Reinforcement learning and rough paths theory

Samy Tindel

Purdue University

Auburn University – 2023
Probability Seminar

Ongoing joint work with
Prakash Chakraborty (Penn State) and Harsha Honnappa (Purdue)

Outline

1. Supervised learning
2. Noisy environment and reinforcement learning
3. Results, perspective, methods
   - Results and perspectives
   - Methods

Outline

A basic classification task

Data:

- Points $\{\mathbf{x}_i; i = 1, \ldots, n\}$ with $\mathbf{x}_i \in \mathbb{R}^d$
- Labels $\{y_i; i = 1, \ldots, n\}$ with $y_i \in \{0, 1\}$
- When labels are known, the learning is supervised

Aim:

- Find a proper separation between labels 0 and labels 1

Linear separation

Separation using hyperplanes:

- We use a classification $\hat{y} = \text{sign}(\mathbf{v} \cdot \mathbf{x})$
- $\mathbf{v}$ optimized
  $\hookrightarrow$ According to our data:

$$\mathbf{v} = \min_{\mathbf{w} \in \mathbb{R}^d} \sum_{i=1}^{n} \left\| \text{sign}\left(\mathbf{w} \cdot \mathbf{x}_i\right) - y_i \right\|^2$$

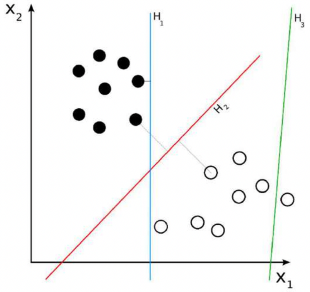


Figure: Separation of 2 subgroups according to H_1, H_2, H_3

Separation using neural networks

Definition of the multilayer neural network:

- Recursion $\mathbf{x}^{k+1} = S(\mathbf{w}^k \mathbf{x}^k + \mathbf{d}^k)$ for $k = 0, \dots, n_{\text{layer}}$
- $\mathbf{w}^k$ matrix-valued, $\mathbf{d}^k$ vector-valued
- S defined componentwise by σ below
- $\mathbf{w}^k$ and $\mathbf{d}^k$ to be optimized

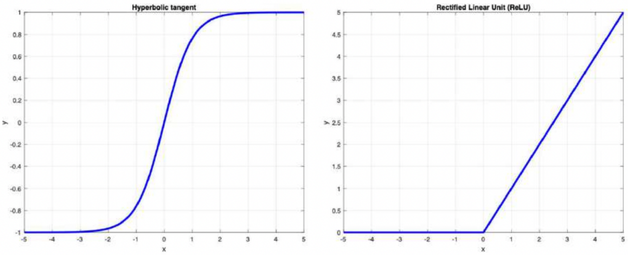


Figure: Sigmoid $\sigma(x) = \frac{2}{\pi}\tanh(x)$ and ReLU $\sigma(x) = \max\{x, 0\}$

Towards a control theory framework (1)

Slight change of notation: We have seen that

- $\mathbf{x}^k$ defined recursively by

$$\mathbf{x}^{k+1} = S\left(\mathbf{w}^k \mathbf{x}^k + \mathbf{d}^k\right) \equiv b\left(\mathbf{x}^k, \mathbf{u}^k\right)$$

- $\mathbf{u}^k = (\mathbf{w}^k, \mathbf{d}^k)$ parameter to be optimized
  $\hookrightarrow$ According to loss function

Example of loss function: With $n = n_{\text{layer}}$ and $y = \text{label}$

$$J(\mathbf{u}) = |\mathbf{y} - \mathbf{x}^n|^2 + \lambda \sum_{k=0}^{n-1} \left|\mathbf{u}^k\right|^2 ,$$

where $\lambda \equiv$ regularization parameter

Towards a control theory framework (2)

Recall: We have seen, for $k = 0, \ldots, n-1$,

$$\begin{cases} \mathbf{x}^{k+1} = b\left(\mathbf{x}^k, \mathbf{u}^k\right) \\ J(\mathbf{u}) = (y - \mathbf{x}^n)^2 + \lambda \sum_{k=0}^{n-1} \left|\mathbf{u}^k\right|^2 \end{cases}$$

Limiting procedure: Take $n \to \infty$ and renormalize. We get

$$\begin{cases} \mathrm{d}x_t = b\left(x_t, u_t\right) \mathrm{d}t\,, & t \in [0, T] \\ J(u) = G(x_T) + \int_0^T r\left(u_t, x_t\right) \mathrm{d}t \end{cases}$$

This is a classical control theory framework.

