Mathematics for Data Science — Lesson 6
Operator and Spectral Norm Exercises

Exercise 1: (Sequences — pedagogical) Work with $(x_n) \subset \mathbb{R}^d$ under any norm $\|\cdot\|$.

a) Show directly from the ε–N definition that $a_n = \dfrac{(-1)^n}{n} \to 0$.

b) Let $v_n = \left(\frac{1}{n}, \frac{(-1)^n}{n}, 1 - \frac{1}{n}\right) \in \mathbb{R}^3$. Prove $v_n \to (0, 0, 1)$ by checking coordinate-wise limits.

Exercise 2: (Asymptotics and rates) Analyse the following growth comparisons.

a) Show that the harmonic series partial sums $H_n = \sum_{k=1}^{n} \frac{1}{k}$ satisfy $H_n = \Theta(\log n)$.

b) Prove $n^\alpha = o(\beta^n)$ for every $\alpha > 0$ and $\beta > 1$.

c) Consider the recursion $x_{k+1} = \rho x_k + \dfrac{c}{k+1}$ with $x_0 \in \mathbb{R}$, $0 < \rho < 1$, $c > 0$. Find the asymptotic rate at which x_k tends to 0.

Exercise 3: (Pedagogical) Consider $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ acting on $(\mathbb{R}^2, \|\cdot\|)$.

a) Compute $\|A\|_1$ and $\|A\|_\infty$ by hand.

b) Compute $A^\top A$ and deduce $\|A\|_2$ from its eigenvalues.

c) Identify a unit vector v such that $\|Av\|_2 = \|A\|_2$ and describe Av.

Exercise 4: (Pedagogical) Let A be linear on $(\mathbb{R}^n, \|\cdot\|)$ and define $\|A\|_{\mathrm{op}} = \sup_{\|x\|=1} \|Ax\|$.

a) Prove $\sup_{\|x\|=1} \|Ax\| = \sup_{\|x\|\le 1} \|Ax\|$.

b) Conclude that $\|A\|_{\mathrm{op}}$ equals the largest radius of $A(\mathbb{S}^{n-1})$ (image of the unit sphere).

Exercise 5: (Intermediate) For a normal matrix A (i.e., $AA^\top = A^\top A$), show that $\|A\|_2 = \max_i |\lambda_i(A)|$.

Exercise 6: (Intermediate) Let $A = U\Sigma V^\top$ be the SVD with $\sigma_1 \ge \cdots \ge \sigma_r > 0$.

a) Show that the best rank-k approximation in spectral norm is $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\top$ and that $\|A - A_k\|_2 = \sigma_{k+1}$.

b) Show that A_k is also optimal for the Frobenius norm and that $\|A - A_k\|_F^2 = \sum_{i>k} \sigma_i^2$.

Exercise 7: (Low-rank linear regression) Let $X \in \mathbb{R}^{n\times d}$, $Y \in \mathbb{R}^{n\times m}$, and fix $k \le \min(d, m)$. Consider

$$\min_{\operatorname{rank}(W)\le k} \|XW - Y\|_F . \tag{$\star$}$$

Assume X has full column rank (if not, replace inverses by pseudoinverses below).

a) Given the SVD $X = P\Sigma Q^\top$ with $\Sigma \in \mathbb{R}^{d\times d}$ invertible, set $Z = Q^\top W$. Show that ($\star$) is equivalent to $\min_{\operatorname{rank}(Z)\le k} \left\|\Sigma Z - P^\top Y\right\|_F$.

b) Prove that an optimal solution is $Z^* = \Sigma^{-1} [\, P^\top Y\,]_k$, where $[\,\cdot\,]_k$ denotes the best rank-k approximation (by SVD truncation).

c) Conclude that $W^* = QZ^*$ solves ($\star$). Comment on how the singular values of $P^\top Y$ influence the error.

Exercise 8: (Pedagogical) Determine convergence and find limits (when they exist) for the following sequences.

a) $a_n = \dfrac{3n^2 + 5n}{2n^2 - 1}$

b) $b_n = \dfrac{2^n + n^{10}}{3 \cdot 2^n - 5}$

c) $c_n = \sin(n\pi) + \dfrac{1}{n}$

d) $d_n = (-0.9)^n$

Exercise 9: (Intermediate) Work with sequences in $\mathbb{R}$ and $\mathbb{R}^d$.

a) Prove that every bounded monotone increasing sequence (a_n) in $\mathbb{R}$ converges.

b) Show that in finite-dimensional $\mathbb{R}^d$, a sequence is Cauchy if and only if it converges.

c) Given two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ on $\mathbb{R}^d$, prove that if $(x_n) \to x$ coordinate-wise, then $(x_n) \to x$ in both norms.

