BATUHAN GÜLER, ALEXIS LAIGNELET, PANOS PARPAS

DEPARTMENT OF COMPUTING
IMPERIAL COLLEGE LONDON

Imperial College London

TOWARDS ROBUST AND STABLE DEEP LEARNING ALGORITHMS FOR FORWARD-BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS

arXiv:1910.11623

PARTIAL DIFFERENTIAL EQUATIONS

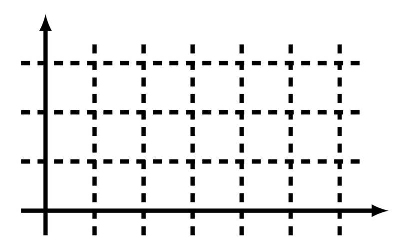
High dimensional PDEs in physics and mathematics:

- mechanics (**Hamilton-Jacobi**, **Schrödinger**, **Navier-Stokes**),
- electromagnetism (**Maxwell**),
- financial mathematics (**Black-Scholes**):
  - derivative pricing,
  - optimal trade execution,
  - risk management of options,
  - optimal asset allocation.

NUMERICAL METHODS

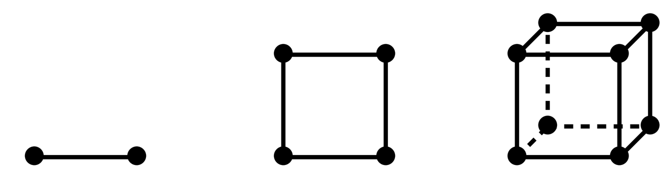
Finite differences

Discretization in space and time.



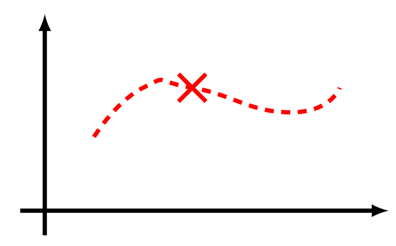
Curse of dimensionnality

$N \sim \mathcal{O}(\epsilon^{-d})$ for $|f - f^N| < \epsilon$



Monte Carlo

Only point evaluation.



DEEP BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS

1. PDEs
2. FBSDEs
3. Discretization
4. Neural network
5. Loss function
6. Results
7. Open problems

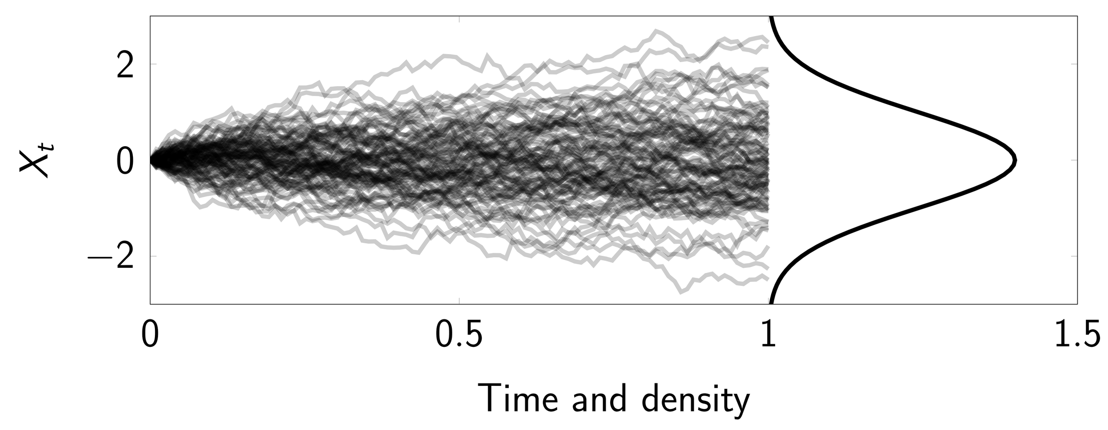
1: Nonlinear PDEs

$$u_t = \varphi(t, x, u, \nabla u) - \mu(t, x, u, \nabla u)^T \nabla u - \frac{1}{2} \mathrm{Tr}\left[\sigma(t, x, u)\sigma(t, x, u)^T \nabla^2 u\right]$$

2: Forward-Backward SDEs

$$\begin{aligned} dX_t &= \mu(t, X_t, Y_t, Z_t)dt + \sigma(t, X_t, Y_t)dW_t, & X_0 &= \xi \\ dY_t &= \varphi(t, X_t, Y_t, Z_t)dt + Z_t^T \sigma(t, X_t, Y_t)dW_t, & Y_T &= g(X_T) \end{aligned}$$

