Mathematical Toolkit **Autumn 2023**

Lecture 8: October 19, 2023

Lecturer: Madhur Tulsiani

1 Singular Value Decomposition

In the previous lecture, we considered linear transformations $\varphi : V \to W$ between inner product spaces V and W, and defined the self-adjoint operators $\varphi^*\varphi : V \to V$ and $\varphi\varphi^* : W \to W$. We proved that these have the same non-zero (positive) eigenvalues.

Proposition 1.1 *Let v be an eigenvector of $\varphi^*\varphi$ with eigenvalue $\lambda \neq 0$. Then $\varphi(v)$ is an eigenvector of $\varphi\varphi^*$ with eigenvalue λ. Similarly, if w is an eigenvector of $\varphi\varphi^*$ with eigenvalue $\lambda \neq 0$, then $\varphi^*(w)$ is an eigenvector of $\varphi^*\varphi$ with eigenvalue λ.*

Using the above properties, we can prove the following useful proposition, which extends the concept of eigenvectors to cases when we have $\varphi : V \to W$ and it might not be possible to define eigenvectors since $V \neq W$ (also φ may not be self-adjoint so we may not get orthonormal eigenvectors).

Proposition 1.2 *Let $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ be the non-zero eigenvalues of $\varphi^*\varphi$, and let $v_1, \ldots, v_r$ be a corresponding orthonormal eigenvectors (since $\varphi^*\varphi$ is self-adjoint, these are a subset of some orthonormal eigenbasis). For $w_1, \ldots, w_r$ defined as $w_i = \varphi(v_i)/\sigma_i$, we have that*

1. *$\{w_1, \ldots, w_r\}$ form an orthonormal set.*
2. *For all $i \in [r]$*

$$\varphi(v_i) = \sigma_i \cdot w_i \quad \text{and} \quad \varphi^*(w_i) = \sigma_i \cdot v_i \,.$$

Proof: For any $i, j \in [r]$, $i \neq j$, we note that

$$\langle w_i, w_j \rangle = \left\langle \frac{\varphi(v_i)}{\sigma_i}, \frac{\varphi(v_j)}{\sigma_j} \right\rangle = \frac{1}{\sigma_i\sigma_j} \cdot \langle \varphi(v_i), \varphi(v_j) \rangle = \frac{1}{\sigma_i\sigma_j} \cdot \langle \varphi^*\varphi(v_i), v_j \rangle$$
$$= \frac{\sigma_i}{\sigma_j} \cdot \langle v_i, v_j \rangle = 0 \,.$$

Thus, the vectors $\{w_1, \ldots, w_r\}$ form an orthonormal set. We also get $\varphi(v_i) = \sigma_i \cdot w_i$ from the definition of w_i. For proving $\varphi^*(w_i) = v_i$, we note that

$$\varphi^*(w_i) \;=\; \varphi^*\left(\frac{\varphi(v_i}{\sigma_i}\right) \;=\; \frac{1}{\sigma_i} \cdot \varphi^*\varphi(v_i) \;=\; \frac{\sigma_i^2}{\sigma_i} \cdot v_i \;=\; \sigma_i \cdot v_i \,,$$

which completes the proof. ■

The values $\sigma_1, \ldots, \sigma_r$ are known as the (non-zero) singular values of φ. For each $i \in [r]$, the vector v_i is known as the right singular vector and w_i is known as the left singular vector corresponding to the singular value σ_i.

Proposition 1.3 *Let r be the number of non-zero eigenvalues of $\varphi^*\varphi$. Then,*

$$\operatorname{rank}(\varphi) \;=\; \dim(\operatorname{im}(\varphi)) \;=\; r\,.$$

Using the above, we can write φ in a particularly convenient form. We first need the following definition.

