UiO : Department of Mathematics
University of Oslo

Pricing options on flow forwards by neural networks in Hilbert space

Fred Espen Benth

Joint work with Nils Detering (UC Santa Barbara) and Luca Galimberti (NTNU)

Problem

- Using a neural network, learn the map

$$x \mapsto \mathbb{E}[\mathscr{P}(X_\tau^{t,x})]$$

$$H \longrightarrow \mathbb{R}$$

- Here, $(X_s^{t,x})_{s\geq t}$ is a stochastic process in a Hilbert space H, starting in $x \in H$ at time t.

- $\mathscr{P} : H \to \mathbb{R}$ is the payoff functional

- Learning on simulated data by exploiting the structure in Hilbert space

Why?

INFINITE DIMENSIONAL OPTION PRICING

Why?

- Pricing options in electricity and gas markets
  - …as well as markets for temperature, wind and freight

- Options written on forwards/futures contracts
  - …where delivery takes place over a period $[T_1, T_2]$

- EEX: calls and puts on flow forwards, $t \leq \tau \leq T_1 < T_2$

$$C(t) = \mathbb{E}[\max(\hat{F}_\tau(T_1, T_2) - K, 0) \mid \mathcal{F}_t]$$

$$\hat{F}_t(T_1, T_2) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} F_t(T)\,dT$$

Why? … some notation

- Musiela parametrization: $\xi := T - t$, ξ is time-to-maturity

$$X_t(\xi) := F_t(\xi + t)$$

- Let H be a Hilbert space of real-valued functions on $\mathbb{R}_+$,
  - where the evaluation functional $\delta_\xi \in H^*$

$$\hat{F}_t(T_1, T_2) = \hat{F}_t(t + (T_1 - t), t + (T_1 - t) + (T_2 - T_1))$$

$$= \frac{1}{T_2 - T_1} \int_0^\infty 1_{[0, T_2 - T_1]}(u - (T_1 - t)) X_t(u) du$$

$$:= \delta_{T_1 - t} \mathcal{D}_{T_2 - T_1}(X_t)$$

- Assume integral operator $\mathcal{D}_\lambda \in L(H)$

Why? … back to options again

- Let $p : \mathbb{R} \to \mathbb{R}$ be the option's payoff function

$$p(\hat{F}_\tau(T_1, T_2)) = p(\delta_{T_1-\tau}\mathcal{D}_{T_2-T_1}(X_\tau)) := \mathcal{P}(X_\tau)$$

- Model for the forward curve: assume H is a Banach algebra
  - Pointwise multiplication of curves

$$X_t := \exp(Y_t)$$

- Log-forward curves given by H-valued OU-process
  - W is Q-Wiener process in H, N is homogeneous Poisson random measure on H with Levy measure $\nu(dz)$

$$dY_t = \partial_\xi Y_t dt + \alpha(t)dt + \eta(t)dW(t) + \int_H \gamma(t,z)\widetilde{N}(dt,dz)$$

Why ?...

- Time-dependent, non-random coefficients, integrable

$$|\alpha(\cdot)|, \quad \|\eta(\cdot)\|_{HS} \in L^2_{loc}(\mathbb{R}_+), \quad \int_H |\gamma(\cdot, z)|^2 \nu(dz) \in L^1(\mathbb{R}_+)$$

- Mild solution
  - shift semigroup $\mathcal{S}_t f := f(\cdot + t)$ strongly continuous with densely defined generator ∂_ξ

$$Y_t = \mathcal{S}_t Y_0 + \int_0^t \mathcal{S}_{t-s}\alpha(s)ds + \int_0^t \mathcal{S}_{t-s}\eta(s)dW(s) + \int_0^t \int_H \mathcal{S}_{t-s}\gamma(s,z)\widetilde{N}(ds,dz)$$

- No-arbitrage drift condition,

$$\alpha(t,\cdot) = -\frac{1}{2}|Q^{1/2}\eta^*(t)(\delta_\cdot^* 1)|^2 - \int_H \{\exp(\gamma(t,z)) - 1 - \gamma(t,z)\}\,\nu(dz)$$

Why?

