

**BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)  
May 2024**

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: Friday, May 10, 2024

Time: 14:00 – 16:30 (BST)

Time Allowed: 2.5 hours

**This paper has 5 Questions.**

**Please Answer All Questions in 1 Answer Booklet**

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.

Allow margins for marking.

Credit will be given for all questions attempted.

Each question carries equal weight.

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

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

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

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

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

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  $C_1([0, 1], \mathbb{R}^d)$  for some  $d \geq 2$ . Assume that their expressions in coordinates  $x = (x^1, \dots, x^d)$  and  $y = (y^1, \dots, y^d)$  are such that  $x_t^1 = y_t^1 = t$ . 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  $x_t^1 = y_t^1 = t$ ,  $x_t^2 = y_t^2 = t^2$  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. Let  $C_{1,0,t} \subset C_1$  be the subset of  $\mathbb{R}^d$ -valued continuous bounded variation paths over  $[0, 1]$ , started at 0, i.e.  $x_0 = 0$ , and with the first coordinate-path corresponding to time, i.e.  $x_t^{(1)} = t$ . For any  $N \in \mathbb{N} \cup \{0\}$ , denote by  $\pi_{\leq N} : T((\mathbb{R}^d)) \rightarrow T^N(\mathbb{R}^d)$  the canonical projection. For any  $f \in T^N(\mathbb{R}^d)^*$ , define the function  $\Phi_f \in \mathbb{R}^{C_{1,0,t}}$  as  $\Phi_f : x \mapsto (f, \pi_{\leq N} \circ S(x)_{0,1})$ .
- (a) Show that the class  $\mathcal{A} = \{\Phi_f : f \in T^N(\mathbb{R}^d)^*, N \in \mathbb{N} \cup \{0\}\}$ :
- (i) is a subalgebra of  $\mathbb{R}^{C_{1,0,t}}$ , (4 marks)
  - (ii) contains the constant functions, (3 marks)
  - (iii) separates points in  $C_{1,0,t}$ . (3 marks)
- (b) Let  $K$  be a compact subset of  $\mathbb{R}^{C_{1,0,t}}$  with respect to the 1-variation topology. Show that  $\mathcal{A}_K := \{\Phi_f|_K : f \in T^N(\mathbb{R}^d)^*, N \in \mathbb{N} \cup \{0\}\}$  is a dense subset of  $C(K)$  endowed with the topology of uniform convergence. (5 marks)

(Total: 15 marks)

4. Let  $V$  be a finite dimensional vector space. Let  $\phi : \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}_+$  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  $T_\phi((V))$ . (2 marks)

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

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

(b) Let  $\pi$  be a real-valued random variable with finite moments. Consider the functions

$$\phi(k) = \mathbb{E}[\pi^k] \quad \text{and} \quad \psi(k) = \mathbb{E}[|\pi|^k]$$

and assume that for any  $C > 0$  the series  $\sum_{k \geq 0} \frac{C^k \psi(k)}{(k!)^2}$  converges.

(i) Show that for any  $x, y \in C_1([0, 1], V)$  the  $\phi$ -signature kernel is well-defined and satisfies

$$k_\phi^{x,y}(s, t) = \mathbb{E}_\pi[k^{\pi x,y}(s, t)] = \mathbb{E}_\pi[k^{x,\pi y}(s, t)] \quad (1)$$

where  $k$  is the unweighted signature kernel. (6 marks)

(ii) Show that if  $\phi(k) = (k/2)!$  then the  $\phi$ -signature kernel satisfies equation (1) where  $\pi \sim \text{Exp}(1)$  is an exponential random variable. (6 marks)

(iii) Suppose that  $x, y \in C_1([0, 1], V)$  are differentiable and let

$$f_\omega(s, t) := \text{Re}[k^{\exp(i\omega)x,y}(s, t)] \quad \text{and} \quad g_\omega(s, t) := \text{Im}[k^{\exp(i\omega)x,y}(s, t)].$$

Show that  $(f_\omega, g_\omega)$  solves a two-dimensional PDE. (6 marks)

(Total: 25 marks)

