Convergence of Deep Solvers for Semilinear PDEs

Christoph Belak
Technische Universität Berlin

Joint work with Oliver Hager, Lotte Schnell, Charlotte Reimers and Maximilian Würschmidt

Stochastic Models and Control Workshop
Lübeck-Travemünde, March 15, 2022

Overview

Derivation of the Deep Solver

We consider a **partial differential equation** of the form

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

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 consider a **partial differential equation** 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

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

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

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

We consider a **partial differential equation** 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

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

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

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

If we define processes

$$
Y_t \triangleq u(t, X_t) \qquad \text{and} \qquad Z_t \triangleq \sigma(t, X_t)^\top \mathrm{D}_x u(t, X_t),
$$

We consider a **partial differential equation** of the form

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

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

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

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

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

If we define processes

$$Y_t \triangleq u(t, X_t) \qquad \text{and} \qquad Z_t \triangleq \sigma(t, X_t)^\top \mathrm{D}_x u(t, X_t),$$

it is straightforward to see that (Y, Z) solves the **backward SDE**

$$Y_t = g(X_T) + \int_t^T f\left(s, X_s, Y_s, Z_s\right)\mathrm{d}s - \int_t^T Z_s^\top \mathrm{d}W_s.$$

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:

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

Idea: If we want to approximate u by a function U, one possibility is to

Minimize the error in the Euler-Maruyama discretization!

Idea: 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**

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

Idea: 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**

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

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: 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**

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

where

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

Note: there are alternative loss functionals / approximation schemes based on similar ideas in the literature; see Beck et al. (2020) for an overview.

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

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

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

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

where

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

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

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

where

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

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

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;
- 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

$$\mathcal{E}(U) \triangleq \max_{i=0,1,\dots,I-1} \mathbb{E}\left[\sup_{t\in[t_i,t_{i+1}]} \left|U(t,X_t) - u(t,X_t)\right|^2\right]$$

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

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

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

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

Similarly, we define

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

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

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

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

Similarly, we define

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

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

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

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

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,

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

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,

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

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.

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

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

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

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

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.
- 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,

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

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,

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

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,

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

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

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

- 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

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

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

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

Then

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

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

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

Then

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

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

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

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

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

Then

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

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

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

Then

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

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

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

Then

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

Now send $\epsilon \downarrow 0$.

Some Final Remarks

Some final **observations**:

- 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.

Some final **observations**:

- 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.
- Using Raissi's loss functional, we would only obtain the bound

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

  which does not converge to zero.

Some final **observations**:

- 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.
- Using Raissi's loss functional, we would only obtain the bound

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

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

Some final **observations**:

- 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.
- Using Raissi's loss functional, we would 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.
- The correction terms

$$\int_{t_i}^{t_{i+1}} \left[\mathrm{D}_x^2 U(t_i, \bar{X}_i)(\bar{X}_t - \bar{X}_i)\right]^\top \sigma(t_i, \bar{X}_i) \mathrm{d}W_t$$

  appearing in our loss functional can in fact be simulated (thanks to an observation due to **Frank Seifried** and **Maximilian Würschmidt**).

Some final **observations**:

- 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.
- Using Raissi's loss functional, we would 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.
- The correction terms

$$\int_{t_i}^{t_{i+1}} \left[\mathrm{D}_x^2 U(t_i, \bar{X}_i)(\bar{X}_t - \bar{X}_i) \right]^\top \sigma(t_i, \bar{X}_i) \mathrm{d}W_t$$

  appearing in our loss functional can in fact be simulated (thanks to an observation due to **Frank Seifried** and **Maximilian Würschmidt**).
- The paper is on SSRN: https://ssrn.com/abstract=3981933