Smoothing the Payoff for Efficient Computation of Option Pricing in Time-Stepping Setting

1 The goal and outline of the project

The first goal of the project was to approximate $\mathrm{E}[f(X(t))]$, using multi-index stochastic collocation(MISC) method proposed in [9], where

- The payoff $f : \mathbb{R}^d \to \mathbb{R}$ has either jumps or kinks. Possible choices of f that we wanted to test are:
  - hockey-stick function, i.e., put or call payoff functions;
  - indicator functions (both relevant in finance (binary option,...) and in other applications of estimation of probabilities of certain events);
  - delta-functions for density estimation (and derivatives thereof for estimation of derivatives of the density).

  More specifically, f should be the composition of one of the above with a smooth function. (For instance, the basket option payoff as a function of the log-prices of the underlying.)
- The process X is simulated via a time-stepping scheme. Possible choices that we wanted to test are
  - The one/multi dimensional discretized Black-Scholes(BS) process where we compare different ways to identify the location of the kink, such as:
    - Exact location of the continuous problem
    - Exact location of the discrete problem found by finding the root of a polynomial in y
    - Newton iteration
  - A relative simple interest rate model or stochastic volatility model, for instance CIR or Heston models: In fact, the impact of the Brownian bridge will disappear in the limit, which may make the effect of the smoothing, but also of the errors in the kink location difficult to identify. For this reason, we suggest to study a more complicated 1-dimensional problem next. We suggest to use a CIR process. To avoid complications at the boundary, we suggest "nice" parameter choices, such that the discretized process is very unlikely to hit the boundary (Feller condition).

– The multi dimensional discretized Black-Scholes(BS) process: Here, we suggest to return to the Black-Scholes model, but in multi-dimensional case. In this case, linearizing the exponential, suggest that a good variable to use for smoothing might be the sum of the final values of the Brownian motion. In general, though, one should probably eventually identify the optimal direction(s) for smoothing via the duals algorithmic differentiation.

To-Do-1: The main motivation of the project Analiticity proof as stated by Raul. To check with Christian!

The expected output was

- An application paper including testing the examples above.
- A theoretical paper including a numerical analysis of the schemes involved, such as Newton iteration, etc.

The numerical outputs were the following:

1. **Example 1**: The basket option with the smoothing trick as in [1] (see Section C): in this example we wanted to check the performance of the MISC without time stepping scheme and also compare the results with reference [1]: **Done**.
2. **Example 2**: The one dimensional call option under discretized BS model (see Section 5.1): In this example, we compared different ways (exact/approximate) to determine the kink (See section 5.1.1). In a second step, we tried to approximate the option price using MISC (with/without Richardson extrapolation). We observed some issues of convergence when using Richardson extrapolation, therefore, we added another case where we use a smooth payoff (with/without Richardson) and there we observed an advantage of using Richardson (see Section E). We do not have any clues why we have such problem when coupling MISC with Richardson for the non-smooth case. I provide in Section R some analysis that may explain why we do not see an improvement the way we are coupling MISC with Richardson extrapolation and then maybe we need to do it differently. (**Under checking**)
3. **Example 3**: The one dimensional binary option under discretized BS model (see Section 5.2): to help understand the issue observed for the previous example, we tested a simpler case which is given by the binary option. Unfortunately, we observed no improvement when using Richardson extrapolation. In Section R, we try to give some explanation of this observed behavior. (**Under checking**).

To-Do-2: What are the desired main contributions of this work.

2 Introduction and literature review

Many option pricing problems require the computation of multivariate integrals. The dimension of these integrals is determined by the number of independent stochastic factors (e.g. the number of time steps in the time discretization or the number of assets under consideration). The high dimension of these integrals can be treated with dimension-adaptive quadrature methods to have the desired convergence behavior.

Unfortunately, in many cases, the integrand contains either kinks and jumps. In fact, an option is normally considered worthless if the value falls below a predetermined strike price. A kink (discontinuity in the gradients) is present when the payoff function is continuous, while a jump (discontinuity in the function) exists when the payoff coreesponds to a binary or other digital options. The existence of kinks or jumps in the integrand heavily degrades the performance of quadrature formulas. In this work, we are interested in solving this problem by using adaptive sparse grids (SG) methods coupled with suitable transformations. The main idea is to find lines or areas of discontinuity and to employ suitable transformations of the integration domain. Then by a pre-integration (smoothing) step with respect to the dimension containing the kink/jump, we end up with integrating only over the smooth parts of the integrand and the fast convergence of the sparse grid method can be regained.

One can ignore the kinks and jumps, and apply directly a method for integration over $\mathbb{R}^d$. Despite the significant progress in SG methods [2] for high dimensional integration of smooth integrands, few works have been done to deal with cases involving integrands with kinks or jumps.

The performance of SG methods is degraded in the presence of kinks and jumps. In [6, 7, 8], an analysis of the performance of Quasi Monte Carlo (QMC) and SG methods has been conducted, in the presence of kinks and jumps. In [6, 7], the authors studied the terms of the ANOVA decomposition of functions with kinks defined on d-dimensional Euclidean space $\mathbb{R}^d$, and showed that under some assumptions all but the the highest order ANOVA term of the 2^d ANOVA terms can be smooth for the case of an arithmetic Asian option with the Brownian bridge construction. Furthermore, [8] extended the work in [6, 7] from kinks to jumps for the case of an arithmetic average digital Asian option with the principal component analysis (PCA). The main findings in [6, 7] was obtained for an integrand of the form $f(\mathbf{x}) = \max(\phi(\mathbf{x}), 0)$ with ϕ being smooth. In fact, by assuming i) the d-dimensional function ϕ has a positive partial derivative with respect to x_j for some $j \in \{1, \ldots, d\}$, ii) certain growth conditions at infinity are satisfied, the authors showed that the ANOVA terms of f that do not depend on the variable x_j are smooth.

To-Do 3: How do we compare to the previous works. What will be our contribution comparing to their work?

This idea of smoothing by pre-integration was also tested in [1] for basket option pricing and also in [10].

3 Problem formulation and setting

3.1 General setting

In this setting, we are interetsed in the basic problem of approximating

$$I_d g := \int_{\mathbb{R}^d} g(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_d) \rho_d(\mathbf{x}) dx_1, \ldots, dx_d, \tag{1}$$

with

$$\rho_d(\mathbf{x}) := \prod_{j=1}^{d} \rho(x_j), \tag{2}$$

where ρ is a continuous and strictly positive probability density function on $\mathbb{R}$ and g is a real-valued function integrable with respect to ρ_d.

In this context, we work mainly with two possible structures of payoff function g. In fact, for the cases of call/put options, the payoff g has a kink and will be of the form

$$g(\mathbf{x}) = \max(\phi(\mathbf{x}), 0). \tag{3}$$

One can also encounter jumps in the payoff when working with binary digital options. In this case, g is given by

$$g(\mathbf{x}) = \mathbf{1}_{(\phi(\mathbf{x})\geq 0)}. \tag{4}$$

We introduce the notation $\mathbf{x} = (x_j, \mathbf{x}_{-j})$, where $\mathbf{x}_{-j}$ denotes the vector of length $d-1$ denoting all the variables other than x_j. Then, if we assume for some $j \in \{1, \ldots, d\}$

$$\frac{\partial \phi}{\partial x_j}(\mathbf{x}) > 0, \ \forall \mathbf{x} \in \mathbb{R}^d \ \textbf{(Monotonicity condition)} \tag{5}$$

$$\lim_{x\to+\infty} \phi(\mathbf{x}) = \lim_{x\to+\infty} \phi(x_j, \mathbf{x}_{-j}) = +\infty, \text{ or } \frac{\partial^2 \phi}{\partial x_j^2}(\mathbf{x}) \ \textbf{(Growth condition)}, \tag{6}$$

then, using Fubini's theorem, we can rewrite (1) as

$$I_d g = \int_{\mathbb{R}^{d-1}} \left(\int_{-\infty}^{\infty} g(x_j, \mathbf{x}_{-j}) \rho(x_j) dx_j \right) \rho_{d-1}(\mathbf{x}_{-j}) d\mathbf{x}_{-j}, \tag{7}$$

where we evaluate the inner integral for each $\mathbf{x}_{-j}$ and which results in a smooth integrand for the outer $(d-1)$-dimensional integral.

We note that conditions ((5) and (6)) imply that for each $\mathbf{x}_{-j}$, the function $\phi(x_j, \mathbf{x}_{-j})$ either has a simple root x_j or is positive for all $x_j \in \mathbb{R}$.

3.2 The context of option pricing using time stepping procedure

In the context of option pricing, we aim at approximating the option price, $E[f(\mathbf{X}(t))]$, using SG method, where $f : \mathbb{R}^d \to \mathbb{R}$ is the payoff function and where the process of the asset prices $\mathbf{X} \in \mathbb{R}^d$ solves

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t a(s, \mathbf{X}(s)) ds + \sum_{\ell=1}^{\ell_0} \int_0^t b^\ell(s, \mathbf{X}(s)) dW^\ell(s) \tag{8}$$

Our proposed method of smoothing the integrand can be seen in two ways, either as a pre-integration or as a conditional sampling step. In the following, we give some details of the pre-integration step.

3.3 Smoothing via pre-integration step

For illustration purposes, let us focus on the one dimensional case, where under the risk-neutral measure the undelying asset follows the geometric Brownian motion (GBM)

$$dX_t = rX_t dt + \sigma X_t dB_t, \tag{9}$$

where r is the risk-free rate, σ is the volatility and B_t is the standard Brownian motion. The analytical solution to (9) is

$$X_t = X_0 \exp((r - \sigma^2)t + \sigma B_t). \tag{10}$$

The Brownian motion B_t can be constructed either sequentially using a standard random walk construction or hierarchically using Brownian bridge (BB) construction. for numerical purposes that we will explain later, we will use the BB construnction since it produces dimensions with different importance for SG, contrary to random walk procedure for which all the dimension of the stochastic space have equal importance.

Let us denote $\{t_i\}_{i=0}^N$ the grid of time steps, then the BB construction [5] consists of the following: given a past value B_{t_i} and a future value B_{t_k}, the value B_{t_j} (with $t_i < t_j < t_k$) can be generated according to the formula:

$$B_{t_j} = (1-\rho)B_{t_i} + \rho B_{t_k} + \sqrt{\rho(1-\rho)(k-i)\Delta t}z,\ z \sim \mathcal{N}(0,1), \tag{11}$$

where $\rho = \frac{j-i}{k-i}$. In particular, if N is a power of 2, then given $B_0 = 0$, BB generates the Brownian motion at times $T, T/2, T/4, 3T/4, \ldots$ according

$$\begin{aligned}
B_T &= \sqrt{T}z_1 \\
B_{T/2} &= \frac{1}{2}(B_0 + B_T) + \sqrt{T/4}z_2 = \frac{\sqrt{T}}{2}z_1 + \frac{\sqrt{T}}{2}z_2 \\
B_{T/4} &= \frac{1}{2}(B_0 + B_{T/2}) + \sqrt{T/8}z_3 = \frac{\sqrt{T}}{4}z_1 + \frac{\sqrt{T}}{4}z_2 + \sqrt{T/8}z_3 \\
&\vdots
\end{aligned} \tag{12}$$

where $\{z_j\}_{j=1}^N$ are independent standard normal variables. In BB construction given by (12), the most important values that determine the large scale structure of Brownian motion are the first components of $\mathbf{z} = (z_1, \ldots, z_N)$.

Let us denote by $\psi : (z_1, \ldots, z_N) \to (B_1, \ldots, B_N)$ the mapping of BB construction and by $\Phi : (B_1, \ldots, B_N) \to X_T$, the mapping consisting of the time-stepping scheme. Then, we can express the option price as

$$\begin{aligned}
\mathrm{E}\left[f(X(T))\right] &= \mathrm{E}\left[f\left(\Phi \circ \psi\right)(z_1, \ldots, z_N)\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(z_1, \ldots, z_N)\rho_N(\mathbf{z}) dz_1, \ldots, dz_N,
\end{aligned} \tag{13}$$

where $g = f \circ \Phi \circ \psi$ and

$$\rho_N(\mathbf{z}) = \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}}. \tag{14}$$

Now, we can easily apply the procedure of pre-integration of section 3.1, where we can assume that the payoff function f can be either the maximum or indicator function and $\phi = \Phi \circ \psi$. The remaining ingredient is to determine with respect to which variable z_j we will integrate.

To-Do-3: We need to justify our choice as explained in Section 3.1. In fact, I think we just need to check the assumptions of $\phi = \Phi \circ \psi$ wrt to z_1.

Claiming that pre-integrating with respect to z_1 is the optimal option then form (13)

$$\begin{aligned}
\mathrm{E}\left[f(X(T))\right] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(z_1, \ldots, z_N)\rho_N(\mathbf{z}) dz_1, \ldots, dz_N \\
&= \int_{\mathbb{R}^{d-1}} \left(\int_{-\infty}^{\infty} g(z_1, \mathbf{z}_{-1})\rho(z_1) dz_1 \right) \rho_{d-1}(\mathbf{z}_{-1}) d\mathbf{z}_{-1} \\
&= \int_{\mathbb{R}^{d-1}} h(\mathbf{z}_{-1})\rho_{d-1}(\mathbf{z}_{-1}) d\mathbf{z}_{-1}, \qquad (15)
\end{aligned}$$

where $h(\mathbf{z}_{-1}) = \int_{-\infty}^{\infty} g(z_1, \mathbf{z}_{-1})\rho(z_1)dz_1 = E\left[g\left(z_1, \ldots, z_N\right) \mid z_1\right]$.

