Short notes on SVD and applications

by Sanand D

"If only closed minds came with closed mouths.."-Anonymous

1 SVD existence and uniqueness

Theorem 1.1 (Existence and uniqueness). *Let $A \in \mathbb{C}^{m\times n}$. Then, there exists unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ such that*

$$U^*AV = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ . & . & \dots & . & . & \dots & . \\ . & . & \dots & . & . & \dots & . \\ 0 & 0 & \dots & \sigma_k & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ . & . & \dots & . & . & \dots & . \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \tag{1}$$

where $\sigma_i \geq \sigma_{i+1}$ for $1 \leq i \leq k-1$ and $\sigma_i \in \mathbb{R}$ for $1 \leq i \leq k$. These σ_is are called singular values of A and the above decomposition is called singular value decomposition (SVD) of A. Furthermore, the singular values σ_i are uniquely determined, and, if A is square and the σ_i are distinct, then columns of U and V are uniquely determined up to multiplication by unit modulus complex numbers $e^{i\theta}$.

Proof. Existence: Note that $\|A\|_2 = \|A\mathbf{v}_1\|_2$ for some $\mathbf{v}_1 \in \mathbb{V}$ such that $\|\mathbf{v}_1\|_2 = 1$. Let $A\mathbf{v}_1 = \sigma_1\mathbf{u}_1$ where $\|\mathbf{u}_1\|_2 = 1$. Extend $\{\mathbf{v}_1\}$ to an orthonormal basis and let columns of $V = [\mathbf{v}_1\ V_1]$ denote this orthonormal basis. Similarly, extend $\mathbf{u}_1$ to an orthonormal basis and let columns of $U = [\mathbf{u}_1\ U_1]$ denote this orthonormal basis. Consider $U^*AV = A_1$.

$$U^*AV = \begin{bmatrix} \mathbf{u}_1^* \\ U_1^* \end{bmatrix} A \begin{bmatrix} \mathbf{v}_1 & V_1 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^* \\ U_1^* \end{bmatrix} \begin{bmatrix} \sigma_1\mathbf{u}_1 & AV_1 \end{bmatrix} = \begin{bmatrix} \sigma_1\mathbf{u}_1^*\mathbf{u}_1 & \mathbf{u}_1^*AV_1 \\ \sigma_1 U_1^*\mathbf{u}_1 & U_1^*AV_1 \end{bmatrix} = \begin{bmatrix} \sigma_1 & \mathbf{w}^* \\ 0 & B \end{bmatrix}. \tag{2}$$

Note that unitary matrices preserve vector norms. Therefore, $\|A_1\|_2 = \|U^*AV\|_2 = \|A\|_2 = \sigma_1$. Observe that $A_1 \begin{bmatrix} \sigma_1^* \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 + \mathbf{w}^*\mathbf{w} \\ B\mathbf{w} \end{bmatrix}$

$$\Rightarrow \sigma_1^2 = \|A_1\|_2^2 \geq \frac{\|A_1 \begin{bmatrix} \sigma_1^* \\ \mathbf{w} \end{bmatrix}\|_2^2}{\|\begin{bmatrix} \sigma_1^* \\ \mathbf{w} \end{bmatrix}\|_2^2} = \frac{(\sigma_1^2 + \|\mathbf{w}\|_2^2)^2 + \|B\mathbf{w}\|_2^2}{\sigma_1^2 + \|\mathbf{w}\|_2^2} \geq (\sigma_1^2 + \|\mathbf{w}\|_2^2)$$

Therefore, $\mathbf{w} = \mathbf{0}$ and $A_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}$. Now by induction, B can be transformed into the canonical form. Let $\sigma_2 = \|B\|_2$. $\|A\|_2 \geq \|B\|_2$. Therefore, $\sigma_1 \geq \sigma_2$ and $\sigma_i \geq \sigma_{i+1}$ for $1 \leq i \leq k-1$.
Uniqueness: Observe that $AA^* = U\Sigma^2U^*$ and $A^*A = V^*\Sigma^2V$. Thus, the singular values of A can be obtained from the square roots of the eigenvalues of AA^* or A^*A implying uniqueness of the singular values. (However, this is not the way singular values are computed numerically.) The statement about the uniqueness of U and V follows from the corresponding result on eigenvectors. (Refer the next section on norms and svd for an alternate proof on uniqueness.) □

