IMPERIAL

MATH70138

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
Summer 2025

This paper is also taken for the relevant examination for the
Associateship of the Royal College of Science

Rough Paths and Applications to Machine Learning

Date: Wednesday, May 14, 2025

Time: Start time 10:00 – End time 12:30 (BST)

Time Allowed: 2.5 hours

This paper has 5 Questions.

Please Answer All Questions in 1 Answer Booklet

This is a closed book examination.

Candidates should start their solutions to each question on a new sheet of paper.

Supplementary books may only be used after the relevant main book(s) are full.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Allow margins for marking.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO DO SO

1. Let $v \in \mathbb{R}^2$ be a fixed vector and denote by X_v the straight line $X_v(t) = tv$ for $t \in [0,1]$. Consider an axis path X defined as the concatenation of n straight lines

$$X = X_{\lambda_1 v_1} * \dots * X_{\lambda_n v_n} : [0,n] \to \mathbb{R}^2$$

where $v_1, \dots, v_n \in \{e_1, e_2\}$ are one of the two standard basis vectors in $\mathbb{R}^2$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ are scalars. A multi-index $(i_1, \dots, i_n) \in \{1,2\}^n$ is called square-free if $i_k \neq i_{k+1}$ for all $k = 1, \dots, n-1$.

(a) Suppose that $v_1 = e_{i_1}, \dots, v_n = e_{i_n}$. Define the following quantity

$$C_X(i_1, \dots, i_n) := \left(e^*_{i_1} \dots e^*_{i_n}, S(X)_{0,n} \right)$$

where S is the signature. Show that $C_X(i_1, \dots, i_n) = \lambda_1 \dots \lambda_n$. (8 marks)

(b) Show that

$$\lambda_k = \frac{2C_X(i_1, \dots, i_{k-1}, i_k, i_k, i_{k+1}, \dots, i_n)}{C_X(i_1, \dots, i_n)}, \quad \text{for } k = 1, \dots, n.$$

(4 marks)

(c) Define two sequences of paths as follows: $Y^1 = (t,0), Z^1 = (0,t)$ and for any $n \geq 1$

$$Y^{n+1} = Y^n * Z^n, \quad Z^{n+1} = Z^n * Y^n$$

Show that Y^{n+1} and Z^{n+1} are distinct axis paths. (2 marks)

(d) Prove that the signatures of Y^n and Z^n coincide up to level n. (6 marks)

(Total: 20 marks)

2. Let x and y be two paths in $C_1\left([0,1],\mathbb{R}^d\right)$ for some $d \geq 2$. Assume that their expressions in coordinates $x = \left(x^1, ..., x^d\right)$ and $y = \left(y^1, ..., y^d\right)$ are such that $x^1_t = y^1_t = t$. Suppose also that $x_0 = y_0 = 0$.

(a) Prove that $S(x)_{0,1} = S(y)_{0,1} \iff x = y$. (8 marks)

(b) Define the restriction of x to a sub-interval $[s,t] \subseteq [0,1]$ as the path $x_{[s,t]} \in C_1([0,1];\mathbb{R}^d)$ with values

$$x_{[s,t]}(r) := \begin{cases} 0 & \text{if } r < s \\ x_r - x_s & \text{if } s \leq r \leq t \\ x_t - x_s & \text{if } r > t \end{cases}$$

Assume that $x^1_t = y^1_t = t$, $x^2_t = y^2_t = t^2$ and $x_0 = y_0 = 0$. Prove that

$$S(x)_{s,t} = S(y)_{u,v} \iff x_{[s,t]} = y_{[u,v]}$$

for any $(s,t),(u,v) \in [0,1]^2$ with $s < t$ and $u < v$. (8 marks)

(c) Explain how 2(a) can be extended to $x, y \in C_p([0,1],\mathbb{R}^d)$ for $p \in [1,2)$. (4 marks)

(Total: 20 marks)

3. Define the half-shuffle product $\prec: T(V^*) \times T(V^*) \mapsto T(V^*)$ by

$$f \prec r = rf \text{ and } r \prec f = 0$$

for any $r \in \mathbb{R}$ and $f \in V$, and then extend it inductively by

$$f \prec g = a \cdot (f_- \prec g + g \prec f_-)$$

for any $f \in V^{\otimes k}$ and $g \in V^{\otimes l}$ of the form $f = a \cdot f_-$, where $a \in V$.