References

Bensoussan, A., Zhou, X. et al (2020).
Machine learning and control theory.
arXiv preprint arXiv:2006.05604.

Chen, R. T., Rubanova, Y., Bettencourt, J., Duvenaud, D. K. (2018).
Neural ordinary differential equations.
Advances in neural information processing systems, 31.

Outline

Generalization 1: noisy environment

A stochastic equation: For the neural network dynamics, take

- A Brownian motion W
- Equation of the form

$$\mathrm{d}x_t = b(x_t, u_t)\,\mathrm{d}t + \sigma(x_t)\,\mathrm{d}W_t$$

Motivation:

- Neural networks are noisy
- Noise stabilizes equations
- Example on the right:
  $\mathrm{d}x_t = x_t\,\mathrm{d}t + \sigma x_t\,\mathrm{d}W_t$

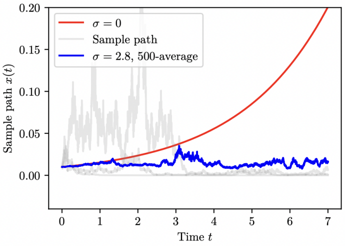


Figure: Stabilization with a multiplicative noise

Generalization 2: reinforcement learning (1)

Equation: So far our problem is

$$\begin{cases} \mathrm{d}x_t = b\left(x_t, u_t\right)\,\mathrm{d}t + \sigma(x_t)\,\mathrm{d}W_t \\ J(u) = \mathbf{E}\left[G(x_T) + \int_0^T r\left(u_t,\, x_t\right)\,\mathrm{d}t\right] \end{cases}$$

RL problematic:

- Problem: In many situations, the dynamics for x is unknown
- RL strategy: Use the control u for both
  1. Optimization of the action
  2. Exploration of different dynamics

Change in the model:

- The control u will be measure-valued
- We add an entropy term to the reward, to favor exploration

Generalization 2: reinforcement learning (2)

Previous version of the model:

$$\begin{cases} x_t = y + \int_0^t b(x_r, u_r)\,\mathrm{d}r + \int_0^t \sigma(x_r)\,\mathrm{d}W_r \\ J(u) = \mathbf{E}\left[G(x_T) + \int_0^T r(u_r,\, x_r)\,\mathrm{d}r\right] \end{cases}$$

Relaxed control: We consider

- $U \subset \mathbb{R}^d$
- Control u_r is replaced by $\gamma_r \in \mathcal{P}(U)$

New version of the model:

$$\begin{cases} x_t = y + \int_0^t \int_U b(x_r, a)\,\gamma_r(\mathrm{d}a)\mathrm{d}r + \int_0^t \sigma(x_r)\,\mathrm{d}W_r \\ J(\gamma) = \mathbf{E}\left[G(x_T) + \int_0^T F(x_r, \gamma_r)\,\mathrm{d}r\right] \end{cases}$$

Generalization 2: reinforcement learning (3)

New version of the model:

$$\begin{cases} x_t = y + \int_0^t \int_U b\,(x_r, a)\; \gamma_r(\mathrm{d}a)\mathrm{d}r + \int_0^t \sigma(x_r)\,\mathrm{d}W_r \\ J(\gamma) = \mathbf{E}\left[G(x_T) + \int_0^T F\,(r, x_r, \gamma_r)\;\mathrm{d}r\right] \end{cases}$$

Recall: We wish to use γ for

1. Optimization of the action
2. Exploration of different dynamics

Typical example of function F: If γ_r has a density $\dot{\gamma}_r$,

$$F(r, x, \gamma) = e^{-\rho r}\left(\int_U r(x, u)\dot{\gamma}_r(u)du - \lambda \int_U \dot{\gamma}_r(u) \log \dot{\gamma}_r(u)du\right)$$

