Computational Methods for High-Dimensional PDEs

Christoph Belak

Technische Universität Berlin

DGVFM Workshop for Young Mathematicians, Günzburg

November 29, 2022

Overview

(1) Partial Differential Equations in Finance

(2) Classical Computational Methods

(3) Neural Networks and Deep Learning

(4) Derivation of the Algorithm

(5) Main Convergence Results

Partial Differential Equations in Finance

▷ **Financial Market Model**: zero interest rate, (discounted) risky asset prices given by

$$\mathrm{d}S_t = \mu(S_t)\mathrm{d}t + \sigma(S_t)\mathrm{d}W_t, \qquad t \in [0, T].$$

E.g. Black-Scholes if $\mu(S_t) = \bar{\mu} S_t$ and $\sigma(S_t) = \bar{\sigma} S_t$.

▷ **Financial Market Model**: zero interest rate, (discounted) risky asset prices given by

$$\mathrm{d}S_t = \mu(S_t)\mathrm{d}t + \sigma(S_t)\mathrm{d}W_t, \qquad t \in [0, T].$$

E.g. Black-Scholes if $\mu(S_t) = \bar{\mu} S_t$ and $\sigma(S_t) = \bar{\sigma} S_t$.

▷ **European Option**: payoff given by

$$g(S_T).$$

E.g. put option if $g(S_T) = (K - S_T)^+$.

▷ **Financial Market Model**: zero interest rate, (discounted) risky asset prices given by

$$\mathrm{d}S_t = \mu(S_t)\mathrm{d}t + \sigma(S_t)\mathrm{d}W_t, \qquad t \in [0, T].$$

E.g. Black-Scholes if $\mu(S_t) = \bar{\mu} S_t$ and $\sigma(S_t) = \bar{\sigma} S_t$.

▷ **European Option**: payoff given by

$$g(S_T).$$

E.g. put option if $g(S_T) = (K - S_T)^+$.

▷ **Pricing Problem**: what is a fair price of this option?

▷ **Risk Neutral Pricing**: an arbitrage-free price is given by

$$\mathrm{E}^{\mathbb{Q}}\big[g(S_T)\big]$$

where $\mathbb{Q}$ is an equivalent martingale measure. Risk neutral pricing also works for more general payoffs/markets.

▷ **Risk Neutral Pricing**: an arbitrage-free price is given by

$$\mathrm{E}^{\mathbb{Q}}\big[g(S_T)\big]$$

where $\mathbb{Q}$ is an equivalent martingale measure. Risk neutral pricing also works for more general payoffs/markets.

▷ **PDE Pricing**: an arbitrage-free price is given by $u(0, S_0)$, where u solves the linear PDE

$$\begin{aligned} u_t(t,s) + \frac{1}{2}\operatorname{tr}\big[\sigma(s)\sigma(s)^\top \mathrm{D}_s^2 u(t,s)\big] &= 0, \\ u(T,s) &= g(s). \end{aligned}$$

▷ **Risk Neutral Pricing**: an arbitrage-free price is given by

$$\mathrm{E}^{\mathbb{Q}}\big[g(S_T)\big]$$

where $\mathbb{Q}$ is an equivalent martingale measure. Risk neutral pricing also works for more general payoffs/markets.

▷ **PDE Pricing**: an arbitrage-free price is given by $u(0, S_0)$, where u solves the linear PDE

$$\begin{aligned} u_t(t,s) + \frac{1}{2}\operatorname{tr}\big[\sigma(s)\sigma(s)^\top \mathrm{D}_s^2 u(t,s)\big] &= 0, \\ u(T,s) &= g(s). \end{aligned}$$

PDE pricing yields a **hedging strategy** given by

$$\mathrm{D}_s u(t, S_t), \qquad t \in [0,T].$$

Classical Computational Methods

Finite Difference Approximation:

- Approximates the PDE solution on a grid $\{t_0, \ldots, t_M\} \times \{s_1, \ldots, s_N\}^d$.

Finite Difference Approximation:

- Approximates the PDE solution on a grid $\{t_0, \ldots, t_M\} \times \{s_1, \ldots, s_N\}^d$.
- Replace derivatives by finite difference approximations, that is, e.g.,

$$u_t(t_m, s_n) \approx \frac{u(t_{m+1}, s_n) - u(t_m, s_n)}{t_{m+1} - t_m}.$$

Finite Difference Approximation:

- Approximates the PDE solution on a grid $\{t_0, \ldots, t_M\} \times \{s_1, \ldots, s_N\}^d$.
- Replace derivatives by finite difference approximations, that is, e.g.,

$$u_t(t_m, s_n) \approx \frac{u(t_{m+1}, s_n) - u(t_m, s_n)}{t_{m+1} - t_m}.$$

- This results in a system of linear equations with $M \times N^d$ unknowns.

Finite Difference Approximation:

- Approximates the PDE solution on a grid $\{t_0, \ldots, t_M\} \times \{s_1, \ldots, s_N\}^d$.
- Replace derivatives by finite difference approximations, that is, e.g.,

$$u_t(t_m, s_n) \approx \frac{u(t_{m+1}, s_n) - u(t_m, s_n)}{t_{m+1} - t_m}.$$

- This results in a system of linear equations with $M \times N^d$ unknowns.

Monte Carlo Integration:

- Simulate independent copies $S_T^1, \ldots, S_T^L$ of S_T under $\mathbb{Q}$.

Finite Difference Approximation:

- Approximates the PDE solution on a grid $\{t_0, \ldots, t_M\} \times \{s_1, \ldots, s_N\}^d$.
- Replace derivatives by finite difference approximations, that is, e.g.,

$$u_t(t_m, s_n) \approx \frac{u(t_{m+1}, s_n) - u(t_m, s_n)}{t_{m+1} - t_m}.$$

- This results in a system of linear equations with $M \times N^d$ unknowns.

Monte Carlo Integration:

- Simulate independent copies $S_T^1, \ldots, S_T^L$ of S_T under $\mathbb{Q}$.
- Compute

$$\mathrm{E}^{\mathbb{Q}}\big[g(S_T)\big] \approx \frac{1}{L} \sum_{\ell=1}^{L} g\big(S_T^\ell\big)$$

Finite Difference Approximation:

- Approximates the PDE solution on a grid $\{t_0, \dots, t_M\} \times \{s_1, \dots, s_N\}^d$.
- Replace derivatives by finite difference approximations, that is, e.g.,

$$u_t(t_m, s_n) \approx \frac{u(t_{m+1}, s_n) - u(t_m, s_n)}{t_{m+1} - t_m}.$$

- This results in a system of linear equations with $M \times N^d$ unknowns.

Monte Carlo Integration:

- Simulate independent copies $S_T^1, \dots, S_T^L$ of S_T under $\mathbb{Q}$.
- Compute

$$\mathrm{E}^{\mathbb{Q}}\big[g(S_T)\big] \approx \frac{1}{L} \sum_{\ell=1}^{L} g\big(S_T^\ell\big)$$

- The error is measured in terms of the empirical variance

$$\mathrm{Var}^{\mathbb{Q}}\Big[\frac{1}{L} \sum_{\ell=1}^{L} g\big(S_T^\ell\big)\Big] = \frac{1}{L} \mathrm{Var}^{\mathbb{Q}}[g(S_T)].$$

Observations:

- Finite Differences suffer from the **curse of dimensionality**. The error in the Monte Carlo method depends on the dimension at most through $\mathrm{Var}^{\mathbb{Q}}[g(S_T)]$.

