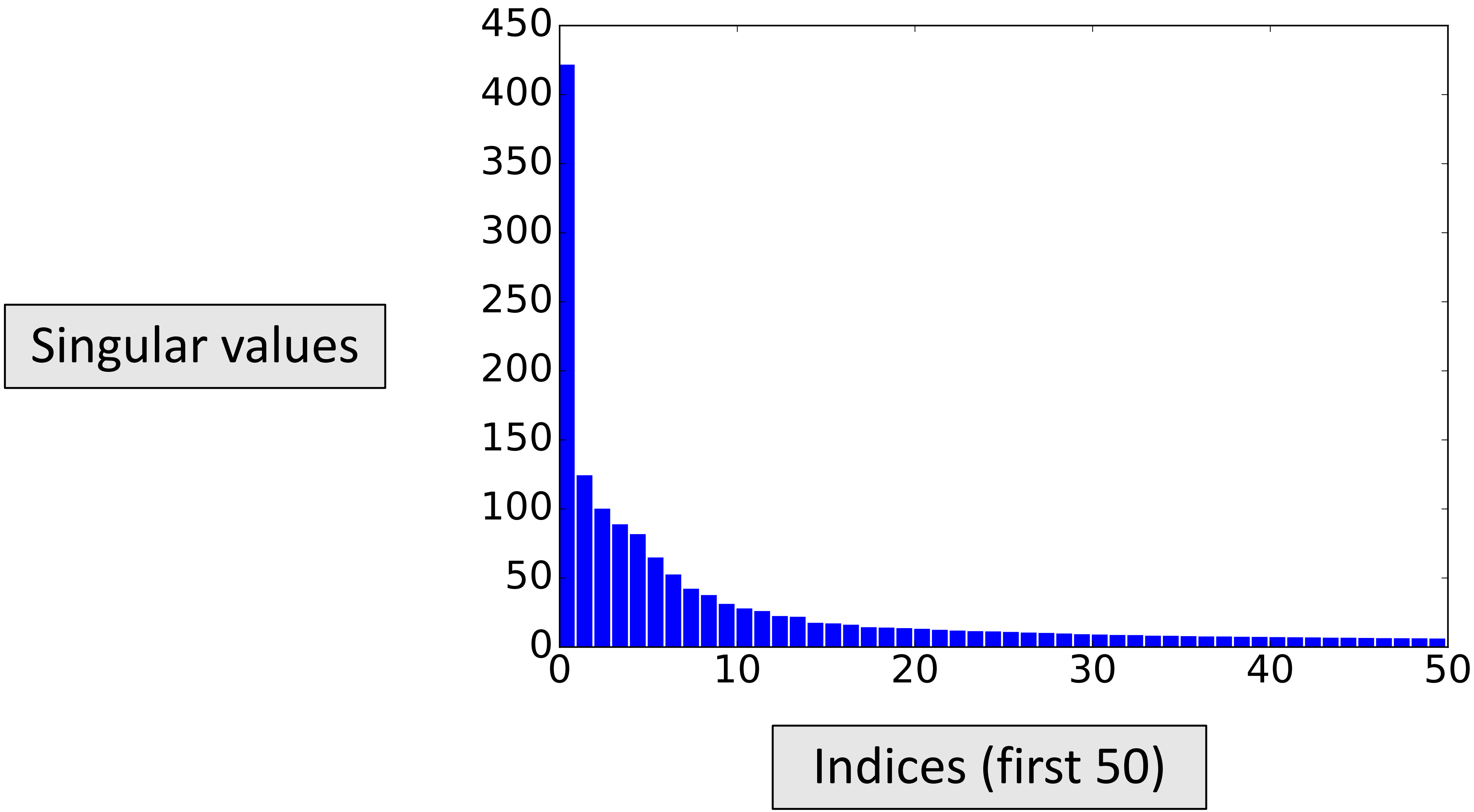
SVD: Example

Singular values



Indices

SVD: Example

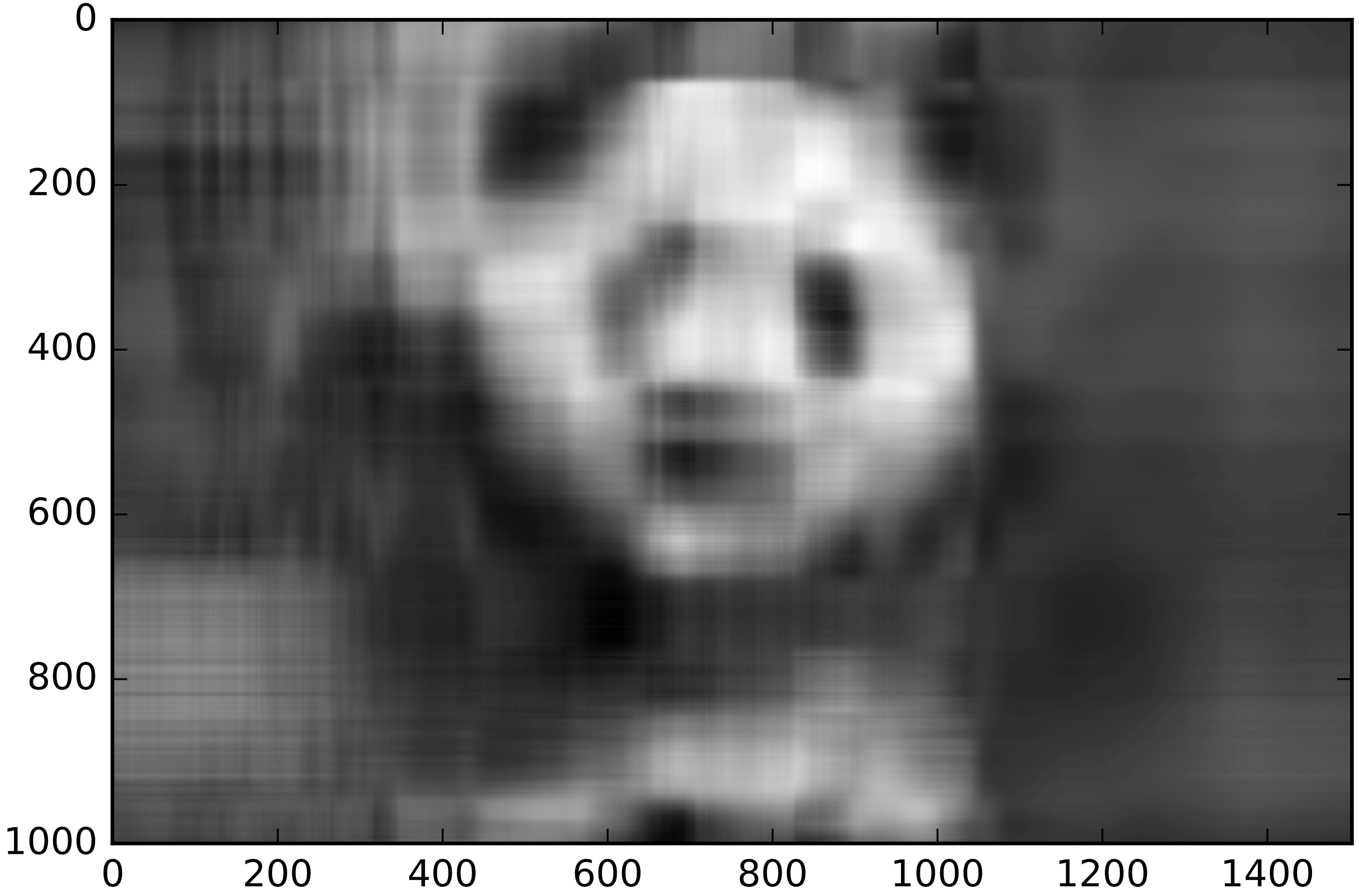

Singular values
450
400
350
300
250
200
150
100
50
0
0
10
20
30
40
50
Indices (first 50)

SVD: Example

Original image (1000×1500)