Example in 3-d folding (1)

Notation: For a protein folding study,

- $s_t \equiv t$-th iteration of the molecule configuration
- $a \equiv$ vector containing bond angle and bond torsion
  $\hookrightarrow$ to be optimized
- γ, κ unknown parameters
- $U(s, a) \equiv$ energy to be minimized, with an entropy term

Dynamics:

$$\begin{cases} \mathrm{d}s_t &= a_t \, \mathrm{d}t \\ \mathrm{d}a_t &= -\gamma \, a_t \, \mathrm{d}t - \kappa \, \mathrm{d}W_t \end{cases}$$

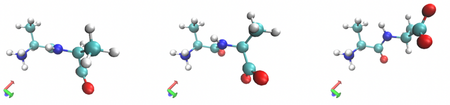


Figure: Sequence of foldings

References

Liu, X., Xiao, T., Si, S., Cao, Q., Kumar, S., Hsieh, C. J. (2019).
Neural sde: Stabilizing neural ode networks with stochastic noise.
arXiv preprint arXiv:1906.02355.

Bajaj, C., Li, C., Nguyen, M. (2022).
Solving the Side-Chain Packing Arrangement of Proteins from Reinforcement Learned Stochastic Decision Making.
arXiv preprint arXiv:2212.03320.

Generalization 3: rough environment (1)

Observation:

- We have assumed that x is driven by a Brownian motion
- In real life, some observations are not Brownian
- In particular
  - The Hölder regularity of $t \mapsto W_t$ is not always $1/2 - \varepsilon$
  - Environments are not always Markovian

Natural generalization:

- Fractional Brownian motion

Fractional Brownian motion

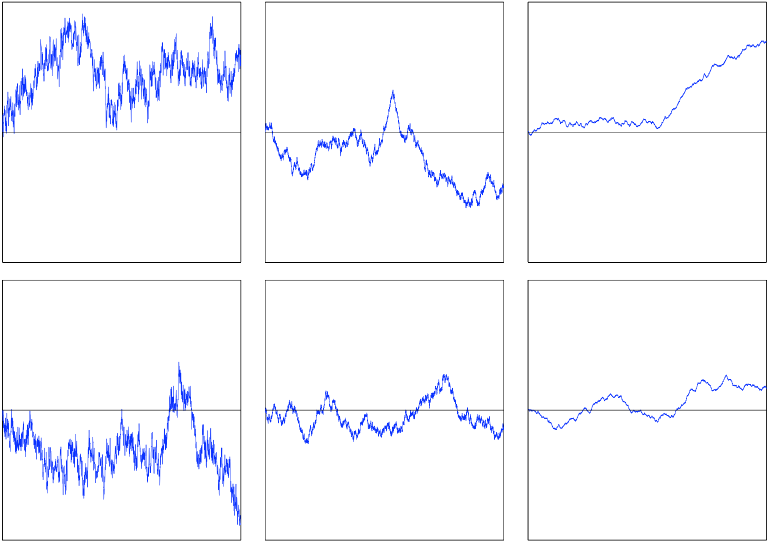
- $H \in (0,1)$
- $B = (B^1, \ldots, B^d)$
- B^j centered Gaussian process, independence of coordinates
- Variance of the increments:

$$\left(\mathbf{E}[|B_t^j - B_s^j|^2]\right)^{1/2} = |t-s|^H$$

- $H^- \equiv$ Hölder-continuity exponent of B
- If $H = 1/2$, $B =$ Brownian motion
- If $H \neq 1/2$ natural generalization of BM