Observations:

- Finite Differences suffer from the **curse of dimensionality**. The error in the Monte Carlo method depends on the dimension at most through $\mathrm{Var}^{\mathbb{Q}}[g(S_T)]$.

- Monte Carlo computes the solution of a PDE in one particular point, namely in $(0, S_0)$. Finite Differences compute the solution of a PDE on an entire grid of points. A fair comparison is hence barely possible.

Observations:

- Finite Differences suffer from the **curse of dimensionality**. The error in the Monte Carlo method depends on the dimension at most through $\mathrm{Var}^{\mathbb{Q}}[g(S_T)]$.

- Monte Carlo computes the solution of a PDE in one particular point, namely in $(0, S_0)$. Finite Differences compute the solution of a PDE on an entire grid of points. A fair comparison is hence barely possible.

- The big advantage of Finite Differences is that they also yield an approximation of a **hedging strategy**.

Observations:

- Finite Differences suffer from the **curse of dimensionality**. The error in the Monte Carlo method depends on the dimension at most through $\mathrm{Var}^{\mathbb{Q}}[g(S_T)]$.

- Monte Carlo computes the solution of a PDE in one particular point, namely in $(0, S_0)$. Finite Differences compute the solution of a PDE on an entire grid of points. A fair comparison is hence barely possible.

- The big advantage of Finite Differences is that they also yield an approximation of a **hedging strategy**.

- Which method is more suitable in practice depends on the situation. There are cases, however, in which both methods have a tough time.

Pricing with **Systemically Important Counterparties**

- We wish to buy an equally weighted basket put option on the S&P500, that is

$$g(S_T) = \left(K - \frac{1}{500}\sum_{i=1}^{500} S_T^i, 0\right)^+.$$

Pricing with **Systemically Important Counterparties**

- We wish to buy an equally weighted basket put option on the S&P500, that is

$$g(S_T) = \left(K - \frac{1}{500}\sum_{i=1}^{500} S_T^i, 0\right)^+.$$

- We buy this option from a systemically important bank (SIB) which is subject to **default risk**.

Pricing with **Systemically Important Counterparties**

- We wish to buy an equally weighted basket put option on the S&P500, that is

$$g(S_T) = \left(K - \frac{1}{500}\sum_{i=1}^{500} S_T^i, 0\right)^+.$$

- We buy this option from a systemically important bank (SIB) which is subject to **default risk**.
- If the bank defaults, the prices of the underlyings experience a shock which results in price jump/crash.

Pricing with **Systemically Important Counterparties**

- We wish to buy an equally weighted basket put option on the S&P500, that is

$$g(S_T) = \left(K - \frac{1}{500}\sum_{i=1}^{500} S_T^i, 0\right)^+.$$

- We buy this option from a systemically important bank (SIB) which is subject to **default risk**.
- If the bank defaults, the prices of the underlyings experience a shock which results in price jump/crash.
- If τ denotes the time of default, the risk neutral price at time t is

$$V_t = \mathrm{E}^{\mathbb{Q}}\Big[g(S_T)\mathbb{1}_{\{\tau>T\}} + h(V_\tau)\mathbb{1}_{\{\tau\leq T\}}\Big].$$

Pricing with **Systemically Important Counterparties**

- We wish to buy an equally weighted basket put option on the S&P500, that is

$$g(S_T) = \left(K - \frac{1}{500}\sum_{i=1}^{500} S_T^i, 0\right)^+.$$

- We buy this option from a systemically important bank (SIB) which is subject to **default risk**.
- If the bank defaults, the prices of the underlyings experience a shock which results in price jump/crash.
- If τ denotes the time of default, the risk neutral price at time t is

$$V_t = \mathrm{E}^{\mathbb{Q}}\Big[g(S_T)\mathbb{1}_{\{\tau>T\}} + h(V_\tau)\mathbb{1}_{\{\tau\leq T\}}\Big].$$

- Here, $h(v) = Rv^+ - v^-$, $R \in [0,1]$, is the recovery value in case of a default and V_τ represents the time-τ price of an identical option with another SIB counterparty that has not defaulted, immediately after the default of the original SIB. In particular, V_τ is based on post-default risky asset prices.

Pricing with **Systemically Important Counterparties**

- We wish to buy an equally weighted basket put option on the S&P500, that is

$$g(S_T) = \left(K - \frac{1}{500}\sum_{i=1}^{500} S_T^i, 0\right)^+.$$

- We buy this option from a systemically important bank (SIB) which is subject to **default risk**.
- If the bank defaults, the prices of the underlyings experience a shock which results in price jump/crash.
- If τ denotes the time of default, the risk neutral price at time t is

$$V_t = \mathrm{E}^{\mathbb{Q}}\Big[g(S_T)\mathbb{1}_{\{\tau>T\}} + h(V_\tau)\mathbb{1}_{\{\tau\le T\}}\Big].$$

- Here, $h(v) = Rv^+ - v^-$, $R \in [0,1]$, is the recovery value in case of a default and V_τ represents the time-τ price of an identical option with another SIB counterparty that has not defaulted, immediately after the default of the original SIB. In particular, V_τ is based on post-default risky asset prices.
- The pricing PDE is no longer linear, but in fact semilinear and nonlocal.

Neural Networks and Deep Learning

▷ We now focus on a general **semilinear PDE** of the form

$$\begin{aligned} u_t + \mathcal{A}[u] + f(\cdot, u, \sigma^\top \mathrm{D}_x u) &= 0 && \text{on } [0,T) \times \mathbb{R}^d, \\ u(T,\cdot) &= g && \text{on } \mathbb{R}^d, \end{aligned}$$

▷ We now focus on a general **semilinear PDE** of the form

$$\begin{aligned} u_t + \mathcal{A}[u] + f(\cdot, u, \sigma^\top \mathrm{D}_x u) &= 0 && \text{on } [0,T) \times \mathbb{R}^d, \\ u(T,\cdot) &= g && \text{on } \mathbb{R}^d, \end{aligned}$$

where $\mathcal{A}$ is the second-order linear **differential operator** given by

$$\mathcal{A}[u] = \mu^\top \mathrm{D}_x u + \frac{1}{2} \operatorname{tr}\big[\sigma\sigma^\top \mathrm{D}_x^2 u\big].$$

▷ We now focus on a general **semilinear PDE** of the form

$$\begin{aligned} u_t + \mathcal{A}[u] + f(\cdot, u, \sigma^\top \mathrm{D}_x u) &= 0 && \text{on } [0,T) \times \mathbb{R}^d, \\ u(T, \cdot) &= g && \text{on } \mathbb{R}^d, \end{aligned}$$

where $\mathcal{A}$ is the second-order linear **differential operator** given by

$$\mathcal{A}[u] = \mu^\top \mathrm{D}_x u + \frac{1}{2} \operatorname{tr}\left[\sigma\sigma^\top \mathrm{D}_x^2 u\right].$$

▷ Our aim is to find a **good approximation** U of the solution u of the PDE, that is

$$U \approx u \qquad \text{on a subset of } [0,T) \times \mathbb{R}^d.$$

U is chosen from a parametric class $\mathcal{U}$ of approximating functions.

