### Basic linear algebra recap. Convergence rates.

#### Seminar

Optimization for ML. Faculty of Computer Science. HSE University

∌ റ ⊘

• Naive matmul  $\mathcal{O}(n^3)$ , naive matvec  $\mathcal{O}(n^2)$ 

- Naive matmul  $\mathcal{O}(n^3)$ , naive matvec  $\mathcal{O}(n^2)$
- All matrices have SVD

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

- Naive matmul  $\mathcal{O}(n^3)$ , naive matvec  $\mathcal{O}(n^2)$
- All matrices have SVD

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

 $\bullet \ \operatorname{tr}(ABCD) = \operatorname{tr}(DABC) = \operatorname{tr}(CDAB) = \operatorname{tr}(BCDA) \ \text{for any matrices ABCD if the multiplication defined}.$ 

- Naive matmul  $\mathcal{O}(n^3)$ , naive matvec  $\mathcal{O}(n^2)$
- All matrices have SVD

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

- tr(ABCD) = tr(DABC) = tr(CDAB) = tr(BCDA) for any matrices ABCD if the multiplication defined.
- $\langle A, B \rangle = \operatorname{tr}(A^T B)$

### **Convergence** rate

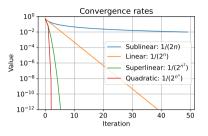


Figure 1: Illustration of different convergence rates

• Linear (geometric, exponential) convergence:

$$r_k \le Cq^k, \quad 0 < q < 1, C > 0$$

**⊕** ດ **ø** 

### **Convergence rate**

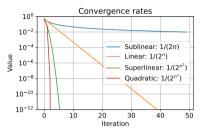


Figure 1: Illustration of different convergence rates

• Linear (geometric, exponential) convergence:

$$r_k \le Cq^k, \quad 0 < q < 1, C > 0$$

 Any convergent sequence, that is slower (faster) than any linearly convergent sequence has sublinear (superlinear) convergence

### **Convergence** rate

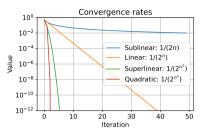


Figure 1: Illustration of different convergence rates

• Linear (geometric, exponential) convergence:

$$r_k \le Cq^k, \quad 0 < q < 1, C > 0$$

- Any convergent sequence, that is slower (faster) than any linearly convergent sequence has sublinear (superlinear) convergence
- The infimum of all  $0 \le q < 1$  such that  $r_k \le Cq^k$  is called the **constant of linear convergence**, and  $q^k$  is called the rate of convergence.

**⊕** ດ **ø** 

Let  $\{r_k\}_{k=m}^\infty$  be a sequence of non-negative numbers, converging to zero, and let

$$q = \lim_{k \to \infty} \sup_{k} \ r_k^{1/k}$$

• If  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.

Let  $\{r_k\}_{k=m}^{\infty}$  be a sequence of non-negative numbers, converging to zero, and let

$$q = \lim_{k \to \infty} \sup_{k} \ r_k^{1/k}$$

- If  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.

Let  $\{r_k\}_{k=m}^\infty$  be a sequence of non-negative numbers, converging to zero, and let

$$q = \lim_{k \to \infty} \sup_{k} \ r_k^{1/k}$$

- If  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.
- If q=1, then  $\{r_k\}_{k=m}^{\infty}$  has sublinear convergence.

♥ n ø

Let  $\{r_k\}_{k=m}^\infty$  be a sequence of non-negative numbers, converging to zero, and let

$$q = \lim_{k \to \infty} \sup_{k} \ r_k^{1/k}$$

- If  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.
- If q=1, then  $\{r_k\}_{k=m}^{\infty}$  has sublinear convergence.
- The case q > 1 is impossible.

♥ റ ഉ

Let  $\{r_k\}_{k=m}^{\infty}$  be a sequence of strictly positive numbers converging to zero. Let

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

• If there exists q and  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.

Let  $\{r_k\}_{k=m}^{\infty}$  be a sequence of strictly positive numbers converging to zero. Let

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- If there exists q and  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.



Let  $\{r_k\}_{k=m}^{\infty}$  be a sequence of strictly positive numbers converging to zero. Let

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- If there exists q and  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.
- If q does not exist, but  $q=\lim_{k\to\infty}\sup_k\frac{r_{k+1}}{r_k}<1$ , then  $\{r_k\}_{k=m}^\infty$  has linear convergence with a constant not exceeding q.



Let  $\{r_k\}_{k=m}^{\infty}$  be a sequence of strictly positive numbers converging to zero. Let

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- If there exists q and  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.
- If q does not exist, but  $q = \lim_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with a constant not exceeding q.
- exceeding q.

   If  $\lim_{k\to\infty}\inf_k\frac{r_{k+1}}{r_k}=1$ , then  $\{r_k\}_{k=m}^\infty$  has sublinear convergence.

Let  $\{r_k\}_{k=m}^{\infty}$  be a sequence of strictly positive numbers converging to zero. Let

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- If there exists q and  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.
- If q does not exist, but  $q=\lim_{k\to\infty}\sup_k\frac{r_{k+1}}{r_k}<1$ , then  $\{r_k\}_{k=m}^\infty$  has linear convergence with a constant not exceeding q.
- If  $\lim_{k\to\infty}\inf_k\frac{r_{k+1}}{r_k}=1$ , then  $\{r_k\}_{k=m}^\infty$  has sublinear convergence.
- The case  $\lim_{k\to\infty}\inf_k\frac{r_{k+1}}{r_{k}}>1$  is impossible.