Remark: FBm widely used in applications

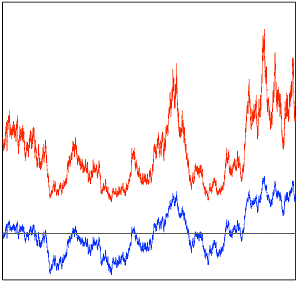
Examples of fBm paths



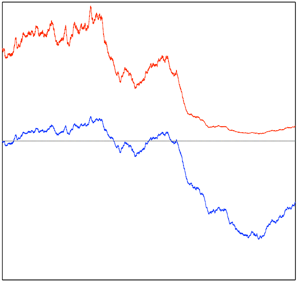
$H = 0.3$ $H = 0.5$ $H = 0.7$

Paths for a linear SDE driven by fBm

$$dY_t = -0.5x_t \,\mathrm{d}t + 2x_t \,\mathrm{d}B_t, \quad x_0 = 1$$



$H = 0.5$



$H = 0.7$

Blue: $(B_t)_{t\in[0,1]}$ Red: $(Y_t)_{t\in[0,1]}$

fBm and integration

Basic facts about fBm: Let B be a fBm with $H \in (0,1)$. Then:

- B is not a finite variation process
- B is not a Markov process
- B is not a martingale

Main step in order to solve equations:

- Define integrals of the form $\int_0^t \sigma(x_s)\, dB_s$

Problem for fBm:

- Itô's theory does not apply to B
- Need for another integration theory $\longrightarrow$ rough paths theory
- This will rely on regularity and Gaussianity of B

Generalization 3: rough environment (2)

Previous version of the model:

$$\begin{cases} x_t = y + \int_0^t \int_U b\left(x_r, a\right)\, \gamma_r(\mathrm{d}a)\mathrm{d}r + \int_0^t \sigma(x_r)\, \mathrm{d}W_r \\ J(\gamma) = \mathbf{E}\left[G(x_T) + \int_0^T F\left(r, x_r, \gamma_r\right)\, \mathrm{d}r\right] \end{cases}$$

New version of the model:

$$\begin{cases} x_t = y + \int_0^t \int_U b\left(x_r, a\right)\, \gamma_r(\mathrm{d}a)\mathrm{d}r + \int_0^t \sigma(x_r)\, \mathrm{d}B_r \\ J(\gamma) = G(x_T) + \int_0^T F\left(r, x_r, \gamma_r\right)\, \mathrm{d}r \end{cases}$$

Typical example of function F: If γ_r has a density $\dot{\gamma}_r$,

$$F(r, x, \gamma) = e^{-\rho r}\left(\int_U r(x, u)\dot{\gamma}_r(u)du - \lambda \int_U \dot{\gamma}_r(u) \log \dot{\gamma}_r(u)du\right)$$

Summary for our setting

We consider:

1. Control point of view on learning
2. Reinforcement learning setting with regularization by entropy
3. Rough environment, possibly driven by a fBm
4. Pathwise optimization

Outline

Outline

A more formal framework (1)

Dynamics: For $0 \le s \le t \le T$ and $x_t \in \mathbb{R}^m$,

$$x_t - x_s = \int_s^t \int_U b(x_r, a)\, \gamma_r(\mathrm{d}a)\mathrm{d}r + \int_s^t \sigma(x_r)\, \mathrm{d}B_r$$

Reward from s to T:

$$J_{sT}(\gamma, y) = \int_s^T F(r, x_r^\gamma, \gamma_r)\, dr + G(x_T^\gamma)$$

State space for γ: Minimal Hölder regularity in Wasserstein distance,

$$\mathcal{V}^\varepsilon([s,T]) = \big\{ \mathcal{C}^\varepsilon([s,T]; \mathcal{P}(U)); W_2(\gamma_r, \gamma_t) \le c|t-r|^\varepsilon \text{ for some } c > 0 \text{ and all } r, t \in [s,T] \big\}.$$

A more formal framework (2)

Recall:

$$J_{sT}(\gamma, y) = \int_s^T F(r, x_r^\gamma, \gamma_r)\, dr + G(x_T^\gamma)$$

Value: We set, for $0 \le s \le T$

$$V(s, y) = \sup\left\{J_{sT}(\gamma, y);\ \gamma \in \mathcal{V}^\varepsilon([s, T])\right\}$$

HJB equation for the value

Theorem 1.

