Matrix Computations
Chapter 6: Singular value, Singular Vectors, and Singular Value Decomposition
Section 6.4 Application of SVD

Jie Lu
ShanghaiTech University

Low-Rank Matrix Approximation

Aim: Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}(\mathbf{A}) = r$ and $k \in \{1, \ldots, r-1\}$, find $\mathbf{B} \in \mathbb{R}^{m \times n}$ s.t. $\mathrm{rank}(\mathbf{B}) \leq k$ and $\mathbf{B}$ best approximates $\mathbf{A}$

- Closely related to the matrix factorization problem in Section 3.4
- Applications: PCA, dimensionality reduction, etc.

Truncated SVD: Denote

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Perform the aforementioned approximation by choosing $\mathbf{B} = \mathbf{A}_k$

Application Example: Image Compression

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix whose (i, j)th entry a_{ij} stores the (i, j)th pixel of an image

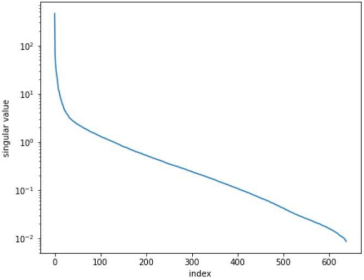
- Memory size for storing $\mathbf{A}$: mn

- Truncated SVD: store $\{\mathbf{u}_i, \sigma_i \mathbf{v}_i\}_{i=1}^{k}$ instead of the full $\mathbf{A}$, and recover the image by $\mathbf{B} = \mathbf{A}_k$

- Memory size for truncated SVD: $(m + n)k$
  - Much less than mn if $k \ll \min\{m, n\}$

Application Example: Image Compression (cont'd)

original image, size: 639 x 853



Application Example: Image Compression (cont'd)

k=10

k=15

k=20

k=30

Low-Rank Matrix Approximation

Truncated SVD provides the best approximation in the LS sense

Theorem (Eckart-Young-Mirsky)

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $k \in \{1, \ldots, r\}$, *The truncated SVD* $\mathbf{A}_k$ *is an optimal solution to*

$$\min_{\mathbf{B} \in \mathbb{R}^{m \times n},\ \mathrm{rank}(\mathbf{B}) \leq k} \|\mathbf{A} - \mathbf{B}\|_F^2$$

The matrix 2-norm version of the Eckart-Young-Mirsky theorem

Theorem

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $k \in \{1, \ldots, r\}$, *the truncated SVD* $\mathbf{A}_k$ *is an optimal solution to*

$$\min_{\mathbf{B} \in \mathbb{R}^{m \times n},\ \mathrm{rank}(\mathbf{B}) \leq k} \|\mathbf{A} - \mathbf{B}\|_2$$

Low-Rank Matrix Approximation

Recall the matrix factorization problem in Section 3.4

$$\min_{\mathbf{A} \in \mathbb{R}^{m \times k}, \mathbf{B} \in \mathbb{R}^{k \times n}} \|\mathbf{Y} - \mathbf{AB}\|_F^2$$

where $k \leq \min\{m, n\}$, $\mathbf{A}$ is a basis matrix, and $\mathbf{B}$ is a coefficient matrix

The matrix factorization problem may be reformulated as

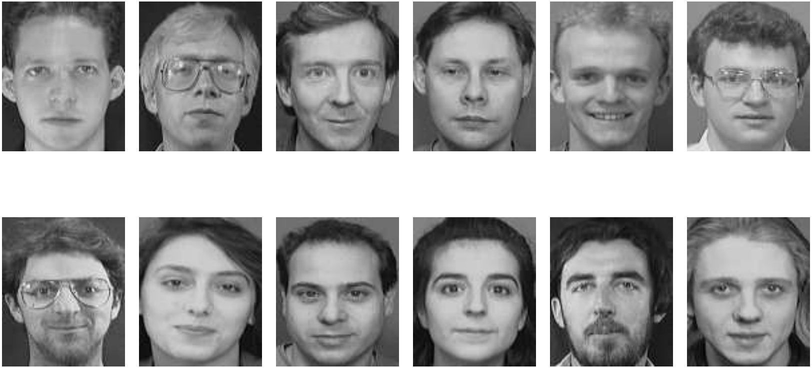
$$\min_{\mathbf{Z} \in \mathbb{R}^{m \times n}, \mathrm{rank}(\mathbf{Z}) \leq k} \|\mathbf{Y} - \mathbf{Z}\|_F^2$$