Given this definition, $\prec$ extends uniquely to an algebra product on $T(V^*) \times T(V^*)$ by linearity.

Similarly, define the product area : $T(V^*) \times T(V^*) \to T(V^*)$ as follows

$$\text{area}(f, g) = f \prec g - g \prec f.$$

Let $f, g, h \in T(V^*)$ such that $\langle f, e \rangle = \langle g, e \rangle = \langle h, e \rangle = 0$, where e is the identity element in $T(V)$. Show the following identities hold:

(a) $f \sqcup\!\sqcup g = f \prec g + g \prec f$; (3 marks)

(b) $f \prec (g \sqcup\!\sqcup h) = (f \prec g) \prec h$; (3 marks)

(c) $f \prec (g \prec h) = (f \prec g) \prec h + (g \prec f) \prec h$; (4 marks)

(d) $h^n_{i,j} = \frac{1}{n+1} \sum_{k=0}^{n} (e_j^*)^{\sqcup\!\sqcup k} \sqcup\!\sqcup a^{n-k}_{i,j}$, where

$$h^n_{i,j} := (...((e_i^* \prec \underbrace{e_j^*) \prec e_j^*) \prec ... \prec e_j^*)}_{n \text{ times}}$$

$$a^n_{i,j} := \text{area}(...\text{area}(\text{area}(e_i^*, \underbrace{e_j^*), e_j^*), ..., e_j^*)}_{n \text{ times}}.$$

for any $i, j \in \{1, ..., d\}$, $i \neq j$ and $n \geq 1$. (10 marks)

(Total: 20 marks)

4. Let V be a finite dimensional vector space. Let $\phi : \mathbb{N} \cup \{0\} \to \mathbb{R}_+$ be a weight function such that for any $C > 0$ the series $\sum_{k \geq 0} \frac{C^k \phi(k)}{(k!)^2}$ converges.

(a) Let $C_1([0,1], V)$ denote the space of continuous bounded variation paths on $[0,1]$ with values in V, and let $\mathcal{S}$ be the image of $C_1([0,1], V)$ by the signature transform.

(i) Define the Hilbert space $T_\phi((V))$. (1 mark)

(ii) Show that $\mathcal{S} \subset T_\phi((V))$. (3 marks)

(iii) For any $x, y \in C_1([0,1], V)$, define the ϕ-signature kernel $k_\phi^{x,y} : [0,1]^2 \to \mathbb{R}$. (1 mark)

(b) Let $\phi : \mathbb{Z} \to \mathbb{C}$ be a complex-valued weight function and define the following kernel function

$$k_\phi^{x,y}(s,t) = \langle S(x)_{a,s}, S(y)_{a,t} \rangle_\phi := \sum_{k=-\infty}^{\infty} \phi(k) \Big\langle S(x)_{a,s}^{|k|}, S(y)_{a,t}^{|k|} \Big\rangle_{V^{\otimes |k|}} \tag{1}$$

Let $x, y \in C_1([a,b], V)$ be two continous paths of bounded variation. Let $\phi : \mathbb{Z} \to \mathbb{C}$ be a complex-valued weight function such that $\{\phi(k) : k \in \mathbb{Z}\}$ are the Fourier coefficients of some bounded integrable function $f : [-\pi, \pi] \to \mathbb{C}$, i.e.

$$f(\omega) = \sum_{k=-\infty}^{\infty} \phi(k) e^{ik\omega}$$

Show that the kernel in (1) satisfies the following relation (15 marks)

$$k_\phi^{x,y}(s,t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Big(k^{\exp(-i\omega)x,y}(s,t) + k^{\exp(i\omega)x,y}(s,t) \Big) f(\omega) d\omega - \phi(0)$$

(Total: 20 marks)

5. Let $p \geq 2$, let $\mathbf{x}$ be a geometric p-rough path on $\mathbb{R}^d$, let $f_\theta \in \mathrm{Lip}^\gamma(\mathbb{R}^e, \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e))$ be a neural network, with $\gamma > p$.

