



Convergence of Deep Solvers for Semilinear PDEs

Christoph Knochenhauer

Key Question

There is an increasing number of examples where machine learning algorithms are able to solve high dimensional PDEs. Can we establish rigorous convergence results to support this observation?

Key Question

There is an increasing number of examples where machine learning algorithms are able to solve high dimensional PDEs. Can we establish rigorous convergence results to support this observation?

More precisely,

- ▶ we focus on *semilinear PDEs* which can be represented in terms of *uncoupled FBSDEs*,

Key Question

There is an increasing number of examples where machine learning algorithms are able to solve high dimensional PDEs. Can we establish rigorous convergence results to support this observation?

More precisely,

- ▶ we focus on *semilinear PDEs* which can be represented in terms of *uncoupled FBSDEs*,
- ▶ and consider a particular deep solver which minimizes the error in the *Euler-Maruyama* discretization of the FBSDE,

Key Question

There is an increasing number of examples where machine learning algorithms are able to solve high dimensional PDEs. Can we establish rigorous convergence results to support this observation?

More precisely,

- ▶ we focus on *semilinear PDEs* which can be represented in terms of *uncoupled FBSDEs*,
- ▶ and consider a particular deep solver which minimizes the error in the *Euler-Maruyama* discretization of the FBSDE,
- ▶ and argue that the optimal *rate of convergence* is of order $1/2$ in the time discretization,

Key Question

There is an increasing number of examples where machine learning algorithms are able to solve high dimensional PDEs. Can we establish rigorous convergence results to support this observation?

More precisely,

- ▶ we focus on *semilinear PDEs* which can be represented in terms of *uncoupled FBSDEs*,
- ▶ and consider a particular deep solver which minimizes the error in the *Euler-Maruyama* discretization of the FBSDE,
- ▶ and argue that the optimal *rate of convergence* is of order $1/2$ in the time discretization,
- ▶ *independent* of the dimension of the state space / forward process

Key Question

There is an increasing number of examples where machine learning algorithms are able to solve high dimensional PDEs. Can we establish rigorous convergence results to support this observation?

More precisely,

- ▶ we focus on *semilinear PDEs* which can be represented in terms of *uncoupled FBSDEs*,
- ▶ and consider a particular deep solver which minimizes the error in the *Euler-Maruyama* discretization of the FBSDE,
- ▶ and argue that the optimal *rate of convergence* is of order $1/2$ in the time discretization,
- ▶ *independent* of the dimension of the state space / forward process
- ▶ but *dependent* on the dimension of the underlying Brownian motion and the approximation capabilities of the class of approximating functions.

Outline:

- ▷ Derivation of the Deep Solver
- ▷ Main Convergence Results
- ▷ Discussion and Outlook

Outline:

- ▷ Derivation of the Deep Solver
- ▷ Main Convergence Results
- ▷ Discussion and Outlook

We consider a *partial differential equation* of the form

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

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

$$\mathcal{L}[u] = \mu^\top D_x u + \frac{1}{2} \text{tr}[\sigma \sigma^\top D_x^2 u].$$

We consider a *partial differential equation* of the form

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

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

$$\mathcal{L}[u] = \mu^\top D_x u + \frac{1}{2} \text{tr}[\sigma \sigma^\top D_x^2 u].$$

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

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

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

$$u(t, X_t) = u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s) ds - \int_t^T D_x u(s, X_s)^\top \sigma(s, X_s) dW_s$$

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

$$\begin{aligned}u(t, X_t) &= u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s) ds \\ &\quad - \int_t^T D_x u(s, X_s)^\top \sigma(s, X_s) dW_s \\ &= g(X_T) + \int_t^T f(s, X_s, u(s, X_s), \sigma(s, X_s)^\top D_x u(s, X_s)) ds \\ &\quad - \int_t^T [\sigma(s, X_s)^\top D_x u(s, X_s)]^\top dW_s.\end{aligned}$$

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

$$\begin{aligned}u(t, X_t) &= u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s) ds \\ &\quad - \int_t^T D_x u(s, X_s)^\top \sigma(s, X_s) dW_s \\ &= g(X_T) + \int_t^T f(s, X_s, u(s, X_s), \sigma(s, X_s)^\top D_x u(s, X_s)) ds \\ &\quad - \int_t^T [\sigma(s, X_s)^\top D_x u(s, X_s)]^\top dW_s.\end{aligned}$$

Now define processes

$$Y_s := u(s, X_s) \quad \text{and} \quad Z_s := \sigma(s, X_s)^\top D_x u(s, X_s),$$