Remark 1.2. *Notice that $U\Sigma = AV$. Thus, $U\Sigma e_i = AVe_i \Rightarrow \sigma_i u_i = Av_i$. These u_i, v_i are called pair of singular vectors associated with the singular value σ_i. Geometrically, when A is full rank, it maps the sphere formed by the unit vectors associated with the columns of V into an ellipse whose axes are given by the columns of U and scaled by singular values of A.*

Remark 1.3. *Note that $A = \sum_{i=1}^{r} \sigma_i(A)\mathbf{u}_i\mathbf{v}_i^*$. Therefore, $Im(A) \subseteq span\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$. But $dim(Im(A)) = r = dim(span\{\mathbf{u}_1, \dots, \mathbf{u}_r\})$, hene, $Im(A) = span\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$. Notice that $ker(\Sigma) = \{\mathbf{e}_{r+1}, \dots, \mathbf{e}_n\}$. Therefore, $ker(A) = span\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$. Kernel and image of A^* follows similarly. It is also clear that the rank of A is equal to the number of nonzero singular values of A.*

Corollary 1.4 (condensed svd)**.** *Let $A \in \mathbb{C}^{m\times n}$ such that $rank(A) = r$. Then, there exists matrices $\hat{U} \in \mathbb{C}^{m\times r}$, $\hat{V} \in \mathbb{C}^{n\times r}$ such that $\hat{U}^*\hat{U} = I$, $\hat{V}^*\hat{V} = I$ and $\hat{\Sigma} = diag(\sigma_1(A), \dots, \sigma_r(A))$, $(\sigma_i > 0,\ 1 \leq i \leq r)$ and $A = \hat{U}\hat{\Sigma}\hat{V}^*$.*

Note that $\hat{U}\hat{U}^*$ gives an orthogonal projector onto $\mathrm{Im}(A)$. Let $V = [\hat{V}\ \bar{V}]$. Then, $\bar{V}\bar{V}^*$ gives an orthogonal projector onto $\ker(A)$. One can similarly construct projector onto the image and the kernel space of A^*.

Example 1.5. *For orthogonal/unitary matrices, all singular values are equal to one. For Hermitian matrices, $U^*AU = D$ where D is a diagonal matrix having eigenvalues of A. We now obtain svd for A. Choose columns of V matrix as follows: Let $v_i = u_i$ if $\lambda_i > 0$ and $v_i = -u_i$ if $\lambda_i < 0$ for $1 \leq i \leq n$. Thus, $U^*AV = \Sigma$ where Σ contains modulus of eigenvalues of A on its diagonal. Therefore, the singular values of a Hermitian matrix are given by the modulus of its eigenvalues. For positive definite matrices P, $U^*AU = D = \Sigma$ and eigenvalues and singular values are one and the same.*

Geometrically, one can think of svds as follows. Consider the action of A on unit vectors i.e. action of A on the unit sphere $\mathbb{S}^n$. $A\mathbb{S}^n = U\Sigma V^*\mathbb{S}^n$. Since V^* is unitary, $V^*\mathbb{S}^n \subseteq \mathbb{S}^n$. Now Σ maps $\mathbb{S}^n$ in to an ellipse and U changes the orientation of this ellipse. Maximum elongation is along u_1 which forms the major axis of the ellipse.

2 Norms and svd

We now consider the relationship between the 2−norm, the Frobenius norm and singular values.

