II.6.SVD

March 6, 2023

1 II.6 Singular Value Decomposition

In this chapter we discuss the *Singular Value Decomposition (SVD)*: a matrix factorisation that encodes how much a matrix "stretches" a random vector. This includes *singular values*, the largest of which dictates the 2-norm of the matirx.

Definition 1 (singular value decomposition) For $A \in \mathbb{C}^{m \times n}$ with rank $r > 0$, the *(reduced) singular value decomposition (SVD)* is

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

where $U \in \mathbb{C}^{m \times r}$ and $V \in \mathbb{C}^{r \times n}$ have orthonormal columns and $\Sigma \in \mathbb{R}^{r \times r}$ is diagonal whose diagonal entries, which which we call *singular values*, are all positive and non-increasing: $\sigma_1 \geq \cdots \geq \sigma_r > 0$. The *full singular value decomposition (SVD)* is

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

where $U \in U(m)$ and $V \in U(n)$ are unitary matrices and $\Sigma \in \mathbb{R}^{m \times n}$ has only diagonal non-zero entries, i.e., if $m > n$,

$$\Sigma = \begin{bmatrix} \sigma_1 \\ & \ddots \\ && \sigma_n \\ && 0 \\ && \vdots \\ && 0 \end{bmatrix}$$

and if $m < n$,

$$\Sigma = \begin{bmatrix} \sigma_1 \\ & \ddots \\ && \sigma_m & 0 & \cdots & 0 \end{bmatrix}$$

where $\sigma_k = 0$ if $k > r$.

In particular, we discuss:

1. Existence of the SVD: we show that an SVD exists by relating it to the eigenvalue Decomposition of $A^\star A$ and $AA^\star$.
2. 2-norm and SVD: the 2-norm of a matrix is defined in terms of the largest singular value.
3. Best rank-k approximation and compression: the best approximation of a matrix by a smaller rank matrix can be constructed using the SVD, which gives an effective way to compress matrices.

[ ]:
```julia
using LinearAlgebra, Plots
```

1.1 1. Existence

To show the SVD exists we first establish some properties of a *Gram matrix* ($A^\star A$):

Proposition 1 (Gram matrix kernel) The kernel of A is the also the kernel of $A^\star A$.

Proof If $A^\star A\mathbf{x} = 0$ then we have

$$0 = \mathbf{x}^\star A^\star A\mathbf{x} = \|A\mathbf{x}\|^2$$

which means $A\mathbf{x} = 0$ and $\mathbf{x} \in \ker(A)$.

Proposition 2 (Gram matrix diagonalisation) The Gram-matrix satisfies

$$A^\star A = Q\Lambda Q^\star \in \mathbb{C}^{n\times n}$$

is a Hermitian matrix where $Q \in U(n)$ and the eigenvalues λ_k are real and non-negative. If $A \in \mathbb{R}^{m\times n}$ then $Q \in O(n)$.

Proof $A^\star A$ is Hermitian so we appeal to the spectral theorem for the existence of the decomposition, and the fact that the eigenvalues are real. For the corresponding (orthonormal) eigenvector $\mathbf{q}_k$,

$$\lambda_k = \lambda_k \mathbf{q}_k^\star \mathbf{q}_k = \mathbf{q}_k^\star A^\star A \mathbf{q}_k = \|A\mathbf{q}_k\|^2 \geq 0.$$

This connection allows us to prove existence:

Theorem 1 (SVD existence) Every $A \in \mathbb{C}^{m\times n}$ has an SVD.