5. Let  $p \geq 2$ , let  $\mathbf{x}$  be a geometric  $p$ -rough path on  $\mathbb{R}^d$ , let  $f_\theta \in \text{Lip}^\gamma(\mathbb{R}^e, \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e))$  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  $\mu_\theta \in \text{Lip}^1(\mathbb{R}^{e+1}, \mathbb{R}^e)$  and  $\sigma_\theta \in \text{Lip}^{2+\epsilon}(\mathbb{R}^{e+1}, \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e))$  be two neural networks. Let  $L : \mathbb{R}^e \rightarrow \mathbb{R}$  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

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

(15 marks)

(Total: 20 marks)

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

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2024

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  $X = X_{\lambda_1 v_1} * \dots * X_{\lambda_n v_n}$ , by Chen's identity we have

unseen ↓

$$\begin{aligned} S(X) &= S(X_{\lambda_1 v_1}) \otimes \dots \otimes S(X_{\lambda_n v_n}) \\ &= \exp(\lambda_1 v_1) \otimes \dots \otimes \exp(\lambda_n v_n) \end{aligned}$$

8, C

Therefore

$$\begin{aligned} S(X)^n &= \sum_{j_1 + \dots + j_n = n} \frac{\lambda_1^{j_1} e_1^{j_1}}{j_1!} \dots \frac{\lambda_n^{j_n} e_n^{j_n}}{j_n!} \\ &= \sum_{j_1 + \dots + j_n = n} \frac{\lambda_1^{j_1} \dots \lambda_n^{j_n}}{j_1! \dots j_n!} e_1^{j_1} \dots e_n^{j_n} \end{aligned}$$

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

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

unseen ↓

- (b) A direct computation gives

4, B

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

Thus

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

unseen ↓

- (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.

2, A

unseen ↓

- (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

6, B

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

Thus

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

For  $k \in \{0, \dots, n-1\}$ , 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  $j = 0, \dots, k$ . 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  $\rho$  is strictly increasing. By making these simplification, we can write

seen  $\Downarrow$

8, D

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

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  $w = (1, \dots, 1, j)$  of length  $k+2$ . Written explicitly, this gives that

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

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

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

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.

unseen  $\Downarrow$

8, D

- (b) The *if* part 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

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

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  $s = s'$ .

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) (i) We know that  $x = y \iff S(x) = S(y)$  when  $x$  and  $y$  are two paths started at 0 and with time as one of their coordinates (lecture notes and previous exercise). Then, that  $\mathcal{A}$  separates points follows from the fact that  $S(x) = S(y)$  if and only if  $(f, S(x)) = (f, S(y))$  for all  $f$  in  $T((V))^*$ .
- (ii) The linear functional  $e = \pi_0$ , i.e.  $e(A) = A^0$  for  $A = (A^0, A^1, \dots) \in T((V))$ , has the property that  $\Phi_{\lambda e} \equiv \lambda$  for any  $\lambda$  in  $\mathbb{R}$  ensuring that  $\mathcal{A}$  contains all constant functions.
- (iii) It is self-evident using linearity that  $\Phi_{\lambda f} = \lambda \Phi_f$  and  $\Phi_{f+g} = \Phi_f + \Phi_g$  for all  $f$  and  $g$  in  $T((V))^*$  and  $\lambda$  in  $\mathbb{R}$ . Finally  $\mathcal{A}$  is closed under taking the pointwise product of its elements since  $\Phi_f \Phi_g = \Phi_{f \sqcup g}$  by the shuffle identity.
- (b) By using an identical argument as in part (a), it is easily seen that  $\mathcal{A}|_K$  is a subalgebra of  $C(K)$  which separates points and contains the constant functions. That  $\mathcal{A}|_K$  is dense then follows from the Stone-Weierstrass Theorem.

unseen ↓

4, B

seen ↓

3, B

seen ↓

3, B

seen ↓

5, A

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

seen ↓

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

2, A

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

seen ↓

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

3, A

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

The summability condition on  $\phi$  guarantees that the series is finite.

