

Robust deep learning methods for coupled FBSDEs

Kristoffer Andersson, Adam Andersson & Cornelis W. Oosterlee

Centrum Wiskunde & Informatica

ICCF, Amsterdam,
April 4, 2024

Table of contents

- 1 Problem formulation and FBSDE - PDE duality
- 2 The deep BSDE method - When and why does it fail to converge?
- 3 Our proposed alternative algorithms

FBSDE

Our starting point is the coupled **FBSDE**

$$\begin{cases} X_t = x_0 + \int_0^t b(s, X_s, Y_s, Z_s) ds + \int_0^t \sigma(s, X_s) dW_s, \\ Y_t = g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s \sigma(s, X_s) dW_s, \quad t \in [0, T]. \end{cases} \quad (1)$$

Solution to (1) is a triple (X, Y, Z) of $\mathcal{F}_t$ -adapted processes.

The associate PDE is then a **semilinear parabolic PDE** on the form

$$\begin{cases} \frac{\partial v}{\partial t}(t, x) + \mathcal{L}v(t, x) + f(t, x, v(t, x), D_x v(t, x)) = 0, & (t, x) \in [0, T) \times \mathbb{R}^d, \\ v(t, x) = g(x), & (t, x) \in \{T\} \times \mathbb{R}^d. \end{cases} \quad (2)$$

Here $\mathcal{L}$ is the infinitesimal generator.

The connection between the FBSDE (1) and the PDE (2) is established through the **nonlinear Feynman-Kac formula**, which states that

$$Y_t = u(t, X_t), \text{ and } Z_t = D_x u(t, X_t).$$

Formulating the FBSDE as a variational problem

The **Deep FBSDE method**¹ uses the following reformulation of a FBSDE

$$\begin{cases} \inf_{y_0, (z_t)_{t \in [0, T]}} \mathbb{E}|Y_T - g(X_T)|^2, & \text{subject to} \\ X_t = x_0 + \int_0^t b(s, X_s, Y_s, z_s)ds + \int_0^t \sigma(s, X_s)dW_s, \\ Y_t = y_0 - \int_0^t f(s, X_s, Y_s, z_s)ds + \int_0^t z_s \sigma(s, X_s)dW_s, \end{cases} \quad (3)$$

compared to

$$\begin{cases} X_t = x_0 + \int_0^t b(s, X_s, Y_s, Z_s)ds + \int_0^t \sigma(s, X_s)dW_s, \\ Y_t = g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T \sigma(s, X_s)Z_s dW_s, \end{cases} \quad (4)$$

Motivation:

- ❶ A solution to (4) solves (3);
- ❷ By wellposedness of the FBSDE, the solution is unique.

¹A. Jentzen et. al. *Solving high-dimensional partial differential equations using deep learning*. Proceedings of the National Academy of Sciences 115.34 (2018): 8505-8510.

Time discretization of the variational problem

Equidistant time grid $\pi := \{0 = t_0, t_1, \dots, t_N = T\}$, with $h = t_{n+1} - t_n$ and Brownian increments $\Delta W_n = W_{n+1} - W_n$.

Time discrete formulation:

$$\begin{cases} \inf_{\bar{y}_0, \{\bar{z}_k\}_{k \in \{0, 1, \dots, N-1\}}} \mathbb{E} |\bar{Y}_N - g(\bar{X}_N)|^2, & \text{subject to} \\ \bar{X}_n = x_0 + \sum_{k=0}^{n-1} b(t_k, \bar{X}_k, \bar{Y}_k, \bar{z}_k)h + \sum_{k=0}^{n-1} \sigma(t_k, \bar{X}_k) \Delta W_k, \\ \bar{Y}_n = \bar{y}_0 - \sum_{k=0}^{n-1} f(t_k, \bar{X}_k, \bar{Y}_k, \bar{z}_k)h + \sum_{k=0}^{n-1} \bar{z}_k \sigma(t_k, \bar{X}_k) \Delta W_k. \end{cases} \quad (5)$$

Aim: Approximate the FBSDE with discrete version of control problem.