- Using algebra-property and semigroup

$$X_\tau = \exp(Y_\tau) = \exp(\mathcal{S}_{\tau-t}Y_t)\exp(Z_{t,\tau}) = (\mathcal{S}_{\tau-t}X(t))\exp(Z_{t,\tau})$$

$$Z_{t,\tau} = \int_t^\tau \mathcal{S}_{\tau-s}\alpha(s)ds + \int_t^\tau \mathcal{S}_{\tau-s}\eta(s)dW(s) + \int_t^\tau\int_H \mathcal{S}_{\tau-s}\gamma(s,z)\widetilde{N}(ds,dz)$$

- Option price (zero risk free rate)

$$V(t) = \mathbb{E}[p(\hat{F}_\tau(T_1,T_2)) \mid \mathcal{F}_t] = \mathbb{E}[\mathcal{P}(X_\tau) \mid \mathcal{F}_t]$$

- Or, $V(t) := V(t, X_t)$, with

$$V(t,x) = \mathbb{E}[\mathcal{P}((\mathcal{S}_{\tau-t}x)\exp(Z_{t,\tau}))]$$

$$H \ni x \mapsto V(t,x) \in \mathbb{R}$$

…and now, really why!

- The option price is *not* a function of $\hat{F}_t(T_1, T_2)$

$$\delta_{T_1-\tau}\mathscr{D}_{T_2-T_1}((\mathcal{S}_{\tau-t}x)\exp(Z_{t,\tau})) = \frac{1}{T_2 - T_1}\int_{T_1-t}^{T_2-t} x(v)e^{Z_{t,\tau}(v-(\tau-t))}dv$$

- The option price is depending on the *term structure* $T \mapsto F_t(T)$

- **Proposition:** Suppose p is Lipschitz. Then V is well-defined and Lipschitz continuous.

  Proof: well-defined comes from exponential moments of Wiener process (Fernique) and exponential integrability in no-arbitrage condition of jumps. Lipschitz follows by direct calculation.

Convenient minimization

- Let μ be a measure on H supported on a compact
- Define for $g \in L^2(\mu)$

$$I(g) = \mathbb{E}[\int_H |\mathscr{P}(\mathcal{S}_{\tau-t} x \exp(Z_{t,\tau})) - g(x)|^2 \mu(dx)]$$

- **Proposition**: $I(V) = \inf_{g \in L^2(\mu)} I(g)$

  Proof: A direct calculation shows

$$I(g) = \int_H \mathrm{Var}(\mathscr{P}(\mathcal{S}_{\tau-t} x \exp(Z_{t,\tau})))\mu(dx) + \int_H |V(t,x) - g(x)|^2 \mu(dx) \geq I(V)$$

BACKGROUND AND OTHER APPLICATIONS

Background

- Jentzen et al.: neural net to learn the map

$$\mathbb{R}^d \ni x \mapsto \mathbb{E}[\mathscr{P}(X_\tau^{t,x})]$$

- $X_\tau^{t,x}$ diffusion process on $\mathbb{R}^d$
- Options on very-high dimensional baskets of assets
  - I.e., x is high dimensional
  - Numerical solution of PDEs in high dimensions

- Cheridito, Teichmann et al.: optimal stopping and American options

- Bayer, Cuchiero, Horvath et al.: implied rough volatility, local volatility

- Weinan E et al.: stochastic control

Other applications

- Hedging volume and price risk by quantos
    - Joint payoff on price and temperature

$$\mathbb{E}[\mathscr{P}(F_\tau^E(T_1, T_2))\mathscr{Q}(F_\tau^T(T_1, T_2))]$$

- Virtual power plants and swing options
    - User-time contracts (volume-flexibility at strikes)
    - Gas and coal-fired power plants (strip of calls)

- Fixed-income: call and puts on SOFR-futures
    - SOFR: secured overnight rates, US substitute for LIBOR
    - CME: trades in SOFR-futures, “flow forwards” on SOFR rates

