Singular Value Decomposition
MATH 578 Mini-Seminar Talk

Kai Yang

McGill University
Montréal, Quebec Canada H3A 1A2

McGill University

October 22, 2020

Introduction

- *Singular Value Decomposition* (SVD) is one of most important tool for mathematical computation
- Contents of this presentation are taken from *Numerical linear algebra* and *Matrix Computations*
- As we'll see later, SVD is connect to eigenvalues; but in applications, eigenvalues are more connected to the behaviour of iterated forms of A, e.g., A^k or e^{tA}; SVD is more related to the behavior of A or A^{-1}
- We'll start with the geometric observation as a motivation for SVD

Geometric Observation

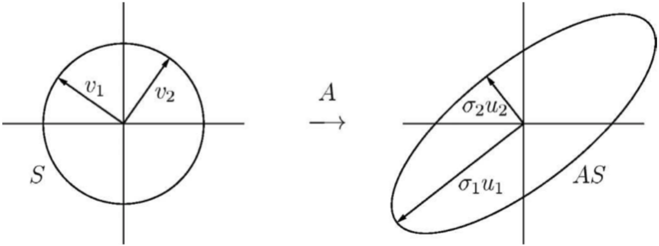

Figure: SVD of 2×2 matrix [1]

- We assume A is a real matrix here for the sake of geometric interpretation
- *The image of unit sphere under any $m \times n$ matrix is a hyperellipse,* – a hyperellipse is a $m-$dimensional generalization of an ellipse.

Geometric Interpretation

- Define *n singular values of A*: these are the lengths of n principal semiaxes of AS – it's conventional to assume that singular values are numbered in a descending order, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$
- Define *n left singular vectors of A*: these are the *unit* vectors $\{u_1, u_2, \ldots, u_n\}$ oriented in the directions of principal semiaxes of AS, numbered corresponding to the singular values
- Define *n right singular vectors of A*: these are the *unit* vectors $\{v_1, v_2, \ldots, v_n\} \in S$ that are preimages of principal semiaxes of AS, number so that $Av_j = \sigma_j u_j$
- For the moment, let's assume A of full column rank n.

Reduced SVD

- Write the linear map as $Av_j = \sigma_j u_j$, $\forall 1 \leq j \leq n$; and express in matrix form:
$$AV = \hat{U}\hat{\Sigma}$$
- $\hat{\Sigma}$ is a $n \times n$ diagonal matrix with diagonal entries $\sigma_1, \sigma_2, \ldots, \sigma_n > 0$
- $\hat{U} \in \mathbb{C}^{m \times n}$ has orthonormal columns $u_1, u_2, \ldots, u_n \in \mathbb{C}^m$
- $V \in \mathbb{C}^{n \times n}$ is a unitary matrix with columns $v_1, v_2, \ldots, v_n$
- Then we have $A = \hat{U}\hat{\Sigma}V^*$
- Note the column vectors of $\hat{U}$ does not form a basis of $\mathbb{C}^n$ – this is why it's called "reduced"

Full SVD

- By adjoining $m - n$ orthonormal columns $\hat{U}_{m-n}$, $\hat{U}$ can be extended to a unitary matrix – let's call the result U
- $\hat{\Sigma}$ also needs to change to accommodate this – by adjoining $m - n$ rows of 0
- As a result, the full SVD is

$$A = \hat{U}\hat{\Sigma}V^* = \begin{bmatrix} \hat{U} & \hat{U}_{m-n} \end{bmatrix} \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix} V^* =: U\Sigma V^*$$

- Now we can discard the initial assumption that A has full (column) rank – all it changes is now not n but only r of the left singular vectors of A are determined by the geometry of hyperellipse, then we'll have $\hat{U} \in \mathbb{C}^{m\times r}$, $\hat{\Sigma} \in \mathbb{C}^{r\times r}$ will be a diagonal matrix with positive diagonal entries $\sigma_1, \sigma_2, \ldots, \sigma_r$, put at the upper-left corner of the otherwise-0 matrix $\Sigma \in \mathbb{C}^{m\times n}$

Formal Definition

- Given $A \in \mathbb{C}^{m \times n}$, a SVD of A is a factorization

$$A = U \Sigma V^*$$

- where $U \in \mathbb{C}^{m \times m}, V \in \mathbb{C}^{n \times n}$ are both unitary; $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal, with diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1}, \ldots, \sigma_p = 0$, where $p = \min\{m, n\}$

- The image of a unit sphere in $\mathbb{R}^n$ under linear map $A = U \Sigma V^*$ will indeed be a hyperellipse in $\mathbb{R}^m$: the unitary map V^* perserves the sphere, then the diagonal matrix Σ stretches the sphere into a hyperellipse aligned with canonical basis, finally unitary map U rotates or reflects the hyperellipse without changing its shape.