Since g can have a kink or jump. Computing $h(\mathbf{z}_{-1})$ in the pre-integration step should be carried carefully to not deteriorate the smoothness of H. This can be done by applying a root finding procedure and then computing the uni-variate integral by summing the terms coming from integrating in each region where f is smooth. In the next sections, we explain those points.

3.4 Description of the Domain Decomposition and Suitable Transformation

The payoff function is not smooth due to the nature of the option. In fact, the holder would not exercise the option if a purchase or sale of the underlying asset would lead to a loss. As a result, the discontinuity of the payoff function carries over to the integrand. In this case, The integrand shows a kink or even a jump with respect to a manifold. Since some (mixed) derivatives are not bounded at these manifolds, the smoothness requirements for the sparse grid method are clearly not fulfilled any more.

The first step consists of identifying the areas of discontinuity or non-differentiability. Then, we decompose the total integration domain Ω into sub-domains Ω_i, $i = 1, \ldots, n$ such that the integrand is smooth in the interior of Ω_i and such that all kinks and jumps are located along the boundary of these areas. This procedure results in integrating several smooth functions, instead of one discontinuous function. The total integral is then given as the sum of the separate integrals, *i.e.*

$$If := \int_{\Omega} f(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^{n} \int_{\Omega_i} f(\mathbf{x}) d\mathbf{x} \tag{16}$$

In this way, the fast convergence of SG can be regained whereas the costs only increase by a constant (the number of terms in the sum), provided the cost required for the decomposition is sufficiently small such that it can be neglected.

In general, such a decomposition is even more expensive than to integrate the function. Nevertheless, for some problem classes, the areas of discontinuity have a particular simple form, which allows to decompose the integration domain with costs that are much smaller than the benefit which results from the decomposition. In this work, we consider those cases.

In the literature, there two classes that have been tackled. In the first one, we have the information that the kink bounds are part of the integration domain where the integrand is zero and can thus be identified by root finding as proposed in [3].

In the second class, we have the information that the discontinuities are located on hyperplanes, which allows a decomposition first into polyhedrons and then into orthants as discussed in [4]. In this work, we start by the first class of problems.

3.5 Root Finding

Without loss of generality, we can assume that the integration domain can be divided into two parts Ω_1, and Ω_2 such that the integrand f is smooth and positive in Ω_1 whereas $f(\mathbf{x}) = 0$ in Ω_2. Therefore,

$$If := \int_{\Omega_1} f(\mathbf{x}) d\mathbf{x} \tag{17}$$

This situation may arise when the integrand is non-differentiable or noncontinuous along the boundary between Ω_1 and Ω_2. For these problems, kinks and jumps can efficiently be identified by a one-dimensional root finding. Then, the kinks and jumps can be transformed to the boundary of integration domain such that they no longer deteriorate the performance of the numerical methods. In fact, we compute the zeros of the integrand with respect to the last dimension. In this dimension, then, e.g., Newton's method or bisection can be used to identify the point which separates Ω_1 and Ω_2.

3.6 Discussion on Smoothing

A crucial element of the smoothing property is that the "location of irregularity" $y^{\text{kink}} = y(\mathbf{z}_{-1})$ such that f is not smooth at the point $\phi(y^{\text{kink}}, \mathbf{z}_{-1})$. Generally, there might be (for given $\mathbf{z}_{-1}$

- no solution, i.e., the integrand in the definition of $h(\mathbf{z}_{-1})$ above is smooth (*best case*);
- a unique solution;
- multiple solutions.

Generally, we need to assume that we are in the first or second case. Specifically, we need that

$$\mathbf{z}_{-1} \mapsto h(\mathbf{z}_{-1}) \text{ and } \mathbf{z}_{-1} \mapsto \widehat{h}(\mathbf{z}_{-1})$$

are smooth, where $\widehat{h}$ denotes the numerical approximation of h based on a grid containing $y(\mathbf{z}_{-1})$. In particular, y itself should be smooth in $\mathbf{z}_{-1}$. This would already be challenging in practice in the third case. Moreover, in the general situation we expect the number of solutions y to increase when the discretization of the SDE gets finer.

To-Do-4: h in our case corresponds to P_1 in [6], I think we can apply theorem 3.2 in [6] to deduce the smoothness of h and also $\widehat{h}$.

To-Do-5: Using the implicit theorem given by Theorem 2.3 in [6], do not we expect to have the unicity of the solution of y independently of the discretization if we fullfill assumptions (5,6)?

In many situations, case 2 (which is thought to include case 1) can be guaranteed by monotonicity (I think we need to add also the growth condition, consequently, we need to check both assumptions 5 and 6). For instance, in the case of one-dimensional SDEs with z_1 representing the terminal value of the underlying Brownian motion (and $\mathbf{z}_{-1}$ representing the Brownian bridge), this can often be seen from the SDE itself. Specifically, if each increment "dX" is increasing in z_1, no matter the value of X, then the solution X_T must be increasing in z_1. This is easily seen to be true in examples such as the Black-Scholes model and the CIR process. (Strictly speaking, we have to distinguish between the continuous and discrete time solutions. In these examples, it does not matter.) On the other hand, it is also quite simple to construct counter examples, where monotonicity fails, for instance SDEs for which the "volatility" changes sign, such as a trigonometric function.[1]

Even in multi-dimensional settings, such monotonicity conditions can hold in specific situations. For instance, in case of a basket option in a multivariate Black Scholes framework, we can choose a linear combination z_1 of the terminal values of the driving Bm, such that the basket is a monotone function of z_1. (The coefficients of the linear combination will depend on the correlations and the weights of the basket.) However, in that case this may actually not correspond to the optimal "rotation" in terms of optimizing the smoothing effect.

3.7 Errors in smoothing

For the analysis it is useful to assume that $\widehat{h}$ is a smooth function of z_{-1}, but in reality this is not going to be true. Specifically, if the true location y of the non-smoothness in the system were available, we could actually guarantee $\widehat{h}$ to be smooth, for instance by choosing

$$\widehat{h}(\mathbf{z}_{-1}) = \sum_{k=-K}^{K} \eta_k f\left(\phi\left(\zeta_k(y(\mathbf{z}_{-1})), \mathbf{z}_{-1}\right)\right),$$

for points $\zeta_k \in \mathbb{R}$ with $\zeta_0 = y$ and corresponding weights η_k.[2] However, in reality we have to numerical approximate y by $\overline{y}$ with error $|y - \overline{y}| \leq \delta$. Now, the actual integrand in $\mathbf{z}_{-1}$ becomes

$$\overline{h}(\mathbf{z}_{-1}) := \sum_{k=-K}^{K} \eta_k g\left(\Phi\left(\zeta_k(\overline{y}(\mathbf{z}_{-1})), \mathbf{z}_{-1}\right)\right),$$

which we cannot assume to be smooth anymore. On the other hand, if $\zeta_k(y)$ is a continuous function of y and y and $\overline{y}$ are continuous in $\mathbf{z}_{-1}$, then *eventually* we will have

$$\left\|\widehat{h} - \overline{h}\right\|_\infty \leq \mathrm{TOL}, \quad \left\|h - \overline{h}\right\|_\infty \leq \mathrm{TOL},$$

[1] Actually, in every such case the simple remedy is to replace the volatility by its absolute value, which does not change the law of the solution. Hence, there does not seem to be a one-dimensional counter-example.

[2] Of course, the points ζ_k have to be chosen in a systematic manner depending on y.

i.e., the smooth functions h and $\widehat{h}$ are close to the integrand $\overline{h}$. (Of course, this may depend on us choosing a good enough quadrature ζ!)

Remark 3.1. If the adaptive collocation used for computing the integral of $\overline{h}$ depends on derivatives (or difference quotients) of its integrand $\overline{h}$, then we may also need to make sure that derivatives of $\overline{h}$ are close enough to derivatives of $\widehat{h}$ or h. This may require higher order solution methods for determining y.

Remark 3.2. In some important cases, f may be trivial (e.g., $\equiv 0$). In these cases, we may be able to make sure that $\overline{y}$ never crosses the "location of non-smoothness". Then even $\overline{h}$ is smooth.

Remark 3.3. We expect that the global error of our procedure will be bounded by the weak error which is in our case of order $O(\Delta t)$. In this case, the overall complexity of our procedure will be of order $O(TOL^{-1})$. We note that this rate can be improved up to $O(TOL^{-\frac{1}{2}})$ if we use **Richardson extrapolation**. Another way that can improve the complexity could be based on **Cubature on Wiener Space** (This is left for a future work). The aimed complexity rate illustrates the contribution of our procedure which outperforms Monte Carlo forward Euler (MC-FE) and multi-level MC-FE, having complexity rates of order $O(TOL^{-3})$ and $O(TOL^{-2}log(TOL)^2)$ respectively.

4 Sparse Grids +Multi-Index Stochastic Collocation (MISC) (TO-DO)

I need to add details for sparse grid quadrature and dimension-adaptive procedure used in MISC in this section.

4.1 Details on the construction

5 Numerical examples

5.1 The discretized 1D Black-Scholes

The second example that we test is the binary and call options under BS model where the process X is the discretized 1D Black-Scholes model and the function f is the indicator or maximum function and which has a kink. Precisely, we are interested in the 1-D lognormal example where the dynamics of the stock are given by

$$dX_t = \sigma X_t dW_t. \tag{18}$$

Assume that $\{W_t, 0 \le t \le T\}$ is a standard one-dimensional Brownian motion. We are interested in simulating the values of $W_{t_1}, \ldots, W_{t_d}$ at d discrete times, where $0 = t_0 < t_1 < \cdots < t_d = T$ and $W_{t_0} = 0$. We also assume that $t_j - t_{j-1} = \Delta t = T/d$, $j = 1, 2, \ldots, d$, and denote $W_i = W_{t_i}$ in the following. Let $\mathbf{W} = (W_1, \ldots, W_d)^T$, then $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, where $\mathcal{N}(\mathbf{0}, \Sigma)$ is a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix Σ.

In the discrete case, we have

$$\begin{aligned} \Delta W_i &= (B_{t_{i+1}} - B_{t_i}) + \Delta t \frac{Y}{\sqrt{T}} \\ &= \Delta B_i + \Delta t \frac{Y}{\sqrt{T}}, \end{aligned} \tag{19}$$

implying that the numerical approximation of $X(T)$ satisfies

$$\overline{X}_T = \Phi(\Delta t, Y, \Delta B_0, \ldots, \Delta B_{N-1}), \tag{20}$$

for some path function Φ.

In the simulation, we generate the stock prices using Brownian bridge (BB) in the following way: first generates the final value W_d , then sample $W_{[d/2]}$ conditional on the values of W_d and W_0, and proceed by progressively filling in intermediate values. Here, $[x]$ denotes the greatest integer less than or equal to x. In particular, if d is a power of 2, then BB generates the Brownian motion as

$$\begin{aligned} W_d &= \sqrt{T} Z_1 \\ W_{d/2} &= \frac{1}{2}(W_0 + W_d) + \sqrt{T/4} Z_2 \\ W_{d/4} &= \frac{1}{2}(W_0 + W_{d/2}) + \sqrt{T/8} Z_3 \\ &\vdots \\ W_{d-1} &= \frac{1}{2}(W_{d-2} + W_d) + \sqrt{T/2d} Z_d, \end{aligned} \tag{21}$$

where Z_j are independent standard normal variables. Referring to (44), we note that $Y = Z_1$.

As explained in Section 3.6, the first step of our approach is determining the location of irregularity (kink). In the following, we want to compare different ways for identifying the location of the kink for this model.

5.1.1 Determining the kink location

Exact location of the kink for the continuous problem

Let us denote y_* an invertible function that satisfies

$$X(T; y_*(x), B) = x. \tag{22}$$

We can easily prove that the expression of y_* for model given by (18) is given by

$$y_*(x) = \left(\log(x/x_0) + T\sigma^2/2\right) \frac{1}{\sqrt{T}\sigma}, \tag{23}$$

and since the kink for Black-Scholes model occurs at $x = K$, where K is the strike price then the exact location of the continuous problem is given by

$$y_*(K) = \left(\log(K/x_0) + T\sigma^2/2\right) \frac{1}{\sqrt{T}\sigma}. \tag{24}$$

Exact location of the kink for the discrete problem

The discrete problem of model (18) is solved by simulating

$$\begin{aligned} \Delta X_{t_i} &= \sigma X_{t_i} \Delta W_i,\ 0 \le i \le N-1 \\ X_{t_{i+1}} - X_{t_i} &= \sigma X_{t_i} \left(W_{t_{i+1}} - W_{t_i}\right),\ 0 < i < N \end{aligned} \tag{25}$$

where $X(T_0) = X_0$ and $X(t_N) = X(T)$.