NEURAL NETWORK IN HILBERT SPACE

Classical neural networks (one layer)

- Given continuous function $f \in C(\mathbb{R}^d, \mathbb{R})$, find neural network that approximates it on a compact $K \subset \mathbb{R}^d$

- Neural network
  - Fix continuous activation function $\sigma : \mathbb{R} \to \mathbb{R}$
  - For $a \in \mathbb{R}^d, \ell, b \in \mathbb{R}$ define a neuron $\mathcal{N}_{\ell,a,b} \in C(\mathbb{R}^d, \mathbb{R})$

$$x \mapsto \ell \sigma(a^\top x + b)$$

- One-layer neural network (NN)

$$\sum_{i=1}^{N} \mathcal{N}_{\ell_i,a_i,b_i}(x) = \sum_{i=1}^{N} \ell_i \sigma(a_i^\top x + b_i)$$

Universal approximation theorem

- For given activation function σ, define linear space generated by neurons

$$\mathfrak{N}(\sigma) := \text{span}\left\{\mathcal{N}_{\ell,a,b} \,:\, \ell, b \in \mathbb{R}, a \in \mathbb{R}^d\right\}$$

- **Universal approximation**: under mild conditions on σ
  - $\mathfrak{N}(\sigma)$ is dense with respect to the topology of uniform convergence on compacts.
  - For every $f \in C(\mathbb{R}^d, \mathbb{R})$ and compact $K \subset \mathbb{R}^d$, given $\epsilon > 0$ there exists $N \in \mathbb{N}$ and $\ell_i, b_i \in \mathbb{R}, a_i \in \mathbb{R}^d$ such that

$$\sup_{x \in K} \left| f(x) - \sum_{i=1}^{N} \mathcal{N}_{\ell_i,a_i,b_i}(x) \right| < \epsilon$$

Neural network in infinite dimensions

- Extend network from $\mathbb{R}^d$ to $\mathbb{R}^\infty$, i.e., to infinite dimensional topological vector space $\mathfrak{X}$
  - Approximate functions $f \in C(\mathfrak{X}, \mathbb{R})$

- Why?
  - Compute efficiently option prices on flow forwards

- $\mathfrak{X} =$ Hilbert space: **exploit structure of basis functions**
  - Train network using data AND structural information!
  - Flexibility on activation function across all dimensions

Neural network

- A neuron is defined by
  - Fixed activation function $\sigma \in C(\mathfrak{X}, \mathfrak{X})$
  - Affine map $x \mapsto Ax + b$, with $A \in L(\mathfrak{X}), b \in \mathfrak{X}$
  - $\ell \in \mathfrak{X}^*$, with $\mathfrak{X}^*$ topological dual of $\mathfrak{X}$

$$\mathcal{N}_{\ell,A,b}(x) = \langle \ell, \sigma(Ax + b) \rangle$$

- Let $\mathfrak{X}$ be a Frechet space
  - Complete metrizable locally convex topological vector space
  - Topology generated by seminorms $(p_k)_{k \in \mathbb{N}}$

- Consider $C(\mathfrak{X}, \mathbb{R})$ with locally convex topology generated by family of seminorms $\{q_K \ : \ K \subset \mathfrak{X}, \text{compact}\}$
  - Riesz representation theorem

Universal approximation

- Activation function σ is called *discriminatory* is for any fixed pair (μ, K)

$$\int_K \langle \ell, \sigma(Ax + b) \rangle \mu(dx) = 0$$

  for all $\ell \in \mathfrak{X}^*, A \in L(\mathfrak{X}), b \in \mathfrak{X}$ implies that $\mu = 0$

- μ is a regular (signed) Borel measure on K

- **Proposition**: Let σ be discriminatory. Then $\mathfrak{N}(\sigma)$ is dense in $C(\mathfrak{X}, \mathbb{R})$

  Proof: Following Cybenko's classical proof using Hahn-Banach. Riesz representation for linear functionals on $C(\mathfrak{X}, \mathbb{R})$

Discriminatory activation function?