Definition 1.4 *Let V, W be inner product spaces and let $v \in V, w \in W$ be any two vectors. The* outer product *of w with v, denoted as $|w\rangle\langle v|$, is a linear transformation from V to W such that*

$$|w\rangle\langle v|\,(u) \;:=\; \langle v, u\rangle \cdot w\,.$$

Note that if $\|v\| = 1$, then $|w\rangle\langle v|\,(v) = w$ and $|w\rangle\langle v|\,(u) = 0$ for all $u \perp v$.

Exercise 1.5 *Show that for any $v \in V$ and $w \in W$, we have*

$$\operatorname{rank}\left(|w\rangle\langle v|\right) \;=\; \dim\left(\operatorname{im}\left(|w\rangle\langle v|\right)\right) \;=\; 1\,.$$

We can then write $\varphi : V \to W$ in terms of outer products of its singular vectors.

Proposition 1.6 *Let V, W be finite dimensional inner product spaces and let $\varphi : V \to W$ be a linear transformation with non-zero singular values $\sigma_1, \ldots, \sigma_r$, right singular vectors $v_1, \ldots, v_r$ and left singular vectors $w_1, \ldots, w_r$. Then,*

$$\varphi \;=\; \sum_{i=1}^{r} \sigma_i \cdot |w_i\rangle\langle v_i|\;.$$

Exercise 1.7 *If $\varphi : V \to V$ is a self-adjoint operator with $\dim(V) = n$, then the real spectral theorem proves the existence of an orthonormal basis of eigenvectors, say $\{v_1, \ldots, v_n\}$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Check that in this case, we can write φ as*

$$\varphi \;=\; \sum_{i=1}^{n} \lambda_i \cdot |v_i\rangle\langle v_i|\;.$$

Note that while the above decomposition has possibly negative coefficients (the λ_is), the singular value decomposition only has positive coefficients (the σ_is). Why is this the case?

2 Singular Value Decomposition for matrices

Using the previous discussion, we can write matrices in convenient form. Let $A \in \mathbb{C}^{m \times n}$, which can be thought of as an operator from $\mathbb{C}^n$ to $\mathbb{C}^m$. Let $\sigma_1, \ldots, \sigma_r$ be the non-zero singular values and let $v_1, \ldots, v_r$ and $w_1, \ldots, w_r$ be the right and left singular vectors respectively. Note that $V = \mathbb{C}^n$ and $W = \mathbb{C}^m$ and $v \in V, w \in W$, we can write the operator $|w\rangle \langle v|$ as the matrix wv^*, there v^* denotes $\overline{v^T}$. This is because for any $u \in V$, $wv^*u = w(v^*u) = \langle v, u\rangle \cdot w$. Thus, we can write

$$A \;=\; \sum_{i=1}^{r} \sigma_i \cdot w_i v_i^* \,.$$

Let $W \in \mathbb{C}^{m \times r}$ be a matrix with $w_1, \ldots, w_r$ as columns, such that i^{th} column equals w_i. Similarly, let $V \in \mathbb{C}^{n \times r}$ be a matrix with $v_1, \ldots, v_r$ as the columns. Let $\Sigma \in \mathbb{C}^{r \times r}$ be a diagonal matrix with $\Sigma_{ii} = \sigma_i$. Then, check that the above expression for A can also be written as

$$A \;=\; W \Sigma V^* \,,$$

where $V^* = \overline{V^T}$ as before.

We can also complete the bases $\{v_1, \ldots, v_r\}$ and $\{w_1, \ldots, w_r\}$ to bases for $\mathbb{C}^n$ and $\mathbb{C}^m$ respectively and write the above in terms of unitary matrices.

