Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS 2022)

Learning Fractional White Noises in Neural Stochastic Differential Equations

Anh Tong

Thanh Nguyen-Tang

Toan Tran

Jaesik Choi

KAIST

JOHNS HOPKINS UNIVERSITY

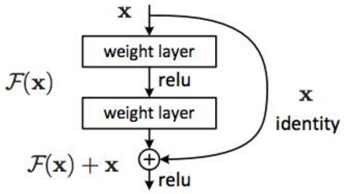
VinAI RESEARCH

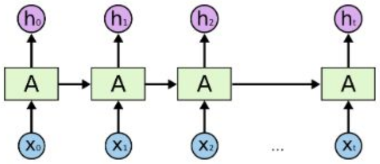
INEEJI

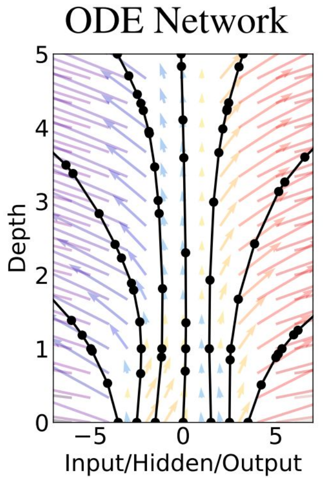
Introduction

- Introduce **fractional noise** modeling for neural SDEs
- Learn **Hurst parameters** of fractional Brownian motions
- **Efficient approximate sampling** noise using sparse Gaussian processes (GPs)
- Convergence guarantee by **rough path theory** and sparse GP bounds.

Neural ODEs







Source: Chen et al, 2018

$$X_{n+1} = X_n + f(X_n)$$

$$\frac{dX_t}{dt} = f_\theta(X_t, t)$$

Layers are continuous as a inspiration of residual nets + recurrent nets

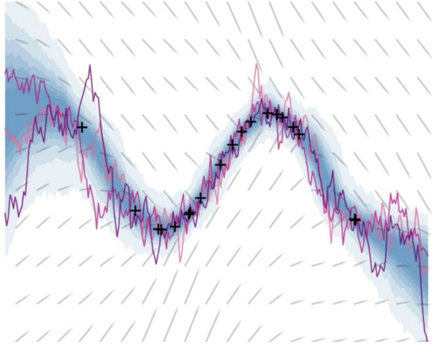
Chen et al, 2018. Neural Ordinary Differential Equations. NeurIPS.

Neural SDEs

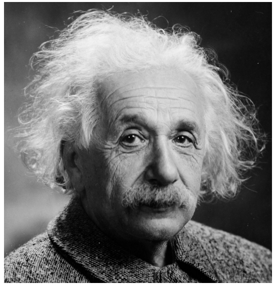
$$dX_t = f_\theta(X_t, t)dt + g_\theta(X_t, t)dW_t$$

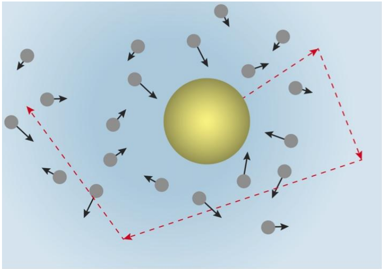
drift

diffusion



Source: Li et al, 2020





Source: https://scottbembenek.com

Physical system

Black-Scholes models

$$dS_t = rS_t dt + \sigma S_t dW_t$$

Stock market crashed (Source: wiki)

Financial modeling

Neural SDEs with multifractional Brownian motion

Driving noise: (as FBM)

$$B_t = \frac{1}{\Gamma\left(h(t) + \frac{1}{2}\right)} \int_0^t (t-s)^{h(t)-\frac{1}{2}} dW_s$$

Hurst exponent:

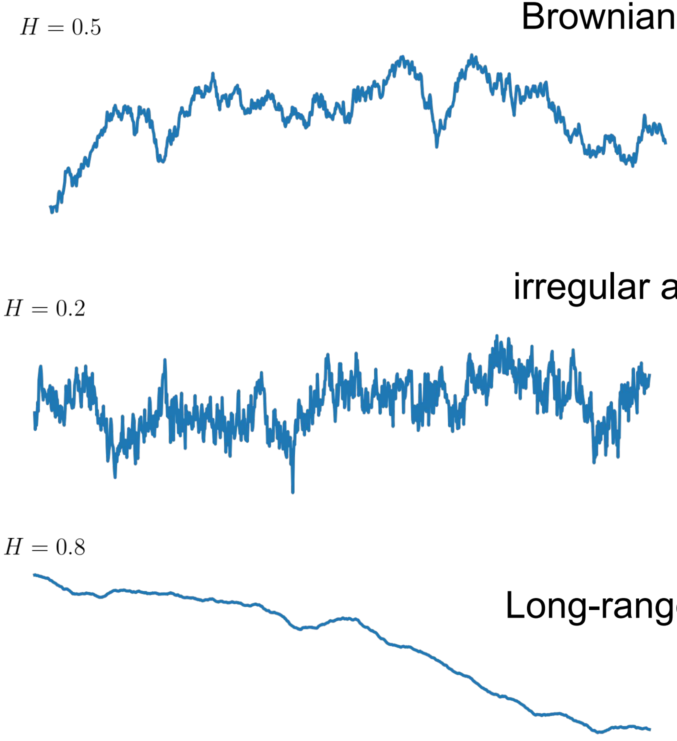
- change over time
- < 0.5: **"rough"**
- > 0.5: **long-range dependency**

Neural SDE: (our model)

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$$

Goal: Learn **Hurst exponent** functions together with neural SDE parameters

Fractional Brownian motions



Brownian motion

irregular and “rough”

Long-range dependence

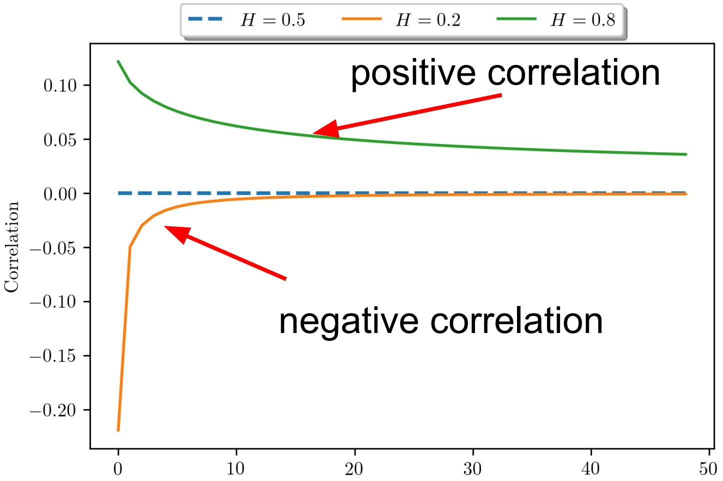
Hurst exponent

$$B_t = \frac{1}{\Gamma(H + 1/2)} \int_0^t (t - s)^{H - 1/2} dW_t$$

- Riemann-Liouville form
- Well-known in Math. Fin.
- Characterize stylized facts in finance market

Gatheral et at, 2018. Volatility is rough. Quant. Fin.
Comte and Renault, 1998. Long memory in continuous-time stochastic volatility. Math. Fin.

Fractional noises have **correlations**



$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$$

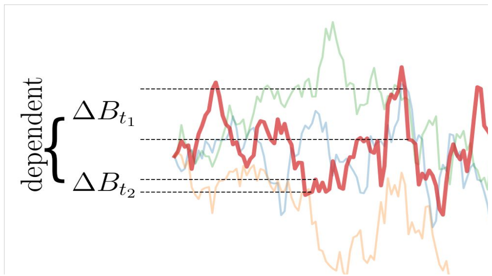
Euler method:

$$X_{t+\Delta t} = X_t + \mu(X_t, t)\Delta t + \sigma(X_t, t)\Delta B_t$$

ΔB_t exists correlations

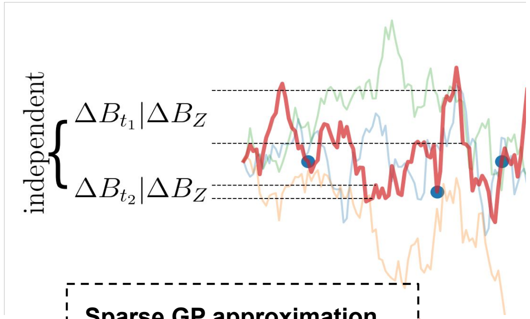
While Brownian noise $\Delta W_t \sim \mathcal{N}(0, \Delta t)$ can be **independently** sampled

Efficient sampling with sparse GP



Fractional Brownian motion (FBM)

Decouple dependency



Sparse GP approximation

Convergence by: Rough path theory [Lyon 1998]

Approx. integral

$$B_t = \int_0^t f(s,t)dW_s \approx \sum_i g_i(t)\varepsilon_t$$

Est. covariance

$$\mathrm{Cov}\left[B'_t B'_s\right] \approx \frac{1}{(\Delta t)^2}\sum_i g_i(t)g_i(s) \qquad B'_t = \frac{\Delta B_t}{\Delta t}$$