- Restrict to activation functions being *von-Neumann bounded*
  - For any $k \in \mathbb{N}$ there exists $c_k > 0$ such that

$$\sup_{y \in \sigma(\mathfrak{X})} p_k(y) \leq c_k$$

- Fix a $\psi \in \mathfrak{X}^* \backslash \{0\}$, define hyperplane Ψ_0 and half-spaces Ψ_+, Ψ_-

$$\Psi_+ := \{x \in \mathfrak{X} \ : \ \langle \psi, x \rangle > 0\}$$

$$\Psi_0 := \mathrm{Ker}(\psi)$$

$$\Psi_- := \{x \in \mathfrak{X} \ : \ \langle \psi, x \rangle < 0\}$$

Discriminatory by separating property

- Activation function σ is *separating* if there exists $\psi \in \mathfrak{X}^*\backslash\{0\}$ and $u_0+, u_-, u_0 \in \mathfrak{X}$ such that either $u_+ \notin$ span$\{u_0, u_-\}$ or $u_- \notin$ span$\{u_0, u_+\}$ and

$$\lim_{\lambda\to\infty} \sigma(\lambda x) = u_*, \text{if } x \in \Psi_*$$

for $* \in \{+,-,0\}$

- **Proposition**: If σ is von Neumann-bounded and separating, then it is discriminatory.

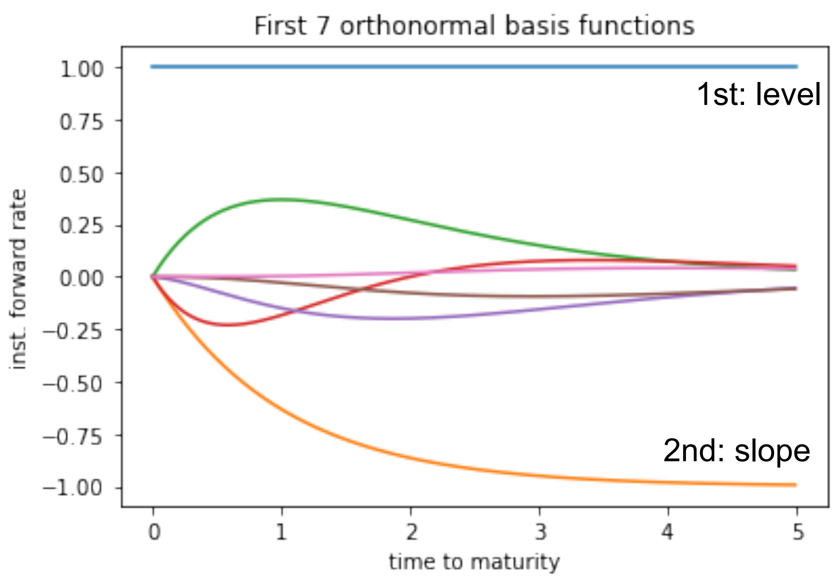
  Proof: Using the Hahn-Jordan decomposition of μ, and playing around wth the flexible choice of the neural network parameters ℓ, A, b and separation of σ.

APPLICATION

Pricing call options on monthly forwards

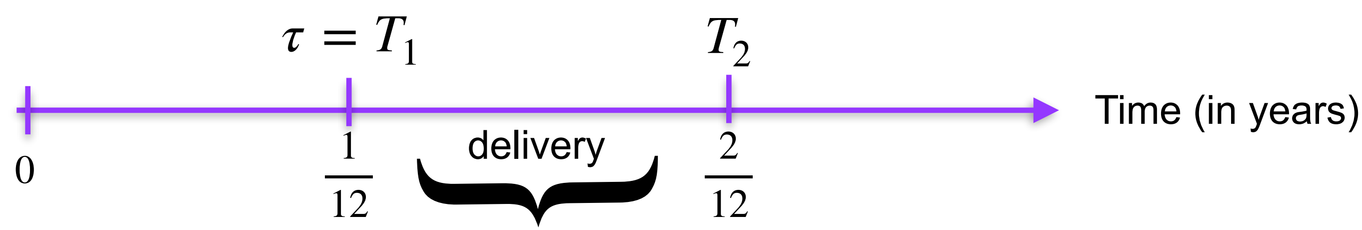
- Focus on Wiener case (no jumps) for X_τ, defined on $\mathfrak{X} = H_w$
  - Volatility is identity operator $\eta = \mathrm{Id}$
  - H_w Hilbert space of absolutely continuous functions on $\mathbb{R}_+$ with weak derivative decaying to zero faster than some monotonly increasing function w
  - Filipovic space, "weighted Sobolev space".
  - Separable and well-suited for forward dynamics