Feedforward Neural Network

We refer to the function

$$F : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}, \quad x \mapsto F(x) \triangleq \Sigma_L \circ W_L \circ \Sigma_{L-1} \circ W_{L-1} \circ \cdots \Sigma_1 \circ W_1(x)$$

as a **feedforward neural network**.

Feedforward Neural Network

Let $L, N_0, N_1, \ldots, N_L \in \mathbb{N}$. For $\ell = 1, \ldots, L$, let us fix $A^\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$, $b^\ell \in \mathbb{R}^{N_\ell}$, as well as a function $\sigma_\ell : \mathbb{R} \to \mathbb{R}$. We set

$$\Sigma_\ell : \mathbb{R}^{N_\ell} \to \mathbb{R}^{N_\ell}, \qquad x \mapsto \Sigma_\ell(x) \triangleq \big(\sigma_\ell(x_1), \ldots, \sigma_\ell(x_{N_\ell})\big)^\top$$

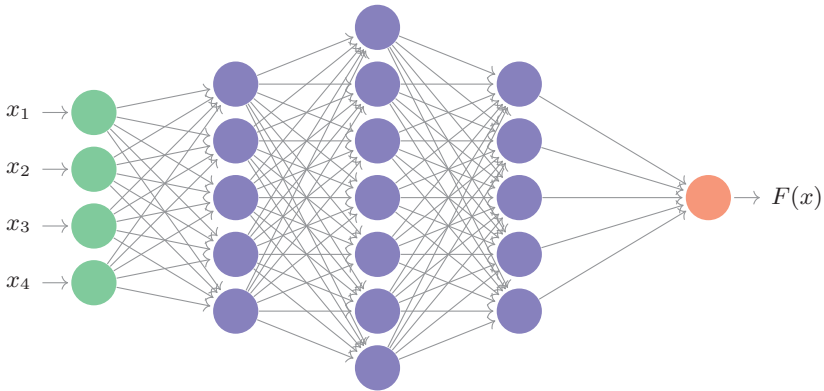
for the component-wise application of σ_ℓ on a vector $x \in \mathbb{R}^{N_\ell}$ and define

$$W_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}, \qquad x \mapsto W_\ell(x) \triangleq A^\ell x + b^\ell.$$

We refer to the function

$$F : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}, \quad x \mapsto F(x) \triangleq \Sigma_L \circ W_L \circ \Sigma_{L-1} \circ W_{L-1} \circ \cdots \Sigma_1 \circ W_1(x)$$

as a **feedforward neural network**.


Input Layer ($d = 4$)
First Hidden Layer ($N_1 = 5$)
Second Hidden Layer ($N_2 = 7$)
Third Hidden Layer ($N_3 = 5$)
Output Layer
x_1
x_2
x_3
x_4
$F(x)$

▷ It is common to choose $\sigma \triangleq \sigma_1 = \cdots = \sigma_L$. A popular choice for σ is

$$\sigma(x) \triangleq x^+, \qquad x \in \mathbb{R},$$

which the machine learning community refers to as the **rectified linear unit** (ReLU).

▷ It is common to choose $\sigma \triangleq \sigma_1 = \cdots = \sigma_L$. A popular choice for σ is

$$\sigma(x) \triangleq x^+, \qquad x \in \mathbb{R},$$

which the machine learning community refers to as the **rectified linear unit** (ReLU).

▷ Neural Networks are **universal approximators**. That is, under mild assumptions on σ, they can approximate any continuous function arbitrarily well on compacts.

▷ It is common to choose $\sigma \triangleq \sigma_1 = \cdots = \sigma_L$. A popular choice for σ is

$$\sigma(x) \triangleq x^+, \qquad x \in \mathbb{R},$$

which the machine learning community refers to as the **rectified linear unit** (ReLU).

▷ Neural Networks are **universal approximators**. That is, under mild assumptions on σ, they can approximate any continuous function arbitrarily well on compacts.

▷ The **parameters** of the neural network are the number of layers L, the number of neurons $N_1, \ldots, N_{L-1}$ per hidden layer, and the affine transformations given in terms of the matrices $A^1, \ldots, A^L$ and the vectors $b^1, \ldots, b^L$.

▷ It is common to choose $\sigma \triangleq \sigma_1 = \cdots = \sigma_L$. A popular choice for σ is

$$\sigma(x) \triangleq x^+, \qquad x \in \mathbb{R},$$

which the machine learning community refers to as the **rectified linear unit** (ReLU).

▷ Neural Networks are **universal approximators**. That is, under mild assumptions on σ, they can approximate any continuous function arbitrarily well on compacts.

▷ The **parameters** of the neural network are the number of layers L, the number of neurons $N_1, \ldots, N_{L-1}$ per hidden layer, and the affine transformations given in terms of the matrices $A^1, \ldots, A^L$ and the vectors $b^1, \ldots, b^L$.

▷ There are efficient algorithms to optimize the parameters of neural networks based on **stochastic gradient descent** and **backpropagation**.

▷ It is common to choose $\sigma \triangleq \sigma_1 = \cdots = \sigma_L$. A popular choice for σ is

$$\sigma(x) \triangleq x^+, \qquad x \in \mathbb{R},$$

which the machine learning community refers to as the **rectified linear unit** (ReLU).

▷ Neural Networks are **universal approximators**. That is, under mild assumptions on σ, they can approximate any continuous function arbitrarily well on compacts.

▷ The **parameters** of the neural network are the number of layers L, the number of neurons $N_1, \ldots, N_{L-1}$ per hidden layer, and the affine transformations given in terms of the matrices $A^1, \ldots, A^L$ and the vectors $b^1, \ldots, b^L$.

▷ There are efficient algorithms to optimize the parameters of neural networks based on **stochastic gradient descent** and **backpropagation**.

▷ We want to approximate the solution of our PDE by a feedforward neural network. How can we do this?

Derivation of the Algorithm

Let us have another look at the **partial differential equation** given by

$$\begin{aligned} u_t + \mathcal{A}[u] + f(\cdot, u, \sigma^\top \mathrm{D}_x u) &= 0 && \text{on } [0,T) \times \mathbb{R}^d, \\ u(T, \cdot) &= g && \text{on } \mathbb{R}^d, \end{aligned}$$

and remind ourselves that $\mathcal{A}$ is defined as

$$\mathcal{A}[u] = \mu^\top \mathrm{D}_x u + \frac{1}{2} \operatorname{tr}\left[\sigma\sigma^\top \mathrm{D}_x^2 u\right].$$

Let us have another look at the **partial differential equation** given by

$$\begin{aligned} u_t + \mathcal{A}[u] + f(\cdot, u, \sigma^\top \mathrm{D}_x u) &= 0 && \text{on } [0,T) \times \mathbb{R}^d, \\ u(T, \cdot) &= g && \text{on } \mathbb{R}^d, \end{aligned}$$

and remind ourselves that $\mathcal{A}$ is defined as

$$\mathcal{A}[u] = \mu^\top \mathrm{D}_x u + \frac{1}{2} \operatorname{tr}\left[\sigma\sigma^\top \mathrm{D}_x^2 u\right].$$