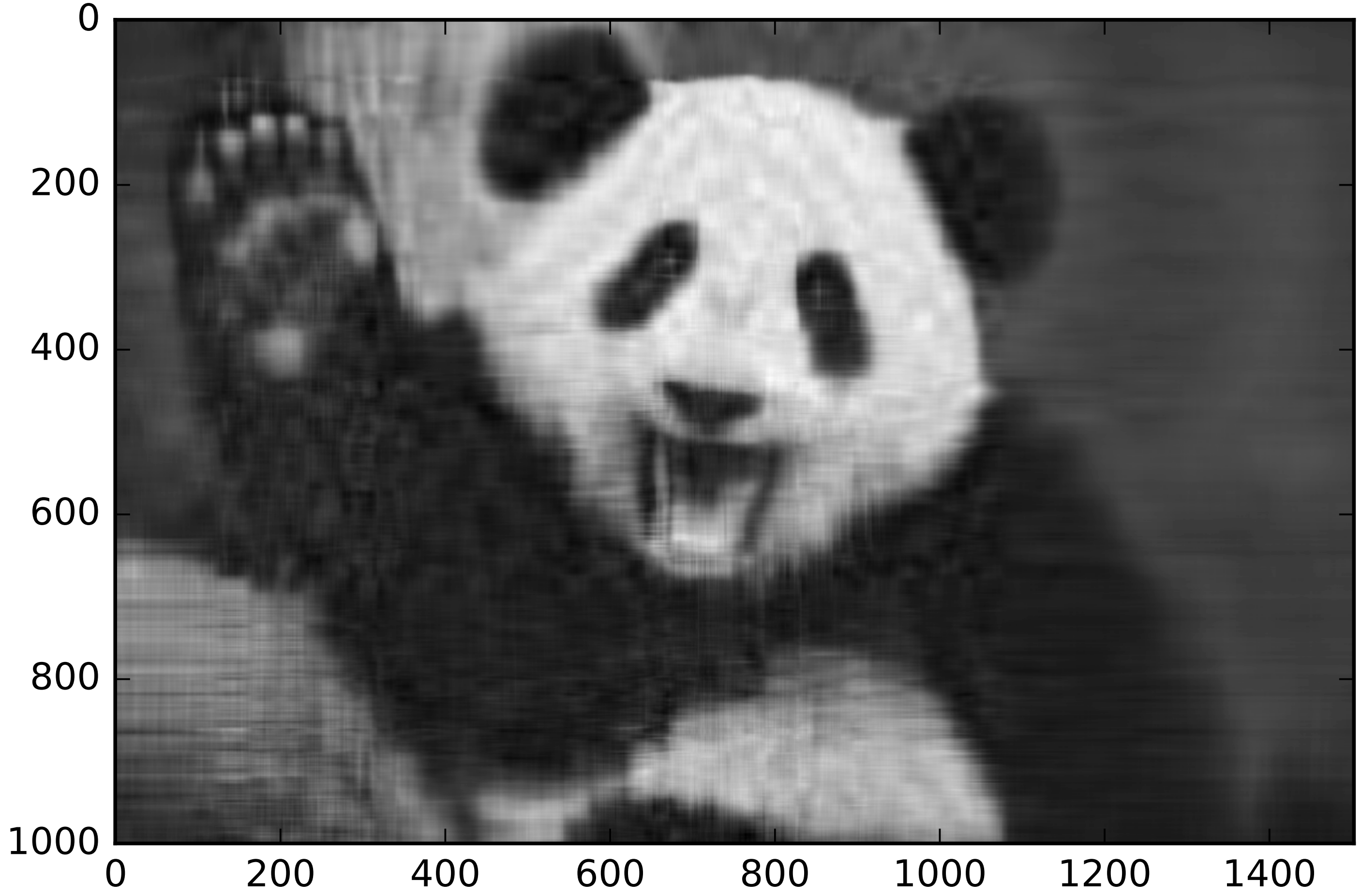
SVD: Example

Rank-5 truncated SVD



SVD: Example

Rank-20 truncated SVD



SVD: Example

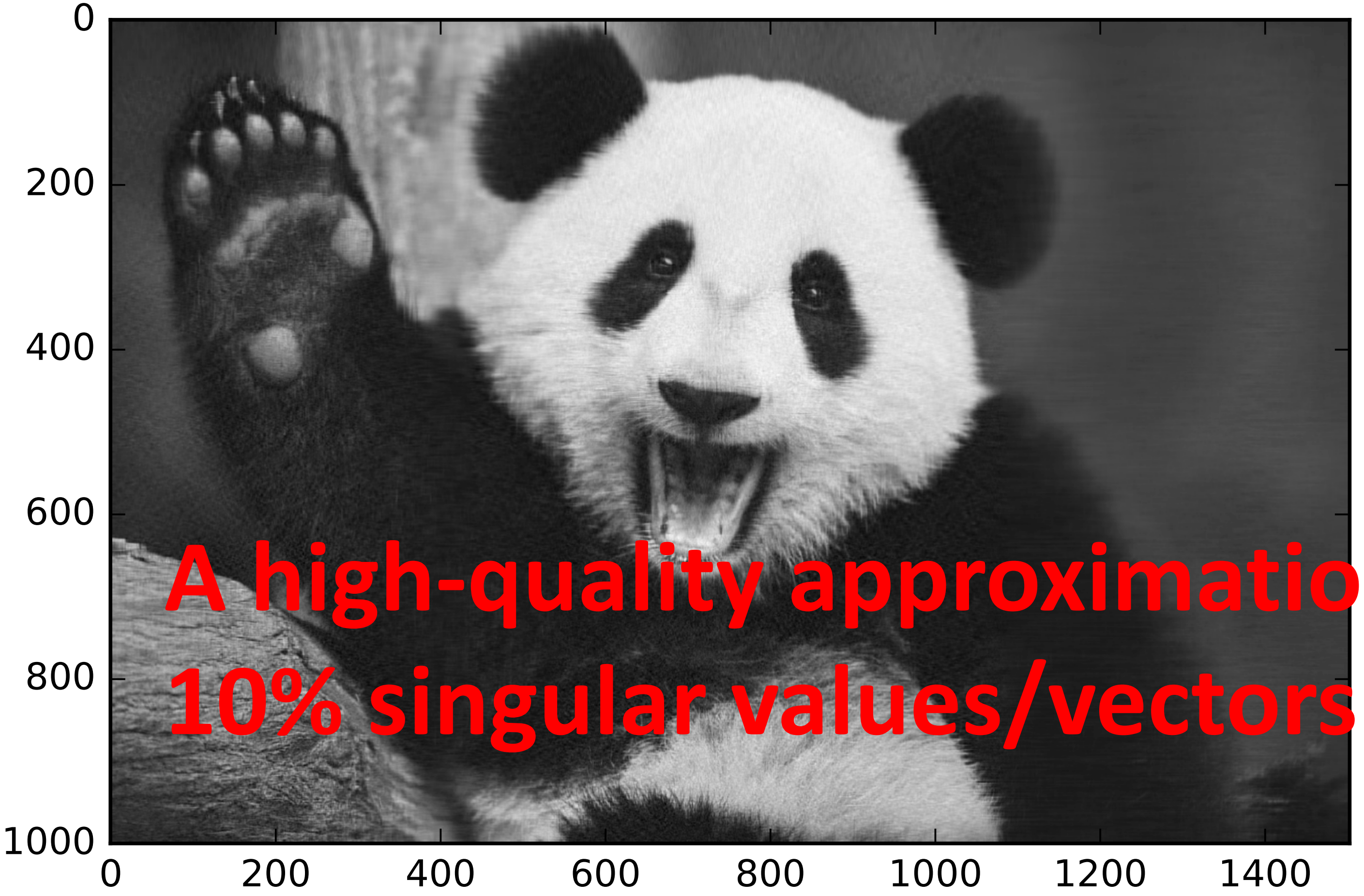
Rank-50 truncated SVD

SVD: Example

Rank-100 truncated SVD

SVD: Example

Rank-100 truncated SVD



A high-quality approximation using 10% singular values/vectors!

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.

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

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.

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

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

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.

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

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.

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

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 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 $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

Merely 2 matrix-vector multiplications.
Very Cheap!

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 $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

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

Power Iteration for Truncated SVD

Goal: Compute the top 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 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 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 $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\|\mathbf{x}_q\right\|_2$.

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.
- $\mathbf{x}_q \propto (\mathbf{A}^T \mathbf{A})^q \, \mathbf{x}_0$

- 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 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.