(a) Define what we mean by solution of the following Neural RDE driven by $\mathbf{x}$

$$dy_t = f_\theta(y_t) d\mathbf{x}_t, \quad y_0 \in \mathbb{R}^e.$$

(5 marks)

(b) Let $\mu_\theta \in Lip^1(\mathbb{R}^{e+1}, \mathbb{R}^e)$ and $\sigma_\theta \in Lip^{2+\epsilon}(\mathbb{R}^{e+1}, \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e))$ be two neural networks. Let $L : \mathbb{R}^e \to \mathbb{R}$ be a continuously differentiable loss function. Show that the adjoint process $a_t := \partial_{y_t} L(y_T)$ of the Neural Stratonovich SDE

$$dy_t = \mu_\theta(t, y_t)dt + \sum_{i=1}^{d} \sigma_\theta^i(t, y_t) \circ dW_t^i, \quad y_0 = a$$

coincides with the solution of the following backwards-in-time linear Stratonovich SDE

$$da_t = -a_t^\top \nabla \mu_\theta(t, y_t)dt - \sum_{i=1}^{d} a_t^\top \nabla \sigma_\theta^i(t, y_t) \circ dW_t, \quad a_T = \nabla L(y_T).$$

(15 marks)

(Total: 20 marks)

s

Setter's signature	Checker's signature	Editor's signature
....	....	....

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2025

This paper is also taken for the relevant examination for the Associateship.

MATH60138/MATH70138

Rough Paths and Applications to Machine Learning (Solutions)

Ss

1. (a) Since $X = X_{\lambda_1 v_1} * ... * X_{\lambda_n v_n}$, by Chen's identity we have

unseen $\Downarrow$

8, C

$$S(X) = S(X_{\lambda_1 v_1}) \otimes ... \otimes S(X_{\lambda_n v_n})$$
$$= \exp(\lambda_1 v_1) \otimes ... \otimes \exp(\lambda_n v_n)$$

Therefore

$$S(X)^n = \sum_{j_1+...+j_n=n} \frac{\lambda_1^{j_1} e_1^{j_1}}{j_1!} ... \frac{\lambda_n^{j_n} e_n^{j_n}}{j_n!}$$
$$= \sum_{j_1+...+j_n=n} \frac{\lambda_1^{j_1} ... \lambda_n^{j_n}}{j_1! ... j_n!} e_1^{j_1} ... e_n^{j_n}$$

For a square-free multi-index $(i_1, ..., i_n)$ we must have $j_1 = ... = j_n = 1$, therefore

$$C_X(i_1, ..., i_n) = S(X)_{0,n}^{(i_1,...,i_n)} = \lambda_1 ... \lambda_n.$$

unseen $\Downarrow$

4, B

(b) A direct computation gives

$$C_X(i_1, ..., i_{k-1}, i_k, i_k, i_{k+1}, ..., i_n) = \left(e_{i_1}^* ... e_{i_{k-1}}^* e_{i_k}^* e_{i_k}^* e_{i_{k+1}}^* ... e_{i_n}^*, e^{\lambda_1 e_{i_1}} ... e^{\lambda_n e_{i_n}}\right)$$
$$= \lambda_1 ... \lambda_{k-1} \lambda_k^2 \lambda_{k+1} ... \lambda_n$$
$$= \frac{C_X(i_1, ..., i_n)\lambda_k}{2}$$

Thus

$$\lambda_k = \frac{2C_X(i_1, ..., i_{k-1}, i_k, i_k, i_{k+1}, ..., i_n)}{C_X(i_1, ..., i_n)}$$

unseen $\Downarrow$

2, B

(c) This is proved by induction. For $n = 1$ the statement clearly holds. Assume it holds for $n \geq 1$. Then $Y^{n+1} \neq Z^{n+1}$ because by induction $Y^n \neq Z^n$. They are both axis paths because the concatenation of two axis paths is still an axis path.

unseen $\Downarrow$

6, B

(d) We prove this by induction. The case $n = 1$ follows by a simple calculation. Assume the statement holds up to some $n \geq 1$. By Chen's identity

$$S(Y^{n+1}) = S(Y^n) \cdot S(Z^n)$$

Thus

$$S(Y^{n+1})^{(k)} = \sum_{j=0}^{k} S(Y^n)^{(j)} \otimes S(Z^n)^{(k-j)}$$

For $k \in \{0, ..., n-1\}$, by the induction hypothesis, we must have

$$S(Y^n)^{(j)} = S(Z^n)^{(j)} \quad \text{and} \quad S(Y^n)^{(k-j)} = S(Z^n)^{(k-j)}$$

for $j = 0, ..., k$. Therefore

$$S(Y^{n+1})^{(k)} = \sum_{j=0}^{k} S(Z^n)^{(j)} \otimes S(Y^n)^{(k-j)} = S(Z^n * Y^n)^{(k)} = S(Z^{n+1})^{(k)}$$