Existence and Uniqueness

Theorem

Every matrix $A \in \mathbb{C}^{m\times n}$ has a SVD. Furthermore, the singular values are uniquely determined, and if A is square and σ_j are distinct, the left and right singular vectors $\left\{u_j\right\}$ and $\left\{v_j\right\}$ are uniquely determined up to the complex signs.

- The existence statement can be proved by induction on the dimension of A, whose induction step is established by: for submatrix B of A,

$$B = U_2\Sigma_2V_2^* \Rightarrow A = U_1\begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^* V_1^*$$

  is a SVD of A.
- The uniqueness statement is justified by the geometric interpretation: should the semiaxis lengths of a hyperellipse are distinct, then the semiaxes are determined by the geometry signs.

SVD v.s. Eigen-Decomposition

- SVD represents a change of bases, i.e., every matrix is diagonal if using proper bases for domain and range spaces – it uses two different bases, the sets of left (for range space) and right (for domain space) singular vectors. Eigen-Decomposition uses one bases – the set of eigenvectors
- SVD uses orthonormal bases; while eigen-decomposition in general does not
- Not all matrices (even square ones) have an eigenvalue decomposition, but all matrices have a SVD

Matrix Properties via SVD I

- $\text{rank}(A) = r$, the number of nonzero singular values. *(note that U, V are full rank)*
- $\text{range}(A) = \langle u_1, \ldots, u_r \rangle$ and $\text{null}(A) = \langle v_{r+1}, \ldots, v_n \rangle$. *(observing the range and null spaces of Σ)*
- $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$. *($U, V$ are unitary, together with the fact that ℓ_2 norm and Frobenius norm is invariant under unitary multiplication)*
- The nonzero singular values of A square roots of nonzero eigenvalues of A^*A or AA^*. *(substitution by SVD)*
- If $A = A^*$, the singular values of A are the absolute values of eigenvalues of A. *(using the fact that the eigenvalues for every Hermitian matrix are real, and the eigenvectors corresponding to distinct eigenvalues are orthogonal)*

Matrix Properties via SVD II

- $\forall A \in \mathbb{C}^{m\times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$. *(substitution by SVD)*
- $\forall A \in \mathbb{C}^{m\times n}$, $E \in \mathbb{C}^{m\times n}$,

$$\begin{aligned}\sigma_{\max}(A+E) &\leq \sigma_{\max}(A) + \|E\|_2 \\ \sigma_{\min}(A+E) &\geq \sigma_{\min}(A) - \|E\|_2\end{aligned}$$

(recall $Av_j = \sigma_j u_j$, which implies $\sigma_{\min}(A) \leq \frac{\|Ax\|_2}{\|x\|_2} \leq \sigma_{\max}(A)$)

- $\forall A \in \mathbb{C}^{m\times n}$, $m > n$, and $\forall z \in \mathbb{C}^m$,

$$\begin{aligned}\sigma_{\max}\left(\begin{bmatrix} A & z \end{bmatrix}\right) &\geq \sigma_{\max}(A) \\ \sigma_{\min}\left(\begin{bmatrix} A & z \end{bmatrix}\right) &\leq \sigma_{\min}(A)\end{aligned}$$

()

Low Rank Approximations

- $A = \sum_{i=1}^{r} \sigma_i u_i v_i^*$. *(trivial)*
- Define $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^*$ for some $0 \leq k \leq r$, and define $\sigma_{k+1} = 0$ if $k = \min\{m, n\}$. Then

$$\min_{B \in \mathbb{C}^{m \times n},\ \mathrm{rank}(B) \leq k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