- Neural network may exploit information from basis functions



Pricing calls….

- Delivery next month, and strike in one month at strike price 1



- Train a neural network for $x \mapsto V(0,x)$, where $x \in K$ compact

- Training using *simulated* data
  - μ probability distribution on K (we use uniform distribution)
  - Draw M samples $(x^{(m)}, z^{(m)})_{m=1}^{M}$, $x^{(m)} \sim \mu$ and $z^{(m)} \sim Z_{0,\tau}$

$$\inf_{N,\ell_i,A_i,b_i} \frac{1}{M} \sum_{m=1}^{M} \left| \mathcal{P}((\mathcal{S}_\tau x^{(m)}) \exp(z^{(m)})) - \sum_{i=1}^{N} \mathcal{N}_{\ell_i,A_i,b_i}(x^{(m)}) \right|^2$$

Choice of activation function

- Consider Lipschitz continuous function $\beta : \mathbb{R} \to \mathbb{R}$

$$\lim_{y\to\infty} \beta(y) = 1, \lim_{y\to-\infty} \beta(y) = 0, \quad \beta(0) = 0$$

- We use

$$\beta(y) = \max\{0,1 - \exp(-y)\}$$

- Let $\psi \in H_w^* \setminus \{0\}$ and $z \in H_w, z \neq 0$, and define

$$\sigma(x) := \beta(\psi(x))z$$

- σ is Lipschitz, **von Neumann-bounded** and **separating**!