Proof Consider

$$A^\star A = Q\Lambda Q^\star.$$

Assume (as usual) that the eigenvalues are sorted in decreasing modulus, and so $\lambda_1, \ldots, \lambda_r$ are an enumeration of the non-zero eigenvalues and

$$V := \left[\mathbf{q}_1 | \cdots | \mathbf{q}_r\right]$$

the corresponding (orthonormal) eigenvectors, with

$$K = \left[\mathbf{q}_{r+1} | \cdots | \mathbf{q}_n\right]$$

the corresponding kernel. Define

$$\Sigma := \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_r} \end{bmatrix}$$

Now define

$$U := AV\Sigma^{-1}$$

which is orthogonal since \$A^ A V = V Σ^2 \$:

$$U^\star U = \Sigma^{-1} V^\star A^\star A V \Sigma^{-1} = I.$$

Thus we have

$$U\Sigma V^\star = AVV^\star = A \underbrace{\left[V | K\right]}_{Q} \underbrace{\begin{bmatrix} V^\star \\ K^\star \end{bmatrix}}_{Q^\star}$$

where we use the fact that $AK = 0$ so that concatenating K does not change the value.

1.2 2. 2-norm and SVD

Singular values tell us the 2-norm:

Corollary 1 (singular values and norm)

$$\|A\|_2 = \sigma_1$$

and if $A \in \mathbb{C}^{n\times n}$ is invertible, then

$$\|A^{-1}\|_2 = \sigma_n^{-1}$$

Proof

First we establish the upper-bound:

$$\|A\|_2 \leq \ \|U\|_2\|\Sigma\|_2\|V^\star\|_2 = \|\Sigma\|_2 = \sigma_1$$

This is attained using the first right singular vector:

$$\|A\mathbf{v}_1\|_2 = \|\Sigma V^\star \mathbf{v}_1\|_2 = \|\Sigma \mathbf{e}_1\|_2 = \sigma_1$$

The inverse result follows since the inverse has SVD

$$A^{-1} = V\Sigma^{-1}U^\star = (VW)(W\Sigma^{-1}W)(WU)^\star$$

is the SVD of A^{-1}, i.e. $VW \in U(n)$ are the left singular vectors and WU are the right singular vectors, where

$$W := P_\sigma = \begin{bmatrix} & & 1 \\ & ⋰ & \\ 1 & & \end{bmatrix}$$

is the permutation that reverses the entries, that is, σ has Cauchy notation

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ n & n-1 & \cdots & 1 \end{pmatrix}.$$

We will not discuss in this module computation of singular value decompositions or eigenvalues: they involve iterative algorithms (actually built on a sequence of QR decompositions).

1.3 3. Best rank-k approximation and compression

One of the main usages for SVDs is low-rank approximation:

Theorem 2 (best low rank approximation) The matrix

$$A_k := \left[\mathbf{u}_1|\cdots|\mathbf{u}_k\right] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \left[\mathbf{v}_1|\cdots|\mathbf{v}_k\right]^\star$$

is the best 2-norm approximation of A by a rank k matrix, that is, for all rank-k matrices B, we have

$$\|A - A_k\|_2 \leq \|A - B\|_2.$$

Proof We have

$$A - A_k = U \begin{bmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{k+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_r \end{bmatrix} V^\star.$$

Suppose a rank-k matrix B has

$$\|A - B\|_2 < \|A - A_k\|_2 = \sigma_{k+1}.$$

For all $\mathbf{w} \in \ker(B)$ we have

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

But for all $\mathbf{u} \in \mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$, that is, $\mathbf{u} = V[:, 1 : k + 1]\mathbf{c}$ for some $\mathbf{c} \in \mathbb{R}^{k+1}$ we have

$$\|A\mathbf{u}\|_2^2 = \|U\Sigma_k\mathbf{c}\|_2^2 = \|\Sigma_k\mathbf{c}\|_2^2 = \sum_{j=1}^{k+1}(\sigma_j c_j)^2 \geq \sigma_{k+1}^2\|c\|^2,$$

i.e., $\|A\mathbf{u}\|_2 \geq \sigma_{k+1}\|c\|$. Thus $\mathbf{w}$ cannot be in this span.

The dimension of the span of $\ker(B)$ is at least $n - k$, but the dimension of $\mathrm{span}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$ is at least $k+1$. Since these two spaces cannot intersect we have a contradiction, since $(n-r)+(r+1) = n + 1 > n$.