The truncated SVD $\mathbf{Y}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$, where $\mathbf{Y} = \mathbf{U} \Sigma \mathbf{V}^T$ denotes the SVD of $\mathbf{Y}$, is an optimal solution by the Eckart-Young-Mirsky Theorem

An optimal solution to the matrix factorization problem is

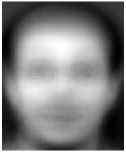
$$\mathbf{A} = [\ \mathbf{u}_1, \ldots, \mathbf{u}_k\ ], \qquad \mathbf{B} = [\ \sigma_1 \mathbf{v}_1, \ldots, \sigma_k \mathbf{v}_k\ ]^T$$

Dimensionality Reduction of a Face Image Dataset



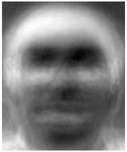
A face image dataset. Image size = 112×92, number of face images = 400. Each $\mathbf{x}_i$ is the vectorization of one face image, leading to $m = 112 \times 92 = 10304$, $n = 400$.

Dimensionality Reduction of a Face Image Dataset

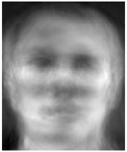


Mean face

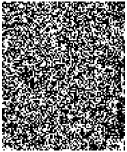
1st principal left singular vector



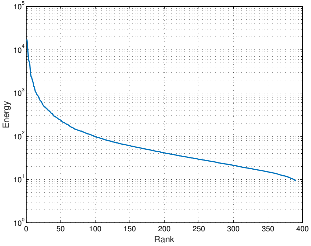
2nd principal left singular vector



3rd principal left singular vector



400th left singular vector



Energy Concentration

Singular Value Inequalities

Similar to variational characterization for eigenvalues of real symmetric matrices, there have been a collection of variational characterization results for singular values

- Courant-Fischer characterization: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. For each $k = 1, \ldots, p = \min\{m, n\}$,

$$\sigma_k(\mathbf{A}) = \min_{\substack{\mathcal{S} \subseteq \mathbb{R}^n: \\ \dim \mathcal{S} = n-k+1}} \max_{\mathbf{x} \in \mathcal{S},\ \|\mathbf{x}\|_2 = 1} \|\mathbf{A}\mathbf{x}\|_2$$

- Weyl's inequality: For any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$,

$$\sigma_{k+\ell-1}(\mathbf{A}+\mathbf{B}) \leq \sigma_k(\mathbf{A}) + \sigma_\ell(\mathbf{B}), \qquad k, \ell \in \{1, \ldots, p\},\ k+\ell-1 \leq p.$$

- Corollaries:
  - $\sigma_k(\mathbf{A}+\mathbf{B}) \leq \sigma_k(\mathbf{A}) + \sigma_1(\mathbf{B}),\ k = 1, \ldots, p$
  - $|\sigma_k(\mathbf{A}+\mathbf{B}) - \sigma_k(\mathbf{A})| \leq \sigma_1(\mathbf{B}),\ k = 1, \ldots, p$

Computing the SVD via the Power Method

Apply the power method to compute the thin SVD

- Assume $m \geq n$ and $\sigma_1 > \sigma_2 > \ldots \sigma_n > 0$
- Apply the power method to $\mathbf{A}^T\mathbf{A}$ to obtain $\mathbf{v}_1$
- Obtain $\mathbf{u}_1 = \mathbf{A}\mathbf{v}_1 / \|\mathbf{A}\mathbf{v}_1\|_2$, $\sigma_1 = \|\mathbf{A}\mathbf{v}_1\|_2$
- Do deflation $\mathbf{A} := \mathbf{A} - \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$, and repeat the above steps until all singular components are found