More specifications in training

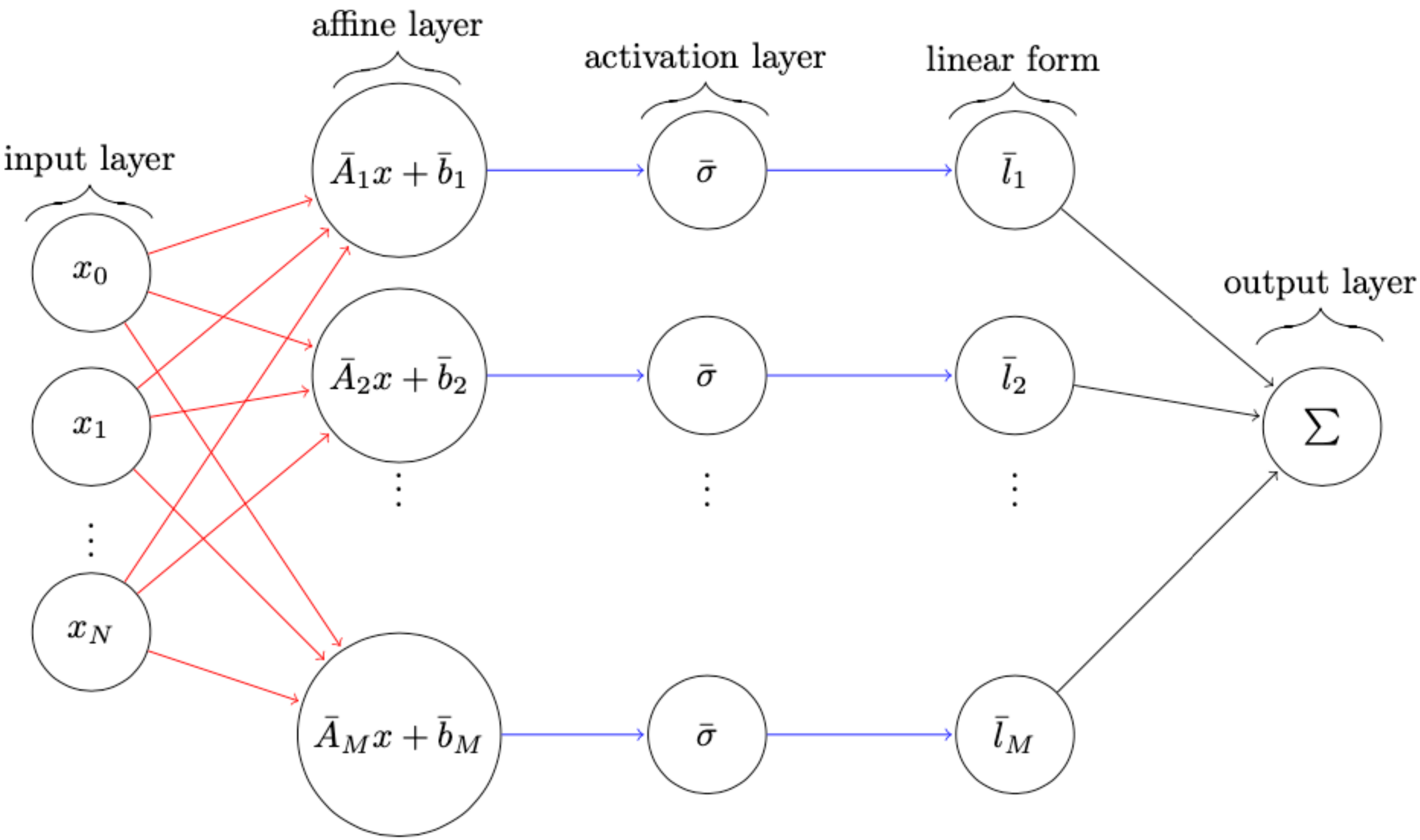
- Compact set K:

$$K_O := \left\{ \sum_{k=1}^{O} x_k e_k \ : \ x_k \in [-r, r] \right\}$$

- Training is done for a finite-dimensional projection of the neural network

- Trained several networks with hidden nodes (neurons) ranging from 1 to 30
    - Stochastic gradient descent in training
    - M=10 million Monte Carlo samples
    - Implemented in Python using TensorFlow and Keras libraries

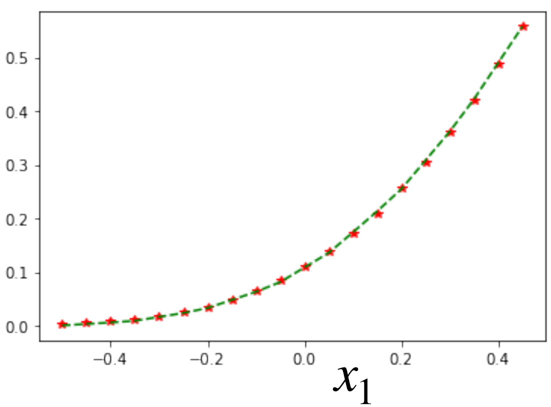
Neural network architecture



Validation

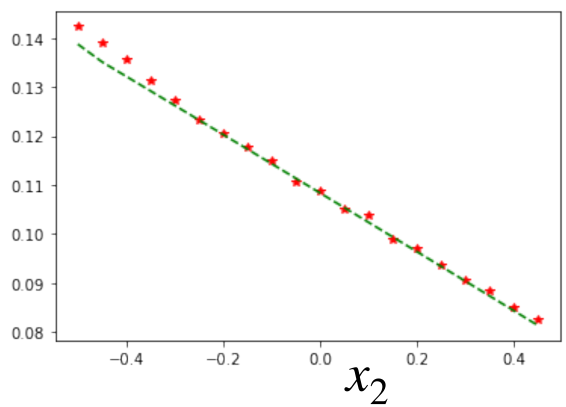
- Validated on 10.000 initial curves x with corresponding “true” option prices
  - x’s are randomly drawn from μ
  - “True” option prices calculated by Monte Carlo simulations, using 100.000 samples