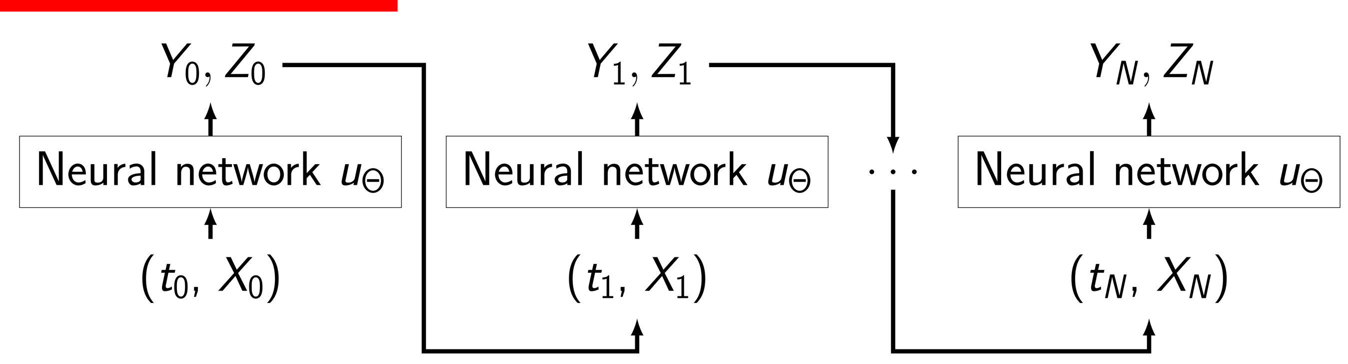
Equivalent to the above PDE, with $Y_t = u(t, X_t)$ and $Z_t = \nabla u(t, X_t)$.



3: Discretization in time

$$\begin{aligned} X_{n+1} &\approx X_n + \mu(t_n, X_n, Y_n, Z_n)\Delta t_n + \sigma(t_n, X_n, Y_n)\Delta W_n \\ Y_{n+1} &\approx Y_n + \varphi(t_n, X_n, Y_n, Z_n)\Delta t_n + Z_n^T \sigma(t_n, X_n, Y_n)\Delta W_n \end{aligned}$$

4: Neural network



- Θ: shared parameters through time,
- Z_n: computed using auto-differentiation.

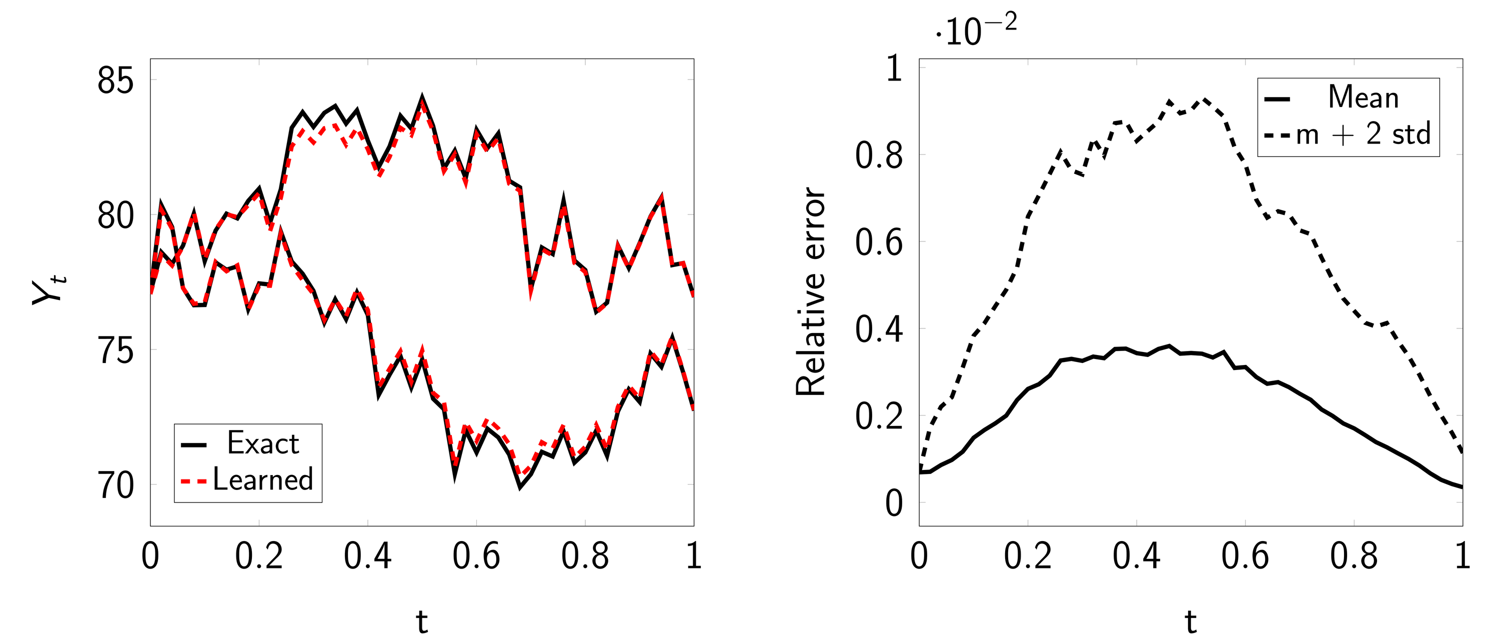
5: Loss function - Minimize approximation error

