Mathematics for New Technologies in Finance

Solution sheet 3

Exercise 3.1 (Hölder-continuous functions as a weighted space) Fix $p \in]0,1]$ and consider the Hölder space $\mathcal{C}^p = \mathcal{C}^p([0,1],\mathbb{R}^d)$ of continuous function which are p-Hölder continuous, equipped with the uniform norm $\|\cdot\|_\infty$. Denote by

$$\|\omega\|_p := |\omega(0)| + \sup_{t\neq s} \frac{|\omega(t)-\omega(s)|}{|t-s|^p}$$

the Hölder norm on this space. Prove that this is a weighted space, under the topology induced by the uniform norm, with weight function

$$\rho(\omega) := 1 + \|\omega\|_p, \quad \forall \omega \in \mathcal{C}^p.$$

Solution 3.1 It is known that Hölder-continuous functions on a compact $[0,1]$ are uniformly continuous. The uniform norm is therefore well-defined and finite for any element in $\mathcal{C}^p$. Let also notice that the case corresponds to $p = 1$ to Lipschitz functions.
We aim to prove that the set $E_R := \{\omega \in \mathcal{C}^p : \rho(w) \leq R\}$ is compact in uniform norm for any $R \geq 1$. More precisely, we shall prove that is set is sequentially compact, which is an equivalent condition in a metric space.
Consider any sequence $\{\omega_n\}_n \subset E_R$. The sequence is clearly uniformly uniformly equicontinuous, since for any $\{x,y\} \in [0,1]^2$

$$|w_n(x) - w_n(y)| \ \leq \|\omega_n\|_p |x-y|^p \leq R|x-y|^p,$$

which is an upper-bound independent from $n \in \mathbb{N}$. Furthermore, the sequence is $\{\omega_n\}_n$ is uniformly bounded, in the sense that

$$|\omega_n(x)| \leq |\omega_n(0)| + |\omega_n(x) - \omega_n(0)| \leq 2\|\omega_n\|_p \leq 2R,$$

where the upper bound is independent from $n \in \mathbb{N}$ and from $x \in [0,1]$. Hence, by a direct application of Arzelà–Ascoli theorem, the sequence admits a subsequence which converges uniformly in $\mathcal{C}([0,1],\mathbb{R}^d)$. Finally, one is left to prove that E_R is closed, a property which (together with the relative compactness just proved) would yield the result. This follows immediately from the lower semi-continuity of $\|\cdot\|_p$, or by the same argument which proves uniformly boundedness.

Exercise 3.2 (Controlled ODEs) Consider the controlled ODE: $X_0 = x \in \mathbb{R}$

$$dX_t^\theta = V^\theta(t, X_t^\theta)dt, \quad t \in [0,T].$$

(a) Let

$$a_t = \frac{\partial X_T^\theta}{\partial X_t^\theta}.$$

Prove that

$$\frac{d}{dt} a_t = -\frac{\partial V^\theta}{\partial x}(t, X_t^\theta) \cdot a_t, \quad a_T = 1,$$

and relate a_t with $J_{t,T}$ in the lecture notebook.

(b) Prove that

$$\frac{d}{dt}(\frac{\partial X_t^\theta}{\partial \theta}a_t) = a_t\frac{\partial V^\theta}{\partial \theta}(t, X_t^\theta),$$

and

$$\frac{\partial X_T^\theta}{\partial \theta} = -\int_T^0 \frac{\partial X_T^\theta}{\partial X_t^\theta}\cdot\frac{\partial V^\theta}{\partial \theta}(t, X_t^\theta)dt.$$

(c) Is every feedforward neural network a discretization of controlled ODE?

Solution 3.2

(a) We know

$$\begin{aligned} a_t = \frac{\partial X_T^\theta}{\partial X_t^\theta} &= \frac{\partial X_T^\theta}{\partial X_{t+\Delta t}^\theta}\cdot\frac{\partial X_{t+\Delta t}^\theta}{\partial X_t^\theta} \\ &= a_{t+\Delta t}\cdot\frac{\partial X_{t+\Delta t}^\theta}{\partial X_t^\theta}. \end{aligned} \tag{1}$$

Also we know

$$X_{t+\Delta t}^\theta = X_t^\theta + \int_t^{t+\Delta t} V^\theta(X_s^\theta, s)ds$$

Taking partial derivative on both side we have

$$\frac{\partial X_{t+\Delta t}^\theta}{\partial X_t^\theta} = 1 + \int_t^{t+\Delta t}\partial_x V^\theta(X_s^\theta, s)ds$$

Plug this into (1) we have

$$\frac{a_t - a_{t+\Delta t}}{a_{t+\Delta t}} = \int_t^{t+\Delta t}\partial_x V^\theta(X_s^\theta, s)ds.$$

Let $\Delta t \to 0$ we obtain

$$\frac{d}{dt}a_t = -\frac{\partial V^\theta}{\partial x}(t, X_t^\theta)\cdot a_t$$

This ODE is known as adjoint equation, and runs backward in time (from T to 0), propagating sensitivity information similarly to backpropagation in feed-forward neural networks. The adjoint equation corresponds to the of $J_{0,T}$, run backwards in time.