The operator $\mathcal{A}$ is the **infinitesimal generator** of a diffusion with dynamics

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t.$$

Suppose that u is a **classical solution** of the PDE. By Itô's formula, we obtain

$$u(t, X_t) = u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s)\mathrm{d}s \\ - \int_t^T \mathrm{D}_x u(s, X_s)^\top \sigma(s, X_s)\mathrm{d}W_s$$

Suppose that u is a **classical solution** of the PDE. By Itô's formula, we obtain

$$\begin{aligned} u(t, X_t) &= u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s)\mathrm{d}s \\ &\qquad - \int_t^T \mathrm{D}_x u(s, X_s)^\top \sigma(s, X_s)\mathrm{d}W_s \\ &= g(X_T) + \int_t^T f\big(s, X_s, u(s, X_s), \sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s)\big)\mathrm{d}s \\ &\qquad - \int_t^T \big[\sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s)\big]^\top \mathrm{d}W_s. \end{aligned}$$

Suppose that u is a **classical solution** of the PDE. By Itô's formula, we obtain

$$\begin{aligned} u(t, X_t) &= u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s)\mathrm{d}s \\ &\qquad - \int_t^T \mathrm{D}_x u(s, X_s)^\top \sigma(s, X_s)\mathrm{d}W_s \\ &= g(X_T) + \int_t^T f\big(s, X_s, u(s, X_s), \sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s)\big)\mathrm{d}s \\ &\qquad - \int_t^T \big[\sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s)\big]^\top \mathrm{d}W_s. \end{aligned}$$

Now define processes

$$Y_s \triangleq u(s, X_s) \qquad \text{and} \qquad Z_s \triangleq \sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s),$$

Suppose that u is a **classical solution** of the PDE. By Itô's formula, we obtain

$$\begin{aligned} u(t, X_t) &= u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s)\mathrm{d}s \\ &\qquad - \int_t^T \mathrm{D}_x u(s, X_s)^\top \sigma(s, X_s)\mathrm{d}W_s \\ &= g(X_T) + \int_t^T f\big(s, X_s, u(s, X_s), \sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s)\big)\mathrm{d}s \\ &\qquad - \int_t^T \big[\sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s)\big]^\top \mathrm{d}W_s. \end{aligned}$$

Now define processes

$$Y_s \triangleq u(s, X_s) \qquad \text{and} \qquad Z_s \triangleq \sigma(s, X_s)^\top \mathrm{D}_x u(s, X_s),$$

and it follows that (Y, Z) is a solution of the **backward SDE**

$$Y_t = g(X_T) + \int_t^T f\big(s, X_s, Y_s, Z_s\big)\mathrm{d}s - \int_t^T Z_s^\top \mathrm{d}W_s, \qquad t \in [0, T].$$

For $I \in \mathbb{N}$, let $h = T/I$ denote the **step size** of the equidistant dissection

$$0 = t_0 < t_1 < \cdots < t_I = T$$

of $[0, T]$. Furthermore, write $\Delta W_{i+1} = W_{t_{i+1}} - W_{t_i}$ for the Brownian increments.

For $I \in \mathbb{N}$, let $h = T/I$ denote the **step size** of the equidistant dissection

$$0 = t_0 < t_1 < \cdots < t_I = T$$

of $[0, T]$. Furthermore, write $\Delta W_{i+1} = W_{t_{i+1}} - W_{t_i}$ for the Brownian increments.

▷ **Euler-Maruyama** discretization of the forward SDE:

$$\begin{aligned} \bar{X}_0 &\triangleq x, \\ \bar{X}_{i+1} &\triangleq \bar{X}_i + \mu(t_i, \bar{X}_i)h + \sigma(t_i, \bar{X}_i)\Delta W_{i+1}, \qquad i = 0, 1, \ldots, I-1. \end{aligned}$$

For $I \in \mathbb{N}$, let $h = T/I$ denote the **step size** of the equidistant dissection

$$0 = t_0 < t_1 < \cdots < t_I = T$$

of $[0, T]$. Furthermore, write $\Delta W_{i+1} = W_{t_{i+1}} - W_{t_i}$ for the Brownian increments.

▷ **Euler-Maruyama** discretization of the forward SDE:

$$\begin{aligned} \bar{X}_0 &\triangleq x, \\ \bar{X}_{i+1} &\triangleq \bar{X}_i + \mu(t_i, \bar{X}_i)h + \sigma(t_i, \bar{X}_i)\Delta W_{i+1}, \qquad i = 0, 1, \ldots, I-1. \end{aligned}$$

▷ **Naive Euler-Maruyama** discretization of the backward SDE:

$$\begin{aligned} \bar{Y}_I &= g(\bar{X}_I), \\ \bar{Y}_{i+1} &= \bar{Y}_i - f\big(t_i, \bar{X}_i, \bar{Y}_i, \bar{Z}_i\big)h + \bar{Z}_i^\top \Delta W_{i+1}, \qquad i = 0, 1, \ldots, I-1. \end{aligned}$$

▷ **Idea**: A function U satisfies $U \approx u$ if and only if $U(t_i, \bar{X}_i) \approx \bar{Y}_i$ since

$$\bar{Y}_i \approx Y_{t_i} = u(t_i, X_{t_i}) \approx u(t_i, \bar{X}_i)$$

▷ **Idea**: A function U satisfies $U \approx u$ if and only if $U(t_i, \bar{X}_i) \approx \bar{Y}_i$ since

$$\bar{Y}_i \approx Y_{t_i} = u(t_i, X_{t_i}) \approx u(t_i, \bar{X}_i) \approx U(t_i, \bar{X}_i).$$

▷ **Idea**: A function U satisfies $U \approx u$ if and only if $U(t_i, \bar{X}_i) \approx \bar{Y}_i$ since

$$U(t_i, \bar{X}_i) \approx \bar{Y}_i \approx Y_{t_i} = u(t_i, X_{t_i}) \approx u(t_i, \bar{X}_i)$$

▷ **Idea**: A function U satisfies $U \approx u$ if and only if $U(t_i, \bar{X}_i) \approx \bar{Y}_i$ since

$$\bar{Y}_i \approx Y_{t_i} = u(t_i, X_{t_i}) \approx u(t_i, \bar{X}_i)$$

Thus, if we want to approximate u by a function U, one possibility is to

Minimize the error in the Euler-Maruyama discretization!

▷ **Idea**: A function U satisfies $U \approx u$ if and only if $U(t_i, \bar{X}_i) \approx \bar{Y}_i$ since

$$\bar{Y}_i \approx Y_{t_i} = u(t_i, X_{t_i}) \approx u(t_i, \bar{X}_i)$$

Thus, if we want to approximate u by a function U, one possibility is to

Minimize the error in the Euler-Maruyama discretization!

This leads Raissi (2018) to propose the **loss functional**

$$\mathbb{L}^{\mathrm{R}}(U) \triangleq \mathrm{E}\Big[\big|U(T, \bar{X}_I) - g(\bar{X}_I)\big|^2\Big] + \sum_{i=0}^{I-1} \mathrm{E}\Big[\big|\mathcal{L}_i^{\mathrm{R}}(U)\big|^2\Big],$$

▷ **Idea**: A function U satisfies $U \approx u$ if and only if $U(t_i, \bar{X}_i) \approx \bar{Y}_i$ since

$$\bar{Y}_i \approx Y_{t_i} = u(t_i, X_{t_i}) \approx u(t_i, \bar{X}_i)$$

Thus, if we want to approximate u by a function U, one possibility is to

Minimize the error in the Euler-Maruyama discretization!