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

$$\begin{aligned}u(t, X_t) &= u(T, X_T) - \int_t^T u_t(s, X_s) + \mathcal{A}[u](s, X_s) ds \\ &\quad - \int_t^T D_x u(s, X_s)^\top \sigma(s, X_s) dW_s \\ &= g(X_T) + \int_t^T f(s, X_s, u(s, X_s), \sigma(s, X_s)^\top D_x u(s, X_s)) ds \\ &\quad - \int_t^T [\sigma(s, X_s)^\top D_x u(s, X_s)]^\top dW_s.\end{aligned}$$

Now define processes

$$Y_s := u(s, X_s) \quad \text{and} \quad Z_s := \sigma(s, X_s)^\top D_x u(s, X_s),$$

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

$$Y_t = g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s^\top dW_s.$$

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

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

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

Euler-Maruyama discretization of the forward SDE:

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

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

$$0 = t_0 < t_1 < \dots < t_l = T$$

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

Euler-Maruyama discretization of the forward SDE:

$$\begin{aligned}\bar{X}_0 &:= x, \\ \bar{X}_{i+1} &:= \bar{X}_i + \mu(t_i, \bar{X}_i)h + \sigma(t_i, \bar{X}_i)\Delta W_{i+1}, \quad i = 0, 1, \dots, l-1.\end{aligned}$$

Naïve Euler-Maruyama discretization of the backward SDE:

$$\begin{aligned}\bar{Y}_l &= g(\bar{X}_l), \\ \bar{Y}_{i+1} &= \bar{Y}_i - f(t_i, \bar{X}_i, \bar{Y}_i, \bar{Z}_i)h + \bar{Z}_i^\top \Delta W_{i+1}, \quad i = 0, 1, \dots, l-1.\end{aligned}$$

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*

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

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*

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

where

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

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*

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

where

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

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*

$$\mathbb{L}(U) := \mathbb{E}\left[|U(T, \bar{X}_T) - g(\bar{X}_T)|^2\right] + \sum_{i=0}^{I-1} \mathbb{E}\left[|\mathcal{L}_i(U)|^2\right],$$

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

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

where

$$\begin{aligned} \mathcal{L}_i(U) := & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f(t_i, \bar{X}_i, U(t_{i+1}, \bar{X}_{i+1}), \sigma(t_i, \bar{X}_i)^\top D_x U(t_i, \bar{X}_i)) h \\ & - D_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i) \Delta W_{i+1} \end{aligned}$$

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

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

where

$$\begin{aligned} \mathcal{L}_i(U) := & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f(t_i, \bar{X}_i, U(t_{i+1}, \bar{X}_{i+1}), \sigma(t_i, \bar{X}_i)^\top D_x U(t_i, \bar{X}_i)) h \\ & - D_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i) \Delta W_{i+1} \\ & - \int_{t_i}^{t_{i+1}} \left[D_x^2 U(t_i, \bar{X}_i) \mu(t_i, \bar{X}_i) h \right. \\ & \left. + D_x^2 U(t_i, \bar{X}_i) \sigma(t_i, \bar{X}_i) (W_t - W_{t_i}) \right]^\top \sigma(t_i, \bar{X}_i) dW_t. \end{aligned}$$

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

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

where

$$\begin{aligned} \mathcal{L}_i(U) := & U(t_{i+1}, \bar{X}_{i+1}) - U(t_i, \bar{X}_i) \\ & + f(t_i, \bar{X}_i, U(t_{i+1}, \bar{X}_{i+1}), \sigma(t_i, \bar{X}_i)^\top D_x U(t_i, \bar{X}_i)) h \\ & - D_x U(t_i, \bar{X}_i)^\top \sigma(t_i, \bar{X}_i) \Delta W_{i+1} \\ & - \int_{t_i}^{t_{i+1}} \left[D_x^2 U(t_i, \bar{X}_i) \mu(t_i, \bar{X}_i) h \right. \\ & \left. + D_x^2 U(t_i, \bar{X}_i) \sigma(t_i, \bar{X}_i) (W_t - W_{t_i}) \right]^\top \sigma(t_i, \bar{X}_i) dW_t. \end{aligned}$$

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

Outline:

- ▷ Derivation of the Deep Solver
- ▷ **Main Convergence Results**
- ▷ Discussion and Outlook

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,

- ▶ $u, u_t, D_x u, D_x^2 u, D_x^3 u, D_x u_t$ 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,

- ▶ $u, u_t, D_x u, D_x^2 u, D_x^3 u, D_x u_t$ exist as continuous and bounded functions;
- ▶ $u, D_x u$ 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,

- ▶ $u, u_t, D_x u, D_x^2 u, D_x^3 u, D_x u_t$ exist as continuous and bounded functions;
- ▶ $u, D_x u$ are globally Lipschitz;
- ▶ $D_x^2 u$ 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,

- ▶ $u, u_t, D_x u, D_x^2 u, D_x^3 u, D_x u_t$ exist as continuous and bounded functions;
- ▶ $u, D_x u$ are globally Lipschitz;
- ▶ $D_x^2 u$ 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,