If $k = n$, we can write

$$\begin{aligned} S(Y^{n+1})^{(n)} &= S(Y^n)^{(n)} + S(Z^n)^{(n)} + \sum_{j=1}^{n-1} S(Z^n)^{(j)} \otimes S(Y^n)^{(n-j)} \\ &= S(Z^n * Y^n)^{(n)} \\ &= S(Z^{n+1})^{(n)}. \end{aligned}$$

2. (a) It is obvious that two identical paths have the same ST. To prove the converse we first notice, by considering $-x$ and $-y$ if necessary, that it is sufficient to prove the result assuming that ρ is strictly increasing. By making these simplification, we can write

$$x_t = \left(t, x_t^-\right) \text{ and } y_t = \left(t, y_t^-\right)$$

where x^- and y^- are paths with values in $\mathbb{R}^{d-1}$. The aim is then to prove that x^- and y^- are equal whenever $S(x) = S(y)$. In the given basis of V we have $S(x)^w = S(y)^w$ where w is any multi-index of the form $w = (1, ..., 1, j)$ of length $k+2$. Written explicitly, this gives that

$$\int_a^b t^{k+1} d\left(x_t^-\right)^j = \int_a^b t^{k+1} d\left(y_t^-\right)^j$$

for every k and every j. For simplicity we drop the index j and refer only to x^-. Integration-by-part then yields that

$$b^{k+1}x_b^- - a^{k+1}x_a^- - (k+1)\int_a^b t^k x_t^- dt = b^{k+1}y_b^- - a^{k+1}y_a^- - (k+1)\int_a^b t^k y_t^- dt$$

Noting that $x_a = y_a$ together with $x_{a,b} = S(x)_{a,b}^{(1)} = S(y)_{a,b}^{(1)} = y_{a,b}$ results in

$$\int_a^b t^k x_t^- dt = \int_a^b t^k y_t^- dt \text{ for every } k = 0, 1, 2...$$

and hence that $x^- - y^-$ as an element of $L^2([a,b])$ is in the orthogonal complement $\mathcal{P}^\perp$ of the space $\mathcal{P}$ of finite degree polynomials. Since $\mathcal{P}$ is dense subspace of $L^2([a,b])$ it follows that $\mathcal{P}^\perp = \{0\}$ and hence $x^- = y^-$ almost everywhere. Using the continuity of x^- and y^- then gives that $x^-$and y^- are identically equal.

seen ⇓

8, D

(b) The *if* part is follows from $S(x)_{s,t} = S(x_{[s,t]})_{0,1}$. For the *only if* part, If $s = s'$ and $t = t'$ the statement holds; this is because if the signatures over the time interval $[s,t]$ of two time-augmented paths are equal, then the two paths must be equal on $[s,t]$. We now show that augmenting the path with t^2 and imposing equality of signatures, implies $s = s'$ and $t = t'$, which will in turn allow us to conclude the proof by the previous remark. Assume $S(y)_{s,t} = S(x)_{s',t'}$, in particular we must have

$$\int_s^t d(r^2) = t^2 - s^2 = (t')^2 - (s')^2 = \int_{s'}^{t'} d(r^2)$$
$$\int_s^t d(r) = t - s = t' - s' = \int_{s'}^{t'} d(r)$$

which reduces to the system

$$\begin{cases} t^2 - s^2 = (t')^2 - (s')^2 \\ t - s = t' - s' \end{cases} \quad \begin{cases} t + s = t' + s' \\ t - s = t' - s' \end{cases} \quad \begin{cases} 2t = 2t' \\ 2s = 2s' \end{cases}$$

Hence it must be true that $t = t'$ and $s = s'$.

unseen ⇓

8, D

seen ⇓

4, C

(c) The case $1 < p < 2$ can be obtained using a simple approximation argument by combining the fact that the closure of $C_1(V)$ in $C_q(V)$ contains $C_p(V)$ for any $p < q$. The argument is then concluded by considering such q with $q < 2$ and then using the joint continuity, in q-variation, of the Young integral map.

