Mathematical Toolkit **Autumn 2025**

Lecture 8: October 23, 2025

Lecturer: Madhur Tulsiani

1 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 $\mathrm{rank}(A) = r$. 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 1.1 $\|A - A_k\|_2 \;=\; \sigma_{k+1}$.

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

[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 $p \neq 2$.

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?) ■

In fact the proof above actually shows the following:

Exercise 1.2 *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 1.3 *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 1.4 *Let $B \in \mathbb{C}^{m \times n}$ have $\mathrm{rank}(B) \leq k$ 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

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

Thus, there exists a $z \in \ker(B) \cap \operatorname{Span}(v_1, \ldots, v_{k+1}) \setminus \{0\}$. 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 $\|A - B\|_2 \geq \sigma_{k+1}$. ■

2 Least squares approximation

We discuss another application of singular value decomposition (SVD) of matrices. Let $a_1, \ldots, a_n \in \mathbb{R}^d$ be points which we want to fit to a low-dimensional subspace. The goal is to find a subspace S of $\mathbb{R}^d$ of dimension at most k to minimize $\sum_{i=1}^n (\operatorname{dist}(a_i, S))^2$, where $\operatorname{dist}(a_i, S)$ denotes the distance of a_i from the closest point in S. As we will see later in the course, this also the same as Principal Component Analysis (PCA).

We first prove the following.

Claim 2.1 *Let $u_1, \ldots, u_k$ be an orthonormal basis for S. Then*

$$(\operatorname{dist}(a_i, S))^2 \;=\; \|a_i\|_2^2 - \sum_{j=1}^k \langle a_i, u_j \rangle^2 \,.$$

Proof: Complete $u_1, \ldots, u_k$ to an orthonormal basis $u_{k+1}, \ldots, u_d$ for all of $\mathbb{R}^d$. For any point $v \in \mathbb{R}^d$, where exist $c_1, \ldots, c_d \in \mathbb{R}$ such that $v = \sum_{j=1}^d c_j \cdot u_j$. To find the distance $\operatorname{dist}(v, S) = \min_{u \in S} \|v - u\|$, we need to find the point $u \in S$, which is closest to v.

Let $u = \sum_{j=1}^k b_j \cdot u_k$ be an arbitrary point in S (any $u \in S$ can be written in this form, since $u_1, \ldots, u_k$ form a basis for S). We have that

$$\|v - u\|^2 \;=\; \left\| \sum_{j=1}^k (c_j - b_j) \cdot u_j + \sum_{j=k+1}^d c_j \cdot u_j \right\|^2 \;=\; \sum_{j=1}^k (c_j - b_j)^2 + \sum_{j=k+1}^d c_j^2 \,,$$

which is minimized when $b_j = c_j$ for all $j \in [k]$. Thus, the cloest point $u \in S$ to $v = \sum_{j=1}^d c_j \cdot u_j$ is given by $u = \sum_{j=1}^k c_j \cdot u_j$, with $v - u = \sum_{j=k+1}^d c_j \cdot u_j$. Moreover, since $u_1, \ldots, u_d$ form an *orthonormal* basis, we have $c_j = \langle u_j, v \rangle$ for all $j \in [d]$, which gives

$$\|v - u\|^2 \;=\; \sum_{j=k+1}^d c_j^2 \;=\; \sum_{j=1}^d c_j^2 - \sum_{j=1}^k c_j^2 \;=\; \|v\|^2 - \sum_{j=1}^k \langle u_j, v \rangle^2 \,.$$

Using the above for each a_i (as the point v) completes the proof. ■

Thus, the goal is to find a set of k orthonormal vectors $u_1, \ldots, u_k$ to maximize the quantity $\sum_{i=1}^{n} \sum_{j=1}^{k} \left\langle a_i, u_j \right\rangle^2$. Let $A \in \mathbb{R}^{n \times d}$ be a matrix with the i^{th} row equal to a_i^T. Then, we need to find orthonormal vectors $u_1, \ldots, u_k$ to maximize $\|Au_1\|_2^2 + \cdots + \|Au_k\|_2^2$. We will prove the following.

Proposition 2.2 *Let $v_1, \ldots, v_r$ be the right singular vectors of A corresponding to singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Then, for all $k \leq r$ and all orthonormal sets of vectors $u_1, \ldots, u_k$*

$$\|Au_1\|_2^2 + \cdots + \|Au_k\|_2^2 \;\leq\; \|Av_1\|_2^2 + \cdots + \|Av_k\|_2^2$$

Thus, the optimal solution is to take $S = \text{Span}\,(v_1, \ldots, v_k)$. We prove the above by induction on k. For $k = 1$, we note that

$$\|Au_1\|_2^2 \;=\; \left\langle A^T A u_1, u_1 \right\rangle \;\leq\; \max_{v \in \mathbb{R}^d \setminus \{0\}} \mathcal{R}_{A^T A}(v) \;=\; \sigma_1^2 \;=\; \|Av_1\|_2^2 \,.$$

To prove the induction step for a given $k \leq r$, define

$$V_{k-1}^{\perp} = \left\{ v \in \mathbb{R}^d \mid \langle v, v_i \rangle = 0 \;\; \forall i \in [k-1] \right\} .$$

First prove the following claim.

Claim 2.3 *Given an orthonormal set $u_1, \ldots, u_k$, there exist orthonormal vectors $u_1', \ldots, u_k'$ such that*

- $u_k' \in V_{k-1}^{\perp}$.
- $\text{Span}\,(u_1, \ldots, u_k) \;=\; \text{Span}\,(u_1', \ldots, u_k')$.
- $\|Au_1\|_2^2 + \cdots + \|Au_k\|_2^2 \;=\; \|Au_1'\|_2^2 + \cdots + \|Au_k'\|_2^2$.

Proof: We only provide a sketch of the proof here. Let $S = \text{Span}\,(\{u_1, \ldots, u_k\})$. Note that $\dim\left(V_{k-1}^{\perp}\right) = d - k + 1$ (why?) and $\dim(S) = k$. Thus,

$$\dim\left(V_{k-1}^{\perp} \cap S\right) \;\geq\; k + (d - k + 1) - d \;=\; 1 \,.$$

Hence, there exists $u_k' \in V_{k-1}^{\perp} \cap S$ with $\|u_k'\| = 1$. Completing this to an orthonormal basis of S gives orthonormal $u_1', \ldots, u_k'$ with the first and second properties. We claim that this already implies the third property (why?). ■

Thus, we can assume without loss of generality that the given vectors $u_1, \ldots, u_k$ are such that $u_k \in V_{k-1}^{\perp}$. Hence,

$$\|Au_k\|_2^2 \;\leq\; \max_{\substack{v \in V_{k-1}^{\perp} \\ \|v\|=1}} \|Av\|_2^2 \;=\; \sigma_k^2 \;=\; \|Av_k\|_2^2 \,.$$

Also, by the inductive hypothesis, we have that

$$\|Au_1\|_2^2 + \cdots + \|Au_{k-1}\|_2^2 \;\leq\; \|Av_1\|_2^2 + \cdots + \|Av_{k-1}\|_2^2 \,,$$

which completes the proof. The above proof can also be used to prove that SVD gives the best rank k approximation to the matrix A in Frobenius norm. We will see this in the next homework.