AMSI Summer School 2027

Optimisation and Decision-Making under Uncertainty

Pre-Enrolment Self-Assessment Quiz

A/Prof Nam Ho-Nguyen and Dr Li Chen
The University of Sydney

Purpose. This is a diagnostic self-assessment, designed to indicate whether your current mathematical background is sufficient for the pace of the subject and, where necessary, to identify topics for revision. **No prior knowledge of optimisation is assumed.** Some suggested background reading is provided at the end.

Quiz

1. Linear algebra and quadratic forms.

Let

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}, \qquad b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^3.$$

(a) State the dimensions of $A^\top A$, $AA^\top$, and $A^\top b$.

(b) Compute $A^\top A$ and $A^\top b$, and solve

$$\left(A^\top A\right) x = A^\top b$$

for $x \in \mathbb{R}^2$.

(c) Recall that a symmetric matrix M is positive definite if $z^\top M z > 0$ for every $z \neq 0$. Show that $A^\top A$ is positive definite. You may use

$$z^\top A^\top A z = \|Az\|_2^2 .$$

2. Multivariable calculus.

Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f\,(x, y) = \frac{1}{2}\,(x - y)^2 + \exp\,(x + 2y)\,.$$

(a) Compute the gradient $\nabla f\,(x, y)$.

(b) Compute the Hessian $\nabla^2 f\,(x, y)$.

(c) Recall that a symmetric matrix H is *positive semidefinite* if $d^\top H d \geq 0$ for every vector d. Show that $\nabla^2 f\,(x, y)$ is positive semidefinite for every (x, y). A useful approach is to express it as a sum of two outer products.

3. Convergence of an iterative sequence.

Let $x_0 \in \mathbb{R}$ and define

$$x_{k+1} = \frac{1}{2} x_k + 1, \qquad k = 0, 1, 2, \ldots .$$

(a) If (x_k) converges to a limit L, determine L.

(b) Show that

$$x_k - L = 2^{-k}\,(x_0 - L)\,,$$

where L is the value found in part (a).

(c) Using your formula from part (b), prove that (x_k) converges to the limit L found in part (a). For $\varepsilon > 0$, give a sufficient number of iterations to ensure $|x_k - L| \le \varepsilon$.

(d) If $h : \mathbb{R} \to \mathbb{R}$ is continuous, what can you conclude about $h\,(x_k)$?

4. A discrete random loss.

A random loss L has the distribution

k	0	2	5
$\mathbb{P}[L = k]$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

(a) Compute $\mathbb{E}[L]$ and $\mathbb{E}\left[L^2\right]$.

(b) Compute $\mathrm{Var}\,(L)$.

(c) Compute $\mathbb{P}[L \ge 2]$.

(d) Let $Y = 3L - 1$. Compute $\mathbb{E}[Y]$ and $\mathrm{Var}\,(Y)$ without first writing out the full distribution of Y.

5. Sample means and a probabilistic convergence bound.

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}\,(X_i) = \sigma^2 < \infty,$$

and let $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$.

(a) Compute $\mathbb{E}\left[\overline{X}_n\right]$.

(b) Compute $\mathrm{Var}\left(\overline{X}_n\right)$ and indicate where independence is used.

(c) You may use Chebyshev's inequality,

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \ge \varepsilon] \le \frac{\mathrm{Var}\,(Z)}{\varepsilon^2}.$$

Show that, for every $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\overline{X}_n - \mu\right| \ge \varepsilon\right] \le \frac{\sigma^2}{n\varepsilon^2},$$

and conclude that this probability tends to zero as $n \to \infty$. This property is called *convergence in probability* of $\overline{X}_n$ to μ.

(d) If $\sigma^2 = 4$, how large must n be for this bound to be at most 0.05 when $\varepsilon = 0.2$?