seen ↓

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

2, A

- (b) (i) The function  $|\phi|$  satisfies the summability condition because  $\psi$  does, therefore the  $\phi$ -signature kernel is well-defined. In addition we have

seen ↓

6, A

$$\begin{aligned} k_\phi^{x,y}(s, t) &= \sum_{i=0}^{\infty} \mathbb{E}[\pi^i] \langle S(x)_{a,s}^i, S(y)_{a,t}^i \rangle_{V^{\otimes i}} \\ &= \mathbb{E}_\pi \left[ \sum_{i=0}^{\infty} \pi^i \langle S(x)_{a,s}^i, S(y)_{a,t}^i \rangle_{V^{\otimes i}} \right] \\ &= \mathbb{E}_\pi \left[ \sum_{i=0}^{\infty} \langle S(\pi x)_{a,s}^i, S(y)_{a,t}^i \rangle_{V^{\otimes i}} \right] = \mathbb{E}_\pi [k^{\pi x, y}(s, t)] \end{aligned}$$

where the first and last equalities follow from the definition of  $\phi$ - and ordinary signature kernels respectively, the second equality holds by Fubini's theorem which is applicable because  $\psi$  satisfies the summability condition and the third follows by definition of the signature.

unseen ↓

- (ii) Since  $\pi \sim \text{Exp}(1)$  we have

6, A

$$\mathbb{E}[\pi^{k/2}] = \int_0^\infty x^{k/2} e^{-x} dx = \Gamma(k/2 + 1) = (k/2)! = \phi(k)$$

Therefore, by (d) (i) we get

$$k_\phi^{x,y}(s, t) = \mathbb{E}_\pi [k^{\pi^{1/2}x, y}(s, t)] = \mathbb{E}_\pi [k^{x, \pi^{1/2}y}(s, t)], \quad s, t \in [0, 1].$$

unseen ↓

- (iii) Note that

6, A

$$f_\omega(s, t) = \sum_{k=0}^{\infty} \cos(k\omega) \langle S(x)_{0,s}^k, S(y)_{0,t}^k \rangle_{V^{\otimes k}}$$

and

$$g_\omega(s, t) = \sum_{k=0}^{\infty} \sin(k\omega) \langle S(x)_{0,s}^k, S(y)_{0,t}^k \rangle_{V^{\otimes k}}$$

and together they solve the following PDE

$$\frac{\partial^2}{\partial s \partial t} \begin{pmatrix} f_\omega(s, t) \\ g_\omega(s, t) \end{pmatrix} = \begin{pmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{pmatrix} \begin{pmatrix} f_\omega(s, t) \\ g_\omega(s, t) \end{pmatrix} \langle \dot{x}_s, \dot{y}_t \rangle_V.$$

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 ↓

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  $f : \mathbb{R}^e \rightarrow \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$  is linear or  $\gamma$ -Lipschitz with  $\gamma > p$ , then  $y$  is unique.

unseen ↓

- (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

$$dy_t^N = \mu(t, y_t^N)dt + \sigma(t, y_t^N)dW_t^N, \quad y_0^N = a \in \mathbb{R}^e. \quad (1)$$

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

$$dy_t = \mu(t, y_t)dt + \sum_{i=1}^d \sigma_i(t, y_t) \circ dW_t^i, \quad y_0 = a. \quad (2)$$

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

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

By letting

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

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

$$da_t^N = -(a_t^N)^\top dz_t^N. \quad (3)$$

Note that since  $\mu_\theta$  and  $\sigma_\theta$  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  $(x^N, y^N)$

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

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

Therefore, by the ULT, there exists  $y \in C^{p-var}([0, T], \mathbb{R}^e)$  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 solution of the Stratonovich SDE

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

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

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

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

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

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

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

converges uniformly to

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

as  $N \rightarrow \infty$ , 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

### Question   Marker's comment

- 3 Question 3 was supposed to be marked out of 15. There is an error in the system (out of 20). I have adjusted the marks accordingly.
- 4 Question 4 was supposed to be marked out of 25. There is an error in the system (out of 20). I have adjusted the marks accordingly.