Theorem 2.1. *$\|A\|_2 = \sigma_1(A)$, $\|A\|_F^2 = \sum_{i=1}^{r} \sigma_i(A)^2$.*

Proof. Recall that $A = U\Sigma V^*$ and $\|A\|_2 = \max_{\|\mathbf{x}\|_2=1}\|A\mathbf{x}\|$. Note that

$$\max_{\|\mathbf{x}\|_2=1}\|A\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|_2=1}\|U\Sigma V^*\mathbf{x}\|_2 = \max_{\|\mathbf{y}\|_2=1}\|\Sigma\mathbf{y}\|_2$$

since U and V are unitary. Therefore, $\|A\|_2 = \sigma_1(A)$. Observe that $\|A\|_F^2 = \mathrm{trace}(A^*A)$. Moreover, $A = U\Sigma V^* \Rightarrow A^*A = V\Sigma^2V^*$ and $\mathrm{trace}(V\Sigma^2V^*) = \sum_i \sigma_i(A)^2$. □

Alternate proof of uniqueness of svd ([1]): It is clear from above theorem that the largest singular value σ_1 is uniquely determined since it is the 2−norm of A. Recall that $A\mathbf{v}_1 = \sigma_1\mathbf{u}_1$. Suppose in addition to $\mathbf{v}_1$, there is another linearly independent vector $\mathbf{z}$ such that $\|\mathbf{z}\|_2 = 1$ and $\|A\mathbf{z}\|_2 = \sigma_1$. Define a unit vector $\mathbf{v}_2$ orthogonal to $\mathbf{v}_1$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{z}$

$$\mathbf{v}_2 = \frac{\mathbf{z} - \langle\mathbf{v}_1, \mathbf{z}\rangle}{\|\mathbf{z} - \langle\mathbf{v}_1, \mathbf{z}\rangle\mathbf{v}_1\|_2}.$$

Since $\|A\|_2 = \sigma_1$, $\|A\mathbf{v}_2\|_2 \leq \sigma_1$; but this must be an equality, for otherwise, since $\mathbf{z} = c\mathbf{v}_1 + s\mathbf{v}_2$ for some constants c and s with $|c|^2 + |s|^2 = 1$, we would have $\|A\mathbf{z}\|_2 < \sigma_1$. This vector $\mathbf{v}_2$ is a second right singular vector of A corresponding to the singular value σ_1. It is clear that $\mathbf{v}_2$ lies in the column span of V_1 (see V_1 in the existence part). Let $\mathbf{y} = [\mathbf{v}_1\ V_1]^*\mathbf{v}_2 = \begin{bmatrix} 0 \\ V_1^*\mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y}_1 \end{bmatrix}$. Notice that $\|\mathbf{y}_1\|_2 = \|\mathbf{v}_2\|_2 = 1$. From Equation (2),

$$A\mathbf{v}_2 = \begin{bmatrix} \mathbf{u}_1 & U_1 \end{bmatrix}\begin{bmatrix} \sigma_1 & \mathbf{0}^* \\ 0 & B \end{bmatrix}\begin{bmatrix} \mathbf{v}_1^* \\ V_1^* \end{bmatrix}\mathbf{v}_2 = \begin{bmatrix} \mathbf{u}_1 & U_1 \end{bmatrix}\begin{bmatrix} \sigma_1 & \mathbf{0}^* \\ 0 & B \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{y}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & U_1 \end{bmatrix}\begin{bmatrix} 0 \\ B\mathbf{y}_1 \end{bmatrix}.$$

Therefore, $\|B\mathbf{y}_1\|_2 = \sigma_1$. This implies that if the singular vector $\mathbf{v}_1$ is not unique, then the corresponding singular value σ_1 is not simple. To complete the uniqueness proof we note that, as indicated above, once σ_1, $\mathbf{v}_1$ and $\mathbf{u}_1$ are determined, the remainder of the SVD is determined by the action of A on the space orthogonal to $\mathbf{v}_1$. Since $\mathbf{v}_1$ is unique up to multiplication by $e^{i\theta}$, this orthogonal space is uniquely defined, and the uniqueness of the remaining singular values and vectors now follows by induction. □