This leads Raissi (2018) to propose the **loss functional**

$$\mathbb{L}^{\mathrm{R}}(U) \triangleq \mathrm{E}\Big[\big|U(T, \bar{X}_I) - g(\bar{X}_I)\big|^2\Big] + \sum_{i=0}^{I-1} \mathrm{E}\Big[\big|\mathcal{L}_i^{\mathrm{R}}(U)\big|^2\Big],$$

where

$$\begin{aligned}\mathcal{L}_i^{\mathrm{R}}(U) \triangleq\ & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f\big(t_i, \bar{X}_i, U(t_i, \bar{X}_i), \sigma(t_i, \bar{X}_i)^\top \mathrm{D}_x U(t_i, \bar{X}_i)\big)h \\ & - \mathrm{D}_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i) \Delta W_{i+1}.\end{aligned}$$

▷ **Idea**: A function U satisfies $U \approx u$ if and only if $U(t_i, \bar{X}_i) \approx \bar{Y}_i$ since

$$\bar{Y}_i \approx Y_{t_i} = u(t_i, X_{t_i}) \approx u(t_i, \bar{X}_i)$$

Thus, if we want to approximate u by a function U, one possibility is to

Minimize the error in the Euler-Maruyama discretization!

This leads Raissi (2018) to propose the **loss functional**

$$\mathbb{L}^{\mathrm{R}}(U) \triangleq \mathrm{E}\Big[\big|U(T, \bar{X}_I) - g(\bar{X}_I)\big|^2\Big] + \sum_{i=0}^{I-1} \mathrm{E}\Big[\big|\mathcal{L}_i^{\mathrm{R}}(U)\big|^2\Big],$$

where

$$\begin{aligned}\mathcal{L}_i^{\mathrm{R}}(U) \triangleq\ & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f\big(t_i, \bar{X}_i, U(t_i, \bar{X}_i), \sigma(t_i, \bar{X}_i)^\top \mathrm{D}_x U(t_i, \bar{X}_i)\big)h \\ & - \mathrm{D}_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i) \Delta W_{i+1}.\end{aligned}$$

If we **minimize** $\mathbb{L}^{\mathrm{R}}(U)$ over a class $\mathcal{U}$ of neural networks, we hope to get a good approximation of u.

▷ It turns out that Raissi's loss functional does not converge fast enough for our purposes, so we also consider the **adjusted loss functional**

$$\mathbb{L}(U) \triangleq \mathrm{E}\Big[\big|U(T, \bar{X}_I) - g(\bar{X}_I)\big|^2\Big] + \sum_{i=0}^{I-1} \mathrm{E}\Big[\big|\mathcal{L}_i(U)\big|^2\Big],$$

▷ It turns out that Raissi's loss functional does not converge fast enough for our purposes, so we also consider the **adjusted loss functional**

$$\mathbb{L}(U) \triangleq \mathrm{E}\Big[\big|U(T, \bar{X}_I) - g(\bar{X}_I)\big|^2\Big] + \sum_{i=0}^{I-1} \mathrm{E}\Big[\big|\mathcal{L}_i(U)\big|^2\Big],$$

where

$$\begin{aligned}\mathcal{L}_i(U) \triangleq\ & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f\big(t_i, \bar{X}_i, U(t_{i+1}, \bar{X}_{i+1}), \sigma(t_i, \bar{X}_i)^\top \mathrm{D}_x U(t_i, \bar{X}_i)\big)h \\ & - \mathrm{D}_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i)\Delta W_{i+1}\end{aligned}$$

▷ It turns out that Raissi's loss functional does not converge fast enough for our purposes, so we also consider the **adjusted loss functional**

$$\mathbb{L}(U) \triangleq \mathrm{E}\Big[\big|U(T, \bar{X}_I) - g(\bar{X}_I)\big|^2\Big] + \sum_{i=0}^{I-1} \mathrm{E}\Big[\big|\mathcal{L}_i(U)\big|^2\Big],$$

where

$$\begin{aligned}\mathcal{L}_i(U) \triangleq\ & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f\big(t_i, \bar{X}_i, U(t_{i+1}, \bar{X}_{i+1}), \sigma(t_i, \bar{X}_i)^\top \mathrm{D}_x U(t_i, \bar{X}_i)\big)h \\ & - \mathrm{D}_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i)\Delta W_{i+1} \\ & - \int_{t_i}^{t_{i+1}} \Big[\mathrm{D}_x^2 U(t_i, \bar{X}_i)\big(\mu(t_i, \bar{X}_i)h + \sigma(t_i, \bar{X}_i)(W_t - W_{t_i})\big)\Big]^\top \sigma(t_i, \bar{X}_i)\mathrm{d}W_t.\end{aligned}$$

▷ It turns out that Raissi's loss functional does not converge fast enough for our purposes, so we also consider the **adjusted loss functional**

$$\mathbb{L}(U) \triangleq \mathrm{E}\Big[\big|U(T, \bar{X}_I) - g(\bar{X}_I)\big|^2\Big] + \sum_{i=0}^{I-1} \mathrm{E}\Big[\big|\mathcal{L}_i(U)\big|^2\Big],$$

where

$$\begin{aligned}\mathcal{L}_i(U) \triangleq\ & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f\big(t_i, \bar{X}_i, U(t_{i+1}, \bar{X}_{i+1}), \sigma(t_i, \bar{X}_i)^\top \mathrm{D}_x U(t_i, \bar{X}_i)\big)h \\ & - \mathrm{D}_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i)\Delta W_{i+1} \\ & - \int_{t_i}^{t_{i+1}} \Big[\mathrm{D}_x^2 U(t_i, \bar{X}_i)\big(\mu(t_i, \bar{X}_i)h + \sigma(t_i, \bar{X}_i)(W_t - W_{t_i})\big)\Big]^\top \sigma(t_i, \bar{X}_i)\mathrm{d}W_t.\end{aligned}$$

The correction term can be thought of as a consequence of a higher-order **Milstein discretization** instead of an Euler-Maruyama discretization of the backward SDE. It can be **simulated exactly**.

Main Convergence Results

We work under **standard assumptions** which guarantee that the PDE admits a unique solution u in the parabolic Hölder space $\mathcal{H}^{3,\gamma}$.