Question: Does the discrete control problem converge to the continuous FBSDE?

- For weakly coupled FBSDEs, satisfying some monotonicity conditions, with Lipschitz coefficients a posteriori error estimates exist;
- We show empirically that even for some simple problems with classical and smooth solutions to the associate PDE, the method fails to converge.

A posteriori error estimate

Theorem

Under certain assumptions, there exists a constant C , independent of the step size h , such that for sufficiently small h (see ^{a,b,c}):

$$\sup_{t \in [0, T]} \mathbb{E}|X_t - \bar{X}_t|^2 + \sup_{t \in [0, T]} \mathbb{E}|Y_t - \bar{Y}_t|^2 + \mathbb{E} \left[\int_0^T |Z_t - \bar{Z}_t|^2 \right] \leq C(\omega(h) + \mathbb{E}|\bar{Y}_N - g(\bar{X}_N)|^2),$$

where for $n \in \{0, 1, \dots, N-1\}$, $\bar{X}_t = \bar{X}_n$, $\bar{Y}_t = \bar{Y}_n$, and $\bar{Z}_t = \bar{Z}_n$, for $t \in [t_n, t_{n+1})$ and ω is an increasing function.

^aHan, J and Long J. *Convergence of the deep BSDE method for coupled FBSDEs*. Probability, Uncertainty and Quantitative Risk 5.1 (2020): 1-33.

^bC. Reisinger, W. Stockinger, and Y. Zhang. *A posteriori error estimates for fully coupled McKean–Vlasov forward-backward SDEs*, IMA Journal of Numerical Analysis, 2023;, drad060

^cB. Negyesi, , Z. Huang, and C. W. Oosterlee. *Generalized convergence of the deep BSDE method: a step towards fully-coupled FBSDEs and applications in stochastic control*. arXiv preprint (2024).

Toy problem with semi-analytical solution with classical smooth solution to the associated PDE:

$$\begin{cases} X_t = x_0 - \int_0^t (X_s + Z_s) ds + W_t, \\ Y_t = X_t^\top X_t + \int_t^T \left(X_s^\top X_s - \frac{1}{2} Z_s^\top Z_s \right) ds - \int_t^T Z_s dW_s, \quad t \in [0, T]. \end{cases} \quad (6)$$

Analytic solution available: $Y_t = X_t^\top P(t)X_t + X_t^\top Q(t) + R(t)$, and $Z_t = 2X_t^\top P(t) + Q(t)$, where (P, Q, R) solves a Riccati ODE.

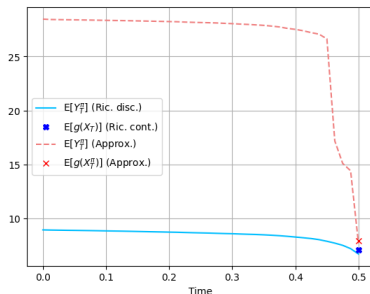
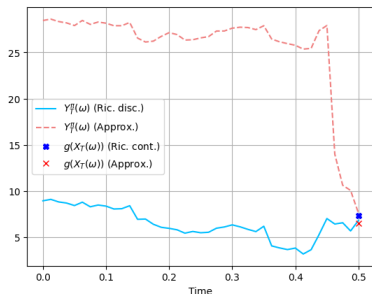


Figure: Comparison of approximation and semi-analytic solution of 5-dimensional FBSDE with an associated classical PDE solution **Left:** A representative path of Y . **Right:** Sample Average of Y .