Orthonormal set and orthonormal projection: Recall that orthogonal projection $\mathbf{x}^*$ of a vector $\mathbf{x}$ onto a subspace $\mathbb{W}$ gives a vector in $\mathbb{W}$ closest to $\mathbf{x}$. (Let $\mathbf{x} = \mathbf{x}^* + \mathbf{x}^\perp$ where $\mathbf{x}^* \perp \mathbf{x}^\perp$. If $\hat{\mathbf{x}} \in \mathbb{W}$, then writing $\hat{\mathbf{x}} = \hat{\mathbf{x}} - \mathbf{x}^* + \mathbf{x}^*$, $\mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - \mathbf{x}^* + \mathbf{x}^* - \hat{\mathbf{x}} = \mathbf{x}^\perp + \mathbf{x}^* - \hat{\mathbf{x}}$ where $\mathbf{x}^\perp \perp \mathbf{x}^* - \hat{\mathbf{x}}$ since $\mathbf{x}^* - \hat{\mathbf{x}} \in \mathbb{W}$. Therefore, $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \geq \|\mathbf{x}^\perp\|_2 = \|\mathbf{x} - \mathbf{x}^*\|_2$.)

Let $A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$ with the Frobenius norm. Note that rank one matrices $\mathbf{u}_i \mathbf{v}_i^*$, $1 \leq i \leq r$ form an orthonormal set since trace$((\mathbf{u}_i\mathbf{v}_i^*)^*(\mathbf{u}_j\mathbf{v}_j^*)) = \delta_{ij}$. Thus, a matrix of rank $r-1$ closest to A w.r.t. the Frobenius norm is obtained by the orthonormal projection of A onto the set $\{\mathbf{u}_i\mathbf{v}_i^*\}$, $1 \leq i \leq r-1$ i.e., $\hat{A} = \sum_{i=1}^{r-1} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$ with $\|A - \hat{A}\|_F^2 = \sigma_r(A)^2$. (This can be related to fourier series expansion using orthonormal sets and truncated fourier series which gives the best approximation of a periodic function using finitely many terms where principle of orthonormality is used.)

Theorem 2.2 (Closest low rank matrix ([2]))**.** *Let $A = U\Sigma V^* = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$. Let $\hat{A} = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$, $k < r$. Then,*

$$min_{rank(B)=k} \|A - B\|_2 = \|A - \hat{A}\|_2 = \sigma_{k+1}(A). \tag{3}$$

Proof. Since $U^*\hat{A}V = \text{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$, $\text{rank}(\hat{A}) = k$ and $U^*(A - \hat{A})V = \text{diag}(0, \ldots, 0, \sigma_{k+1}, \ldots, \sigma_r)$. Therefore, $\|A - \hat{A}\|_2 = \sigma_{k+1}$.

Suppose $\text{rank}(B) = k$ for some $B \in \mathbb{C}^{m\times n}$. We can find orthonormal vectors $\mathbf{x}_1, \ldots, \mathbf{x}_{n-k}$ such that $\ker(B) = \text{span}\ \{\mathbf{x}_1, \ldots, \mathbf{x}_{n-k}\}$. A dimension argument shows that $\text{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_{n-k}\} \cap$ span $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\} \neq \{0\}$. Let $\mathbf{z}$ be a unit 2−norm vector in this intersection. Now $B\mathbf{z} = 0$ and $A\mathbf{z} = \sum_{i=1}^{k+1} \sigma_i \mathbf{u}_i (\mathbf{v}_i^* \mathbf{z})$. Therefore,

$$\|A - B\|_2^2 \geq \|(A - B)\mathbf{z}\|_2^2 = \|A\mathbf{z}\|_2^2 \geq \sigma_{k+1}^2.$$

This proves the theorem. □

Thus, the smallest singular value of A is the 2−norm distance to the nearest singular matrix. It also follows that the set of full rank matrices is both open and dense.