- ▶ $u, u_t, D_x u, D_x^2 u, D_x^3 u, D_x u_t$ exist as continuous and bounded functions;
- ▶ $u, D_x u$ are globally Lipschitz;
- ▶ $D_x^2 u$ 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,

- ▶ $u, u_t, D_x u, D_x^2 u, D_x^3 u, D_x u_t$ exist as continuous and bounded functions;
- ▶ $u, D_x u$ are globally Lipschitz;
- ▶ $D_x^2 u$ 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

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

A natural measure for the *approximation error* is

$$\begin{aligned}\mathcal{E}(U) &:= \max_{i=0,1,\dots,l-1} \mathbb{E} \left[\sup_{t \in [t_i, t_{i+1}]} |U(t, X_t) - u(t, X_t)|^2 \right] \\ &= \max_{i=0,1,\dots,l-1} \mathbb{E} \left[\sup_{t \in [t_i, t_{i+1}]} |U(t, X_t) - Y_t|^2 \right].\end{aligned}$$

A natural measure for the *approximation error* is

$$\begin{aligned}\mathcal{E}(U) &:= \max_{i=0,1,\dots,l-1} \mathbb{E} \left[\sup_{t \in [t_i, t_{i+1}]} |U(t, X_t) - u(t, X_t)|^2 \right] \\ &= \max_{i=0,1,\dots,l-1} \mathbb{E} \left[\sup_{t \in [t_i, t_{i+1}]} |U(t, X_t) - Y_t|^2 \right].\end{aligned}$$

Similarly, we define

$$\mathcal{D}(U) := \mathbb{E} \left[\sum_{i=0}^{l-1} \int_{t_i}^{t_{i+1}} |\sigma(t, X_t)^\top (D_x U(t, X_t) - D_x u(t, X_t))|^2 dt \right]$$

A natural measure for the *approximation error* is

$$\begin{aligned}\mathcal{E}(U) &:= \max_{i=0,1,\dots,l-1} \mathbb{E} \left[\sup_{t \in [t_i, t_{i+1}]} |U(t, X_t) - u(t, X_t)|^2 \right] \\ &= \max_{i=0,1,\dots,l-1} \mathbb{E} \left[\sup_{t \in [t_i, t_{i+1}]} |U(t, X_t) - Y_t|^2 \right].\end{aligned}$$

Similarly, we define

$$\begin{aligned}\mathcal{D}(U) &:= \mathbb{E} \left[\sum_{i=0}^{l-1} \int_{t_i}^{t_{i+1}} |\sigma(t, X_t)^\top (D_x U(t, X_t) - D_x u(t, X_t))|^2 dt \right] \\ &= \mathbb{E} \left[\sum_{i=0}^{l-1} \int_{t_i}^{t_{i+1}} |\sigma(t, X_t)^\top D_x U(t, X_t) - Z_t|^2 dt \right].\end{aligned}$$

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,

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

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,

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

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,

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

Remarks:

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

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,

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

Remarks:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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,

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

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,

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

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,

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

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 $\|U - u\|_{C^{1,2}}$ 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,

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

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 $\|U - u\|_{C^{1,2}}$ 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 $\mathbb{L}^R(U)$.

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,

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

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 $\|U - u\|_{C^{1,2}}$ 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 $\mathbb{L}^R(U)$.
- ▶ 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

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

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

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

One can show that

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

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

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

One can show that

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

Conclusion:

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

Example: Consider the PDE

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

The unique solution is given by

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

One can show that

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

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^* be a minimizer of

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

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,

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

Convergence Theorem III

Let U_h^* be a minimizer of

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

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,

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

Remarks:

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

Convergence Theorem III

Let U_h^* be a minimizer of

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

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,

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

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

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

Convergence Theorem III

Let U_h^* be a minimizer of

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

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,

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

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

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

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

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

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

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

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

Then

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

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

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

Then

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

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

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

Then

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

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

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

Then

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

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

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

Then

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

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

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

Then

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

Now send $\varepsilon \downarrow 0$.

Outline:

- ▷ Derivation of the Deep Solver
- ▷ Main Convergence Results
- ▷ Discussion and Outlook

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. The dependence on n might be an artifact of our proof though.

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. The dependence on n might be an artifact of our proof though.
- ▶ Using Raissi's loss functional, we only obtain the bound

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

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. The dependence on n might be an artifact of our proof though.
- ▶ Using Raissi's loss functional, we only obtain the bound

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

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. The dependence on n might be an artifact of our proof though.
- ▶ Using Raissi's loss functional, we only obtain the bound

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

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*:

- ▶ 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

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

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, ...
- ▶ Enjoy the paper: <https://ssrn.com/abstract=3981933>