Definition 2.1 *A matrix $U \in \mathbb{C}^{n \times n}$ is known as a* unitary matrix *if the columns of U form an orthonormal basis for $\mathbb{C}^n$.*

Proposition 2.2 *Let $U \in \mathbb{C}^{n \times n}$ be a unitary matrix. Then $UU^* = U^*U = \mathrm{id}$, where* id *denotes the identity matrix.*

Let $\{v_1, \ldots, v_n\}$ be a completion of $\{v_1, \ldots, v_r\}$ to an orthonormal basis of $\mathbb{C}^n$, and let $V_n \in \mathbb{C}^{n \times n}$ be a unitary matrix with $\{v_1, \ldots, v_n\}$ as columns. Similarly, let $W_m \in \mathbb{C}^{m \times m}$ be a unitary matrix with a completion of $\{w_1, \ldots, w_r\}$ as columns. Let $\Sigma' \in \mathbb{C}^{m \times n}$ be a matrix with $\Sigma'_{ii} = \sigma_i$ if $i \leq r$, and all other entries equal to zero. Then, we can also write

$$A = W_m \Sigma' V_n^* \,.$$

3 Low-rank approximation for matrices

Given a matrix $A \in \mathbb{C}^{m \times n}$, we want to find a matrix B of rank at most k which "approximates" A. For now we will consider the notion of approximation in spectral norm i.e., we want to minimize $\|A - B\|_2$, where

$$\|(A - B)\|_2 \;=\; \max_{v \neq 0} \frac{\|(A - B)v\|_2}{\|v\|_2} \,.$$

Here, $\|v\|_2 = \sqrt{\langle v, v\rangle}$ denotes the norm defined by the standard inner product on $\mathbb{C}^n$. The 2 in the notation $\|\cdot\|_2$ comes from the express from the expression we get by expressing v in the orthonormal basis of the coordinate vectors. If $v = (c_1, \ldots, c_n)^T$, then $\|v\|_2 = \left(\sum_{i=1}^n |c_i|^2\right)^{1/2}$ which is simply the Euclidean norm we are familiar with [1]. Note that while the norm here seems to be defined in terms of the coefficients, which indeed depend on the choice of the orthonormal basis, the value of the norm is in fact $\sqrt{\langle v, v\rangle}$ *which is just a function of the vector itself and not of the basis we are working with*. The basis and the coefficients merely provide a convenient way of computing the norm.

SVD also gives the optimal solution for another notion of approximation: minimizing the Frobenius norm $\|A - B\|_F$, which equals $(\sum_{ij}(A_{ij} - B_{ij})^2)^{1/2}$. We will see this later. Let $A = \sum_{i=1}^r w_i v_i^*$ be the singular value decomposition of A and let $\sigma_1 \geq \cdots \geq \sigma_r > 0$. If $k \geq r$, we can simply use $B = A$ since $\text{rank}(A) = r$. If $k < r$, we claim that $A_k = \sum_{i=1}^k \sigma_i w_i v_i^*$ is the optimal solution. If is easy to check the following.

Proposition 3.1 $\|A - A_k\|_2 = \sigma_{k+1}$.

Proof: Complete $v_1, \ldots, v_k$ to an orthonormal basis $v_1, \ldots, v_n$ for $\mathbb{C}^n$. Given any $v \in \mathbb{C}^n$, we can uniquely express it as $\sum_{i=1}^n c_i \cdot v_i$ for appropriate coefficients $c_1, \ldots, c_n$. Thus, we have

$$(A - A_k)v = \left(\sum_{j=k+1}^r \sigma_j \cdot w_j v_j*\right)\left(\sum_{i=1}^n c_i \cdot v_i\right) = \sum_{j=k+1}^r \sum_{i=1}^n c_i\sigma_j \cdot \langle v_j, v_i\rangle \cdot w_j = \sum_{j=k+1}^r c_j\sigma_j \cdot w_j\,,$$

where the last equality uses the orthonormality of $\{v_1, \ldots, v_n\}$. We can also complete $w_1, \ldots, w_r$ to an orthonormal basis $w_1, \ldots, w_m$ for $\mathbb{C}^m$. Since $(A - A_k)$ is already expressed in this basis above, we get that

$$\|(A - A_k)v\|_2^2 = \left\|\sum_{j=k+1}^r c_j\sigma_j \cdot w_j\right\|_2^2 = \left\langle \sum_{j=k+1}^r c_j\sigma_j \cdot w_j, \sum_{j=k+1}^r c_j\sigma_j \cdot w_j \right\rangle = \sum_{j=k+1}^r \left|c_j\right|^2 \cdot \sigma_j^2\,.$$