Using Brownian bridge construction given by (44), we have

$$\begin{aligned} X_{t_1} &= X_{t_0} \left[1 + \frac{\sigma}{\sqrt{T}} Y \Delta t + \sigma \Delta B_0\right] = X_{t_0} \left[1 + \sigma \Delta W_0\right] \\ X_{t_2} &= X_{t_1} \left[1 + \frac{\sigma}{\sqrt{T}} Y \Delta t + \sigma \Delta B_1\right] = X_{t_1} \left[1 + \sigma \Delta W_1\right] \\ &\vdots = \vdots = \vdots \\ X_{t_N} &= X_{t_{N-1}} \left[1 + \frac{\sigma}{\sqrt{T}} Y \Delta t + \sigma \Delta B_{N-1}\right] = X_{t_{N-1}} \left[1 + \sigma \Delta W_{N-1}\right], \end{aligned} \tag{26}$$

implying that

$$\overline{X}(T) = X_0 \prod_{i=0}^{N-1} \left[1 + \frac{\sigma}{\sqrt{T}} Y \Delta t + \sigma \Delta B_i\right]. \tag{27}$$

Therefore, in order to determine y_*, we need to solve

$$x = \overline{X}(T; y_*, B) = X_0 \prod_{i=0}^{N-1} \left[1 + \frac{\sigma}{\sqrt{T}} y_*(x) \Delta t + \sigma \Delta B_i\right], \tag{28}$$

which implies that the location of the kink point for the approximate problem is equivalent to finding the roots of the polynomial $P(y_*(K))$, given by

$$P(y_*(K)) = \prod_{i=0}^{N-1} \left[1 + \frac{\sigma}{\sqrt{T}} y_*(K)\Delta t + \sigma \Delta B_i\right] - \frac{K}{X_0}. \tag{29}$$

The exact location of the kink can be obtained exactly by solving exactly $P(y_*(K)) = 0$.

Approximate location of the discrete problem

Here, we try to find the roots of polynomial $P(y_*(K))$, given by (29), by using **Newton iteration method**. In this case, we need the expression $P' = \frac{dP}{dy_*}$. If we denote $f_i(y) = 1 + \frac{\sigma}{\sqrt{T}} y\Delta t + \sigma \Delta B_i$, then we can easily show that

$$P'(y) = \frac{\sigma \Delta t}{\sqrt{T}} \left(\prod_{i=0}^{N-1} f_i(y)\right) \left[\sum_{i=0}^{N-1} \frac{1}{f_i(y)}\right] \tag{30}$$

5.2 Results for the binary option example

In this case, the integrand $H(B)$ is given by

$$\begin{aligned} H(B) &= \int \mathbf{1}_{X(T;y,B)>K} \exp(-y^2/2) dy \\ &= \sqrt{2\pi} P(Y > y_*(K)), \end{aligned} \tag{31}$$

where $y_*(x)$, is an invertible function that satisfies

$$X(T; y_*(x), B) = x \tag{32}$$

We get the kink point by running Newton iteration with a precision of 10^{-10}.

The paramters that we used in our numerical experiments are: $T = 1$, $\sigma = 0.4$ and $S_0 = K = 100$. The exact value of this case is 0.42074.

5.2.1 Comparing relative errors

In the following, we compare the relative errors for the binary option example under Black-Scholes model (see Tables (1, 4)). We report the results for 2 scenarios: i) Without using Richardson extrapolation, ii) Using level 1 Richardson extrapoaltion. You may see appendix A for the values of binary option prices.

Given the normalized bias computed by MC method (See Section 5.2.2) (reported as bold values in the tables), we report in red in each table the smallest tolerance that MISC required to get below that relative bias (I do not put values for smaller tolerances, once the required bias is reached).

From the tables (1, 4)), we may observe that to get a relative error below 1%, we need around 16 time steps for the case without Richardson extrapolation compared to only using 1 time step in the coarse level for the case of level 1 Richardson extraplation.

Method \Steps	2	4	8	16
MISC ($TOl = 5.10^{-1}$)	0.0982	0.0468	0.0219	0.0102
MC method ($M = 10^6$)	**0.0999**	**0.0470**	**0.0224**	**0.0099**

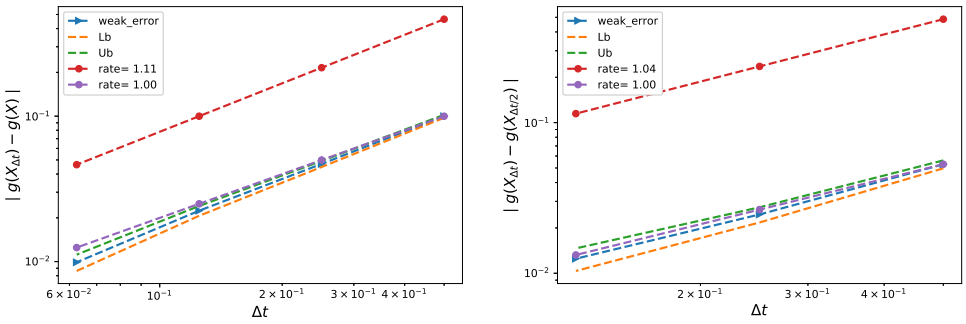
Table 1: Relative error of the binary option price of the different tolerances for different number of time steps, without Richardson extrapolation

Method \Steps	$1-2$	$2-4$	$4-8$	$8-16$
MISC ($TOl = 5.10^{-1}$)	0.0076	0.0045	0.0031	0.0017
MISC ($TOl = 10^{-2}$)	–	0.0036	0.0031	0.0014
MISC ($TOl = 10^{-3}$)	–	–	0.0021	0.0005
MC method ($M = 10^6$)	**0.0073**	**0.0038**	**0.0022**	**0.0006**

Table 2: Relative error of the binary option price of the different tolerances for different number of time steps, using Richardson extrapolation (level 1)

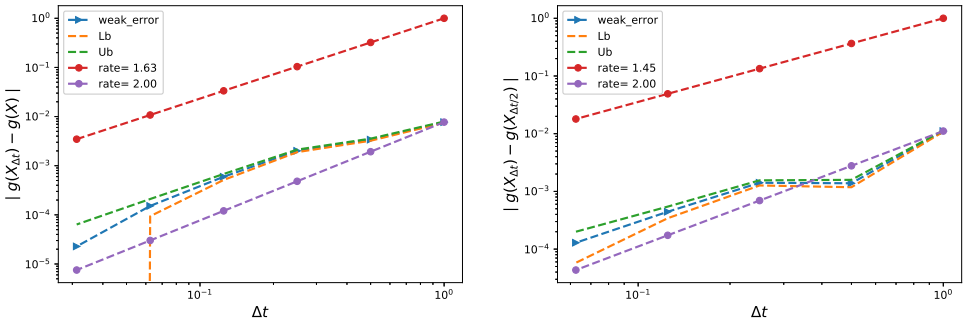
5.2.2 Weak error plots

From figure 1, we see that we get a weak error of order Δt. From figure 2, we observe no improvement in the rate when using level 1 of Richardson extrapolation but it is clear that we get a significant improvement in the constant. The upper and lower bounds are 95% confidence interval.



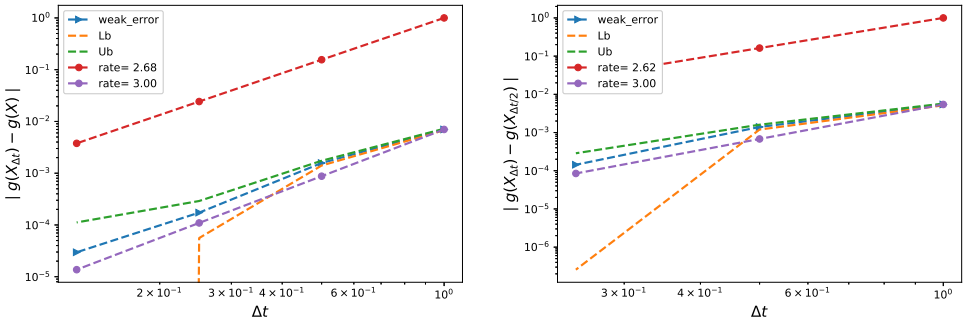
(a) (b)

Figure 1: The rate of convergence of the weak error for the binary option, without Richardson extraploation, using MC with $M = 10^4$: a) $|\mathrm{E}[g(X_{\Delta t})] - g(X)|$ b) $\left|\mathrm{E}\left[g(X_{\Delta t}) - g(X_{\Delta t/2})\right]\right|$



(a) (b)

Figure 2: The rate of convergence of the weak error for the binary option, with Richardson extraploation, using MC with $M = 10^6$: a) $\left|\mathrm{E}\left[2g(X_{\Delta t/2}) - g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\mathrm{E}\left[3g(X_{\Delta t/2}) - g(X_{\Delta t}) - 2g(X_{\Delta t/4})\right]\right|$



(a) (b)

Figure 3: The rate of convergence of the weak error for the binary option, with Richardson extraploation (level 2), using MC with $M = 5.10^6$: a) $\left|\frac{1}{3}\mathrm{E}\left[8g(X_{\Delta t/4}) - 6g(X_{\Delta t/2}) + g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\frac{1}{3}\mathrm{E}\left[-8g(X_{\Delta t/8}) + 14g(X_{\Delta t/4}) - 7(X_{\Delta t/2}) + g(X_{\Delta t})\right]\right|$

5.3 Results for the Call option example

In this case, the integrand $H(B)$ is given by

$$H(B) = \int_{\Omega} \max\left(X(T; y, B) - K, 0\right) \exp(-y^2/2) dy \tag{33}$$

We get the kink point by running Newton iteration with a precision of 10^{-10}. We decompose the total integration domain Ω into sub-domains Ω_i, $i = 1, 2$ such that the integrand is smooth in the interior of Ω_i and such that the kink is located along the boundary of these areas. The total

integral is then given as the sum of the separate integrals, *i.e.*

(34)
$$H(B) := \int_{\Omega} \max\left(X(T;y,B) - K, 0\right) \exp(-y^2/2) dy = \sum_{i=1}^{2} \int_{\Omega_i} \max\left(X(T;y,B) - K, 0\right) \exp(-y^2/2) dy,$$

where we use Gauss-laguerre quadrature with 10 points to get each part.

The paramters that we used in our numerical experiments are: $T = 1$, $\sigma = 0.4$ and $S_0 = K = 100$. The exact value of this case is 15.8519.

5.3.1 Comparing relative errors

In the following, we compare the relative errors for the call option example under Black-Scholes model(see Tables (3, 4, 5)). We report the results for 3 scenarios: i) Without using Richardson extrapolation, ii) Using level 1 Richardson extrapoaltion, iii) Using level 2 Richardson extrapoaltion. You may see appendix B for the values of call option prices.

Given the normalized bias computed by MC method (See Section 5.3.2) (reported as bold values in the tables), we report in red in each table the smallest tolerance that MISC required to get below that relative bias (I do not put values for smaller tolerances, once the required bias is reached). In case I do not reach those bias I put the best value that I get with MISC in red.

From the tables (3, 4, 5)), we may observe that to get a relative error below 0.5%, we need more than 16 time steps for the case without Richardson extrapolation compared to only using 4 time step in the coarse level for the case of level 1 Richardson extraplation, and only using 1 time step in the coarse level for the case of level 2 Richardson extraplation.

Method \Steps	2	4	8	16
MISC ($TOl = 5.10^{-1}$)	0.0229	0.0179	0.0111	0.0068
MISC ($TOl = 10^{-3}$)	–	0.0177	–	0.0066
MC method ($M = 10^5$)	**0.0231**	**0.0175**	**0.0111**	**0.0064**

Table 3: Relative error of the call option price of the different tolerances for different number of time steps, without Richardson extrapolation

Method \Steps	$1-2$	$2-4$	$4-8$	$8-16$
MISC ($TOl = 5.10^{-1}$)	0.0372	0.0129	0.0043	0.0025
MISC ($TOl = 10^{-3}$)	–	0.0126	0.0042	0.0023
MC method ($M = 10^5$)	**0.0374**	**0.0116**	**0.0027**	**0.0022**

Table 4: Relative error of the call option price of the different tolerances for different number of time steps, using Richardson extrapolation (level 1)

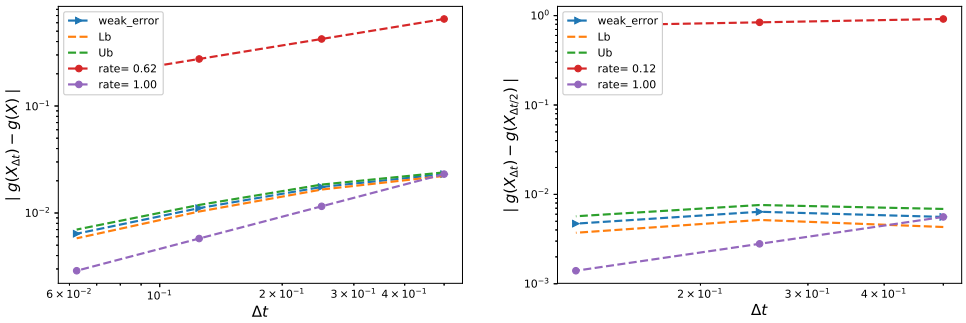
Method \Steps	$1-2-4$	$2-4-8$	$4-8-16$
MISC ($TOl = 5.10^{-1}$)	0.0047	0.0015	0.0019
MISC ($TOl = 10^{-3}$)	0.0043	0.0013	0.0017
MC method ($M = 10^6$)	**0.0041**	**0.0013**	**0.0016**

Table 5: Relative error of the call option price of the different tolerances for different number of time steps, using Richardson extrapolation (level 2)

5.3.2 Weak error plots

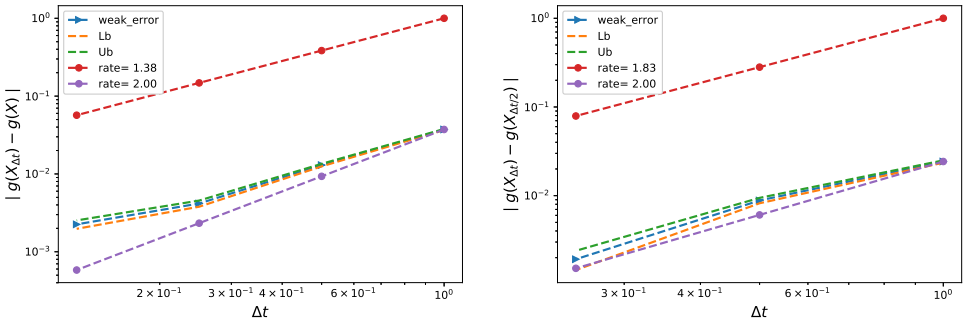
$\beta = 10$

From figure 4, we see that we get a weak error of order Δt. From figure 5, we observe that we get a weak error of order Δt^2 (if eliminate the last point having a wide confidence interval point). From figure 6, we observe that we get an almost constant weak error, with that constant being the smallest cpmpared to without and with level 1 Ricardson extrapolation. The upper and lower bounds are 95% confidence interval.