$$C_p(V) \times C_p(V) \ni (f, g) \mapsto \int f dg.$$

3. (a) It suffices to show the identity holds on basis elements because shuffle and half shuffle products are both associative and distributive. On basis elements we get

unseen ⇓

3, B

$$(e^*_{i_1}...e^*_{i_m}) \prec (e^*_{j_1}...e^*_{j_n}) = e_{i_1}((e^*_{i_2}...e^*_{i_m}) \sqcup\!\sqcup (e^*_{j_1}...e^*_{j_n})),$$

$$(e^*_{j_1}...e^*_{j_n}) \prec (e^*_{i_1}...e^*_{i_m}) = e_{j_1}((e^*_{j_2}...e^*_{j_n}) \sqcup\!\sqcup (e^*_{i_1}...e^*_{i_m})).$$

By definition of the shuffle product we have that

$$(e^*_{i_1}...e^*_{i_m}) \sqcup\!\sqcup (e^*_{j_1}...e^*_{j_n}) = e^*_{i_1}((e^*_{i_2}...e^*_{i_m}) \sqcup\!\sqcup (e^*_{j_1}...e^*_{j_n})) + e^*_{j_1}((e^*_{i_1}...e^*_{i_m}) \sqcup\!\sqcup (e^*_{j_2}...e^*_{j_n})).$$

The result follows from commutativity of the shuffle product $\sqcup\!\sqcup$.

unseen ⇓

3, A

(b) Similarly to (i) it suffices to show the identity holds on basis elements. Expanding the left and right hand side using the definition of shuffle and half shuffle product and showing that the two expression are equal yields the result.

unseen ⇓

4, A

(c) A direct application of (ii) yields

$$\begin{aligned}(u \prec v) \prec w &= u \prec (v \sqcup\!\sqcup w)\\ &= u \prec (v \prec w + w \prec v)\\ &= u \prec (v \prec w) + u \prec (w \prec v).\end{aligned}$$

by distributivity of the half-shuffle product $\prec$.

unseen ⇓

10, A

(d) By a simple induction on n it is easy to see that

$$h^n_{i,j} = n! e^*_i (e^*_j)^{\otimes n} \quad \text{and} \quad a^n_{i,j} = n!(e^*_i (e^*_j)^{\otimes n} - e^*_j e^*_i (e^*_j)^{\otimes (n-1)}).$$

Hence, we deduce the following expression

$$a^n_{i,j} + n e^*_j \sqcup\!\sqcup h^{n-1}_{i,j} = (n+1) h^n_{i,j}.$$

The asserted expression follows from an induction on n.

4. (a) (i) Define the ϕ-inner product $\langle \cdot, \cdot \rangle_\phi$ on the tensor algebra $T(V)$ as seen $\Downarrow$

1, B

$$\langle v, w \rangle_\phi := \sum_{k=0}^{\infty} \phi(k) \langle v_k, w_k \rangle_{V^{\otimes k}}$$

for any $v = (v_0, v_1, ...), w = (w_0, w_1, ...,)$ in $T(V)$. Then, $T_\phi((V))$ the Hilbert space obtained by completing $T(V)$ with respect to $\langle \cdot, \cdot \rangle_\phi$.

seen $\Downarrow$

(ii) For any path $x \in C_1$, the factorial decay of the ST yields

3, B

$$\big|\big|S(x)_{a,b}\big|\big|_\phi^2 = \sum_{k=0}^{\infty} \phi(k) \big|\big|S(x)_{a,b}^{(k)}\big|\big|_{V^{\otimes k}}^2 \leq \sum_{k=0}^{\infty} \phi(k) \frac{\|x\|_{1,[a,b]}^{2k}}{(k!)^2}$$

The summability condition on ϕ guarantees that the series is finite.

seen $\Downarrow$

(iii) The ϕ-signature kernel $k_\phi : C_1 \times C_1 \to \mathbb{R}$ is $k_\phi(x, y) = \langle S(x), S(y) \rangle_\phi$.

1, B

(b) For any (s, t) and $\omega \in (-\pi, \pi)$ one has

unseen $\Downarrow$

15, A

$$k^{\exp(\pm i\omega)x,y}(s, t) = \sum_{k=0}^{\infty} e^{\pm ik\omega} \big\langle S(x)_{a,s}^k, S(y)_{a,t}^k \big\rangle_{V^{\otimes k}}. \tag{1}$$

