# Midterm Exam

CS265 / CME309 Autumn 2026

110 minutes 100 points + 5 bonus points

**Instructions**

- You may use results stated in those classes as black boxes unless a question explicitly asks you to prove them.
- The exam is non-collaborative and closed-book, except for up to **three double-sided sheets of standard-sized paper** that you prepared yourself. No electronic devices or internet access.
- A reference sheet appears at the end. You may detach that sheet; keep all other pages together.
- You may use the conclusion of an earlier part even if you have not proved it.
- There is an extra-work page near the end. Clearly label any work placed there and indicate on the relevant question page where to find it.

**Advice.** If a part is taking too long, move on and return to it later. Point values indicate the intended weight of each part, not necessarily its difficulty.

Name: _____

SUNet ID: _____

## 1. Quick checks **28 points**

**(a) (4 pt.)** Suppose a Las Vegas algorithm always returns the correct answer and has expected running time at most 8 seconds on a fixed input. If we stop it after 80 seconds and return FAIL if it has not finished, what is the strongest upper bound on the probability of returning FAIL that holds for an such algorithm?

(A) 0 (B) 1/10 (C) 1/2 (D) 1

**(b) (4 pt.)** Suppose we divide a class of $n$ into two rooms by independently flipping a fair coin for each student. Let $X, Y$ denote the number of students assigned to each of the two rooms. Which of the following is true (circle all true statements):

(A) $X$ and $Y$ are independent.

(B) $X$ and $Y$ are dependent.

(C) The marginal distributions of $X$ and $Y$ are each distributed like a binomial random variable.

**(c) (4 pt.)** Suppose that the number of students who show up, $S$, is distributed like a Poisson random variable of expectation $n$. And we divide the $S$ students who show up into two rooms by independently flipping a fair coin for each student. Let $X, Y$ denote the number of students assigned to each of the two rooms. Which of the following is true (circle all true statements):

(A) $X$ and $Y$ are independent.

(B) $X$ and $Y$ are dependent.

(C) The marginal distributions of $X$ and $Y$ are each distributed like Poisson random variables.

(D) The distribution of $X$ conditioned on $S$ is distributed like a binomial random variable.

## 1. Quick checks (continued)

**(d) (4 pt.)** Contract any non-loop edge of a connected undirected multigraph with at least three vertices, deleting self-loops and keeping parallel edges. Which statement about the minimum-cut value is always true?

(A) It cannot decrease.

(B) It strictly decreases.

(C) It cannot increase.

(D) It is unchanged.

**(e) (4 pt.)** Suppose $C, D$ are strictly positive random variables over some finite number of possible outcomes (over some common probability space, in the sense that one can take about events like $Pr[C = 17 \text{ and } D = 3]$). Suppose $\mathbb{E}[C] = 2, \quad \mathbb{E}[D] = 1$. Which statement must hold?

(A) $\mathbb{E}[C/D] \leq 2$.

(B) $Pr[C/D \leq 2] > 0$.

(C) $C/D \leq 2$ for every outcome.

**(f) (4 pt.)** Let $x \in \mathbb{R}^d$ be nonzero and let $y = 3x$. Choose a $k \times d$ matrix $A$ with independent $N(0, 1)$ entries, where $k \geq 1$, and define the length $k$ vector $h(z) = \text{sign}(Az)$ coordinatewise (ie the $i$th coordinate is $sign(A_i z)$ where $A_i$ is the $i$th row of $A$). What is $\Pr[h(x) = h(y)]$?

(A) 0 (B) $2^{-k}$ (C) 1/2 (D) 1

**(g) (4 pt.)** Let $A : \mathbb{R}^d \to \mathbb{R}^k$ be a random Gaussian Johnson–Lindenstrauss map, with $1 \leq k < d$. What is the probability that **there exists** a unit vector $x$, possibly chosen after seeing $A$, for which $Ax = 0$?

(A) 0 (B) 0.01 (C) 0.99 (D) 1

## 2. Short arguments **32 points**

For each part, justify your answer in at most three sentences.

**(a) (4 pt.)** Let $U$ be a finite nonempty set, and let $S_1, \ldots, S_n \subseteq U$ be fixed subsets, each containing at least half of $U$. Draw $t = \log_2(100n)$ independent uniform elements of $U$, with replacement. Show that, with probability at least 0.99, every $S_i$ contains at least one sampled element. The subsets can overlap arbitrarily.