The value V solves the following first order rough PDE:

$$\Big[\partial_t v(t,y) + \sup_{\gamma\in\mathcal{P}(U)} H(t,y,\gamma,\nabla v(t,y))\Big]\mathrm{d}t$$
$$+\sigma(t,y)\cdot\nabla v(t,y)\mathrm{d}B_t = 0$$

for $(t,y)\in[0,T]\times\mathbb{R}^n$, with final condition

$$v(T,y) = G(y)\,.$$

The Hamiltonian H above is defined by

$$H(t,y,\gamma,p) = p\cdot\int_U b(y,a)\,\gamma(\mathrm{d}a) + F(t,y,\gamma)$$

Notes on Theorem 1

Remarks:

- V should be considered as a rough viscosity solution
- Extra care due to the fact that γ_t is a measure

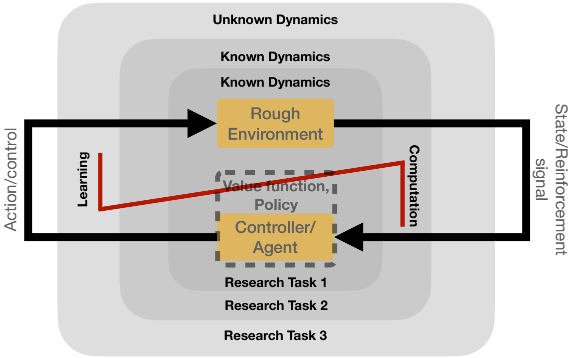
More results related to Theorem 1:

1. Definition of test functions for rough viscosity solutions
   $\hookrightarrow$ This is usually not done in other related works
2. Definition of related jets
3. Regularity of V
4. Existence of a minimizer γ^*
5. Transformation:
   rough PDEs $\longrightarrow$ PDE with random coefficients

Perspectives

Program:

- Numerical schemes for HJB
  $\hookrightarrow$ Policy iteration
- Actor-critic scheme for the reinforcement part



Outline

The rough paths problem (1)

The main integration problem: Give a meaning to integrals like

$$\int_s^t \sigma(r,y)\cdot\nabla v(r,y)\mathrm{d}x_r\,,$$

where $x \equiv$ fBm or other γ-Hölder path

Related toy model: With 1-d notations, define integrals like

$$\int_s^t V(x_r)\mathrm{d}x_r$$

The rough paths problem (2)

Easy expansion: We have

$$\int_s^t V(x_r)\mathrm{d}x_r = V(x_s)\mathbf{x}^1_{st} + V'(x_s)\mathbf{x}^2_{st} + \int_s^t \int_s^r \left(V'(x_u) - V'(x_s)\right) \mathrm{d}x_u \mathrm{d}x_r$$

Explanation of terms: In the above expansion we have

1. $\mathbf{x}^1_{st} = x_t - x_s$, well-defined
2. Assumption: $\mathbf{x}^2_{st} = \int_s^t \int_s^r \mathrm{d}x_u \mathrm{d}x_r$ well-defined
   $\hookrightarrow$ Main rough paths assumption
3. $\int_s^t \int_s^r \left(V'(x_u) - V'(x_s)\right) \mathrm{d}x_u \mathrm{d}x_r$
   $\hookrightarrow$ defined as a "Young" integral if $x \in \mathcal{C}^\gamma$

Rough paths assumptions

Context: Consider a Hölder path x and

- For $n \geq 1$, $x^n \equiv$ linearization of x with mesh $1/n$
  $\hookrightarrow$ x^n piecewise linear.
- For $0 \leq s < t \leq 1$, set

$$\mathbf{x}^{\mathbf{2},n,i,j}_{st} \equiv \int_{s<u<v<t} dx^{n,i}_u \, dx^{n,j}_v$$

Rough paths assumption 1:

- x is a $\mathcal{C}^\gamma$ function with $\gamma > 1/3$.
- The process $\mathbf{x}^{\mathbf{2},n}$ converges to a process $\mathbf{x}^{\mathbf{2}}$ as $n \to \infty$
  $\hookrightarrow$ in a $\mathcal{C}^{2\gamma}$ space.

Rough paths assumption 2:

- Vector fields $V_0, \ldots, V_j$ in $\mathcal{C}^\infty_b$.