Finally, as in the computation with Rayleigh quotients, we have that for any $v \neq 0$ expressed as $v = \sum_{i=1}^n c_i \cdot v_i$,

$$\frac{\|(A - A_k)v\|_2^2}{\|v\|_2^2} = \frac{\sum_{j=k+1}^r \left|c_j\right|^2 \cdot \sigma_j^2}{\sum_{i=1}^n |c_i|^2} \leq \frac{\sum_{j=k+1}^r \left|c_j\right|^2 \cdot \sigma_{k+1}^2}{\sum_{i=1}^n |c_i|^2} \leq \sigma_{k+1}^2\,.$$

This gives that $\|A - A_k\|_2 \leq \sigma_{k+1}$. Check that it is in fact equal to σ_{k+1} (why?) ∎

[1] In general, one can consider the norm $\|v\|_p := \left(\sum_{i=1}^n |c_i|^p\right)^{1/p}$ for any $p \geq 1$. While these are indeed valid notions of distance satisfying a triangle inequality for any $p \geq 1$, they do not arise as a square root of an inner product when $p \neq 2$.

In fact the proof above actually shows the following:

Exercise 3.2 *Let $M \in \mathbb{C}^{m \times n}$ be any matrix with singular values $\sigma_1 \geq \cdots \sigma_r > 0$. Then, $\|M\|_2 = \sigma_1$ i.e., the spectral norm of a matrix is actually equal to its largest singular value.*

Thus, we know that the error of the best approximation B is at most σ_{k+1}. To show the lower bound, we need the following fact.

Exercise 3.3 *Let V be a finite-dimensional vector space and let S_1, S_2 be subspaces of V. Then, $S_1 \cap S_2$ is also a subspace and satisfies*

$$\dim(S_1 \cap S_2) \;\geq\; \dim(S_1) + \dim(S_2) - \dim(V)\,.$$

We can now show the following.

Proposition 3.4 *Let $B \in \mathbb{C}^{m \times n}$ have $\operatorname{rank}(B) \leq k$ and let $k < r$. Then $\|A - B\|_2 \geq \sigma_{k+1}$.*

Proof: By rank-nullity theorem $\dim(\ker(B)) \geq n - k$. Thus, by the fact above

$$\dim\left(\ker(B) \cap \operatorname{Span}\left(v_1, \ldots, v_{k+1}\right)\right) \;\geq\; (n-k) + (k+1) - n \;\geq\; 1\,.$$

Thus, there exists a $z \in \ker(B) \cap \operatorname{Span}\left(v_1, \ldots, v_{k+1}\right) \setminus \{0\}$. Then,

$$\begin{aligned} \|(A-B)z\|_2^2 \;=\; \|Az\|_2^2 \;=\; \langle z, A^*Az\rangle \;&=\; \mathcal{R}_{A^*A}(z) \cdot \|z\|_2^2 \\ &\geq\; \left(\min_{y \in \operatorname{Span}(v_1,\ldots,v_{k+1}) \setminus \{0\}} \mathcal{R}_{A^*A}(y)\right) \cdot \|z\|_2^2 \\ &\geq\; \sigma_{k+1}^2 \cdot \|z\|_2^2\,. \end{aligned}$$

Thus, there exists a $z \neq 0$ such that $\|(A-B)z\|_2 \geq \sigma_{k+1} \cdot \|z\|_2$, which implies $\|A-B\|_2 \geq \sigma_{k+1}$. ■