Investigate convergence of the deep FBSDE method

To investigate convergence:

- 1 Fix $\bar{y}_0 \in \mathbb{R}$;
- 2 Minimize the objective in (5) (only over $\bar{z}_0, \bar{z}_1, \dots, \bar{z}_{N-1}$, since $\bar{y}_0$ fixed);
- 3 Explore the values of the objective function for different $\bar{y}_0$ (one optimization per $\bar{y}_0$).

$$\left\{ \begin{array}{ll} \inf_{\{\bar{z}_k\}_{k \in \{0,1,\dots,N-1\}}} \mathbb{E}|\bar{Y}_N - g(\bar{X}_N)|^2, & \text{subject to} \\ \bar{X}_n = x_0 + \sum_{k=0}^{n-1} b(t_k, \bar{X}_k, \bar{Y}_k, \bar{z}_k)h + \sum_{k=0}^{n-1} \sigma(t_k, \bar{X}_k)\Delta W_k, \\ \bar{Y}_n = \bar{y}_0 - \sum_{k=0}^{n-1} f(t_k, \bar{X}_k, \bar{Y}_k, \bar{z}_k)h + \sum_{k=0}^{n-1} \sigma(t_k, \bar{X}_k)\bar{z}_k\Delta W_k. \end{array} \right. \quad (7)$$

If convergence, then for small h , $\bar{y}_0 \approx Y_0$ should yield the smallest value of the objective function.

Is this what we observe empirically? No, there is large class of FBSDEs for which the discretized FBSDE "almost" satisfies the terminal condition, without being accurate approximations!

For this problem, we cannot expect convergence

$$\text{MSE}(y_0) := \inf_{\zeta} \mathbb{E} [|\bar{Y}_N^{y_0, \zeta} - g(\bar{X}_N^{y_0, \zeta})|^2]. \quad (8)$$

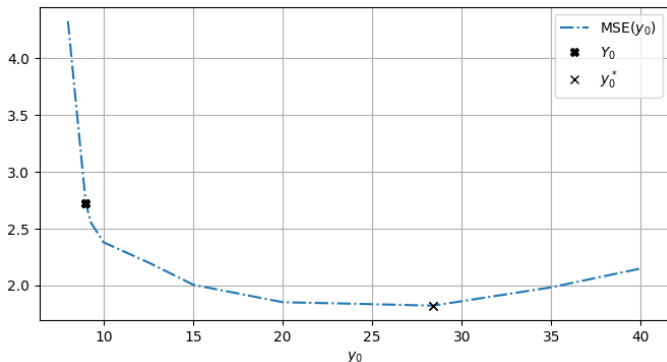


Figure: MSE plotted for different values of y_0 .

Observations

Important observations:

- For each y_0 , the optimal strategy $\{\zeta_n\}_{n \in \{0, \dots, N-1\}}$ is very different;
- For $y_0 \approx Y_0$, the original FBSDE is accurately approximated:
- Although clear minimum of the mean-squared error can be discerned at y_0^* , for a wide range of initial values, the mean-squared error achieves low values. "Too many degrees of freedom" in the system of equations?;

Conclusions:

- If we would, a priori, have access to Y_0 , then the deep FBSDE method would converge for a significantly broader class of FBSDEs;
- How to tackle "too many degrees of freedom"? - Either: (1) optimize over fewer entities or (2) adding constraints. Analogy to an undetermined system of linear equations.

Alternative formulation 1 - For problems stemming from stochastic optimal control

Formulation 1: Optimize over fewer entities²

$$\left\{ \begin{array}{l} \inf_{(z_t)_{t \in [0, T]}} \mathbb{E}[|Y_T - g(X_T)|^2] + \lambda \mathbb{E}[\hat{y}_0(z)], \quad \text{subject to} \\ \hat{y}_0(z) = g(X_T) + \int_0^T f(t, X_t, z_t) dt - \int_0^T z_t dW_t, \quad \mathcal{Y}_0 = \mathbb{E}[\hat{y}_0(z)], \\ X_t = x_0 + \int_0^t b(s, X_s, z_s) ds + \int_0^t \sigma(s, X_s) dW_s, \\ Y_t = \mathcal{Y}_0 - \int_0^t f(s, X_s, z_s) ds + \int_0^t z_s dW_s. \end{array} \right. \quad (9)$$

compared to

$$\left\{ \begin{array}{l} X_t = x_0 + \int_0^t b(s, X_s, Z_s) ds + \int_0^t \sigma(s, X_s) dW_s, \\ Y_t = g(X_T) + \int_t^T f(s, X_s, Z_s) ds - \int_t^T Z_s dW_s, \end{array} \right. \quad (10)$$

Motivation:

- ① Using the fact that Y_0 is not only the initial value of the BSDE, but also the minimum of a stochastic optimal control problem;
- ② Instead of optimizing over $\mathcal{Y}_0$, we let $\mathcal{Y}_0$ be a consequence of the optimization procedure - one fewer entity to optimize over.