Theorem 2.3 ([2])**.** *Let $A \in \mathbb{R}^{n\times m}$, $n \geq m$, $rank(A) = m$ with singular values $\sigma_1 \geq \ldots \geq \sigma_m > 0$. Then, $\|(A^\top A)^{-1}\|_2 = \sigma_m^{-2}$, $\|(A^\top A)^{-1}A^\top\|_2 = \sigma_m^{-1}$, $\|A(A^\top A)^{-1}\|_2 = \sigma_m^{-1}$, $\|A(A^\top A)^{-1}A^\top\|_2 = 1$.*

Proof. Follows from svd decomposition. □

3 Applications

3.1 Condition number

Define the condition number of a matrix as $\kappa_2(A) = \|A\|_2\|A^{-1}\|_2$. It is clear that $\|A\|_2 = \sigma_1(A)$ and $\|A^{-1}\|_2 = \sigma_n^{-1}$. Therefore, $\kappa_2(A) = \frac{\sigma_1(A)}{\sigma_n(A)}$. Note that $\kappa_2(A) \geq 1$. If $\kappa_2(A)$ is small, we say A is well conditioned, otherwise, it is ill conditioned.

Consider $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b}$ represents measurement and $\mathbf{x}$ is to be estimated. Suppose there is a measurement error (due to noise) say $\delta\mathbf{b}$ in $\mathbf{b}$ and let $A(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$. Suppose A is invertible. Therefore $\delta\mathbf{x} = A^{-1}\delta\mathbf{b}$ and $\|\delta\mathbf{x}\| = \|A^{-1}\|\|\delta\mathbf{b}\|$. Now, the relative error can be written as

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A\|\|A^{-1}\|\frac{\|\delta\mathbf{y}\|}{\|\mathbf{y}\|} = \kappa_2(A)\frac{\|\delta\mathbf{y}\|}{\|\mathbf{y}\|}. \tag{4}$$

Thus, we can bound the relative error is estimating $\mathbf{x}$ using condition number and relative measurement error. Condition number of a matrix has a lot of applications in numerical linear algebra, particularly in sensitivity analysis and backward stability of certain algorithms.

3.2 Numerical rank

Given an $\epsilon > 0$, the numerical rank of a matrix is the number of its singular values which are greater than ϵ.

If a matrix does not have full rank, any small perturbation is almost certain to transform it to a matrix that does have full rank. It follows that in the presence of uncertainty in the data, it is impossible to calculate the (exact, theoretical) rank of a matrix or even detect that it is rank deficient. It is reasonable to call a matrix numerically rank deficient if it is very close to a rank-deficient matrix, since it could have been rank deficient except for a small perturbation,

3.3 Least squares solution, least norm solution and pseudoinverse using svd

Consider the least squares problem $A\mathbf{x} = \mathbf{b}$ where $A \in \mathbb{R}^{n\times m}$, $\mathbf{b} \in \mathbb{R}^n$. Suppose $n > m$. We seek $\mathbf{x}$ such that $\|\mathbf{b} - A\mathbf{x}\|_2$ is minimized. If $\text{rank}(A) = m$, then the solution is unique. Observe that

$$\|\mathbf{b} - A\mathbf{x}\|_2 = \|U^\mathsf{T}\mathbf{b} - \Sigma(V^\mathsf{T}\mathbf{x})\|_2.$$

Let $\mathbf{c} = U^\mathsf{T}\mathbf{b}$ and $\mathbf{y} = V^\mathsf{T}\mathbf{x}$. Thus, find $\mathbf{y}$ such that $\|\mathbf{c} - \Sigma\mathbf{y}\|_2$ is minimized. Then $\mathbf{x}$ can be recovered from $\mathbf{y}$. Clearly, the solution is given by $y_i = \frac{c_i}{\sigma_i}$, $i = 1, \ldots, m$. If $\text{rank}(A) = r < m$, then one can ask for a norm minimizing solution. In this case, $y_i = \frac{c_i}{\sigma_i}$, $i = 1, \ldots, r$ and $y_i = 0$ for $i = r + 1, \ldots, m$.

Now consider the case when $n < m$ and $\text{rank}(A) = n$. Again, $U^\mathsf{T}\mathbf{b} = \Sigma(V^\mathsf{T}\mathbf{x})$. It follows that the norm minimizing solution is given by $y_i = \frac{c_i}{\sigma_i}$, $i = 1, \ldots, n$ and $y_i = 0$ for $i = n + 1, \ldots, m$.