Brief summary of rough paths theory

Main rough paths theorem (Lyons): Under previous assumptions
↪ Consider y^n defined by

$$y_t^n = \sum_{j=1}^{d} \int_0^t V_j(x_u^n)\, dx_u^{n,j}.$$

Then

- y^n converges to a function Y in $\mathcal{C}^\gamma$.
- Y can be seen as the integral path $Y_t = \sum_{j=1}^{d} \int_0^t V_j(x_u)\, dx_u^j$.

Brief summary of rough paths theory

Main rough paths theorem (Lyons): Under previous assumptions
$\hookrightarrow$ Consider y^n defined by

$$y_t^n = \sum_{j=1}^{d} \int_0^t V_j(x_u^n)\, dx_u^{n,j}.$$

Then

- y^n converges to a function Y in $\mathcal{C}^\gamma$.
- Y can be seen as the integral path $Y_t = \sum_{j=1}^{d} \int_0^t V_j(x_u)\, dx_u^j$.

Rough paths theory

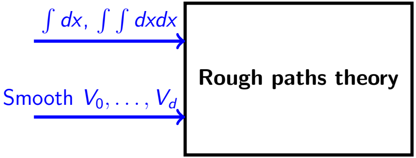
Brief summary of rough paths theory

Main rough paths theorem (Lyons): Under previous assumptions
$\hookrightarrow$ Consider y^n defined by

$$y_t^n = \sum_{j=1}^{d} \int_0^t V_j(x_u^n)\, dx_u^{n,j}.$$

Then

- y^n converges to a function Y in $\mathcal{C}^\gamma$.
- Y can be seen as the integral path $Y_t = \sum_{j=1}^{d} \int_0^t V_j(x_u)\, dx_u^j$.



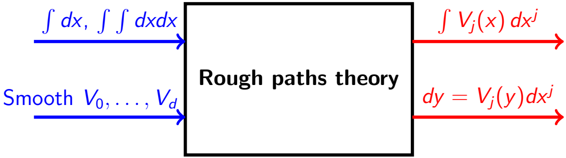
Brief summary of rough paths theory

Main rough paths theorem (Lyons): Under previous assumptions
↪ Consider y^n defined by

$$y_t^n = \sum_{j=1}^{d} \int_0^t V_j(x_u^n)\, dx_u^{n,j}.$$

Then

- y^n converges to a function Y in $\mathcal{C}^\gamma$.
- Y can be seen as the integral path $Y_t = \sum_{j=1}^d \int_0^t V_j(x_u)\, dx_u^j$.



Rough viscosity solutions

Recall: Our equation is

$$\left[\partial_t v(t,y) + \sup_{\gamma\in\mathcal{P}(U)} H(t,y,\gamma,\nabla v(t,y))\right]\mathrm{d}t + \sigma(t,y)\cdot\nabla v(t,y)\mathrm{d}B_t = 0$$

Problem:
∇v above is ill-defined. The solution is not smooth in general

Viscosity solution idea:
Transfer derivatives on test functions

Changes in the rough paths setting:
Test functions should also be rough!

Rough viscosity solutions: test functions

Definition 2.

We are given

- σ smooth enough
- A rough path x (example $x = B$ fBm with $H > 1/3$)
- Drift term $\psi^t \in \mathcal{C}^{\varepsilon,1}([0,T] \times \mathbb{R}^m)$
- $\psi : [0,T] \times \mathbb{R}^m \mapsto \mathbb{R}^m$

Then ψ is a test function in $\mathcal{T}_\sigma$ if ψ satisfies:

$$\delta\psi_{s_1 s_2}(y) = \int_{s_1}^{s_2} \psi_r^t(y)\,dr - \int_{s_1}^{s_2} \sigma(r,y) \cdot \nabla\psi_r(y)\,dx_r$$

Rough viscosity solutions: Definition

Definition 3.

Consider

- x rough path
- v path whose increments are controlled by x

We say that v is a rough viscosity supersolution of HJB equation if

1. $v_T(y) \geq G(y)$
2. If $\psi \in \mathcal{T}_\sigma$ is such that
   $v - \psi$ admits a local minimum at (s, y), then

$$\psi_s^t(y) \leq - \sup_{\gamma \in \mathcal{K}} H(s, y, \gamma).$$