²K. Andersson, A. Andersson, and C. W. Oosterlee. *Convergence of a robust deep FBSDE method for stochastic control*. SIAM Journal on Scientific Computing 45.1 (2023): A226-A255.

Alternative formulation 2 - For general FBSDEs

Formulation 2: Add constraints - Relies on a change of measure of the forward SDE.

Assume that $\psi: [0, T] \times \mathbb{R}^d \times \mathbb{R}^\ell \rightarrow \mathbb{R}^d$ is sufficiently regular, then it holds by the Itô formula for $Y_t^\psi = u(t, X_t^\psi)$ and $Z_t^\psi = D_x u(t, X_t^\psi)$ that

$$\begin{cases} X_t^\psi = x_0 + \int_0^t (b(s, X_s^\psi, Y_s^\psi, Z_s^\psi) - \psi(s, X_s^\psi, Y_s^\psi, Z_s^\psi)) ds + \int_0^t \sigma(s, X_s^\psi) dW_s, \\ Y_t^\psi = g(X_T^\psi) + \int_t^T (f(s, X_s^\psi, Y_s^\psi, Z_s^\psi) - \langle \psi(s, X_s^\psi, Y_s^\psi, Z_s^\psi), Z_s^\psi \rangle) ds \\ \quad - \int_t^T Z_s^\psi \sigma(s, X_s^\psi) dW_s, \quad t \in [0, T]. \end{cases} \quad (11)$$

With $X_0^\psi = X_0 = x_0$, we deduce that $u(0, x_0) = Y_0^\psi = Y_0$ (and $D_x u(0, x_0) = Z_0^\psi = Z_0$).

Apply the deep FBSDE method to (11) instead of the original FBSDE?

- For each ψ , the terminal condition is "almost satisfied", but the FBSDE is not well approximated. Very different approximations of y_0 and $\zeta_n(\cdot)$ for different ψ ;
- Force y_0 and $\zeta_n(\cdot)$ to coincide by using the deep FBSDE method simultaneously for different ψ .

Formulation 2: Add constraints (equations) ψ .

$$\left\{ \begin{array}{l} \text{minimize}_{y_0, \zeta} \sum_{k=1}^K \mathbb{E} |Y_T^{\Psi_k} - g(X_T^{\Psi_k})|^2, \quad \text{where for } k=1,2,\dots,K \\ X_t^{\Psi_k, \zeta} = x_0 + \int_0^t (b(s, X_s^{\Psi_k}, Y_s^{\Psi_k}, Z_s^{\Psi_k}) \\ \quad - \psi(s, X_s^{\Psi_k}, Y_s^{\Psi_k}, Z_s^{\Psi_k})) ds + \int_0^t \sigma(s, X_s^{\Psi_k}) dW_s, \\ Y_t^{\Psi_k} = y_0 - \int_0^t (f(s, X_s^{\Psi_k}, Y_s^{\Psi_k}, Z_s^{\Psi_k}) \\ \quad - \langle \psi(s, X_s^{\Psi_k}, Y_s^{\Psi_k}, Z_s^{\Psi_k}), Z_s^{\Psi_k} \rangle) ds + \int_0^t Z_s^{\Psi_k} \sigma(s, X_s^{\Psi_k}) dW_s, \\ Z_t^{\Psi_k} = \zeta(t, X_t^{\Psi_k}), \quad t \in [0, T]. \end{array} \right. \quad (12)$$

Motivation:

- ❶ Using the fact that Y_0 and ζ are identical for each choice of ψ ;
- ❷ Simultaneously solving for multiple choices of ψ adds equations to the "underdetermined system".