(b)

$$\begin{aligned} \frac{d}{dt}(\frac{\partial X_t^\theta}{\partial \theta}a_t) &= \frac{d}{dt}(\frac{\partial X_t^\theta}{\partial \theta})\cdot a_t + \frac{da_t}{dt}\cdot(\frac{\partial X_t^\theta}{\partial \theta}) \\ &= \frac{\partial}{\partial \theta}V^\theta(X_t^\theta, t)\cdot a_t - \frac{\partial V^\theta}{\partial x}(t, X_t^\theta)\cdot a_t\cdot(\frac{\partial X_t^\theta}{\partial \theta}) \\ &= a_t\frac{\partial V^\theta}{\partial \theta}(t, X_t^\theta). \end{aligned}$$

The last equation is because:

$$\frac{\partial}{\partial \theta}V^\theta(X_t^\theta, t) = \frac{\partial V^\theta}{\partial x}(t, X_t^\theta)\cdot(\frac{\partial X_t^\theta}{\partial \theta}) + \frac{\partial V^\theta}{\partial \theta}(t, X_t^\theta).$$

(c) Yes. A feedforward neural network is a discretization of a controlled ODE because each layer represents a step in a discretized time-evolution equation, approximating a continuous transformation in the limit when the number L of layers goes to infinity. Neural ODEs explicitly formulate this continuous perspective. Similarly, the sensitivity analysis of the adjoint equation can be seen as a continuous-time counterpart of backpropagation in neural networks.

Exercise 3.3 (Solution of CODEs as features) This exercise exemplifies how, for any input curve u, the solution of the associated controlled ODE (CODE) is a (non-trivial) feature of u. Consider any control $u \in \mathcal{C}^1([0,1],\mathbb{R})$ and define for $t \in [0,1]$ the linear CODE

$$\mathrm{d}X_t = (\lambda X_t + u_t)\mathrm{d}t, \quad X_0 = x$$

where $\lambda \in \mathbb{R}$.

(a) Solve the system for a generic $u \in \mathcal{C}^1([0,1],\mathbb{R})$.

(b) Explicitly compute the solution for $u = \sin$.

Solution 3.3

(a) Multiplying on both sides by $e^{-\lambda t}$, we can rewrite the equation as

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^{-\lambda t}X_t) = e^{-\lambda t}u_t, \quad X_0 = x.$$

Integrating, we obtain

$$e^{-\lambda t}X_t = X_0 + \int_0^t e^{-\lambda s}u_s\mathrm{d}s.$$

Hence, the explicit solution to the equation is give by

$$X_t = xe^{-\lambda t} + \int_0^t e^{\lambda(t-s)}u_s\mathrm{d}s$$

From a more abstract perspective, we have describe the map $u \to X_t^\lambda(u)$. Interestingly, this map does not depend directly on the feature u, but on its Laplace transform, which is indeed used frequently to approximate a path as a feature.

(b) In case $u = \sin$, we can directly compute the integral via integration by parts, obtaining

$$\int_0^t e^{\lambda(t-s)}u_s\mathrm{d}s = \frac{e^{\lambda t} - \lambda\sin(t) - \cos(t)}{\lambda^2+1} = \frac{e^{\lambda t} - \lambda u_t - u_t'}{\lambda^2+1}.$$

Exercise 3.4 (Dependence of a non-linear ODE on the starting value) For $t \in [0,1]$, consider the ODE

$$\mathrm{d}X_t = \sin(X_t)\mathrm{d}t, \quad X_0 = x$$

Compute the value of $\partial_x X_t$ in two different ways, respectively

(a) by direct calculation;

(b) using the operator $J_{s,t}$ introduced in the lecture notebook.

Solution 3.4

(a) The ODE admits a unique solution, which is also explicit. Notably, by the method of separation of variables, we may compute

$$X_t = 2\arctan\left(e^t \left|\tan\left(\frac{x}{2}\right)\right|\right)$$

Directly computing its derivative with respect to x, one obtains (making use of some trigonometric equalities),

$$\partial_x X_t = \frac{e^t \sec^2\left(\frac{x}{2}\right)}{1+e^{2t}\tan^2\left(\frac{x}{2}\right)}.$$

(b) We proceed instead with the approach outline in the lecture notebook. The evolution operator $J_{s,t}$ corresponds to the derivative of X_t with respect to $X_s = x$. It satisfies the equation

$$\mathrm{d}J_{s,t} = \sin'(X_t)J_{s,t}\mathrm{d}t = \cos(X_t)J_{s,t}\mathrm{d}t, \quad J_{s,s} = 1.$$

Using the explicit expression of X_t, we may recast this equation as

$$\mathrm{d}J_{s,t} = \frac{1-e^{2t}\tan^2(x/2)}{1+e^{2t}\tan^2(x/2)}J_{s,t}\mathrm{d}t, \quad J_{s,s} = 1.$$

This is a separable equation, which can be explicitly solved. In the case $s = 0$ previously considered, the solution of the separable equation is

$$\partial_x X_t = J_{0,t} = \frac{e^t \sec^2\left(\frac{x}{2}\right)}{1+e^{2t}\tan^2\left(\frac{x}{2}\right)},$$

which is consistent with the direct approach of the previous point.
Notably, this exercise shows that a (typically hard) operation as computing the sensitivity with respect to a parameter can be recasted in solving a linear equation. Exactly as in a neutral network, the first step is to solve first the equation forward and compute X_t.