**(b) (3 pt.)** Let $Z_1, \ldots, Z_n$ be independent, each uniformly distributed on $\{-1, +1\}$, and let $S = \sum_i Z_i$. Prove that $\Pr[S \geq 2\sqrt{n}] \leq 1/4$..

**(c) (3 pt.)** For $S$ as defined above, prove that $\Pr[S \geq 2\sqrt{n}] \leq 1/8$..

## 2. Short arguments (continued)

**(d) (6 pt.)** For $S$ as defined above, prove that $\Pr[S \geq n/2] \leq e^{-n/8}$. You must prove this 'from scratch' without appealing to Chernoff bounds. Feel free to assume Markov's inequality, and you may also assume that

$$\mathbb{E}[e^{tZ_i}] \leq e^{t^2/2} \qquad \text{for every real } t.$$

If you leverage the above fact, do state the value of $t$ you use.

**(e) (3 pt.)** A map $f$ embeds a finite metric space $(X, d)$ into Euclidean space with distortion at most $D$. A second map $g$ satisfies, for all $x, y \in X$,

$$\|f(x) - f(y)\|_2 \leq \|g(f(x)) - g(f(y))\|_2 \leq 1.1\,\|f(x) - f(y)\|_2.$$

What upper bound on the distortion of $g \circ f$ follows?

**(f) (3 pt.)** In the setting of the previous part, If $f$ is Bourgain's embedding into $\ell_2$ and $g$ is a Johnson–Lindenstrauss map, what are the properties of $g \circ f$ in terms of distortion?

## 2. Short arguments (continued)

**(g) (3 pt.)** Describe a practical/efficient algorithm for sampling a uniformly random 1024-bit prime number (e.g. to use as a secret cryptographic key). [Feel free to use a randomized primality testing algorithm as a black-box, no need to describe how that works.]

**(h) (3 pt.)** Describe the runtime of your algorithm above, in big-Oh notation, assuming you'll adapt it to sample a uniformly random $n$-bit prime number [ie give your answer in terms of $n$.]. Justify your answer in a few sentences.

**(i) (4 pt.)** Consider the four-point metric space $X = \{a, b, c, d\}$. The distance from $a$ to each of $b, c, d$ is 1, and the distance between any two distinct points in $\{b, c, d\}$ is 2.

Give an embedding of this metric into the Euclidean plane $\mathbb{R}^2$ with distortion at most 2. You may give coordinates or a clearly labeled drawing. Briefly justify the distortion bound.

## 3. Checking a very large matrix power **20 points**

All matrices in this problem are real $n \times n$ matrices, where $n \geq 2$. Use exact arithmetic and count arithmetic operations. A matrix-vector product takes $O(n^2)$ operations. All matrices are fixed before any test randomness is chosen and stay fixed throughout the test.

**(a) (6 pt.)** Let $D \neq 0$ be a fixed matrix. Choose $r \in \{0,1\}^n$ with independent uniform coordinates. Prove that

$$\Pr[Dr = 0] \leq \frac{1}{2}.$$

*Hint: Focus on a nonzero entry in one row, and condition on all but one coordinate of $r$.*

**(b) (6 pt.)** Given $A, B, C$, design an $O(n^2)$-time randomized test for the claim $AB = C$. It must always accept a correct claim and accept an incorrect claim with probability at most $1/2$. Explain the running time and both guarantees.

## 3. Checking a very large matrix power (continued)

**(c) (8 pt.)** Someone claims to have computed $A^{2^t}$ for an integer $t \geq 1$. They supply a sequence

$$B_0 = A,\ B_1, \ldots, B_t,$$

claiming $B_i = B_{i-1}^2$ for every $i = 1, \ldots, t$.

Design an $O(tn^2)$-time test using **one random vector $r$ shared across all $t$ checks**. Your test must:

- always accept when every claimed equality is correct; and
- accept with probability at most $1/2$ whenever $B_t \neq A^{2^t}$.

Explain carefully why sharing $r$ is sufficient.

## 4. A large sum-free subset **20 + 5 bonus points**

Let $A$ be a set of $n \geq 1$ positive integers. A subset $B$ is **sum-free** if there are no $x, y, z \in B$ satisfying $x + y = z$.