For $K = 3$ we can expect convergence.

$$\text{MSE}(y_0) := \inf_{\zeta} \mathbb{E} [|\bar{Y}_N^{y_0, \zeta} - g(\bar{X}_N^{y_0, \zeta})|^2]. \quad (13)$$

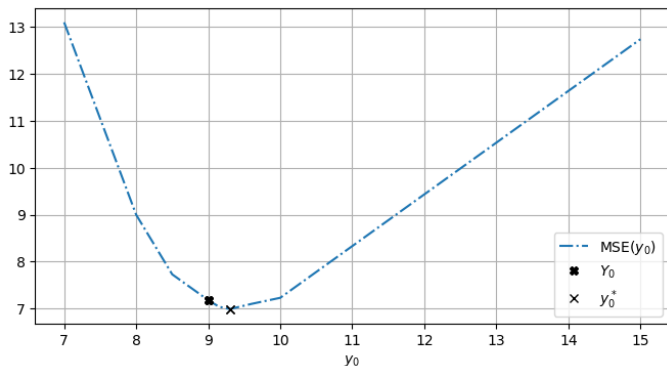


Figure: MSE plotted for different values of y_0 for $K = 3$.

Compared to $K = 1$ when cannot expect convergence.

$$\text{MSE}(y_0) := \inf_{\zeta} \mathbb{E} \left[\left| \bar{Y}_N^{y_0, \zeta} - g(\bar{X}_N^{y_0, \zeta}) \right|^2 \right]. \quad (14)$$

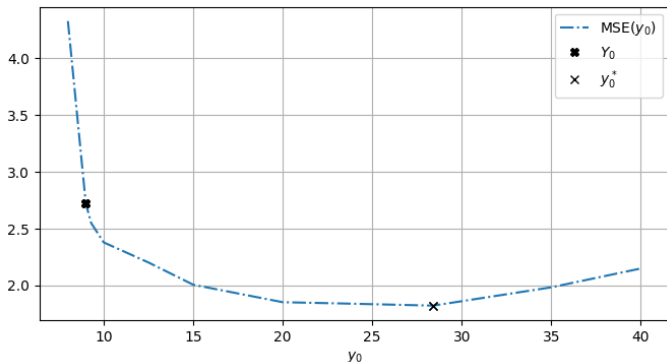


Figure: MSE plotted for different values of y_0 for $K = 1$ i.e., the original deep FBSDE method.

Summary and conclusion

- We have shown empirically that a straight forward extension of the deep BSDE method to strongly coupled FBSDEs fails to converge even for some simple problems;
- We have introduced one method tailored for FBSDEs stemming from stochastic optimal control. This method does not have the initial condition of the BSDE as a free parameter to be optimized. Instead the initial condition is a consequence of the optimization procedure;
- For general FBSDEs, we have introduced a method which uses additional constraints to reduce the degrees of freedom in the problem formulation. By adding these additional equations, we have shown that the loss surface has a distinct minima at the true initial condition.

Stochastic control problem

Dynamical system described by an SDE referred to as the **State equation**

$$X_t^u = x_0 + \int_0^t \bar{b}(t, X_t^u, u_t) dt + \int_0^t \sigma(t, X_t^u) dW_t.$$

$X^u = (X_t^u)_{t \in [0, T]}$ state of the system, $u = (u_t)_{t \in [0, T]}$ control of the system, taking on values in $\mathbb{R}^d$ and $U \subset \mathbb{R}^\ell$, respectively.

To measure performance of the control, a **cost functional** is used

$$J^u(t, x) = \mathbb{E}_{t, x} \left[\int_t^T \bar{f}(t, X_t^u, u_t) dt + g(X_T^u) \right].$$

The **control problem** is to find a control $u \in \mathcal{U}_{[0, T]}$ ($\mathcal{U}_{[0, T]}$:= set of admissible controls) such that the cost functional is minimized.