- $\mathbf{x}_q \propto \boxed{(\mathbf{A}^T\mathbf{A})^q}\,\mathbf{x}_0 = \boxed{\left(\sum_{i=1}^{r} \sigma_i^{2q} \, \mathbf{v}_i \, \mathbf{v}_i^T\right)}\left(\sum_{j=1}^{n} \alpha_j \, \mathbf{v}_j\right)$.

- 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 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 $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left|\left|\mathbf{x}_q\right|\right|_2$.

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.
- $\mathbf{x}_q \propto (\mathbf{A}^T\mathbf{A})^q \, \mathbf{x}_0 = \left(\sum_{i=1}^{r} \sigma_i^{2q} \, \mathbf{v}_i \, \mathbf{v}_i^T\right)\left(\sum_{j=1}^{n} \alpha_j \, \mathbf{v}_j\right) = \sum_{i=1}^{r} \alpha_i \, \sigma_i^{2q} \, \mathbf{v}_i$.

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 $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left|\left|\mathbf{x}_q\right|\right|_2$.

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 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 $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left|\left|\mathbf{x}_q\right|\right|_2$.

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.
- $\mathbf{x}_q \propto \sum_{i=1}^{r} \alpha_i \, \sigma_i^{2q} \, \mathbf{v}_i$.
- $\mathbf{x}_q \propto \sum_{i=1}^{r} \alpha_i \left(\frac{\sigma_i}{\sigma_1}\right)^{2q} \mathbf{v}_i = \alpha_1 \mathbf{v}_1 + \sum_{i=2}^{r} \alpha_i \left(\frac{\sigma_i}{\sigma_1}\right)^{2q} \mathbf{v}_i$.

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 $\mathbf{x}_q \leftarrow \mathbf{x}_q / \left\| \mathbf{x}_q \right\|_2$.

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

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: $\mathbf{X}_0 \leftarrow \mathrm{orth}(\mathbf{X}_0)$;
3. Repeat the power iteration:
   - *i.* $\mathbf{X}_q \leftarrow \mathbf{A}^T\mathbf{A}\,\mathbf{X}_{q-1}$;
   - *ii.* $\mathbf{X}_q \leftarrow \mathrm{orth}(\mathbf{X}_q)$.

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.