We will prove that $A$ contains a sum-free subset with at least $n/3$ elements, and strengthen this bound in the bonus part.

Choose a real number $\alpha$ uniformly at random from $[0, 1)$. For each $x \in A$, define

$$r_\alpha(x) = \alpha x - \lfloor \alpha x \rfloor.$$

Thus, $r_\alpha(x)$ is the fractional part of $\alpha x$, a number in $[0, 1)$. For example, if $x = 10$ and $\alpha = 0.123$, then $r_\alpha(x) = 0.23$ which is the fractional part of $10 \cdot 0.123 = 1.23$.

For a given $\alpha$, define the set $B_\alpha \subset A$ to consist of those $x \in A$ for which

$$\frac{1}{3} < r_\alpha(x) < \frac{2}{3}.$$

**(a) (3 pt.)** For a fixed $x \in A$, show that $r_\alpha(x)$ is uniformly distributed on $[0, 1)$.

**(b) (5 pt.)** Using the previous part (even if you didnt prove it), compute the expected number of elements in $B_\alpha$.

## 4. A large sum-free subset (continued)

**(c) (5 pt.)** Prove that $B_\alpha$ is a "sum-free" set for every choice of $\alpha$.

**(d) (7 pt.)** Prove that $A$ contains a sum-free subset of size at least $n/3$.

**(e) (5 bonus points)** Strengthen the conclusion of part (d): prove that $A$ contains a sum-free subset with **strictly more than** $n/3$ **elements**.

*Hint: What happens to $B_\alpha$ when $\alpha$ is sufficiently small?*

**End of the questions.**

## Extra work

Keep this page attached. Label the question and part clearly, and indicate on the relevant question page that your work continues here.

## Reference sheet

These statements may be used as black boxes. Check their hypotheses when applying them.

**Expectation and variance.**

$$\mathbb{E}\sum_i a_i X_i = \sum_i a_i \mathbb{E}X_i, \qquad \mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2.$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y, \qquad \mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}(X_i) + 2\sum_{i<j}\mathrm{Cov}(X_i, X_j).$$

For $X \sim \mathrm{Ber}(p)$, $\mathbb{E}X = p$ and $\mathrm{Var}(X) = p(1-p)$.

**Markov and Chebyshev.** For nonnegative $X$ and $a > 0$, $\Pr[X \geq a] \leq \mathbb{E}X/a$. For a real random variable of finite variance,

$$\Pr[|X - \mathbb{E}X| \geq a] \leq \frac{\mathrm{Var}(X)}{a^2}.$$

**Exponential moments.** If $X_i$ are independent, then

$$\mathbb{E}e^{t\sum_i X_i} = \prod_i \mathbb{E}e^{tX_i}.$$

For $t > 0$, $\Pr[X \geq a] \leq e^{-ta}\mathbb{E}e^{tX}$ whenever the expectation is finite.

**Chernoff.** If $X$ is a sum of independent 0/1 variables and $\mu = \mathbb{E}X$, then for $0 < \delta \leq 1$,

$$\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}, \qquad \Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}.$$

**Poissonization.** If $T \sim \mathrm{Poi}(\lambda)$ and each of $T$ objects is independently placed uniformly into one of $b$ bins, the bin counts are independent $\mathrm{Poi}(\lambda/b)$ variables. For $Y \sim \mathrm{Poi}(a)$,

$$\Pr[Y = j] = e^{-a}a^j/j!, \qquad \mathbb{E}Y = \mathrm{Var}(Y) = a.$$

**Metric distortion.** An embedding $f$ has distortion at most $D$ if there is a common scale $\beta > 0$ such that

$$\beta d(x,y) \leq \|f(x) - f(y)\| \leq D\beta d(x,y) \quad \text{for all } x, y.$$

Bourgain's construction embeds an $n$-point metric into $\ell_2$ (also into $\ell_1$) in $O(\log^2 n)$ dimensions with distortion $O(\log n)$, with high probability. Gaussian JL maps preserve all distances in a fixed $n$-point Euclidean set within factors $1 \pm \varepsilon$ with probability at least 0.99, using $O(\varepsilon^{-2}\log n)$ dimensions.

**Basic facts.** $\Pr[\bigcup_i E_i] \leq \sum_i \Pr[E_i]$, with no independence needed; $1 - x \leq e^{-x}$ for all real $x$; a finite collection of real numbers contains one at least as large as its average.