If infimum is attainable, the **value function**, V , satisfies

$$V(t, x) \in \inf_{u \in \mathcal{U}_{[t, T]}} J^u(t, x).$$

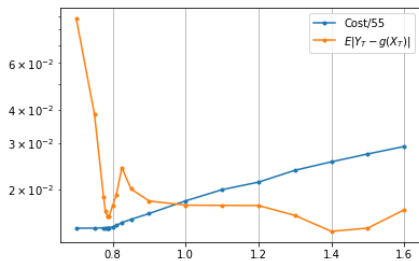
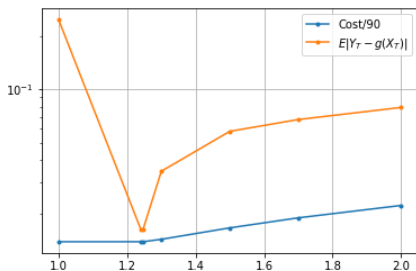


Figure: **Left:** Simple linear problem, $Y_0 = 1.27$. **Right:** Non-linear problem, $Y_0 = 0.79$.

Left: Y_0 is close to the global minimum of the objective function $\rightarrow$ convergence!

Right: Y_0 is close to a local minimum of the objective function. But **not** close to global minimum $\rightarrow$ **No** convergence!

Discussion - Alternative formulation 1

Why should this method work better?

- Using mathematical structure from the specific problem leads to fewer entities to approximate;
- Loss surface of objective function seems to be nice and monotonic $\rightarrow$ easy to optimize;

Disadvantage of this method:

- While deep BSDE method is (at least conceptually) applicable for all FBSDEs, our algorithm is applicable only for FBSDEs stemming from stochastic control problems.

Setup of numerical experiments

- All optimization problems are approximated with the help of ANNs, but any function approximator efficient enough could be used;
- In the following examples, we have analytical solutions available to compare with;
- For each solution component, X , Y and Z we compare to analytical counterpart in strong and weak sense;
- One problem with control in each spatial dimension ($d = \ell$) $\rightarrow$ Algorithm 1;
- One problem with control in lower dimensions than the state space ($d < \ell$) $\rightarrow$ Algorithm 2.

Strong and weak approximations of X for $\ell = d = 2$

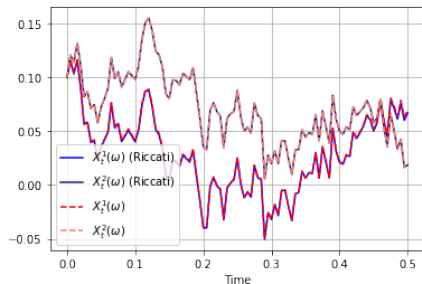
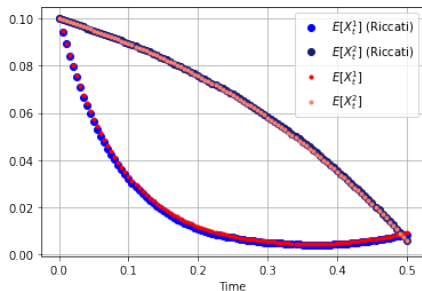


Figure: Average of solutions and a single solution path compared to their analytical counterparts.

Strong and weak approximations of Y for $\ell = d = 2$

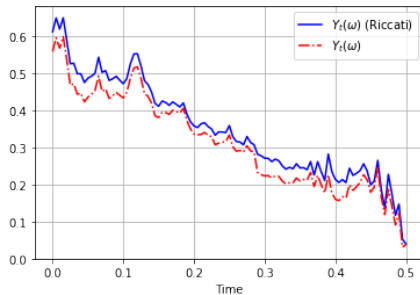
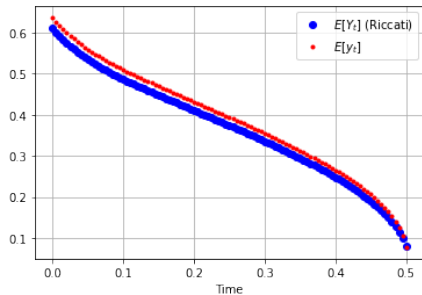


Figure: Average of solutions and a single solution path compared to their analytical counterparts.