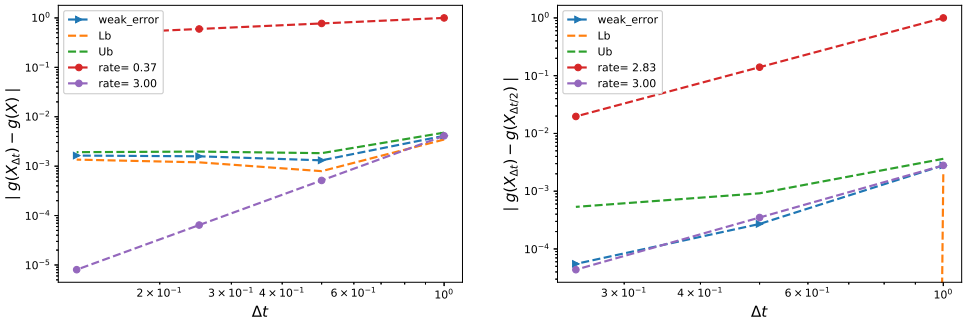
(a) (b)

Figure 4: The rate of convergence of the weak error for the call option, without Richardson extraploation, using MC with $M = 10^5$: a) $|\mathrm{E}[g(X_{\Delta t})] - g(X)|$ b) $\left|\mathrm{E}\left[g(X_{\Delta t}) - g(X_{\Delta t/2})\right]\right|$



(a) (b)

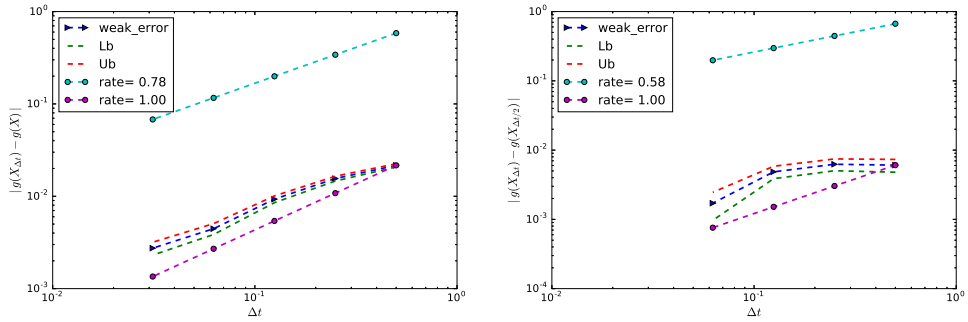
Figure 5: The rate of convergence of the weak error for the call option with Richardson extraploation, using MC with $M = 10^6$: a) $\left|\mathrm{E}\left[2g(X_{\Delta t/2}) - g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\mathrm{E}\left[3g(X_{\Delta t/2}) - g(X_{\Delta t}) - 2g(X_{\Delta t/4})\right]\right|$



(a) (b)

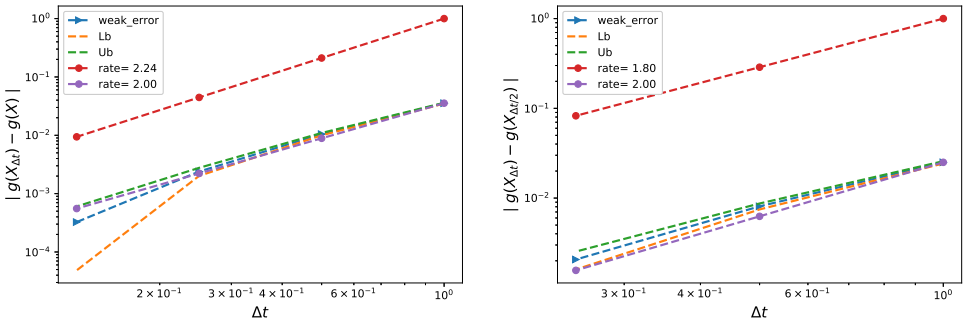
Figure 6: The rate of convergence of the weak error for the call option with Richardson extraploation (level 2), using MC with $M = 10^6$: a) $\left|\frac{1}{3}\mathrm{E}\left[8g(X_{\Delta t/4}) - 6g(X_{\Delta t/2}) + g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\frac{1}{3}\mathrm{E}\left[-8g(X_{\Delta t/8}) + 14g(X_{\Delta t/4}) - 7(X_{\Delta t/2}) + g(X_{\Delta t})\right]\right|$

$\beta = 32$



(a) (b)

Figure 7: The rate of convergence of the weak error for the call option, without Richardson extraploation, using MC with $M = 10^5$: a) $|\mathrm{E}\left[g(X_{\Delta t})\right] - g(X)|$ b) $\left|\mathrm{E}\left[g(X_{\Delta t}) - g(X_{\Delta t/2})\right]\right|$



(a) (b)

Figure 8: The rate of convergence of the weak error for the call option with Richardson extraploation (level 1), using MC with $M = 5.10^6$: a) $\left|\mathrm{E}\left[2g(X_{\Delta t/2}) - g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\mathrm{E}\left[3g(X_{\Delta t/2}) - g(X_{\Delta t}) - 2g(X_{\Delta t/4})\right]\right|$

References Cited

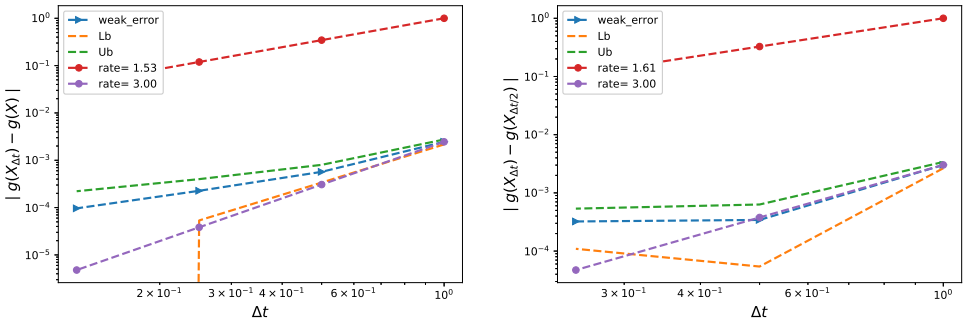
[1] CHRISTIAN BAYER, MARKUS SIEBENMORGEN, and RAUL TEMPONE. Smoothing the payoff for efficient computation of basket option pricing.

[2] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta numerica*, 13:147–269, 2004.

[3] Thomas Gerstner. Sparse grid quadrature methods for computational finance.

[4] Thomas Gerstner and Markus Holtz. Valuation of performance-dependent options. *Applied Mathematical Finance*, 15(1):1–20, 2008.

[5] Paul Glasserman. *Monte Carlo methods in financial engineering*. Springer, New York, 2004.



(a) (b)

Figure 9: The rate of convergence of the weak error for the call option with Richardson extraploation (level 2), using MC with $M = 5.10^6$: a) $\left|\frac{1}{3}\mathrm{E}\left[8g(X_{\Delta t/4}) - 6g(X_{\Delta t/2}) + g(X_{\Delta t})\right] - g(X)\right|$ b) $\left|\frac{1}{3}\mathrm{E}\left[-8g(X_{\Delta t/8}) + 14g(X_{\Delta t/4}) - 7(X_{\Delta t/2}) + g(X_{\Delta t})\right]\right|$

[6] Michael Griebel, Frances Kuo, and Ian Sloan. The smoothing effect of integration in ˆ{} and the anova decomposition. *Mathematics of Computation*, 82(281):383–400, 2013.

[7] Michael Griebel, Frances Kuo, and Ian Sloan. Note on the smoothing effect of integration in ˆ{} and the anova decomposition. *Mathematics of Computation*, 86(306):1847–1854, 2017.

[8] Andreas Griewank, Frances Y Kuo, Hernan Leövey, and Ian H Sloan. High dimensional integration of kinks and jumps–smoothing by preintegration. *arXiv preprint arXiv:1712.00920*, 2017.

[9] Abdul-Lateef Haji-Ali, Fabio Nobile, Lorenzo Tamellini, and Raul Tempone. Multi-index stochastic collocation for random pdes. *Computer Methods in Applied Mechanics and Engineering*, 306:95–122, 2016.

[10] Chengfeng Weng, Xiaoqun Wang, and Zhijian He. Efficient computation of option prices and greeks by quasi–monte carlo method with smoothing and dimension reduction. *SIAM Journal on Scientific Computing*, 39(2):B298–B322, 2017.

A Prices for different methods for binary option

Method \Steps	2	4	8	16
MISC ($TOl = 5.10^{-1}$)	0.4620	0.4404	0.4299	0.4250

Table 6: Binary option price of the different methods for different number of time steps, without Richardson extrapolation.

Method \Steps	$1-2$	$2-4$	$4-8$	$8-16$
MISC ($TOl = 5.10^{-1}$)	0.4239	0.4188	0.4194	0.4200
MISC ($TOl = 10^{-2}$)	–	0.4192	0.4194	0.4201
MISC ($TOl = 10^{-3}$)	–	–	0.4198	0.4205

Table 7: Binary option price of the different methods for different number of time steps, with Richardson extrapolation (level 1).

B Call prices for different methods

Method \Steps	2	4	8	16
MISC ($TOl = 5.10^{-1}$)	16.2145	16.1352	16.0275	15.9594
MISC ($TOl = 10^{-3}$)	–	16.1328	–	–

Table 8: Call option price of the different methods for different number of time steps, without Richardson extrapolation.

Method \Steps	$1-2$	$2-4$	$4-8$	$8-16$
MISC ($TOl = 5.10^{-1}$)	16.4419	16.0558	15.9205	15.8915
MISC ($TOl = 10^{-3}$)	–	16.0510	15.9177	15.8880

Table 9: Call option price of the different methods for different number of time steps, with Richardson extrapolation (level 1).

Method \Steps	$1-2-4$	$2-4-8$	$4-8-16$
MISC ($TOl = 5.10^{-1}$)	15.9269	15.8756	15.8815
MISC ($TOl = 10^{-3}$)	15.9205	15.8732	15.8789

Table 10: Call option price of the different methods for different number of time steps, with Richardson extrapolation (level 2).

C The basket option with the smoothing trick as in [1]

The first experiment that we consider is the pricing of a European basket call option in a Black-Scholes model. The basket is composed of d assets ($d = 3, 8, 25$) and we use the same trick of smoothing the integrand that was proposed in [1]. In this case, the dimension of the parameter space $N = d-1$. The interpolation over the parameter space is based on the tensorized Lagrangian interpolation technique with Gaussian points.

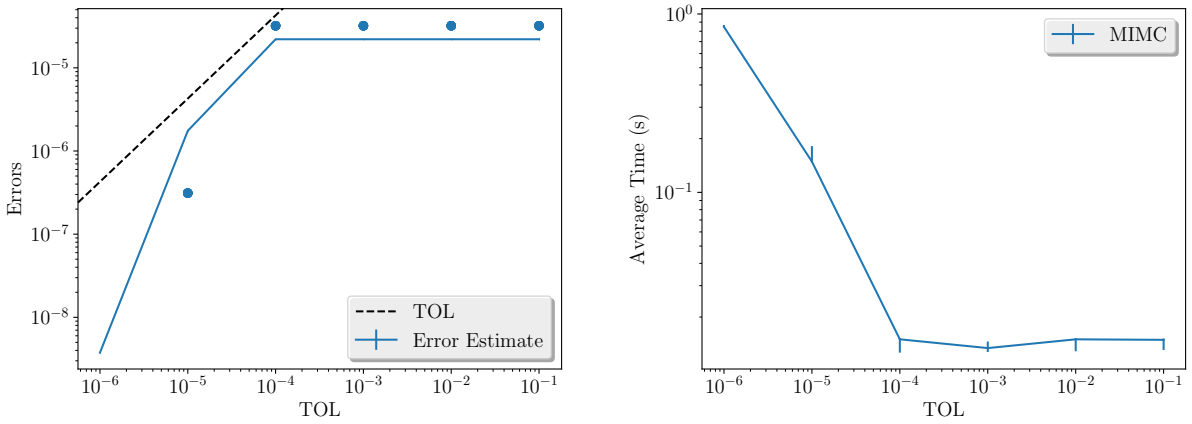
C.0.1 Results using MISC

In table 11, we summarize the observed complexity rates for different tested settings for the basket example. From this table, we can check that even with the 25 dimensional case, the complexity rate in terms of the elapsed time is at least order 1, which is better than MC, which is 2. Detailed plots for each case are given by figures (10, 11) for $d = 3$, figures (12, 13) for $d = 8$ and figures (14, 15) for $d = 25$. Mainly, from the plots, we checked that we achive the prescribed tolerance using MISC, the convergence rates of mixed differences which is a basic assumption for using MISC (we observe exponential decay of error rates wrt to the number of quadrature points) and finally the complexity rates.