The pseudoinverse of $A = \hat{U}\hat{\Sigma}\hat{V}^\mathsf{T}$ is given by $A^\dagger = \hat{V}\hat{\Sigma}^{-1}\hat{U}^\mathsf{T}$. This can also be written as $A^\dagger = V^\mathsf{T}\Sigma^\dagger U$ where $\Sigma^\dagger = \text{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)$. One can check that $\mathbf{x} = A^\dagger\mathbf{b}$ gives the norm minimizing solution of $A\mathbf{x} = \mathbf{b}$.

3.4 Polar decomposition

All complex numbers z can be represented in polar form $z = re^{i\theta}$. On similar lines, every real square matrix can be factored into $A = QS$, where Q is orthogonal and S is symmetric positive semidefinite. If A is invertible then S is positive definite. This can be seen as follows

$$A = U\Sigma V^\mathsf{T} = (UV^\mathsf{T})(V\Sigma V^\mathsf{T}), \tag{5}$$

where $(UV^\mathsf{T})^\mathsf{T}(UV^\mathsf{T}) = I$ and $V\Sigma V^\mathsf{T} \geq 0$. Polar decomposition has applications in mechanics and engineering.

3.5 Probability and statistics

For a random vector $\mathbf{x}$, with mean $\bar{\mathbf{x}} = E[\mathbf{x}]$, the covariance matrix is given by $C_{\mathbf{xx}} = E[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^\mathsf{T}] = E[\mathbf{x}\mathbf{x}^\mathsf{T}] - \bar{\mathbf{x}}\bar{\mathbf{x}}^\mathsf{T}$ which is a positive semidefinite matrix. The diagonal entries of the covariance matrix are variances of random variables x_i, $i = 1, \ldots, n$. The eigenvalues/singular values of the covariance matrix represents the variance of random variables in new coordinates obtained after a change of basis where the new basis is given by eigenvectors of $C_{\mathbf{xx}}$.

3.6 Data compression/image processing

Suppose a satellite takes a picture, and wants to send it to Earth. The picture may contain 1000 by 1000 "pixels"-a million little squares, each with a definite color. We can code the colors, and send back $1,000,000$ numbers. It is better to find the essential information inside the 1000 by 1000 matrix, and send only that. Suppose we know the SVD. The key is in the singular values (in Σ). Typically, some σ's are significant and others are extremely small. If we keep 20 and throw away 980, then we send only the corresponding 20 columns of U and V. The other 980 columns are multiplied in $U\Sigma V^\top$ by the small σ's that are being ignored. We can do the matrix multiplication as columns times rows:

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top.$$

Any matrix is the sum of r matrices of rank 1. If only 20 terms are kept, we send 20 times 2000 numbers instead of a million (25 to 1 compression).

It is clear that SVD is invariant under unitary transformations. Note that Fourier transform/discrete Fourier transform are unitary transformations. Therefore, the SVD of data and its Fourier transform are the same. In general, svd has many applications in big data analysis.

3.7 Control theory and svd

Let $G(s) = C(sI - A)^{-1}B + D$ be the transfer function of a linear dynamical system $\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}$, $\mathbf{y} = C\mathbf{x} + D\mathbf{u}$. The $2-$norm of $G(s)$ is defined as

$$\|G\|_2 := (\int_{-\infty}^{\infty} \text{trace}(G^*(j\omega)G(j\omega))d\omega)^{\frac{1}{2}} = (\int_{-\infty}^{\infty} \sum_{i=1}^{r} \sigma_i(G(j\omega))^2 d\omega)^{\frac{1}{2}} \tag{6}$$

and the infinity norm is defined as

$$\|G\|_\infty := \sup_\omega \sigma_1(G(j\omega)). \tag{7}$$

These norms on $G(s)$ allows one to capture the effect of noise and disturbance on the measured output of the system. Using a state feedback of the form say $\mathbf{u} = -K\mathbf{x} + \mathbf{r}$, one can modify the transfer function to meet the design specifications e.g., the effect of disturbance/noise on the output. The modified transfer function must have a norm less than some nonzero positive constant. This norm is related to svds as mentioned in the definition. In control literature, this is the well studied theory of H_2 and H_∞ control.