6. Expected squared error under a discrete distribution.

Let X be a discrete random variable taking values $x_1, \ldots, x_m$ with probabilities $p_1, \ldots, p_m$, where $p_i \ge 0$ and $\sum_{i=1}^{m} p_i = 1$. For a constant prediction $a \in \mathbb{R}$, define its expected squared error by

$$R\,(a) = \mathbb{E}\left[(X - a)^2\right] = \sum_{i=1}^{m} p_i\,(x_i - a)^2\,.$$

(a) Show that

$$R\,(a) = \mathbb{E}\left[X^2\right] - 2a\mathbb{E}[X] + a^2.$$

(b) Compute $R'(a)$ and $R''(a)$, and prove that $R(a)$ has the unique minimiser

$$a^* = \mathbb{E}[X].$$

(c) Show the decomposition

$$R(a) = \mathrm{Var}(X) + (a - \mathbb{E}[X])^2,$$

and hence identify the minimum value of $R(a)$.

(d) Apply the result to the random loss L in Question 4: give the best constant prediction and its minimum expected squared error.

7. Python/NumPy readiness check.

Without running the code, state what it prints and briefly explain the meanings of `.T`, `@`, `.shape`, and `mean()`.

```
import numpy as np

A = np.array([[1., 0.],
              [1., 1.],
              [0., 1.]])
xi = np.array([1., 2., 2., 5.])

M = A.T @ A
xbar = xi.mean()
loss = np.mean((xbar - xi)**2)

print(A.shape, M.shape, xbar, loss)
```

Solutions

1. Linear algebra and quadratic forms.

Since $A \in \mathbb{R}^{3\times 2}$, we have

$$A^\top A \in \mathbb{R}^{2\times 2}, \qquad AA^\top \in \mathbb{R}^{3\times 3}, \qquad A^\top b \in \mathbb{R}^2.$$

Direct calculation gives

$$A^\top A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \qquad A^\top b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Thus

$$2x_1 + x_2 = 1, \qquad x_1 + 2x_2 = 1.$$

Subtracting the two equations gives $x_1 = x_2$, and hence $3x_1 = 1$. Therefore

$$x = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}.$$

For any $z \in \mathbb{R}^2$,

$$z^\top A^\top A z = (Az)^\top (Az) = \|Az\|_2^2 \geq 0.$$

Moreover, the two columns of A are linearly independent, so $Az = 0$ implies $z = 0$. Therefore $\|Az\|_2^2 > 0$ whenever $z \neq 0$, proving that $A^\top A$ is positive definite.

2. Multivariable calculus.

Let $s = \exp(x + 2y)$. Differentiating gives

$$\nabla f(x, y) = \begin{bmatrix} x - y + s \\ y - x + 2s \end{bmatrix}.$$

Differentiating once more,

$$\nabla^2 f(x, y) = \begin{bmatrix} 1 + s & -1 + 2s \\ -1 + 2s & 1 + 4s \end{bmatrix}.$$

This Hessian can be written as

$$\nabla^2 f(x, y) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} + s \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}.$$

Therefore, for any $d = [d_1, d_2]^\top$,

$$d^\top \nabla^2 f(x, y)\, d = (d_1 - d_2)^2 + s\,(d_1 + 2d_2)^2 \geq 0.$$

Hence the Hessian is positive semidefinite. In fact it is positive definite because the two vectors $[1, -1]^\top$ and $[1, 2]^\top$ are linearly independent and $s > 0$.

3. Convergence of an iterative sequence.

If $x_k \to L$, then also $x_{k+1} \to L$, so passing to the limit in the recurrence gives

$$L = \frac{1}{2}L + 1,$$

and therefore $L = 2$.
Since $L = \frac{1}{2}L + 1$, subtracting this fixed-point identity from the recurrence gives

$$x_{k+1} - L = \frac{1}{2}\left(x_k - L\right).$$

Iterating this identity yields

$$x_k - L = 2^{-k}\left(x_0 - L\right),$$

which proves the stated formula. Consequently,

$$|x_k - L| = 2^{-k}\,|x_0 - L| \to 0,$$

so $x_k \to L$. If $|x_0 - L| \le \varepsilon$, then $k = 0$ already suffices. Otherwise it is enough to choose

$$k \ge \left\lceil \log_2\left(\frac{|x_0 - L|}{\varepsilon}\right)\right\rceil.$$

Finally, continuity of h implies preservation of limits, so $h\left(x_k\right) \to h\left(L\right) = h\left(2\right)$.