# assets \	3	8	25
rate	$-1/3$	$-9/20$	$-16/25$

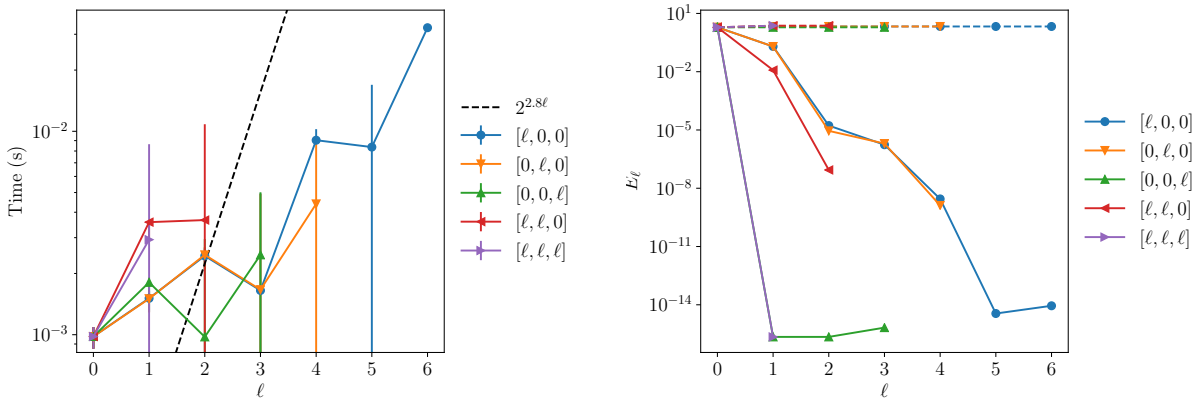
Table 11: Complexity rates of the different experiments for the basket option using BS model

Case of 3-dimensional Basket



(a) Error estimate (b) Average running time as a function of TOL

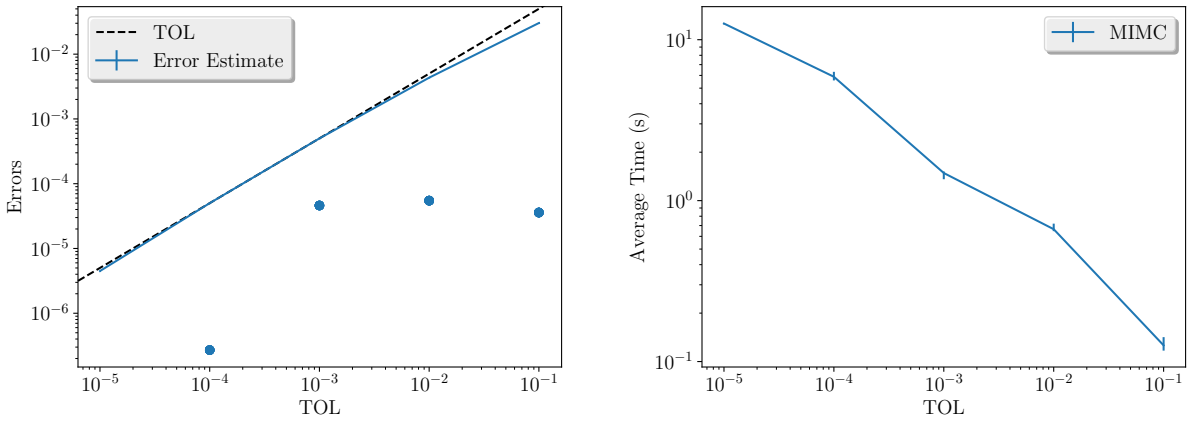
Figure 10: Convergence and complexity results for the 3-dimensional basket option using BS model.



(a) Average Computational time per level. (b) The convergence rate of mixed differences per level.

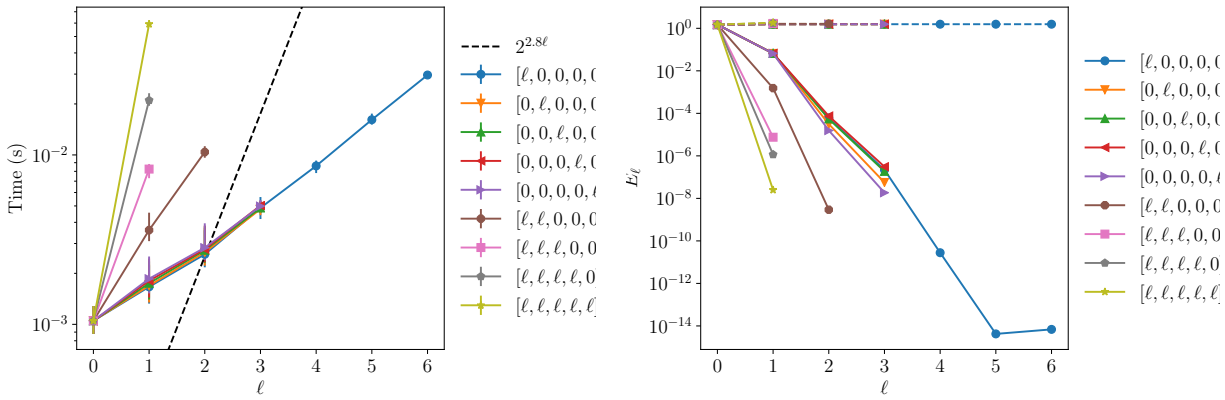
Figure 11: Convergence and work rates for discretization levels for the 3-dimensional basket option using BS model.

Case of 8-dimensional Basket



(a) Error estimate (b) Average running time as a function of TOL

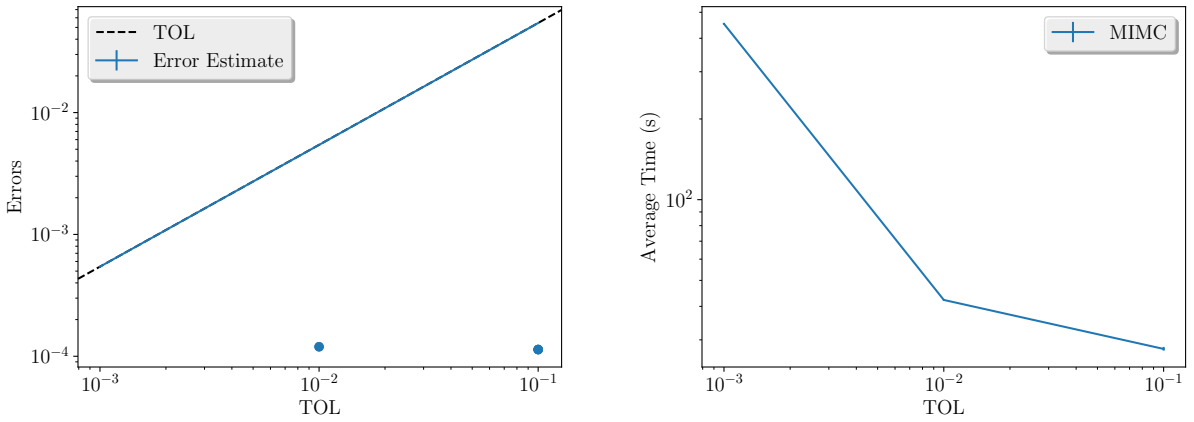
Figure 12: Convergence and complexity results for the 8-dimensional basket option using BS model.



(a) Average Computational time per level. (b) The convergence rate of mixed differences per level.

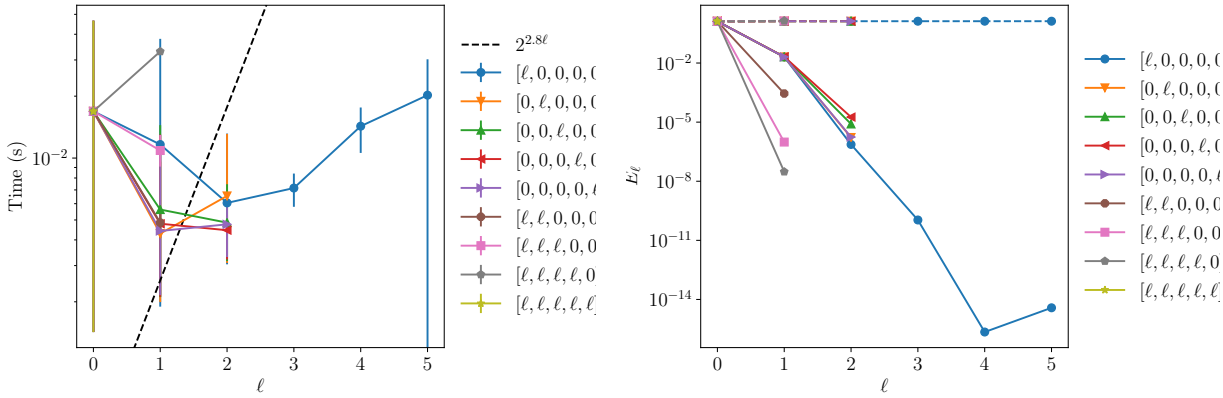
Figure 13: Convergence and work rates for discretization levels for the 8-dimensional basket option using BS model.

Case of 25-dimensional Basket



(a) Error estimate (b) Average running time as a function of TOL

Figure 14: Convergence and complexity results for the 25-dimensional basket option using BS model.



(a) Average Computational time per level. (b) The convergence rate of mixed differences per level.

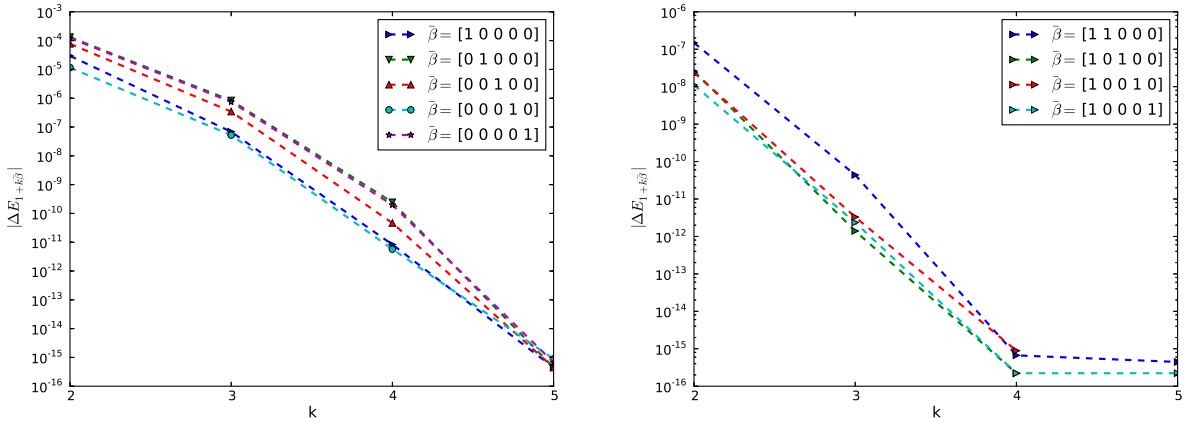
Figure 15: Convergence and work rates for discretization levels for the 25-dimensional basket option using BS model.

D Results of the basket example

D.0.1 Verifying bounds on work and error contributions

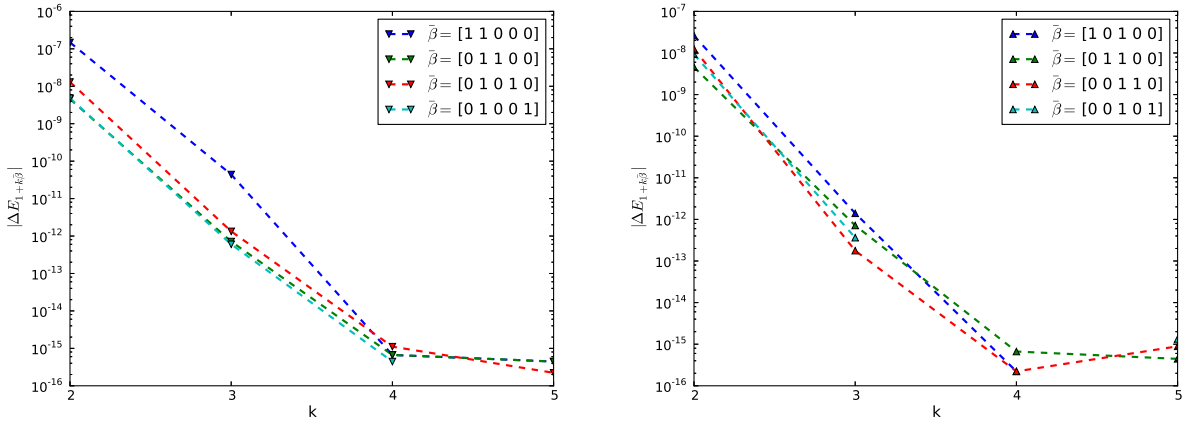
In this subsection, we discuss the validity of Assumptions 2-4 in [9], upon which the MISC convergence theorem is based. In this case, basket option case we do not use a deterministic solver since the we have a close for of the integrand and then we analyze only the properties of the collocation method applied to the problem.