The factorial decay of the terms in the signature ensures that the series (1) converges uniformly and hence that

$$\begin{aligned}
\frac{1}{2\pi} \int_{-\pi}^{\pi} k^{\exp(\pm i\omega)x,y}(s, t) f(\omega) d\omega &= \sum_{k=0}^{\infty} \big\langle S(x)_{a,s}^k, S(y)_{a,t}^k \big\rangle_{V^{\otimes k}} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{\pm ik\omega} f(\omega) d\omega \\
&= \sum_{k=0}^{\infty} \phi(\pm k) \big\langle S(x)_{a,s}^k, S(y)_{a,t}^k \big\rangle_{V^{\otimes k}}
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
&\frac{1}{2\pi} \int_{-\pi}^{\pi} (k^{\exp(-i\omega)x,y}(s, t) + k^{\exp(i\omega)x,y}(s, t)) f(\omega) d\omega \\
&= \sum_{k=-\infty}^{\infty} \phi(k) \big\langle S(x)_{a,s}^{|k|}, S(y)_{a,t}^{|k|} \big\rangle_{V^{\otimes |k|}} + \phi(0) \big\langle S(x)_{a,s}^0, S(y)_{a,t}^0 \big\rangle_{V^{\otimes 0}} \\
&= k_\phi^{x,y}(s, t) + \phi(0)
\end{aligned}$$

as required.

5. (a) Let $\mathbf{x} \in G\Omega_p(\mathbb{R}^d)$ be a geometric p-rough path. We say that $y \in C^{p-var}([0,T],\mathbb{R}^e)$ is a solution to the RDE seen $\Downarrow$ 5, M

$$dy_t = f(y_t)d\mathbf{x}_t, \quad y_0 = a \in \mathbb{R}^e$$

if y belongs to the set of uniform limit points constructed in the ULT (seen in the lecture notes). In particular, if $f : \mathbb{R}^e \to \mathcal{L}(\mathbb{R}^d,\mathbb{R}^e)$ is linear or γ-Lipschitz with $\gamma > p$, then y is unique.

unseen $\Downarrow$

(b) Let $\{W^N\}_{N\geq 1}$ be a sequence of piecewise linear paths converging to the Stratonovich enhanced Brownian motion $\mathbf{W}$. Let $\{y^N\}_{N\geq 1}$ be the sequence of solutions to the following CDE 15, M

$$dy_t^N = \mu(t,y_t^N)dt + \sigma(t,y_t^N)dW_t^N, \quad y_0^N = a \in \mathbb{R}^e. \tag{2}$$

Then, by the ULT, y^N converges uniformly to a process $y : [0,T] \to \mathbb{R}^e$ which coincides almost surely with the strong solution of the Stratonovich SDE

$$dy_t = \mu(t,y_t)dt + \sum_{i=1}^d \sigma_i(t,y_t) \circ dW_t^i, \quad y_0 = a. \tag{3}$$

By results seen in lecture, each adjoint process $a_t^N := \partial_{y_t^N} L(y_T^N)$ satisfies the backwards-in-time linear CDE

$$da_t^N = -(a^N)_t^\top \nabla\mu_\theta(y_t^N)dt - (a^N)_t^\top \nabla\sigma_\theta(y_t^N)dW_t^N, \quad a_T^N = \nabla L(y_T^N).$$

By letting

$$z_t^N := \int_0^t \nabla\mu_\theta\left(y_s^N\right) ds + \int_0^t \nabla\sigma_\theta\left(y_s^N\right) d(W^N)_s^{(i)} \in \mathcal{L}\left(\mathbb{R}^e,\mathbb{R}^e\right)$$

we can rewrite the adjoint equation for a^N as follows

$$da_t^N = -(a^N)_t^\top dz_t^N. \tag{4}$$

Note that since μ_θ and σ_θ are bounded, the path $t \mapsto z_t^N$ is of bounded variation. Consider the solutions $(z^N, \tilde{z}^N)$ of the following CDEs driven by (x^N, y^N)

$$\begin{aligned} dz_t^N &= \nabla\mu_\theta\left(\tilde{z}_t^N\right) dt + \nabla\sigma_\theta\left(\tilde{z}_t^N\right) d(W^N)_t^{(i)} \\ d\tilde{z}_t^N &= dy_t^N \end{aligned}$$

Since $f_\theta \in \mathrm{Lip}^\gamma(\mathbb{R}^e,\mathcal{L}(\mathbb{R}^d,\mathbb{R}^e))$, $\nabla f_\theta \in \mathrm{Lip}^{\gamma-1}(\mathbb{R}^e,\mathcal{L}(\mathbb{R}^d,\mathcal{L}(\mathbb{R}^e,\mathbb{R}^e)))$. Thus, by the ULT, the sequence (z^N) is such that the sequence $(\pi_{\leq\lfloor p\rfloor} \circ S(z^N))$ converges in p-variation to a geometric rough path $\mathbf{z}$.