Let  $\{r_k\}_{k=m}^{\infty}$  be a sequence of strictly positive numbers converging to zero. Let

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- If there exists q and  $0 \le q < 1$ , then  $\{r_k\}_{k=m}^{\infty}$  has linear convergence with constant q.
- In particular, if q=0, then  $\{r_k\}_{k=m}^{\infty}$  has superlinear convergence.
- If q does not exist, but  $q=\lim_{k\to\infty}\sup_k\frac{r_{k+1}}{r_k}<1$ , then  $\{r_k\}_{k=m}^\infty$  has linear convergence with a constant not exceeding q.
- If  $\lim_{k\to\infty}\inf_k\frac{r_{k+1}}{r_k}=1$ , then  $\{r_k\}_{k=m}^\infty$  has sublinear convergence.
- The case  $\lim_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} > 1$  is impossible.
- In all other cases (i.e., when  $\lim_{k\to\infty}\inf_k\frac{r_{k+1}}{r_k}<1\leq\lim_{k\to\infty}\sup_k\frac{r_{k+1}}{r_k}$ ) we cannot claim anything concrete about the convergence rate  $\{r_k\}_{k=m}^\infty$ .

Suppose, you have the following expression

$$b = A_1 A_2 A_3 x,$$

where the  $A_1,A_2,A_3\in\mathbb{R}^{3\times 3}$  - random square dense matrices and  $x\in\mathbb{R}^n$  - vector. You need to compute b.

Which one way is the best to do it?

1.  $A_1A_2A_3x$  (from left to right)

Check the simple **?** code snippet after all.



Suppose, you have the following expression

$$b = A_1 A_2 A_3 x,$$

where the  $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$  - random square dense matrices and  $x \in \mathbb{R}^n$  - vector. You need to compute b.

Which one way is the best to do it?

- 1.  $A_1A_2A_3x$  (from left to right)
- 2.  $(A_1(A_2(A_3x)))$  (from right to left)

Check the simple **d** code snippet after all.

Problems



Suppose, you have the following expression

$$b = A_1 A_2 A_3 x,$$

where the  $A_1,A_2,A_3\in\mathbb{R}^{3\times 3}$  - random square dense matrices and  $x\in\mathbb{R}^n$  - vector. You need to compute b.

Which one way is the best to do it?

- 1.  $A_1A_2A_3x$  (from left to right)
- 2.  $(A_1(A_2(A_3x)))$  (from right to left)
- 3. It does not matter

Check the simple **?** code snippet after all.

Problems

େ ଚ

Suppose, you have the following expression

$$b = A_1 A_2 A_3 x,$$

where the  $A_1,A_2,A_3\in\mathbb{R}^{3\times 3}$  - random square dense matrices and  $x\in\mathbb{R}^n$  - vector. You need to compute b.

Which one way is the best to do it?

- 1.  $A_1A_2A_3x$  (from left to right)
- 2.  $(A_1(A_2(A_3x)))$  (from right to left)
- 3. It does not matter
- 4. The results of the first two options will not be the same.

Check the simple **?** code snippet after all.



# Problem 2. Connection between Frobenius norm and singular values.

Let  $A \in \mathbb{R}^{m \times n}$ , and let  $q := \min\{m, n\}$ . Show that

$$||A||_F^2 = \sum_{i=1}^q \sigma_i^2(A),$$

where  $\sigma_1(A) \ge \ldots \ge \sigma_q(A) \ge 0$  are the singular values of matrix A. Hint: use the connection between Frobenius norm and scalar product and SVD.

# Problem 3. Know your inner product.

Simplify the following expression:

$$\sum_{i=1}^{n} \langle S^{-1} a_i, a_i \rangle,$$

where  $S = \sum\limits_{i=1}^{n} a_i a_i^T, a_i \in \mathbb{R}^n, \det(S) 
eq 0$ 

⊕ ი ⊘

• 
$$r_k = \frac{1}{3^k}$$

- $r_k = \frac{1}{3^k}$   $r_k = \frac{4}{3^k}$



- $r_k = \frac{1}{3^k}$   $r_k = \frac{4}{3^k}$   $r_k = \frac{1}{k^{10}}$

- $r_k = \frac{1}{3^k}$   $r_k = \frac{4}{3^k}$   $r_k = \frac{1}{k^{10}}$   $r_k = 0.707^k$

- $r_k = \frac{1}{3^k}$   $r_k = \frac{4}{3^k}$
- $r_k = \frac{5}{k^{10}}$
- $r_k = 0.707^k$   $r_k = 0.707^{2^k}$

### Problem 5. One test is simpler, than another

$$r_k = \frac{1}{k^k}$$

# Problem 6. Super but not quadratic.

Show, that the following sequence does not have a quadratic convergence.

$$r_k = \frac{1}{3^{k^2}}$$

# LoRA: Low-Rank Adaptation of Large Language Models (arXiv:2106.09685)

Since current LLMs are too big to fit into the memory of the average user, we need to use some tricks to make them smaller. One of the most popular tricks is LoRA (Low-Rank Adaptation of Large Language Models).

Suppose we have matrix  $W \in \mathbb{R}^{d \times k}$  and we want to perform the following update:

$$W = W_0 + \Delta W$$
.

The main idea of LoRA is to decompose the update  $\Delta W$  into two low-rank matrices:

$$W = W_0 + \Delta W = W_0 + BA, \quad B \in \mathbb{R}^{d \times r}, A \in \mathbb{R}^{r \times k},$$
  
 $rank(A) = rank(B) = r \ll \min\{d, k\}.$ 

Check the  $\clubsuit$  notebook for the example implementation of LoRA.

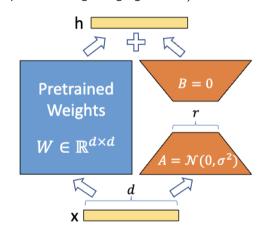


Figure 2: Illustration of LoRA