Below, we show in figures (16,17,18) the rate of convergence of first and mixed differences for different orders and we check that the error is decaying exponentially wrt number of points used in the quadrature.



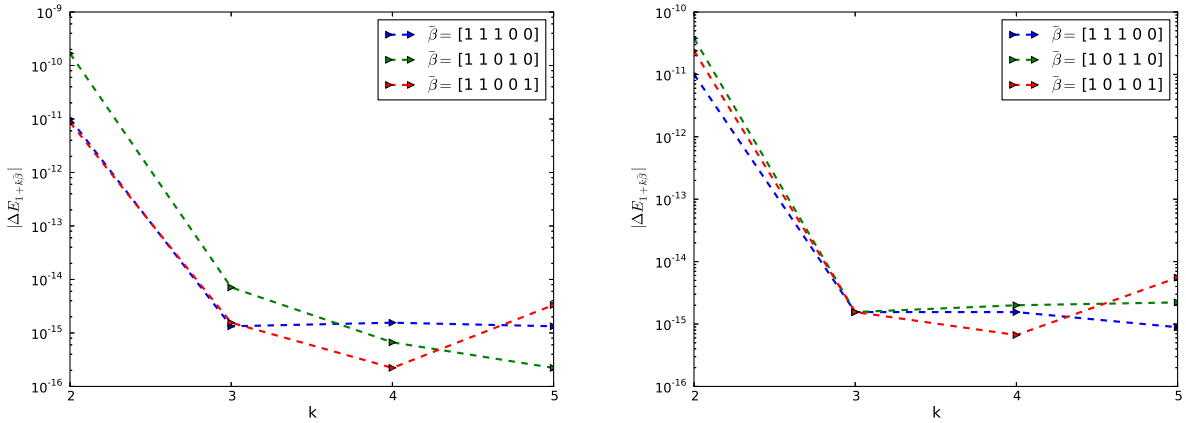
(a) (b)

Figure 16: The rate of convergence of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$): a) shows the rate of convergence of first order differences. b) shows the rate of convergence of second order mixed differences.



(a) (b)

Figure 17: The rate of convergence of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$)): a) and b) show the rate of convergence of second order mixed differences for different settings.



(a) (b)

Figure 18: The rate of convergence of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$)): a) and b) show the rate of convergence of third order mixed differences for different settings.

E Implementation details and MISC results

The first step, when testing this example is to fix $\Delta t_{\min}$ for the time discretization parameter and do not introduce a hierarchy of levels in the deterministic dimension. Therefore, we will only construct a hierarchy of levels in the stochastic dimension associated with the Brownian increments. Then, the second step is to avoid this restriction of minimum time discretization parameter (somehow like letting $\Delta t_{\min} \to 0$) and in that case, the problem will be close to do the integration in ∞-dimension, we need to see how the convergence rates will behave. The third step, is to use Richardson extrapolation which we expect bring the order of convergence for the Euler scheme from $o(\Delta t)$ to $o(\Delta t^2)$, and then we may expect that we will need fewer time steps to converge than the first case, and which means that fewer stochastic directions will be activated. One technical difference here is that for 1-dimensional, we will have two kink points instead of one but the procedure will be the same. We note that we used Laguerre quadrature for integrating with respect to the dimension containing the kink.

Since we observed some problems for the results of the non-smooth payoff with Richardson extrapolation, we tried to add another case where we use a smoothed version of the call payoff to figure out the source of the problem. Using a smooth payoff allow us to separate the effect of infinite dimension and non-smoothness. As a possible choice, we use the Black Scholes call option price formula, given by

$$\mathrm{C(S,t)} = \mathrm{N(d_1)S} - \mathrm{N(d_2)Ke}^{-rt} \tag{35}$$

$$\mathrm{d_1} = \frac{1}{\sigma\sqrt{\mathrm{t}}}\left[\ln\left(\frac{S}{K}\right) + t\left(r + \frac{\sigma^2}{2}\right)\right] \tag{36}$$

$$\mathrm{d_2} = d_1 - \sigma\sqrt{t} \tag{37}$$

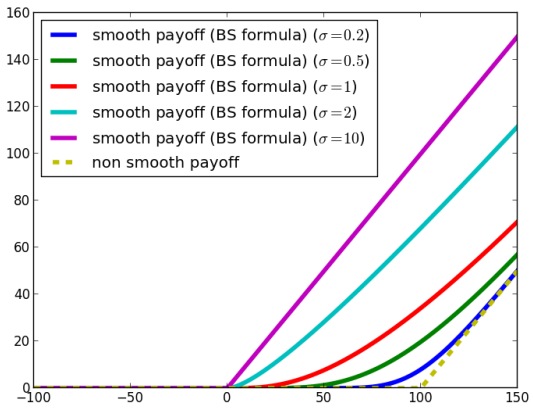
$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{1}{2}z^2} dz \tag{38}$$

, which can be seen as a smoothed version of the true payoff (See figure 19a). The two differences with respect to the non-smooth case is that the dimension of the parameter space (stochastic space) now is increased by one due to including the terminal value of the driving Brownian motion. Numerically, we observed weird results , in fact the quadrature performed worse than with the non-smooth case, even for various values of volatilities. We still need to understand why this choice is not a good one.

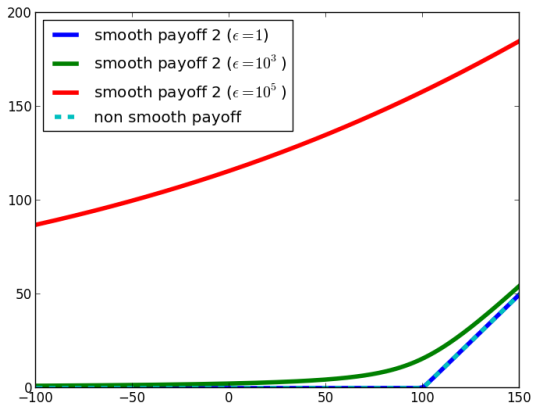
Another possible choice that we worked with, is by using the following smooth approximation of the max call payoff (See figure 19b)

$$\max_\epsilon(S-K,0) \approx 0.5(S-K+\sqrt{(S-K)^2+\epsilon}), \tag{39}$$

where $\epsilon > 0$.



(a) Black Scholes call option price for different values of σ



(b) Smooth approximation of the max call payoff for different values of ϵ.

In table 12, we summarize the results that we get in terms of complexity rates for different methods using different number of time steps. We clarify that the results for the Richardson extrapolation are given for just one level, where the associated steps in the table corresponds to those of level 1.

Method \Steps	2	4	8	16
Non-smooth payoff	$-1/5$	$-1/5$	$-2/5$	$-7/10$
Non-smooth payoff with Richardson extrapolation	–	–	–	–
Smooth payoff(given by (39),$\epsilon = 10^{-5}$)	$-1/3$	$-1/3$	$-3/5$	-1
Smooth payoff with Richardson extrapolation	-	$-1/10$	$-4/10$	–

Table 12: Complexity rates of the different methods for different number of time steps

Detailed plots for each case are given in appendix F. Mainly, from the plots, we checked that we achive the prescribed tolerance using MISC, the convergence rates of mixed differences which is a basic assumption for using MISC (we observe exponential decay of error rates wrt to the number of quadrature points) and finally the complexity rates.

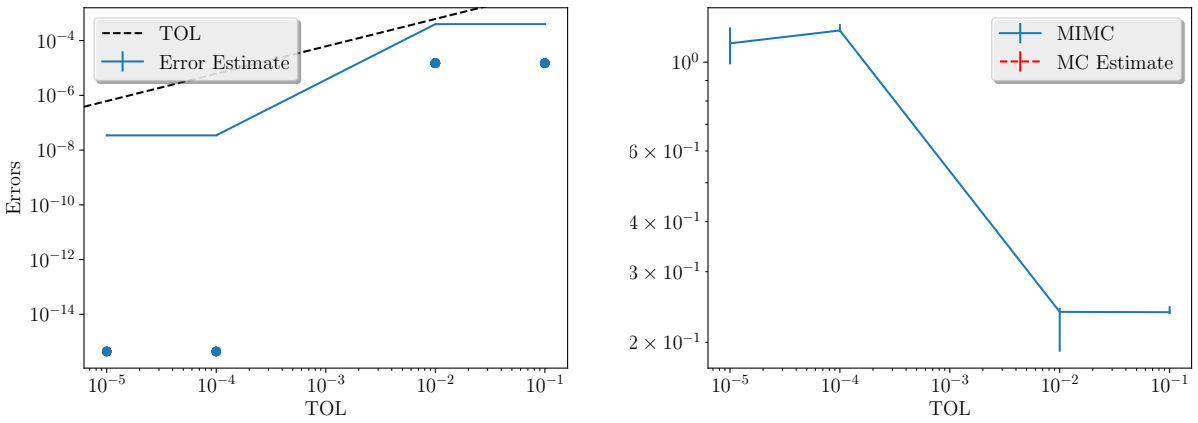
F MISC Results for the call option under discretized BS model

F.1 Case of 2 time steps with non-smooth payoff

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-1/5$ for 2 time steps.

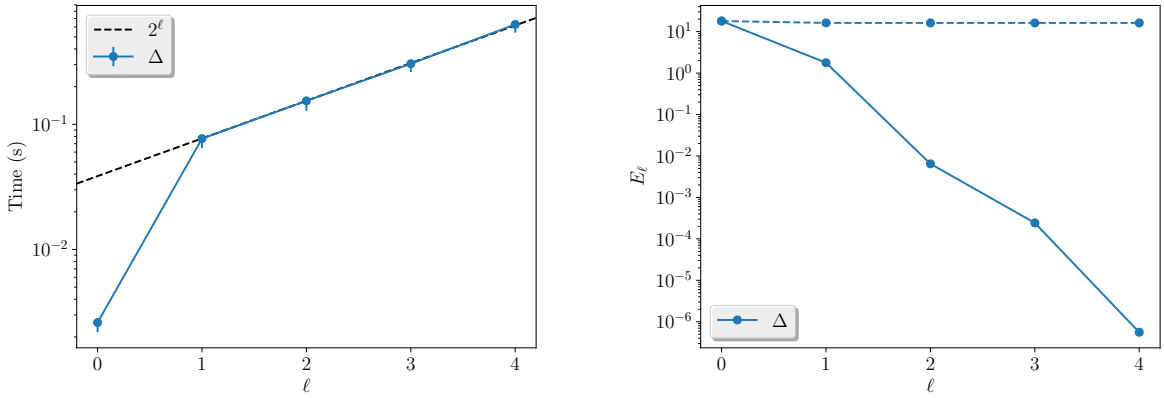
F.2 Case of 2 time steps with smooth payoff (given by (39))

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average



(a) Error estimate (b) Average running time as a function of TOL

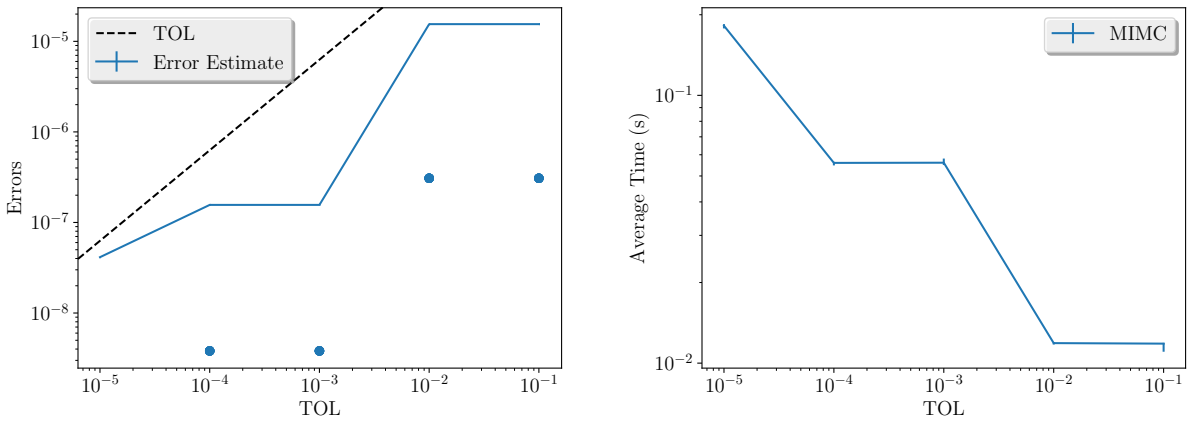
Figure 20: Convergence and complexity results for for the $1D$ BS model with the non-smooth call payoff.



(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 21: Convergence and work rates for discretization levels for the $1D$ BS model with the non-smooth call payoff.