We work under **standard assumptions** which guarantee that the PDE admits a unique solution u in the parabolic Hölder space $\mathcal{H}^{3,\gamma}$. In particular,

- u, u_t, $\mathrm{D}_x u$, $\mathrm{D}_x^2 u$, $\mathrm{D}_x^3 u$, $\mathrm{D}_x u_t$ exist as continuous and bounded functions;

We work under **standard assumptions** which guarantee that the PDE admits a unique solution u in the parabolic Hölder space $\mathcal{H}^{3,\gamma}$. In particular,

- u, u_t, $\mathrm{D}_x u$, $\mathrm{D}_x^2 u$, $\mathrm{D}_x^3 u$, $\mathrm{D}_x u_t$ exist as continuous and bounded functions;
- u, $\mathrm{D}_x u$ are globally Lipschitz;

We work under **standard assumptions** which guarantee that the PDE admits a unique solution u in the parabolic Hölder space $\mathcal{H}^{3,\gamma}$. In particular,

- u, u_t, $\mathrm{D}_x u$, $\mathrm{D}_x^2 u$, $\mathrm{D}_x^3 u$, $\mathrm{D}_x u_t$ exist as continuous and bounded functions;
- u, $\mathrm{D}_x u$ are globally Lipschitz;
- $\mathrm{D}_x^2 u$ is Lipschitz in x, uniformly in t;

We work under **standard assumptions** which guarantee that the PDE admits a unique solution u in the parabolic Hölder space $\mathcal{H}^{3,\gamma}$. In particular,

- u, u_t, $\mathrm{D}_x u$, $\mathrm{D}_x^2 u$, $\mathrm{D}_x^3 u$, $\mathrm{D}_x u_t$ exist as continuous and bounded functions;
- u, $\mathrm{D}_x u$ are globally Lipschitz;
- $\mathrm{D}_x^2 u$ is Lipschitz in x, uniformly in t;
- u_t is $1/2$-Hölder in t, uniformly in x.

We work under **standard assumptions** which guarantee that the PDE admits a unique solution u in the parabolic Hölder space $\mathcal{H}^{3,\gamma}$. In particular,

- u, u_t, $\mathrm{D}_x u$, $\mathrm{D}_x^2 u$, $\mathrm{D}_x^3 u$, $\mathrm{D}_x u_t$ exist as continuous and bounded functions;
- u, $\mathrm{D}_x u$ are globally Lipschitz;
- $\mathrm{D}_x^2 u$ is Lipschitz in x, uniformly in t;
- u_t is $1/2$-Hölder in t, uniformly in x.

The precise assumptions can be found in the article and are derived from classical results; see Ladyzhenskaia et al. (1968).

We work under **standard assumptions** which guarantee that the PDE admits a unique solution u in the parabolic Hölder space $\mathcal{H}^{3,\gamma}$. In particular,

- u, u_t, $\mathrm{D}_x u$, $\mathrm{D}_x^2 u$, $\mathrm{D}_x^3 u$, $\mathrm{D}_x u_t$ exist as continuous and bounded functions;
- u, $\mathrm{D}_x u$ are globally Lipschitz;
- $\mathrm{D}_x^2 u$ is Lipschitz in x, uniformly in t;
- u_t is $1/2$-Hölder in t, uniformly in x.

The precise assumptions can be found in the article and are derived from classical results; see Ladyzhenskaia et al. (1968).

In light of the regularity of u, we assume that the **approximating functions** are chosen from a set

$$\mathcal{U} \subset \mathcal{H}^{3,\gamma} \text{ bounded.}$$

A natural measure for the **approximation error** is

$$\mathcal{E}(U) \triangleq \max_{i=0,1,\ldots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t) - u(t,X_t)\big|^2\Big]$$

A natural measure for the **approximation error** is

$$\begin{aligned}\mathcal{E}(U) &\triangleq \max_{i=0,1,\ldots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-u(t,X_t)\big|^2\Big] \\ &= \max_{i=0,1,\ldots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-Y_t\big|^2\Big].\end{aligned}$$

A natural measure for the **approximation error** is

$$\begin{aligned}\mathcal{E}(U) &\triangleq \max_{i=0,1,\dots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-u(t,X_t)\big|^2\Big] \\ &= \max_{i=0,1,\dots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-Y_t\big|^2\Big].\end{aligned}$$

Similarly, we define

$$\mathcal{D}(U) \triangleq \mathrm{E}\Big[\sum_{i=0}^{I-1}\int_{t_i}^{t_{i+1}} \big|\sigma(t,X_t)^\top\big(\mathrm{D}_x U(t,X_t)-\mathrm{D}_x u(t,X_t)\big)\big|^2 \mathrm{d}t\Big]$$

A natural measure for the **approximation error** is

$$\begin{aligned}\mathcal{E}(U) &\triangleq \max_{i=0,1,\dots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-u(t,X_t)\big|^2\Big] \\ &= \max_{i=0,1,\dots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-Y_t\big|^2\Big].\end{aligned}$$

Similarly, we define

$$\begin{aligned}\mathcal{D}(U) &\triangleq \mathrm{E}\Big[\sum_{i=0}^{I-1}\int_{t_i}^{t_{i+1}} \big|\sigma(t,X_t)^\top\big(\mathrm{D}_x U(t,X_t)-\mathrm{D}_x u(t,X_t)\big)\big|^2\mathrm{d}t\Big] \\ &= \mathrm{E}\Big[\sum_{i=0}^{I-1}\int_{t_i}^{t_{i+1}} \big|\sigma(t,X_t)^\top\mathrm{D}_x U(t,X_t)-Z_t\big|^2\mathrm{d}t\Big].\end{aligned}$$

A natural measure for the **approximation error** is

$$\begin{aligned}\mathcal{E}(U) &\triangleq \max_{i=0,1,\dots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-u(t,X_t)\big|^2\Big] \\ &= \max_{i=0,1,\dots,I-1} \mathrm{E}\Big[\sup_{t\in[t_i,t_{i+1}]} \big|U(t,X_t)-Y_t\big|^2\Big].\end{aligned}$$

Similarly, we define

$$\begin{aligned}\mathcal{D}(U) &\triangleq \mathrm{E}\Big[\sum_{i=0}^{I-1}\int_{t_i}^{t_{i+1}} \big|\sigma(t,X_t)^\top\big(\mathrm{D}_x U(t,X_t)-\mathrm{D}_x u(t,X_t)\big)\big|^2\mathrm{d}t\Big] \\ &= \mathrm{E}\Big[\sum_{i=0}^{I-1}\int_{t_i}^{t_{i+1}} \big|\sigma(t,X_t)^\top\mathrm{D}_x U(t,X_t)-Z_t\big|^2\mathrm{d}t\Big].\end{aligned}$$

Thus, we measure the distance between u and U along the paths of the forward process.

Convergence Theorem I

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U) + \mathcal{D}(U) \leq Ch^{-1}\mathbb{L}(U) + Ch, \qquad U \in \mathcal{U}.$$

Convergence Theorem I

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U) + \mathcal{D}(U) \leq Ch^{-1}\mathbb{L}(U) + Ch, \qquad U \in \mathcal{U}.$$

Remarks:

- Heuristically, the result implies that small loss means good approximation.

Convergence Theorem I

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U) + \mathcal{D}(U) \leq Ch^{-1}\mathbb{L}(U) + Ch, \qquad U \in \mathcal{U}.$$

Remarks:

- Heuristically, the result implies that small loss means good approximation.
- The same result is true for Raissi's loss functional $\mathbb{L}^{\mathrm{R}}(U)$.

Convergence Theorem I

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U) + \mathcal{D}(U) \leq Ch^{-1}\mathbb{L}(U) + Ch, \qquad U \in \mathcal{U}.$$

Remarks:

- Heuristically, the result implies that small loss means good approximation.
- The same result is true for Raissi's loss functional $\mathbb{L}^{\mathrm{R}}(U)$.
- The estimate is "sharp".