Therefore, by the ULT, there exists $y \in C^{p\text{-var}}([0,T],\mathbb{R}^e)$ such that the sequence (a^N) converges uniformly to the solution of a linear RDE run backwards-in-time which coincides almost surely with the soluton of the Stratonovich SDE

$$da_t = -a_t^\top \nabla\mu_\theta(t,y_t)dt - \sum_{i=1}^d a_t^\top \nabla\sigma_\theta^i(t,y_t) \circ dW_t, \quad a_T = \nabla L(y_T).$$

It remains to show that $a_t = \partial_{y_t} L(y_T)$.

Recall that the Jacobians $J_t^{s,y_s^N,x,N} = J_t^{s,N}$ satisfy the CDEs

$$dJ_t^{s,N} = \nabla\mu_\theta\left(y_t^N\right)\cdot J_t^{s,N}dt + \nabla\sigma_\theta\left(y_t^N\right)\cdot J_t^{s,N}dW_t^N, \quad J_s^N = I_e.$$

Considering the augmented path $(J^{s,N}, y^N)$, we can invoke again the ULT to show that the sequence $(J^{s,N})$ converges uniformly to the solution in $C^{p\text{-var}}([0,T],\mathcal{L}(\mathbb{R}^e,\mathbb{R}^e))$ of the following Stratonovich SDE

$$dJ_t^s = \nabla\mu_\theta\left(y_t\right)\cdot J_t^s dt + \nabla f_\theta\left(y_t\right)\cdot J_t^s \circ dW_t, \quad J_s^s = I_e.$$

Similarly, by a lemma seen in lectures, the inverses $M^{s,N} := (J^{s,N})^{-1}$ exist and satisfy the CDEs

$$dM_t^{s,N} = -M_t^{s,N}\cdot dz_t^N, \quad M_s^{s,N} = I_e.$$

Therefore, by the ULT, there exists $M^s \in C^{p\text{-var}}([0,T],\mathcal{L}(\mathbb{R}^e,\mathbb{R}^e))$ such that the sequence $(M^{s,N})$ converges uniformly to the solution of the linear RDE

$$dM_t^s = -M_t^s\cdot d\mathbf{z}_t, \quad M_s^s = I_e.$$

Because L is continuously differentiable, we have that the sequence

$$a_t^N = \nabla L\left(y_T^N\right)^\top\cdot J_T^{t,N} = \nabla L\left(y_T^N\right)^\top\cdot J_T^{0,N}\cdot\left(J_t^{0,N}\right)^{-1} = \nabla L\left(y_T^N\right)^\top\cdot J_T^{0,N}\cdot M_t^{0,N}$$

converges uniformly to

$$\nabla L\left(y_T\right)^\top\cdot J_T^0\cdot M_t^0 = \partial_{y_t}L(y_T)$$

as $N\to\infty$, which concludes the proof.

Review of mark distribution:

Total A marks: 32 of 32 marks
Total B marks: 20 of 20 marks
Total C marks: 12 of 12 marks
Total D marks: 16 of 16 marks
Total marks: 100 of 80 marks
Total Mastery marks: 20 of 20 marks

MATH70138 Rough Paths and Applications to Machine Learning Markers Comments

Question 1	Questions (a), (b), and (c) were answered correctly overall. Question (d), which required a slightly more involved induction argument, proved more challenging for some students.
Question 2	Question (a) was a lemma covered in the lecture notes, so most students were able to reproduce the proof correctly. Question (b) asked for a similar result, but with a subtle twist; only a few students managed to answer it correctly. Question (c) involved a standard argument that had been seen at least twice in class, and most students answered it correctly.
Question 3	Question 3 was entirely unseen and proved to be the most challenging question on the exam. Most students answered parts (a) and (c) correctly but struggled with parts (b) and (d).
Question 4	Question 4 was arguably the easiest on the exam, as most of the material had been covered previously. Overall, students managed to answer all parts correctly.
Question 5	The mastery question asked students to reproduce a complex proof of a key theorem covered in Chapter 3. Most students demonstrated a solid understanding of the result and were able to answer the question correctly.