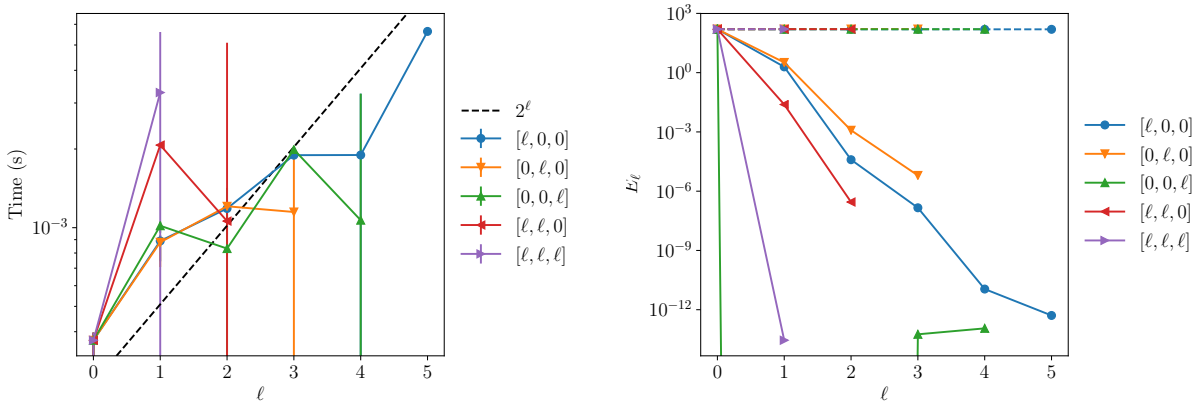
running time) approximately of order $-1/3$ for 2 time steps.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 22: Convergence and complexity results for for the $1D$ BS model with the smooth call payoff given by (39).



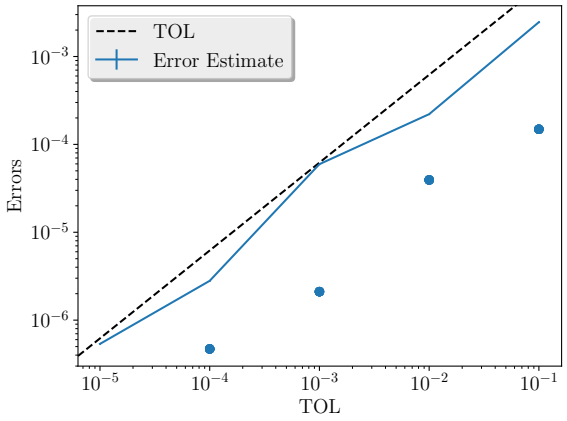
(a) Average Computational time per level

(b) The convergence rate of first differences per level

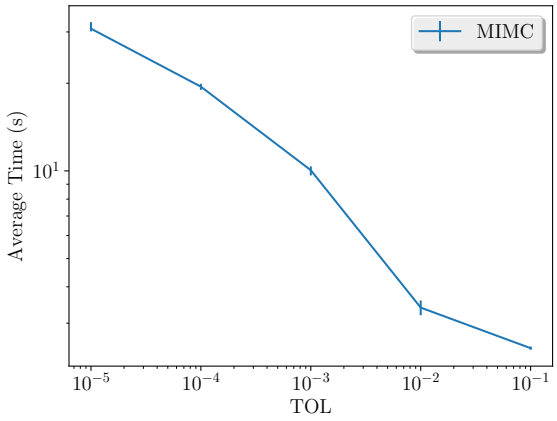
Figure 23: Convergence and work rates for discretization levels for the $1D$ BS model with the smooth call payoff given by (39).

F.3 Case of 4 time steps with non-smooth payoff

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-1/5$ for 4 time steps.

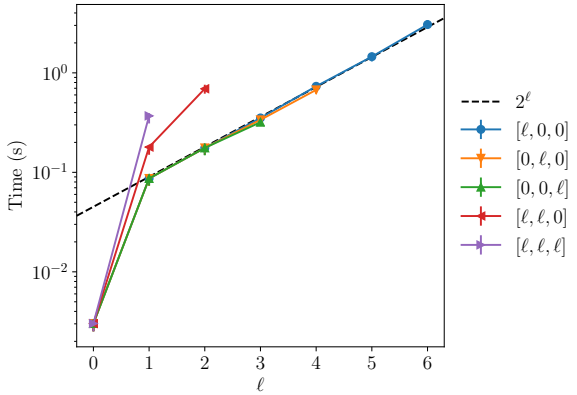


(a) Error estimate

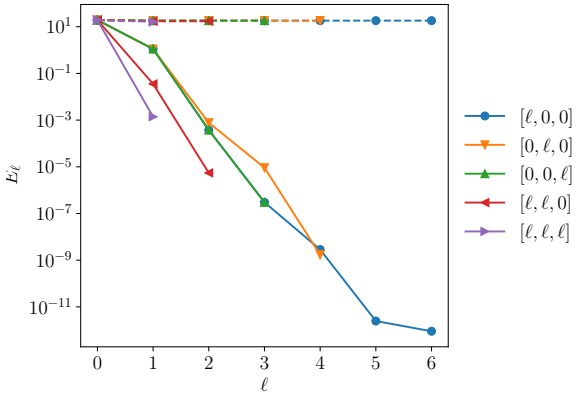


(b) Average running time as a function of TOL

Figure 24: Convergence and complexity results for for the $1D$ BS model with the non-smooth call payoff .



(a) Average Computational time per level

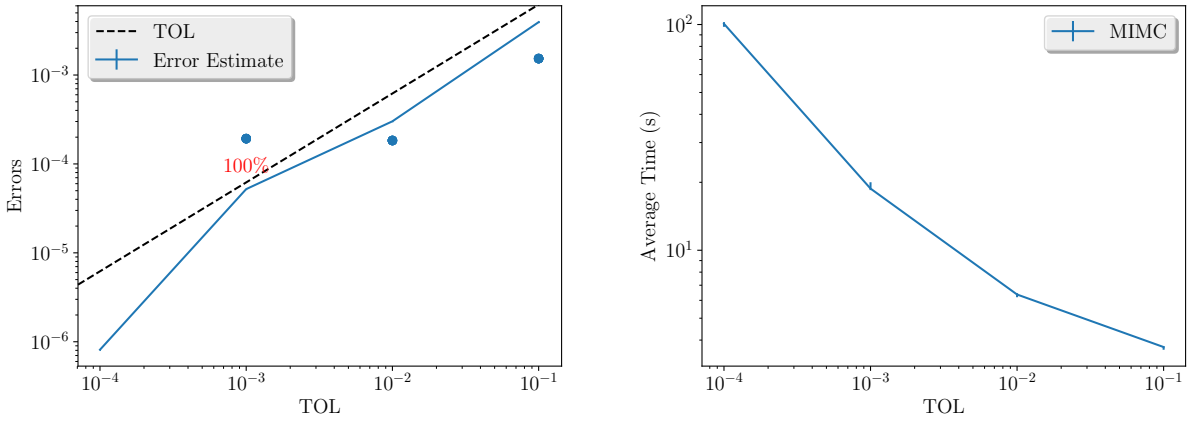


(b) The convergence rate of first differences per level

Figure 25: Convergence and work rates for discretization levels for the $1D$ BS model with the non-smooth call payoff.

F.4 Case of non smooth payoff with Richardson extrapolation (level 0: 2 time steps, level 1: 4 time steps) (Ignoring the middle part of integration)

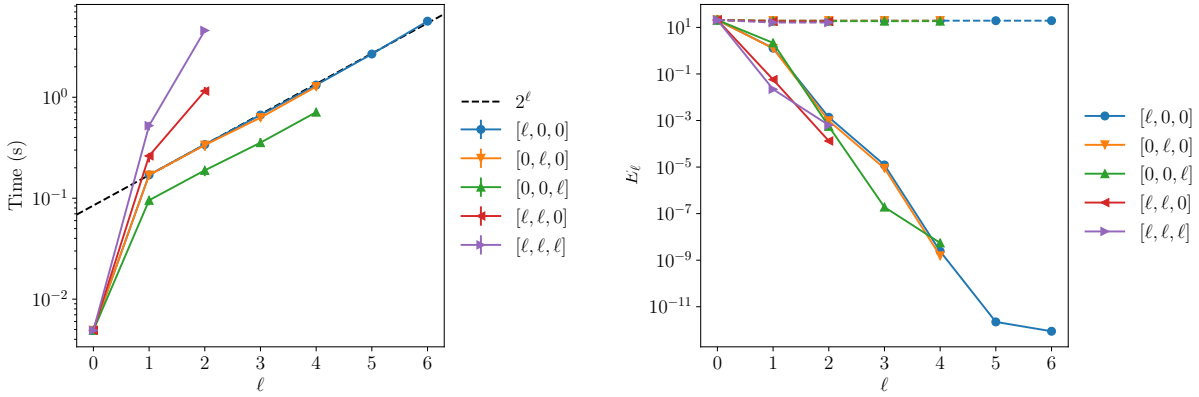
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-1/2$.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 26: Convergence and complexity results for for the $1D$ BS model with the non smooth call payoff with Richardson extrapolation (Ignoring the middle part of integration).



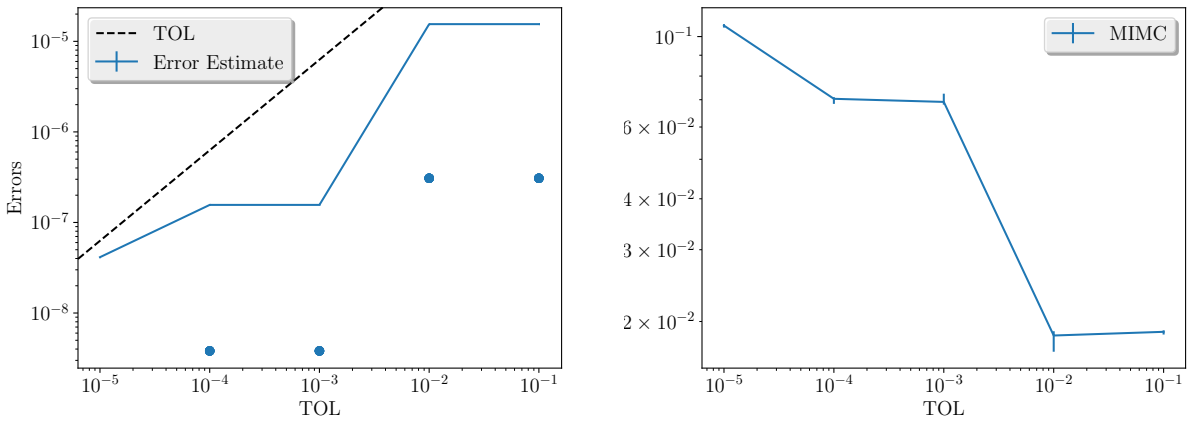
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 27: Convergence and work rates for discretization levels for the $1D$ BS model with the non smooth call payoff with Richardson extrapolation (Ignoring the middle part of integration).

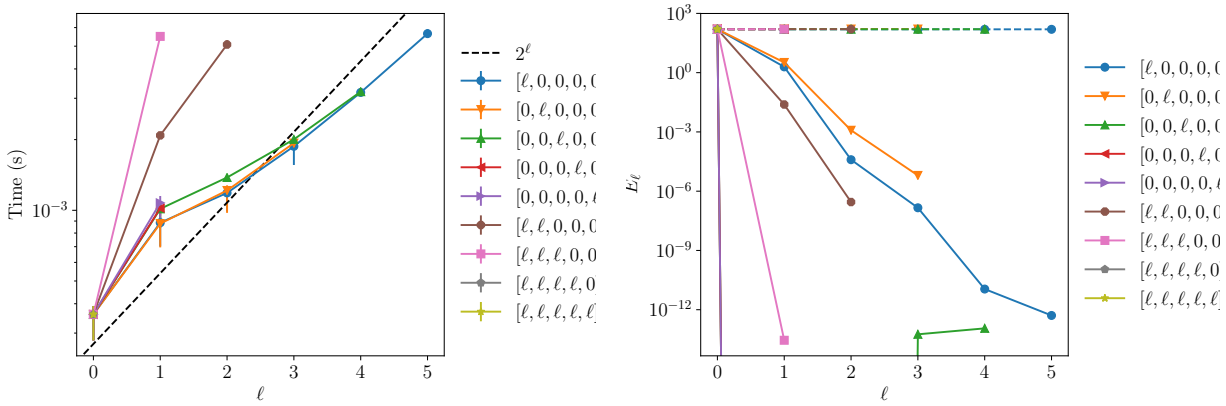
F.5 Case of 4 time steps with smooth payoff (given by (39))

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-1/3$ for 4 time steps.



(a) Error estimate (b) Average running time as a function of TOL

Figure 28: Convergence and complexity results for for the $1D$ BS model with the smooth call payoff given by (39).

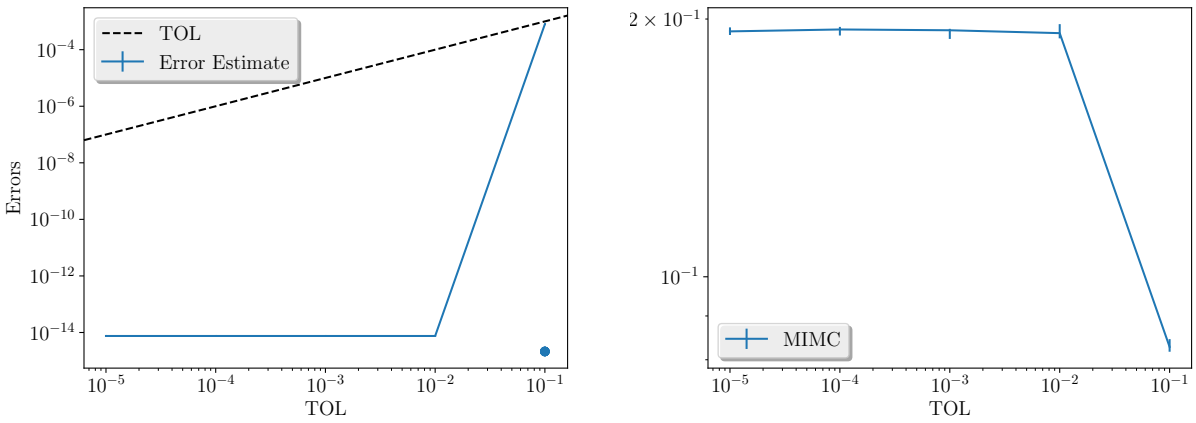


(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 29: Convergence and work rates for discretization levels for the $1D$ BS model with the smooth call payoff given by (39).