4. A discrete random loss.

The first two moments are

$$\mathbb{E}\left[L\right] = 0 \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} + 5 \cdot \frac{1}{6} = \frac{3}{2},$$

and

$$\mathbb{E}\left[L^2\right] = 0^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{3} + 5^2 \cdot \frac{1}{6} = \frac{11}{2}.$$

Hence

$$\mathrm{Var}\left(L\right) = \mathbb{E}\left[L^2\right] - \mathbb{E}\left[L\right]^2 = \frac{11}{2} - \left(\frac{3}{2}\right)^2 = \frac{13}{4}.$$

Also,

$$\mathbb{P}\left[L \ge 2\right] = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

For $Y = 3L - 1$, linearity of expectation and the variance scaling rule give

$$\mathbb{E}\left[Y\right] = 3\mathbb{E}\left[L\right] - 1 = \frac{7}{2}, \qquad \mathrm{Var}\left(Y\right) = 3^2\,\mathrm{Var}\left(L\right) = \frac{117}{4}.$$

5. Sample means and a probabilistic convergence bound.

By linearity of expectation,

$$\mathbb{E}\left[\overline{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] = \mu.$$

For the variance,

$$\mathrm{Var}\left(\overline{X}_n\right) = \frac{1}{n^2}\,\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right).$$

Independence makes all pairwise covariance terms equal to zero, so

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) = n\sigma^2.$$

Therefore

$$\mathrm{Var}\left(\overline{X}_n\right) = \frac{\sigma^2}{n}.$$

Applying Chebyshev's inequality to $Z = \overline{X}_n$ gives

$$\mathbb{P}\left[\left|\overline{X}_n - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{n\varepsilon^2}.$$

The right-hand side converges to zero for every fixed $\varepsilon > 0$. Thus

$$\mathbb{P}\left[\left|\overline{X}_n - \mu\right| \geq \varepsilon\right] \to 0 \qquad \text{for every } \varepsilon > 0.$$

By definition, this means that $\overline{X}_n$ converges in probability to μ.
When $\sigma^2 = 4$ and $\varepsilon = 0.2$,

$$\frac{\sigma^2}{n\varepsilon^2} = \frac{4}{n\left(0.2\right)^2} = \frac{100}{n}.$$

To make this at most 0.05, it is sufficient that $100/n \leq 0.05$, or

$$n \geq 2000.$$

This is a conservative guarantee because Chebyshev's inequality uses only the variance.

6. Expected squared error under a discrete distribution.

Expanding the square and using $\sum_{i=1}^{m} p_i = 1$ gives

$$\begin{aligned} R\left(a\right) &= \sum_{i=1}^{m} p_i\left(x_i^2 - 2ax_i + a^2\right) \\ &= \sum_{i=1}^{m} p_i x_i^2 - 2a\sum_{i=1}^{m} p_i x_i + a^2\sum_{i=1}^{m} p_i \\ &= \mathbb{E}\left[X^2\right] - 2a\mathbb{E}\left[X\right] + a^2. \end{aligned}$$

Therefore

$$R'\left(a\right) = 2\left(a - \mathbb{E}\left[X\right]\right), \qquad R''\left(a\right) = 2.$$

The only stationary point is $a^* = \mathbb{E}\left[X\right]$. Since $R''\left(a\right) > 0$ for every a, this stationary point is the

unique minimiser.
Let $m = \mathbb{E}[X]$. Using $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - m^2$, we obtain

$$\begin{aligned} R(a) &= \mathbb{E}\left[X^2\right] - 2am + a^2 \\ &= \left(\mathbb{E}\left[X^2\right] - m^2\right) + (a-m)^2 \\ &= \mathrm{Var}(X) + (a - \mathbb{E}[X])^2 . \end{aligned}$$