Example 1 (Hilbert matrix) Here we show an example of a simple low-rank approximation using the SVD. Consider the Hilbert matrix:

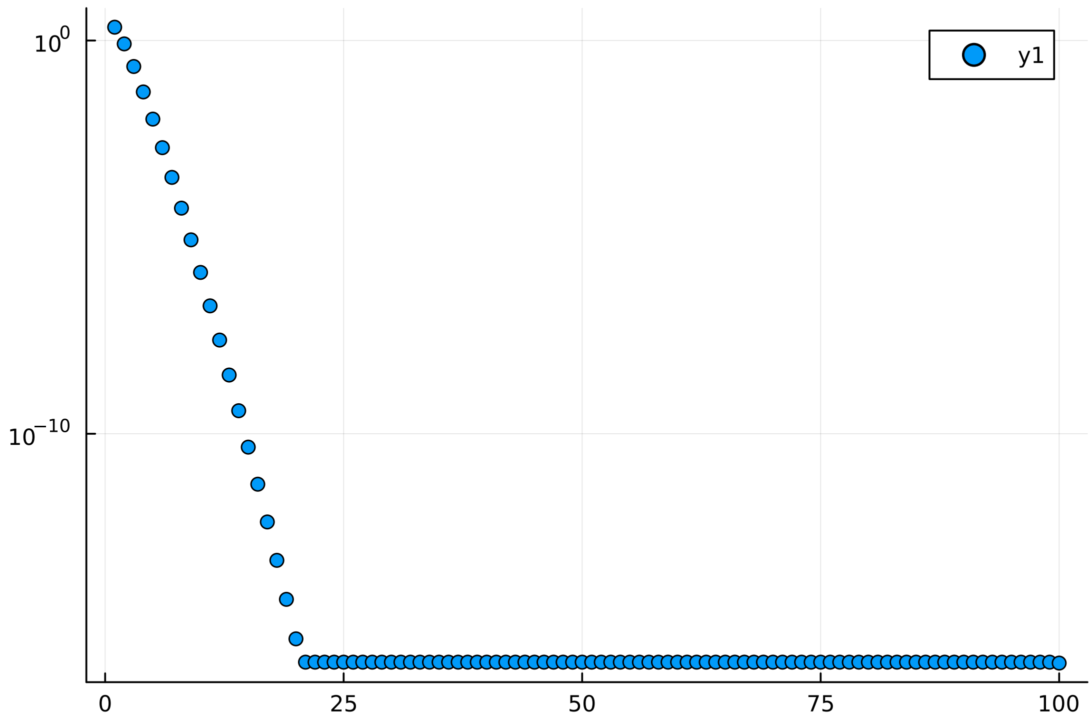
[ ]:
```
hilbertmatrix(n) = [1/(k+j-1) for j = 1:n, k=1:n]
hilbertmatrix(5)
```

[ ]:
```
5×5 Matrix{Float64}:
 1.0       0.5       0.333333  0.25      0.2
 0.5       0.333333  0.25      0.2       0.166667
 0.333333  0.25      0.2       0.166667  0.142857
 0.25      0.2       0.166667  0.142857  0.125
 0.2       0.166667  0.142857  0.125     0.111111
```

That is, the $H[k, j] = 1/(k + j - 1)$. This is a famous example of matrix with rapidly decreasing singular values:

[ ]:
```
H = hilbertmatrix(100)
U, ,V = svd(H)
scatter( ; yscale=:log10)
```

[ ]:



Note numerically we typically do not get a exactly zero singular values so the rank is always treated as $\min(m, n)$. Because the singular values decay rapidly we can approximate the matrix very well with a rank 20 matrix:

[ ]:
```
k = 20 # rank
Σ_k = Diagonal(σ[1:k])
U_k = U[:,1:k]
V_k = V[:,1:k]
opnorm(U_k * Σ_k * V_k' - H)
```

[ ]: 8.20222266307798e-16

Note that this can be viewed as a *compression* algorithm: we have replaced a matrix with $100^2 = 10,000$ entries by two matrices and a vector with $4,000$ entries without losing any information. In the problem sheet we explore the usage of low rank approximation to smooth functions and to compress images.