- This exam consists of four problems.
- You have 3 hours to complete the exam.
- You are allowed to bring a calculator.
- **Include intermediate steps and justify your answers.**

Problem 1. This problem consists of two parts.

(1a) Let $f: \mathbb{R}^2 \to \mathbb{R}$ be a function given by $f(x, y) = \log(1 + e^{2x-4y+3})$. Compute the gradient of f.

(1b) Let $\mathbf{g}: \mathbb{R}^3 \to \mathbb{R}^2$ be a function given by

$$\mathbf{g}(x, y, z) = \begin{bmatrix} 5\log(1 + e^{-x+z}) \\ -\log(1 + e^{y+2x}) + z \end{bmatrix}.$$

Compute the Jacobian of $\mathbf{g}$.

Problem 2. This problem consists of two parts.

(2a) Find all critical points of the function $f: \mathbb{R}^d \to \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|^2 - \|\mathbf{x}\|^4$.

(2b) Let $\mathbf{a} \in \mathbb{R}^d$ be fixed. Find all critical points of the function $g: \mathbb{R}^d \to \mathbb{R}$ given by $g(\mathbf{x}) = \exp(-3\|\mathbf{x}\|^2 + \mathbf{a} \cdot \mathbf{x})$.

Problem 3. Let $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ be a data vector, $\mathbf{B} \in \mathbb{R}^{n \times d}$ a matrix, and $\lambda > 0$ a scalar. Find the coefficients $\mathbf{a}^* \in \mathbb{R}^d$ that minimise the loss function

$$L(\mathbf{a}) = \|\mathbf{y} - \mathbf{B}\mathbf{a}\|^2 + \lambda\|\mathbf{a}\|^2.$$

Your answer can involve matrix products, inverses, and other operations of linear algebra.

Problem 4. A fair 6-sided die is rolled twice. Let X be a random variable that represents the maximum of the two rolls.

(4a) Write out the probability mass function of X.

(4b) Compute the expected value of X.

[*It is not necessary to simplify the expressions or compute their decimal forms.*]

The reverse of this sheet contains a collection of potentially useful formulae.

Useful Formulae

$\frac{\mathrm{d}}{\mathrm{d}x}x^n = nx^{n-1}$ $\quad \frac{\mathrm{d}}{\mathrm{d}x}c = 0$ $\quad \frac{\mathrm{d}}{\mathrm{d}x}\log(x) = \frac{1}{x}$

$\frac{\mathrm{d}}{\mathrm{d}x}e^x = e^x$ $\quad \frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x)$ $\quad \frac{\mathrm{d}}{\mathrm{d}x}\cos(x) = -\sin(x)$

$\frac{\mathrm{d}}{\mathrm{d}x}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$ $\quad \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$

$\frac{\mathrm{d}}{\mathrm{d}x}f(g(x)) = f'(g(x))g'(x)$ $\quad f(x) \approx L(x) = f(x_0) + f'(x_0)(x - x_0)$

$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$ $\quad F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x$

$\int x^n\,\mathrm{d}x = \frac{1}{n+1}x^{n+1}$ $\quad \int x^{-1}\,\mathrm{d}x = \log(x)$ $\quad \int e^x\,\mathrm{d}x = e^x$

$\int \sin(x)\,\mathrm{d}x = -\cos(x)$ $\quad \int \cos(x)\,\mathrm{d}x = \sin(x)$ $\quad \int_{g(a)}^{g(b)} f(x)\,\mathrm{d}x = \int_a^b f(g(y))g'(y)\,\mathrm{d}y$

$\nabla f(\mathbf{x}) = \begin{bmatrix} \partial_1 f(\mathbf{x}) \\ \vdots \\ \partial_d f(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^d$ $\quad \mathbf{J_f}(\mathbf{x}) = \begin{bmatrix} \partial_1 f_1(\mathbf{x}) & \cdots & \partial_d f_1(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(\mathbf{x}) & \cdots & \partial_d f_m(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^{m \times d}$