Exercise 10: (Advanced) Explore deeper results about sequences and series.

a) State and prove the Monotone Convergence Theorem: every bounded monotone sequence in $\mathbb{R}$ converges. Explain how this extends to functions.

b) Use the integral test to show that $\sum_{k=1}^{\infty} \frac{1}{k^p}$ converges if and only if $p > 1$.

c) Prove the Bolzano–Weierstrass theorem: every bounded sequence in $\mathbb{R}$ has a convergent subsequence.

Exercise 11: (Limits and Continuity — Pedagogical) Work with limits and continuity in $\mathbb{R}$.

a) Prove using the ε–η definition that $\lim\limits_{x\to 2}(3x + 1) = 7$.

b) Use the squeeze theorem to show $\lim\limits_{x\to 0} \dfrac{\sin x}{x} = 1$.

c) Compute the one-sided limits $\lim\limits_{x\to 1^-} h(x)$ and $\lim\limits_{x\to 1^+} h(x)$ for the piecewise function

$$h(x) = \begin{cases} x^2, & x < 1 \\ 2x, & x \ge 1. \end{cases}$$

Does $\lim\limits_{x\to 1} h(x)$ exist? Is h continuous at $x = 1$?

d) Classify the discontinuity at $x = 1$ for $g(x) = \dfrac{x^2 - 1}{x - 1}$ (removable) and at $x = 0$ for the step function $s(x) = \begin{cases} -1, & x < 0 \\ 1, & x \geq 0 \end{cases}$ (jump).

Exercise 12: (Continuity Concepts — Pedagogical) Explore deeper continuity properties.

a) Determine whether each expression represents an indeterminate form or a definite answer:

(i) $\displaystyle\lim_{x \to 2} \frac{x^2 - 4}{x - 2}$

(ii) $\displaystyle\lim_{x \to 0} \frac{\sin x}{x^2}$

(iii) $\displaystyle\lim_{x \to 1} \frac{2}{x - 1}$

b) Prove that if (x_n) is a sequence with $x_n \to a$ and f is continuous at a, then $f(x_n) \to f(a)$ (sequential characterization of continuity).

c) Show by example that if $\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x)$ does not guarantee f is continuous at a unless this common value also equals $f(a)$.

Exercise 13: (Derivatives and Linearization — Intermediate) Analyze derivatives and first-order approximations.

a) Compute $f'(x)$ for $f(x) = e^{-0.2x^2} + 0.2\sin(2x)$ using the chain rule.

b) Construct the linear model $\ell_1(x) = f(x_0) + f'(x_0)(x - x_0)$ at $x_0 = 1$.

c) Numerically validate: compute $f(1 \pm 0.1)$ and $\ell_1(1 \pm 0.1)$, then find the absolute errors.

d) Error analysis: at what distance $|h|$ does the relative error $\frac{|f(1+h) - \ell_1(1+h)|}{|f(1)|}$ first exceed 5%?

Exercise 14: (ML-Flavored Applications — Intermediate) Apply first-order calculus to optimization.

a) For a differentiable function $f : \mathbb{R} \to \mathbb{R}$ and small $\eta > 0$, use the first-order Taylor model to show

$$f(x - \eta f'(x)) \approx f(x) - \eta (f'(x))^2.$$

Explain why this guarantees monotone decrease when $f'(x) \neq 0$ and η is sufficiently small.

b) Suppose $|f'(x)| \leq L$ for all x in some interval. Derive a step-size bound $\eta < \eta_{\max}$ that ensures the approximation error $\left| f(x - \eta f'(x)) - [f(x) - \eta(f'(x))^2] \right|$ is small relative to the predicted decrease.

c) Given the quadratic loss $f(w) = \|Xw - y\|_2^2$ where $X \in \mathbb{R}^{n \times d}$ and $y \in \mathbb{R}^n$, compute $\nabla f(w)$ and explain how $\left| \frac{\partial f}{\partial w_i}(w) \right|$ measures the sensitivity of the loss to changes in the i-th feature weight.