Strong and weak approximations of Z for $\ell = d = 2$

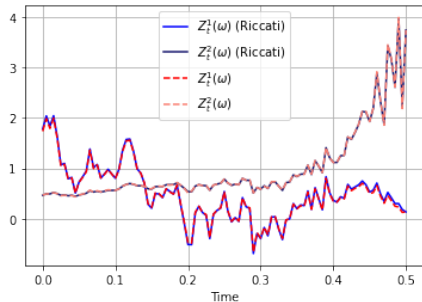
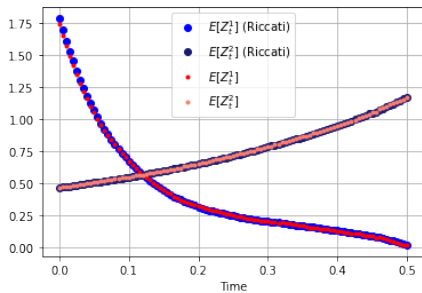


Figure: Average of solutions and a single solution path compared to their analytical counterparts.

Strong and weak approximations of X for $\ell = 2$ and $d = 6$

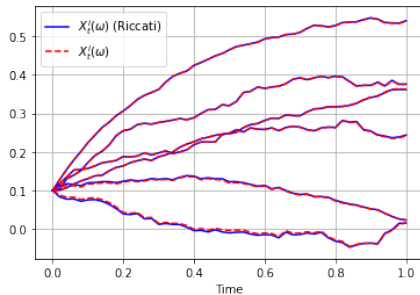
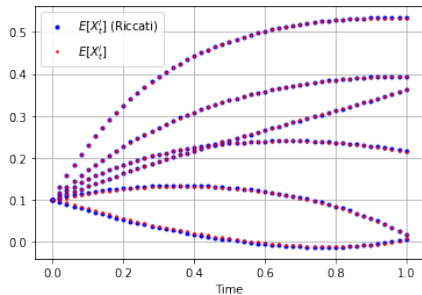


Figure: Average of solutions and a single solution path compared to their analytical counterparts.

Strong and weak approximations of Y for $\ell = 2$ and $d = 6$

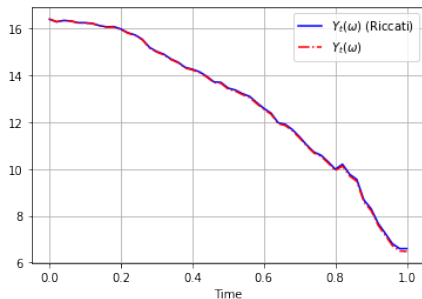
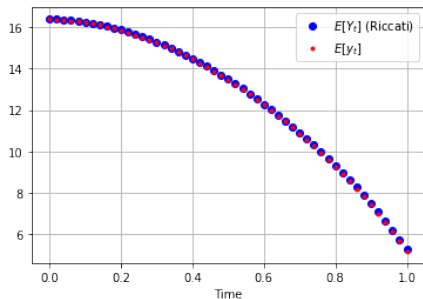


Figure: Average of solutions and a single solution path compared to their analytical counterparts.

Strong and weak approximations of Z for $\ell = 2$ and $d = 6$

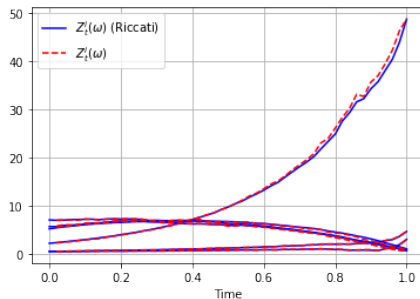
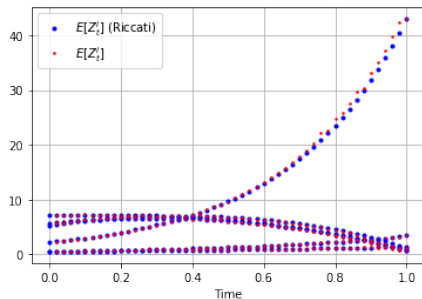


Figure: Average of solutions and a single solution path compared to their analytical counterparts.