$\nabla[f(\mathbf{x})g(\mathbf{x})] = g(\mathbf{x})\nabla f(\mathbf{x}) + f(\mathbf{x})\nabla g(\mathbf{x})$ $\quad \nabla[f(\mathbf{g}(\mathbf{x}))] = \mathbf{J_g}(\mathbf{x})^\mathsf{T}\nabla f(\mathbf{g}(\mathbf{x}))$

$\nabla[f(g(\mathbf{x}))] = f'(g(\mathbf{x}))\nabla g(\mathbf{x})$ $\quad \nabla[\mathbf{a} \cdot \mathbf{x}] = \mathbf{a}$

$\nabla[\mathbf{b}^\mathsf{T}\mathbf{A}\mathbf{x}] = \mathbf{A}^\mathsf{T}\mathbf{b}$ $\quad \nabla[\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}] = (\mathbf{A} + \mathbf{A}^\mathsf{T})\mathbf{x}$

$\nabla[\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}] = 2\mathbf{A}\mathbf{x}$ if $\mathbf{A}$ is symmetric $\quad \nabla\|\mathbf{x}\|^2 = 2\mathbf{x}$

$\nabla\|\mathbf{x} - \mathbf{a}\| = (\mathbf{x} - \mathbf{a})/\|\mathbf{x} - \mathbf{a}\|$ $\quad \mathbf{J_f}(\mathbf{x}) = \mathbf{A}$ if $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$

$\frac{\mathrm{d}}{\mathrm{d}x}\mathbf{A}(x)^{-1} = -\mathbf{A}(x)^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}\mathbf{A}(x)\right)\mathbf{A}(x)^{-1}$ $\quad \frac{\mathrm{d}}{\mathrm{d}x}\log(\det(\mathbf{A}(x))) = \mathrm{tr}\left(\mathbf{A}(x)^{-1}\frac{\mathrm{d}}{\mathrm{d}x}\mathbf{A}(x)\right)$

$\mathbf{H}_f(\mathbf{x}) = \mathbf{J}_{\nabla f}(\mathbf{x}) = \begin{bmatrix} \partial_1\partial_1 f(\mathbf{x}) & \cdots & \partial_1\partial_d f(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \partial_d\partial_1 f(\mathbf{x}) & \cdots & \partial_d\partial_d f(\mathbf{x}) \end{bmatrix}$ $\quad f(\mathbf{x}) \approx L(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$

$\mathbf{f}(\mathbf{x}) \approx \mathbf{L}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + \mathbf{J_f}(\mathbf{x})(\mathbf{x} - \mathbf{x}_0)$ $\quad T_n(\mathbf{x}) = \sum_{|\boldsymbol{\alpha}| \le n} \frac{\mathrm{D}^{\boldsymbol{\alpha}} f(\mathbf{x}_0)}{\boldsymbol{\alpha}!}(\mathbf{x} - \mathbf{x}_0)^{\boldsymbol{\alpha}}$

$\int_{\mathbf{g}(A)} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_A f(\mathbf{g}(\mathbf{y}))|\det(\mathbf{J_g}(\mathbf{y}))|\,\mathrm{d}\mathbf{y}$ $\quad \int_{\mathbb{R}^d} f(\mathbf{A}\mathbf{x} - \mathbf{b})\,\mathrm{d}\mathbf{x} = |\det(\mathbf{A})|^{-1}\int_{\mathbb{R}^d} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$

$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ if A_i are disjoint $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ $\quad P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1})$

$P(A) = \sum_{i=1}^{n} P(B_i)P(A \mid B_i)$ if B_i form a partition of Ω $\quad P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$

$P_X(A) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\})$ $\quad p_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$

$F_X(x) = P(X \le x) = P(\{\omega \in \Omega : X(\omega) \le x\})$ $\quad X$ unif. if $X(\Omega) = \{x_1, \ldots, x_k\}$ and $p_X(x_k) = \frac{1}{n}$

$X \sim \mathrm{Poisson}(\lambda)$ if $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k \in \mathbb{N} \cup \{0\}$ $\quad \mathbb{E}(X) = \sum_{k=1}^{\infty} x_k p_X(x_k)$ when $X(\Omega) = \{x_1, x_2, \ldots\}$

$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$ $\quad \mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$