Example: Consider the PDE

$$\begin{aligned} u_t(t,x) + u_x(t,x) + 1 &= 0, & (t,x) &\in [0,T) \times \mathbb{R}, \\ u(T,x) &= 1, & x &\in \mathbb{R}. \end{aligned}$$

The unique solution is given by

$$u(t,x) = T - t + 1, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

Example: Consider the PDE

$$\begin{aligned} u_t(t,x) + u_x(t,x) + 1 &= 0, & (t,x) &\in [0,T) \times \mathbb{R}, \\ u(T,x) &= 1, & x &\in \mathbb{R}. \end{aligned}$$

The unique solution is given by

$$u(t,x) = T - t + 1, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

Consider the approximating function $U \equiv 1$. Then

$$\mathbb{L}(U) = Ih^2 = Th$$

Example: Consider the PDE

$$\begin{aligned} u_t(t,x) + u_x(t,x) + 1 &= 0, & (t,x) &\in [0,T) \times \mathbb{R}, \\ u(T,x) &= 1, & x &\in \mathbb{R}. \end{aligned}$$

The unique solution is given by

$$u(t,x) = T - t + 1, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

Consider the approximating function $U \equiv 1$. Then

$$\mathbb{L}(U) = Ih^2 = Th \qquad \text{and} \qquad \mathcal{E}(U) = T^2 = Th^{-1}\mathbb{L}(U).$$

Example: Consider the PDE

$$\begin{aligned} u_t(t,x) + u_x(t,x) + 1 &= 0, & (t,x) &\in [0,T) \times \mathbb{R}, \\ u(T,x) &= 1, & x &\in \mathbb{R}. \end{aligned}$$

The unique solution is given by

$$u(t,x) = T - t + 1, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

Consider the approximating function $U \equiv 1$. Then

$$\mathbb{L}(U) = Ih^2 = Th \qquad \text{and} \qquad \mathcal{E}(U) = T^2 = Th^{-1}\mathbb{L}(U).$$

Conclusion: The fact that the previous theorem is sharp is a consequence of

- us not imposing **additional structure** on the set $\mathcal{U}$ of approximating functions
- and the fact that our **estimate is uniform** in $\mathcal{U}$.

Convergence Theorem II

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathbb{L}(U) \leq C\|U - u\|^2_{\mathrm{C}^{1,2}} + Ch^2, \qquad U \in \mathcal{U}.$$

Convergence Theorem II

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathbb{L}(U) \leq C\|U - u\|^2_{C^{1,2}} + Ch^2, \qquad U \in \mathcal{U}.$$

Remarks:

- If there are functions U which can approximate u well, their loss functional $\mathbb{L}(U)$ will be small and hence the error $\mathcal{E}(U) + \mathcal{D}(U)$ of this approximating function will be small as well.

Convergence Theorem II

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathbb{L}(U) \leq C\|U - u\|^2_{\mathrm{C}^{1,2}} + Ch^2, \qquad U \in \mathcal{U}.$$

Remarks:

- If there are functions U which can approximate u well, their loss functional $\mathbb{L}(U)$ will be small and hence the error $\mathcal{E}(U) + \mathcal{D}(U)$ of this approximating function will be small as well.
- In the paper we prove a version where the uniform distance $\|U - u\|^2_{\mathrm{C}^{1,2}}$ is only considered on a closed subset $K \subseteq \mathbb{R}^d$. This requires us to also use the probability of leaving K as another control.

Convergence Theorem II

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathbb{L}(U) \leq C\|U - u\|^2_{\mathrm{C}^{1,2}} + Ch^2, \qquad U \in \mathcal{U}.$$

Remarks:

- If there are functions U which can approximate u well, their loss functional $\mathbb{L}(U)$ will be small and hence the error $\mathcal{E}(U) + \mathcal{D}(U)$ of this approximating function will be small as well.
- In the paper we prove a version where the uniform distance $\|U - u\|^2_{\mathrm{C}^{1,2}}$ is only considered on a closed subset $K \subseteq \mathbb{R}^d$. This requires us to also use the probability of leaving K as another control.
- The same result is not true for Raissi's loss functional $\mathbb{L}^{\mathrm{R}}(U)$.

Convergence Theorem II

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathbb{L}(U) \leq C\|U - u\|^2_{\mathrm{C}^{1,2}} + Ch^2, \qquad U \in \mathcal{U}.$$

Remarks:

- If there are functions U which can approximate u well, their loss functional $\mathbb{L}(U)$ will be small and hence the error $\mathcal{E}(U) + \mathcal{D}(U)$ of this approximating function will be small as well.
- In the paper we prove a version where the uniform distance $\|U - u\|^2_{\mathrm{C}^{1,2}}$ is only considered on a closed subset $K \subseteq \mathbb{R}^d$. This requires us to also use the probability of leaving K as another control.
- The same result is not true for Raissi's loss functional $\mathbb{L}^{\mathrm{R}}(U)$.
- The estimate is "sharp".

Example: Consider the PDE

$$u_t(t,x) + \frac{1}{2}u_{xx}(t,x) = 0, \qquad (t,x) \in [0,T) \times \mathbb{R},$$
$$u(T,x) = \sin(x), \qquad x \in \mathbb{R}.$$

The unique solution is given by

$$u(t,x) = \sin(x)e^{-(T-t)/2}, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

Example: Consider the PDE

$$u_t(t,x) + \frac{1}{2}u_{xx}(t,x) = 0, \qquad (t,x) \in [0,T) \times \mathbb{R},$$
$$u(T,x) = \sin(x), \qquad x \in \mathbb{R}.$$

The unique solution is given by

$$u(t,x) = \sin(x)e^{-(T-t)/2}, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

One can show that

$$\mathbb{L}(u) \geq \frac{T}{12}e^{-T}h^2 \qquad \text{and} \qquad \mathbb{L}^{\mathrm{R}}(u) \geq \frac{T}{4}e^{-T}h.$$

Example: Consider the PDE

$$\begin{aligned} u_t(t,x) + \frac{1}{2}u_{xx}(t,x) &= 0, & (t,x) &\in [0,T) \times \mathbb{R}, \\ u(T,x) &= \sin(x), & x &\in \mathbb{R}. \end{aligned}$$

The unique solution is given by

$$u(t,x) = \sin(x)e^{-(T-t)/2}, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

One can show that

$$\mathbb{L}(u) \geq \frac{T}{12}e^{-T}h^2 \qquad \text{and} \qquad \mathbb{L}^{\mathrm{R}}(u) \geq \frac{T}{4}e^{-T}h.$$

Conclusion:

- Even if $u \in \mathcal{U}$, the result cannot be improved.

Example: Consider the PDE

$$u_t(t,x) + \frac{1}{2}u_{xx}(t,x) = 0, \qquad (t,x) \in [0,T) \times \mathbb{R},$$
$$u(T,x) = \sin(x), \qquad x \in \mathbb{R}.$$

The unique solution is given by

$$u(t,x) = \sin(x)e^{-(T-t)/2}, \qquad (t,x) \in [0,T] \times \mathbb{R}.$$

One can show that

$$\mathbb{L}(u) \geq \frac{T}{12}e^{-T}h^2 \qquad \text{and} \qquad \mathbb{L}^{\mathrm{R}}(u) \geq \frac{T}{4}e^{-T}h.$$

Conclusion:

- Even if $u \in \mathcal{U}$, the result cannot be improved.
- Raissi's loss functional has a lower order of convergence.