$$\min_{\Theta} \sum_{m=1}^{M} \sum_{n=0}^{N-1} |Y_{n+1}^m(\Theta) - Y_n^m(\Theta) - \varphi(t_n, X_n^m, Y_n^m(\Theta), Z_n^m(\Theta))\Delta t_n - (Z_n^m(\Theta))^T \sigma(t_n, X_n^m, Y_n^m(\Theta))\Delta W_n^m|^2 + \sum_{m=1}^{M} |Y_N^m(\Theta) - g(X_N^m)|^2$$

- M: number of trajectories (batch size),
- N: number of time steps.

6: Results for Black-Scholes equation in 100 dimensions

Eg: $g(x) = \|x\|^2$ leading to the closed-form solution $u(x, t) = e^{(r+\sigma^2)(T-t)}\|x\|^2$.



7: Open problems

- **Generalisation**: is approximation sensitive to initial value?
- **Stability parameters**: are same parameters close to optimal?
- **Computational efficiency**: how to speed-up the process?

ARCHITECTURE

ResNet

Defined by:

$$x(k+1) = x(k) + f(x(k), \theta(k))$$

- $x(k)$: output of the k^{th} layer,
- $\theta(k)$: parameters of the k^{th} layer,
- f: a nonlinear transformation.

NAIS-Net

A non-autonomous input-output stable neural network:

$$x(k+1) = x(k) + h\sigma(Ax(k) + Bu + C)$$

- A, B, C: trainable parameters,
- u: makes the system non-autonomous, and the output input-dependent.

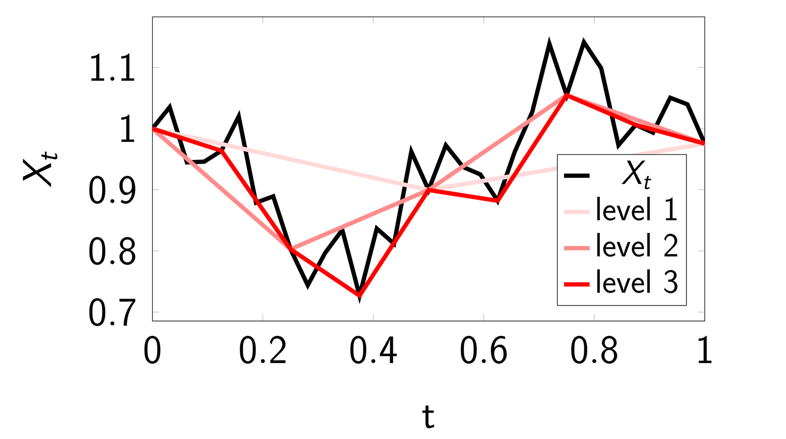
COMPUTATIONAL EFFICIENCY: MULTILEVEL

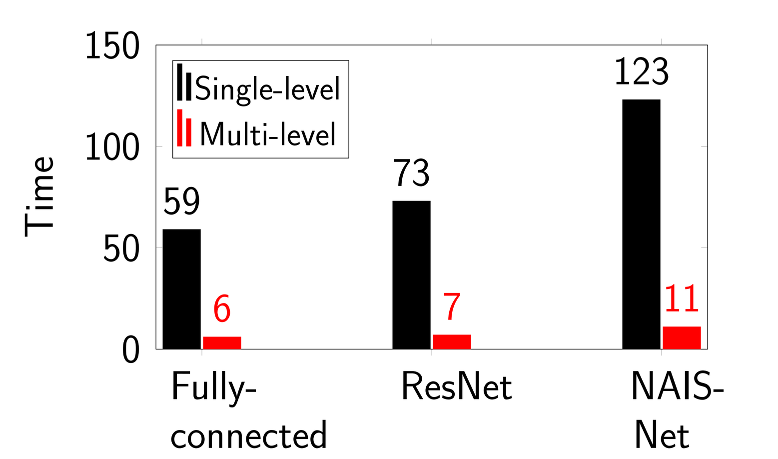
The discretization scheme, at level l, is:

$$X_{n+1} = X_n + a(t_n, X_n)h + b(t_n, X_n)\Delta W_n$$

where $h_l = h_0 M^{-l}$ with h_0 the initial step size and M the ratio between two levels.