Sparse GP sampling

$$\Delta B_t = m(t)\Delta t + v(t)\Delta t \xi_t$$

mean — variance

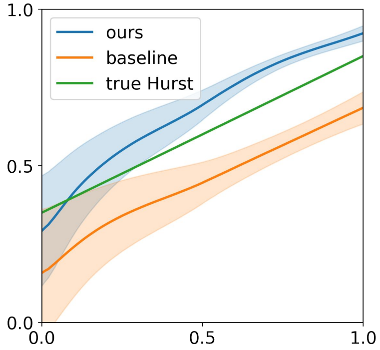
Sparse Gaussian process

$$m(x) = k(x,\, Z)k(Z,Z)^{-1}u$$

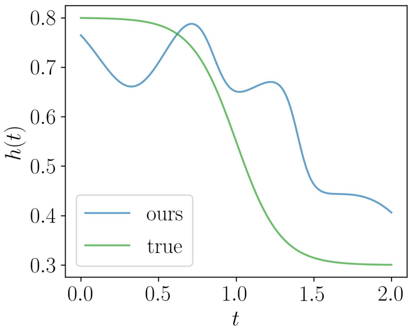
$$v(x) = k(x,x) - k(x,Z)k(Z,Z)^{-1}k(Z,x)$$

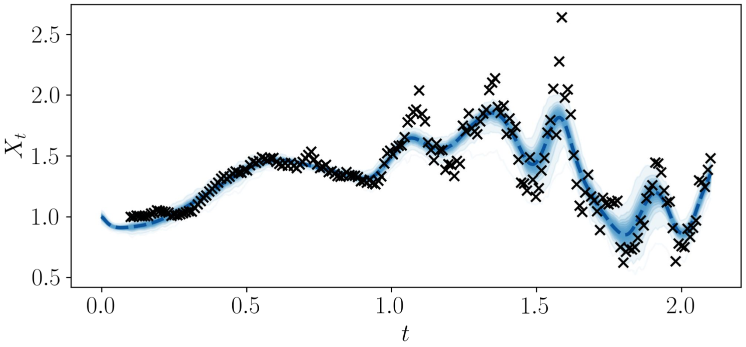
Experimental results

Approximated vs true samples



Learn dynamic of Hurst exponents

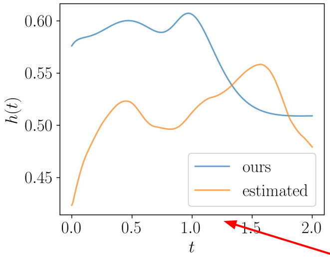




Fit well with varying Hurst over time

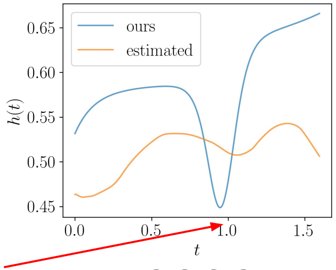
Experimental results

- Discover Hurst for financial data



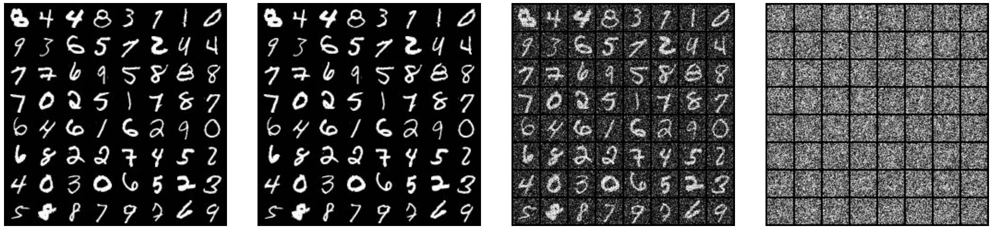
AMZN

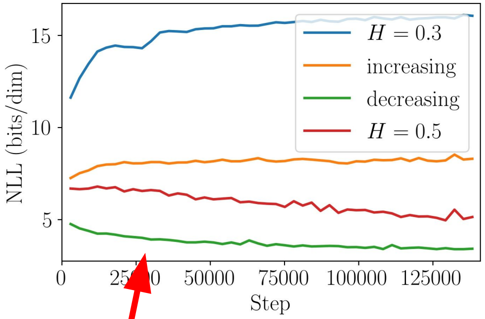
Quarterly report



GOOGL

- Novel scored-based generative model with FBM (**decreasing Hurst is prefered**)





better NLL

Thank you!