Convergence Theorem III

Let U_h^* a minimizer of

$$U \mapsto \mathbb{L}(U), \qquad U \in \mathcal{U}.$$

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1} \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathrm{C}^{1,2}} + Ch.$$

Convergence Theorem III

Let U_h^* a minimizer of

$$U \mapsto \mathbb{L}(U), \qquad U \in \mathcal{U}.$$

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1} \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathrm{C}^{1,2}} + Ch.$$

Remarks:

- The best rate of convergence that we can achieve is of order $1/2$.

Convergence Theorem III

Let U_h^* a minimizer of

$$U \mapsto \mathbb{L}(U), \qquad U \in \mathcal{U}.$$

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1} \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathrm{C}^{1,2}} + Ch.$$

Remarks:

- The best rate of convergence that we can achieve is of order $1/2$.
- Given h, the set $\mathcal{U}$ should be chosen such that

$$\inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathrm{C}^{1,2}} \leq Ch^2.$$

Convergence Theorem III

Let U_h^* a minimizer of

$$U \mapsto \mathbb{L}(U), \qquad U \in \mathcal{U}.$$

There exists a constant $C > 0$ which does not depend on the dimension d of the forward process X such that, for all h small enough,

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1} \inf_{U \in \mathcal{U}} \|U - u\|_{\mathrm{C}^{1,2}}^2 + Ch.$$

Remarks:

- The best rate of convergence that we can achieve is of order $1/2$.
- Given h, the set $\mathcal{U}$ should be chosen such that

$$\inf_{U \in \mathcal{U}} \|U - u\|_{\mathrm{C}^{1,2}}^2 \leq Ch^2.$$

- We can enlarge the set $\mathcal{U}$ as long as the bound in $\mathcal{H}^{3,\gamma}$ remains unchanged.

Proof. For $\epsilon > 0$, let $U_\epsilon \in \mathcal{U}$ such that

$$\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} \leq \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + \epsilon h.$$

Proof. For $\epsilon > 0$, let $U_\epsilon \in \mathcal{U}$ such that

$$\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} \leq \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + \epsilon h.$$

Then

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*)$$

Proof. For $\epsilon > 0$, let $U_\epsilon \in \mathcal{U}$ such that

$$\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} \leq \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + \epsilon h.$$

Then

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1}\mathbb{L}(U_h^*) + Ch$$

Proof. For $\epsilon > 0$, let $U_\epsilon \in \mathcal{U}$ such that

$$\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} \leq \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + \epsilon h.$$

Then

$$\begin{aligned} \mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) &\leq Ch^{-1}\mathbb{L}(U_h^*) + Ch \\ &\leq Ch^{-1}\mathbb{L}(U_\epsilon) + Ch \end{aligned}$$

Proof. For $\epsilon > 0$, let $U_\epsilon \in \mathcal{U}$ such that

$$\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} \leq \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + \epsilon h.$$

Then

$$\begin{aligned}
\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) &\leq Ch^{-1}\mathbb{L}(U_h^*) + Ch \\
&\leq Ch^{-1}\mathbb{L}(U_\epsilon) + Ch \\
&\leq Ch^{-1}\left[C\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} + Ch^2\right] + Ch
\end{aligned}$$

Proof. For $\epsilon > 0$, let $U_\epsilon \in \mathcal{U}$ such that

$$\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} \leq \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + \epsilon h.$$

Then

$$\begin{aligned}
\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) &\leq Ch^{-1}\mathbb{L}(U_h^*) + Ch \\
&\leq Ch^{-1}\mathbb{L}(U_\epsilon) + Ch \\
&\leq Ch^{-1}\left[C\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} + Ch^2\right] + Ch \\
&\leq C^2 h^{-1} \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + (C^2 + C)h + \epsilon C^2.
\end{aligned}$$

Proof. For $\epsilon > 0$, let $U_\epsilon \in \mathcal{U}$ such that

$$\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} \leq \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + \epsilon h.$$

Then

$$\begin{aligned}
\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) &\leq Ch^{-1}\mathbb{L}(U_h^*) + Ch \\
&\leq Ch^{-1}\mathbb{L}(U_\epsilon) + Ch \\
&\leq Ch^{-1}\left[C\|U_\epsilon - u\|^2_{\mathbb{C}^{1,2}} + Ch^2\right] + Ch \\
&\leq C^2 h^{-1} \inf_{U \in \mathcal{U}} \|U - u\|^2_{\mathbb{C}^{1,2}} + (C^2 + C)h + \epsilon C^2.
\end{aligned}$$

Now send $\epsilon \downarrow 0$.

Some Final Remarks

Some final **observations**:

- There are many variations of the algorithm in the literature. However, most of them approximate the solution of the PDE only in one point (in time).

Some final **observations**:

- There are many variations of the algorithm in the literature. However, most of them approximate the solution of the PDE only in one point (in time).
- While the constants do not depend on the dimension d of the state space, they do depend on the dimension n of the Brownian motion. The dependence on n might be an artifact of our proof though.

Some final **observations**:

- There are many variations of the algorithm in the literature. However, most of them approximate the solution of the PDE only in one point (in time).
- While the constants do not depend on the dimension d of the state space, they do depend on the dimension n of the Brownian motion. The dependence on n might be an artifact of our proof though.
- Using Raissi's loss functional, we only obtain the bound

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1} \inf_{U \in \mathcal{U}} \|U - u\|_{\mathrm{C}^{1,2}}^2 + C,$$

  which does not converge to zero. This does not mean that Raissi's deep solver does not converge.

Some final **observations**:

- There are many variations of the algorithm in the literature. However, most of them approximate the solution of the PDE only in one point (in time).
- While the constants do not depend on the dimension d of the state space, they do depend on the dimension n of the Brownian motion. The dependence on n might be an artifact of our proof though.
- Using Raissi's loss functional, we only obtain the bound

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1} \inf_{U \in \mathcal{U}} \|U - u\|_{\mathrm{C}^{1,2}}^2 + C,$$

  which does not converge to zero. This does not mean that Raissi's deep solver does not converge.
- There are many open problems: optimizing the loss functional, choice of approximating functions, Monte Carlo error, more general PDEs, ...

Some final **observations**:

- There are many variations of the algorithm in the literature. However, most of them approximate the solution of the PDE only in one point (in time).
- While the constants do not depend on the dimension d of the state space, they do depend on the dimension n of the Brownian motion. The dependence on n might be an artifact of our proof though.
- Using Raissi's loss functional, we only obtain the bound

$$\mathcal{E}(U_h^*) + \mathcal{D}(U_h^*) \leq Ch^{-1} \inf_{U \in \mathcal{U}} \|U - u\|_{\mathrm{C}^{1,2}}^2 + C,$$

  which does not converge to zero. This does not mean that Raissi's deep solver does not converge.
- There are many open problems: optimizing the loss functional, choice of approximating functions, Monte Carlo error, more general PDEs, ...
- Interested? Check the paper on SSRN and/or talk to me!