One can define a Hankel operator for linear systems whose singular values are called Hankel singular values which are invariants of the system. Hankel singular values are important in control applications.

Model order reduction: Find $\hat{G}$ of lower size such that $\|G - \hat{G}\|$ is minimized. Model order reduction is an important topic in control theory which involves applications of svd.

Moreover, in linear control theory, there are positive definite matrices called controllability/observability Gramians which are positive definite for controllable/observable systems. They are used to compute energy required for a state transfer and also to find the least energy input. The eigenvalues/singular values of this matrix gives indication of energy required to control in eigen directions. The inverse of the minimum singular value indicates the worst energy to go from the origin to the points on the unit sphere. The sum of invereses of singular values indicate the average energy required for control. The product of singular values i.e., the determinant indicates the volume of the ellipsoid that can be reached with unit energy inputs.

3.8 Tensor decomposition

One can extend the idea of writing matrices as a sum of rank one tensors. For data represented by higher order tensors instead of matrices, one can have a similar decomposition of higher order tensors known as higher order svd. This has applications in computer graphics, machine learning, signal processing etc.

3.9 Limitations

Low rank approximation using SVD fails when one wants to obtain a structured low rank approximation of matrices where certain structure of the underlying matrix needs to be preserved (e.g., Hankel, Toeplitz, banded matrices etc.). However, the structured low rank approximation is a hard problem in general.

4 Computation algorithms

The eigenvalue problem can be made much easier if we first reduce the matrix to a condensed form, such as tridiagonal or Hessenberg. The same is true of the SVD problem. The eigenvalue problem requires that the reduction be done via similarity transformations. For the singular value decomposition, it is clear that similarity transformations are not required, but the transforming matrices should be orthogonal. One can reduce any matrix $A \in \mathbb{R}^{n\times m}$ to bidiagonal form by an orthogonal equivalence transformation, in which each of the transforming matrices is a product of m or fewer reflectors.

A matrix $B \in \mathbb{R}^{n\times m}$ is said to be bidiagonal if $b_{ij} = 0$ whenever $i > j$ or $i < j - 1$.

Theorem 4.1 ([2])**.** *Let $A \in \mathbb{R}^{n\times m}$ with $n > m$. Then there exist orthogonal $\hat{U} \in \mathbb{R}^{n\times n}$ and $\hat{V} \in \mathbb{R}^{m\times m}$, both products of a finite number of reflectors, and a bidiagonal $B \in \mathbb{R}^{n\times m}$ such that $A = \hat{U}B\hat{V}^{\mathsf{T}}$. There is a finite algorithm to calculate $\hat{U}$, $\hat{V}$ and B.*

Proof. Watkins, Theorem 5.9.21. □

This is called Householder reduction to bidiagonal form. Thus, we are reduced to computing the svd of B. This is same as computing eigenvalues of tridiagonal matrices $B^{\mathsf{T}}B$ or BB^{T}. There are numerous algorithms that can perform this task e.g., the QR algorithm. However, one prefers not to form products $B^{\mathsf{T}}B$ or BB^{T}. There are algorithms which operate directly on B (Watkins, Golub).

References

[1] L. N. Trefethen, D. Bau III, *Numerical linear algebra*, SIAM, 1997.

[2] D. Watkins, *Fundamentals of matrix computations*, second edition, Wiley-Interscience Series of Texts, Monographs, 2002.

[3] G. Golub, C. Van Loan, *Matrix computations*, third edition, John-Hopkins University Press, 1996.

[4] G. Strang, *Linear algebra and its applications*, fourth edition, Cengage Learning, 2018.

[5] A. Antoulas, *Approximation of Large scale dynamical systems*, SIAM, 2005.

"Don't take yourself too seriously, no one else does.."-Anonymous