The second term is nonnegative and equals zero only at $a = \mathbb{E}[X]$. Hence

$$\min_{a \in \mathbb{R}} R(a) = \mathrm{Var}(X) .$$

For the random loss L in Question 4, $\mathbb{E}[L] = 3/2$ and $\mathrm{Var}(L) = 13/4$. Thus the best constant prediction under squared error is

$$a^* = \frac{3}{2},$$

and its minimum expected squared error is

$$R(a^*) = \frac{13}{4}.$$

This illustrates how an expectation can arise as the solution of an optimisation problem under uncertainty.

7. Python/NumPy readiness check.

The code prints, up to routine formatting,

```
(3, 2) (2, 2) 2.5 2.25
```

Here `A.shape` reports that A has 3 rows and 2 columns. The attribute `.T` forms the transpose, and `@` performs matrix multiplication, so `A.T @ A` is a 2×2 matrix. The method `mean()` computes the arithmetic mean. The expression `(xbar - xi)**2` squares each component, and `np.mean(...)` averages the resulting squared deviations.

Pre-reading and preparation

Core references.

- **[BV]** Stephen Boyd and Lieven Vandenberghe, *Convex Optimization* (free PDF; publisher).
- **[KKN]** Fatma Kılınç-Karzan and Arkadi Nemirovski, *Essential Mathematics for Convex Optimization*—the more advanced reference (free PDF; publisher).
- **[MML]** Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong, *Mathematics for Machine Learning*—a supplementary reference (free PDF; publisher).

Essential refresher

1. Linear algebra and quadratic forms.

- **[BV], Appendix A.5:** range, nullspace, eigenvalue decomposition, and definiteness.
- **[KKN], Appendix A, Sections A.1–A.3; Appendix D, Sections D.1–D.2:** vectors, matrices, linear mappings, symmetric matrices, eigenvalues, and quadratic forms.

2. Multivariable and matrix calculus.

- **[BV], Appendix A.3–A.4, especially Sections A.4.1–A.4.4:** functions, continuity, Jacobians, gradients, chain rules, and Hessians.
- **[KKN], Appendix C, Sections C.1–C.2; Section 10.2.1:** derivatives, higher-order derivatives, and their connection with convexity.

3. Probability and elementary statistics.

- **[MML], Chapter 6, Sections 6.1–6.4:** random variables and distributions; sum, product, and Bayes' rules; expectation, variance, covariance, independence, sample means, and sample covariance.

4. Sequences and convergence.

- **[BV], Sections A.2.1 and A.3.2:** sequences, limits, and continuity.
- **[KKN], Sections B.1.2 and B.2.1–B.2.3:** convergence of sequences and preservation of limits.

Optimisation

Both [BV] and [KKN] cover the fundamentals of convex optimisation. This material is not a pre-requisite and will be covered in the course, but both books are useful references.

Programming

The course uses the standard scientific Python stack (NumPy, SciPy, and pandas), commonly used in data-science workflows, together with state-of-the-art optimisation solvers Gurobi and/or MOSEK through their Python APIs.

- **Anaconda Distribution:** Python, conda, Jupyter, and commonly used scientific-computing packages.
- **IBM: Python Basics for Data Science (edX):** free through the audit track; Python fundamentals, Jupyter, pandas, and NumPy.
- **Gurobi with Python:** install `gurobipy`, activate a licence, then work through the first-model tutorial.
- **MOSEK Fusion with Python:** install `mosek`, add the licence file, verify `from mosek.fusion import *`, then work through the linear optimisation tutorial.

Licensing. Gurobi and MOSEK provide free academic licences for eligible students, faculty, and staff. If you cannot obtain an academic licence, contact the instructors.