Option price as a function of initial curve



Level, first coordinate

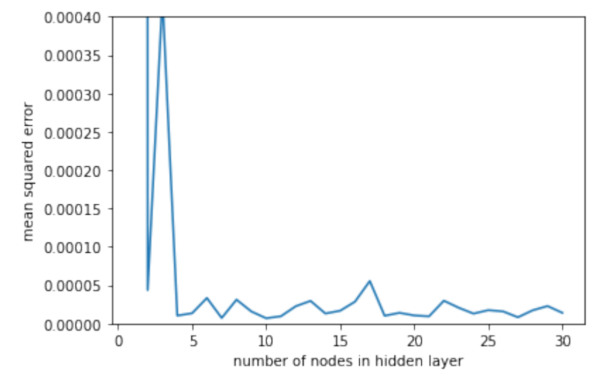
$x_2 = x_3 = \ldots = x_7 = 0$



Slope, second coordinate

$x_1 = x_3 = \ldots = x_7 = 0$

MSE

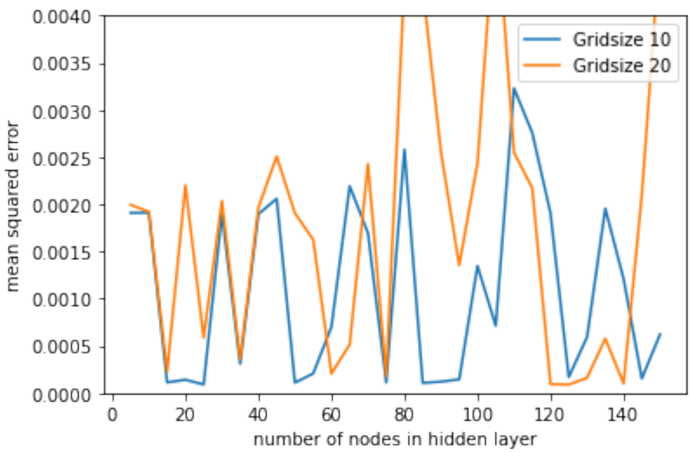


Nodes

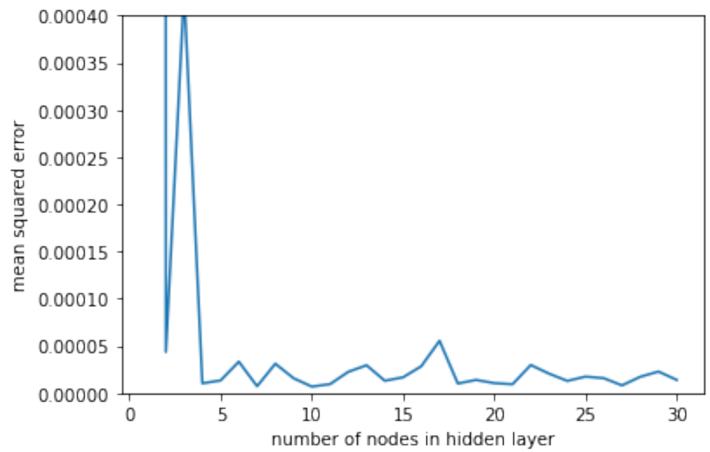
Comparison

- Comparing with classical neural network
  - Samples the initial forward curve $\xi \mapsto x(\xi)$
- Use networks with similar complexity
  - 10 and 20 discrete sample points of curves
  - 60-1800 parameters for 5-50 nodes in classical network, vs. 63-1890 parameters for 1-30 nodes in Hilbert network

MSE classical



MSE Hilbert



References

- Benth, Detering, Galimberti: Neural networks in Frechet spaces. *Arxiv: 2109.13512*

- Benth, Detering, Galimberti: Pricing options on flow forward by neural networks in Hilbert space. *Arxiv: 2202.11606*