(second equality is trivial; prove first equality by assuming $\|A - B\|_2 < \|A - A_k\|_2 = \sigma_{k+1}$, then $\forall w \in null(B)$ (dim $\geq n - k$),

$$\|Aw\|_2 = \|(A - B)\,w\|_2 \leq \|A - B\|_2 \, \|w\|_2 < \sigma_{k+1} \, \|w\|_2$$

on the other hand, the first $k+1$ right singular vectors of A span a $k+1$-dimensional space s.t. for all w in it, $\|Aw\|_2 \geq \sigma_{k+1} \, \|w\|_2$, which leads to contradiction)

- Similar to above statement *(and proof)*, we also have

$$\min_{B \in \mathbb{C}^{m \times n},\ \mathrm{rank}(B) \leq k} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$$

Some Basic Facts

- For linear system $Ax = b$, applying SVD, we have that

$$x = A^{-1}b = \left(U\Sigma V^T\right)^{-1} b = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i$$

  which means should σ_n be small, a small changes in A or b can induce relatively large changes in x
- One way to compute SVD of A is, form A^*A and take eigen-decomposition of $A^*A = V\Gamma V^*$, then $\Sigma \in \mathbb{R}^{m\times n}$ will have diagonal square root of Γ, and we can solve $U\Sigma = AV$ for unitary U. But this algorithm is *unstable* – eigen-decomposition of A^*A will be much more sensitive to perturbations
- One can, however, reduce the SVD to an eigenvalue problem by taking a different approach

A Different Approach

- Construct the Hermitian matrix $H = \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} \in \mathbb{C}^{2m \times 2m}$ for square matrix A, then based on SVD of A we have

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} \begin{bmatrix} V & V \\ U & -U \end{bmatrix} = \begin{bmatrix} V & V \\ U & -U \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}$$

- Then singular values of A are the absolute values of the eigenvalues of H, and singular vectors vectors of A can be extracted from eigenvectors of H

- And this allows us to convert SVD problem of A to eigen-decomposition of H, which is stable

Two Phrases

- Golub, Kahan, et al. proposed the two-phrase method to obtain SVD: the first phrase is to convert the matrix to a bi-diagonal form (diagonal and first super-diagonal); and the second phrase is to diagonalize the bi-diagonal matrix

- The first phrase of bi-diagonalization is called *Golub-Kahan bi-diagonalization*

- The second phrase of diagonalization was conventionally solved by a variant of QR algorithm; and more recently, divide-and-conquer algorithms were also developed for the second phrase.

Golub-Kahan Bi-Diagonalization

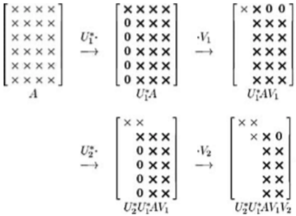

Figure: Golub-Kahan bi-diagonalization for a 6×4 matrix [1]

- Golub-Kahan bi-diagonalization applies Householder reflectors alternatively to left and right
- Left reflection introduces a column of 0s below diagonal
- Right reflection introduces a row of 0s to the right of first superdiagonal
- Both left and right reflections will preserve zeros introduced before

Faster Methods for the First Phrase

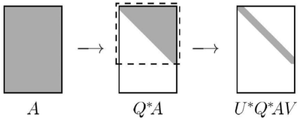

Figure: LHC bi-diagonalization [1]

- Golub-Kahan method *per se* is not efficient for $m \gg n$. Lawson, Hanson and Chan discussed a much more efficient way: use a single QR factorization step to introduce zeros everywhere below diagonal, then apply Golub-Kahan on the upper $n \times n$ matrix only – indeed, this will destroy some zeros introduced by the QR step
- LHC procedure is advantageous only when $m > \frac{5}{3}n$; note the Golub-Kahan process will make the matrix thinner as it proceeds, then one can apply QR factorization step when adequate
- And it was discussed that QR step should be performed when the matrix reaches an aspect ratio of 2

Bibliography

[1] Lloyd Trefethen. *Numerical linear algebra.* Philadelphia, Pa: Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104, 1997. ISBN: 0898719577.

[2] Gene H. Golub. *Matrix Computations.* J. Hopkins Uni. Press, Jan. 7, 2013. ISBN: 1421407949. URL: https://www.ebook.de/de/product/20241149/gene_h_golub_matrix_computations.html.