Results: The convergence is on average 10x faster.

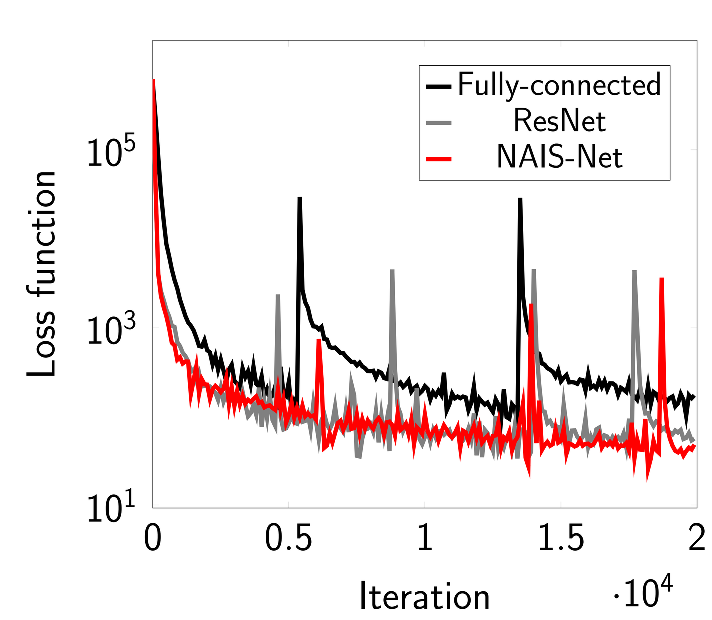




STABILITY & GENERALISATION

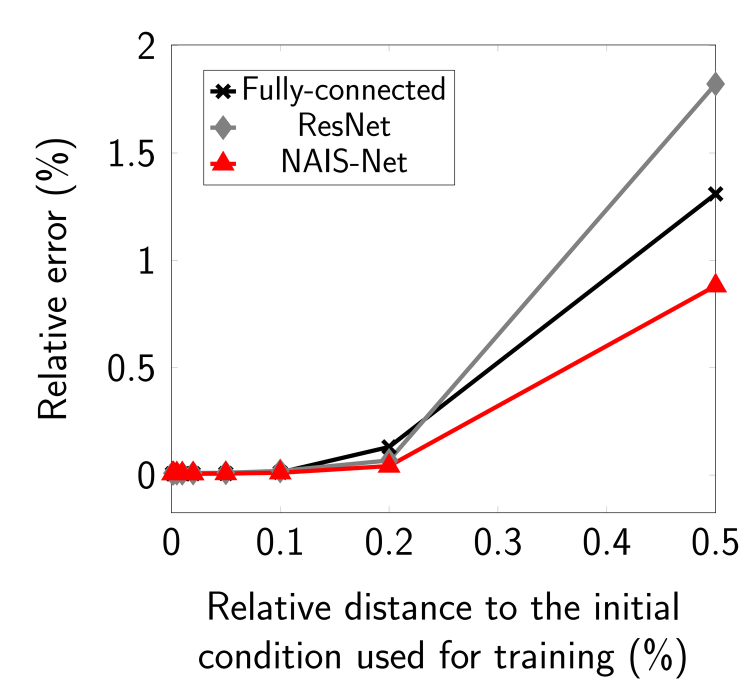
Smoother loss functions

Results: Loss functions are smoother for ResNet and NAIS-Net.



Generalisation

Results: Improved generalisation with NAIS-Net.



REFERENCES

- Raissi, M. (2018). Forward-backward stochastic neural networks: Deep learning of high-dimensional partial differential equations. *arXiv preprint arXiv:1804.07010.*
- Henry-Labordere, P. (2017). Deep primal-dual algorithm for BSDEs: Applications of machine learning to CVA and IM. *Available at SSRN 3071506.*
- Gobet, E. (2016). *Monte-Carlo methods and stochastic processes: from linear to non-linear.* Chapman and Hall/CRC.

This research was funded in part by JPMorgan Chase & Co. Any views or opinions expressed herein are solely those of the authors listed, and may differ from the views and opinions expressed by JPMorgan Chase & Co. or its affiliates. This material is not a product of the Research Department of J.P. Morgan Securities LLC. This material does not constitute a solicitation or offer in any jurisdiction. The work of the authors was partly funded by Engineering & Physical Sciences Research Council grants EP/M028240/1.