F.6 Case of smooth payoff (given by (39)) with Richardson extrapolation (level 0: 2 time steps, level 1: 4 time steps)

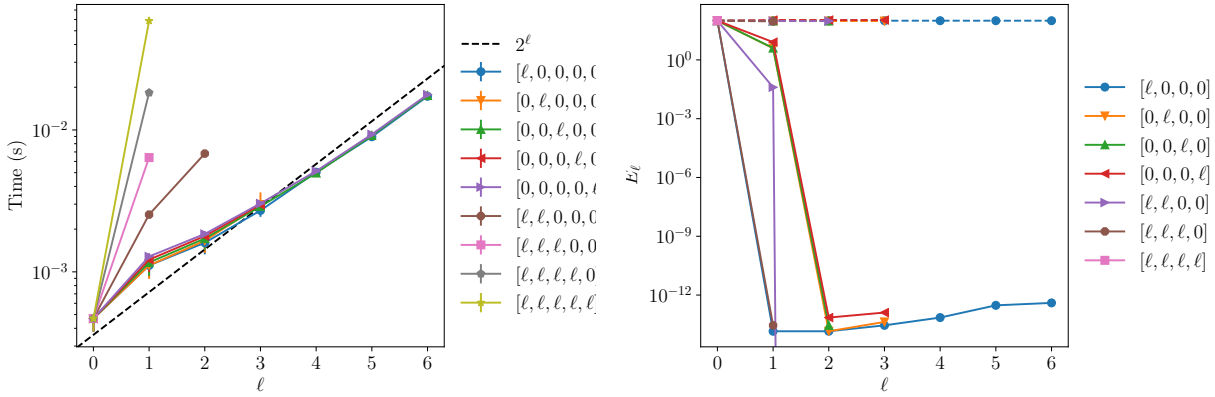
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-1/10$.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 30: Convergence and complexity results for for the $1D$ BS model with the smooth call payoff given by (39) with Richardson extrapolation.



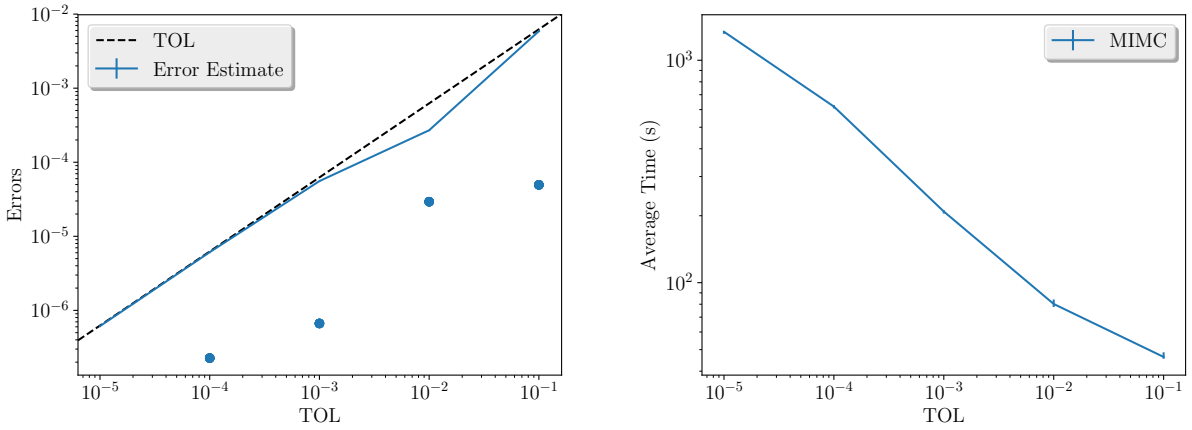
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 31: Convergence and work rates for discretization levels for the $1D$ BS model with the smooth call payoff given by (39) with Richardson extrapolation.

F.7 Case of 8 time steps with non-smooth payoff

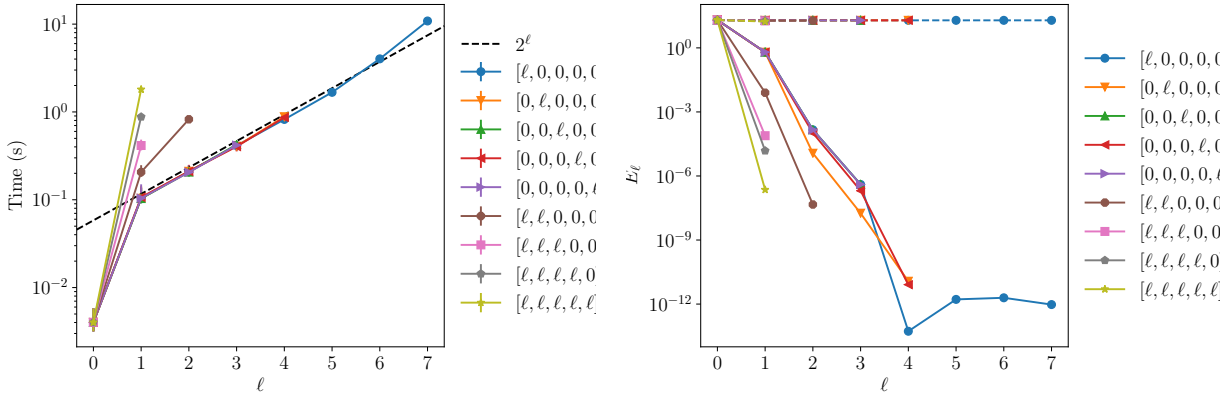
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-2/5$ for 8 time steps.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 32: Convergence and complexity results for for the $1D$ BS model with the non-smooth call payoff.



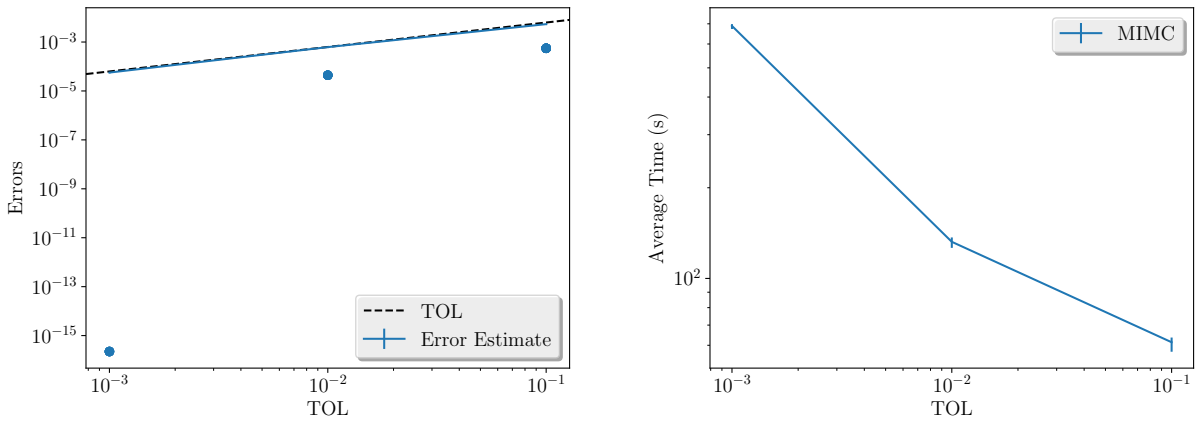
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 33: Convergence and work rates for discretization levels for the $1D$ BS model with the non-smooth call payoff.

F.8 Case of non smooth payoff with Richardson extrapolation (level 0: 4 time steps, level 1: 8 time steps) (Ignoring the middle part of integration)

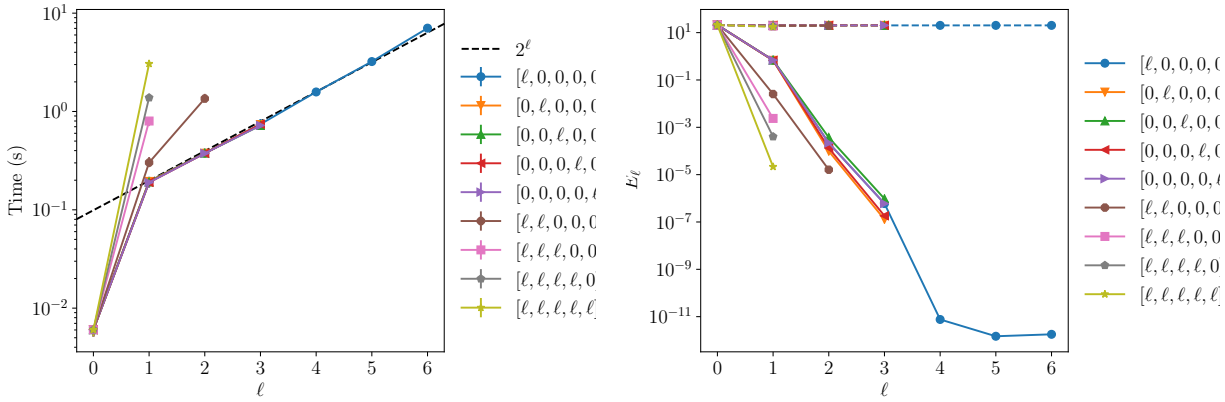
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-1/2$.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 34: Convergence and complexity results for for the $1D$ BS model with the non smooth call payoff with Richardson extrapolation (Ignoring the middle part of integration).



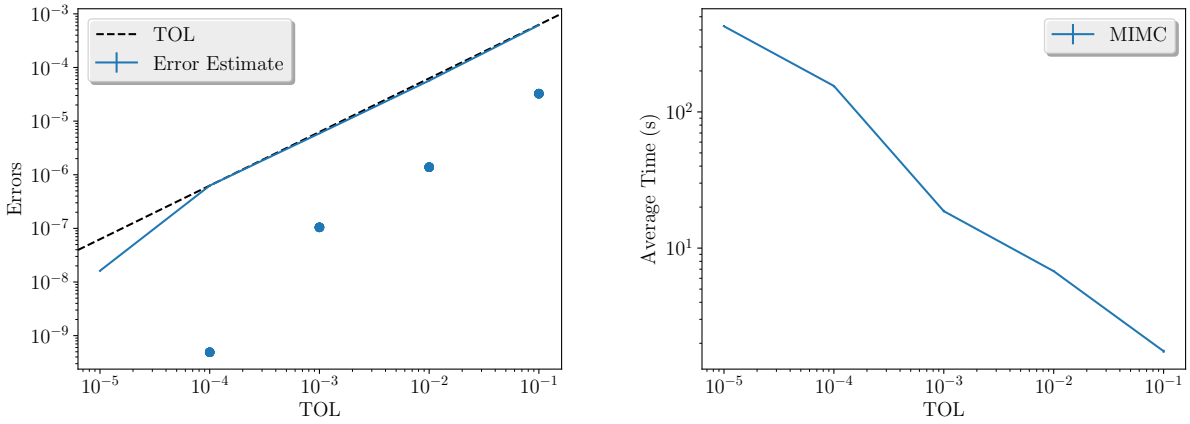
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 35: Convergence and work rates for discretization levels for the $1D$ BS model with the non smooth call payoff with Richardson extrapolation (Ignoring the middle part of integration).

F.9 Case of 8 time steps with smooth payoff (given by (39))

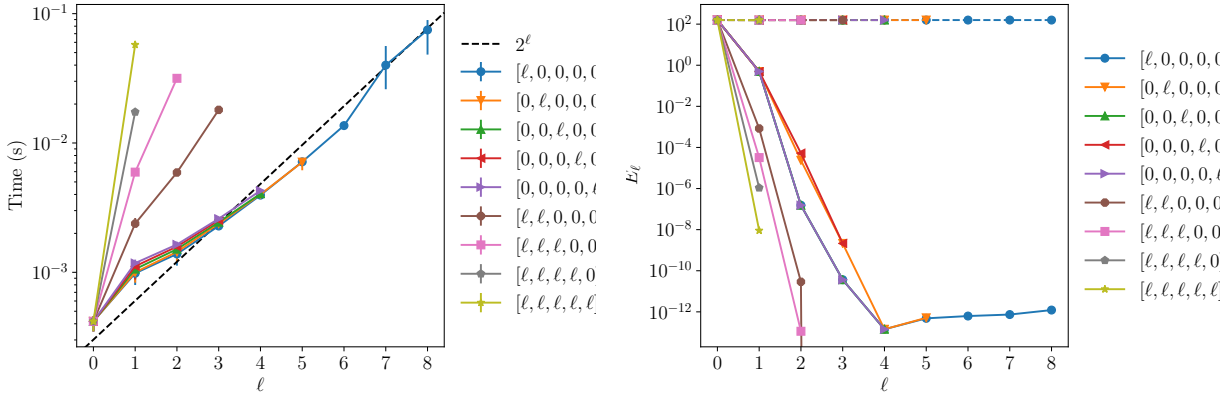
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-3/5$ for 8 time steps.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 36: Convergence and complexity results for for the $1D$ BS model with the smooth call payoff given by (39).



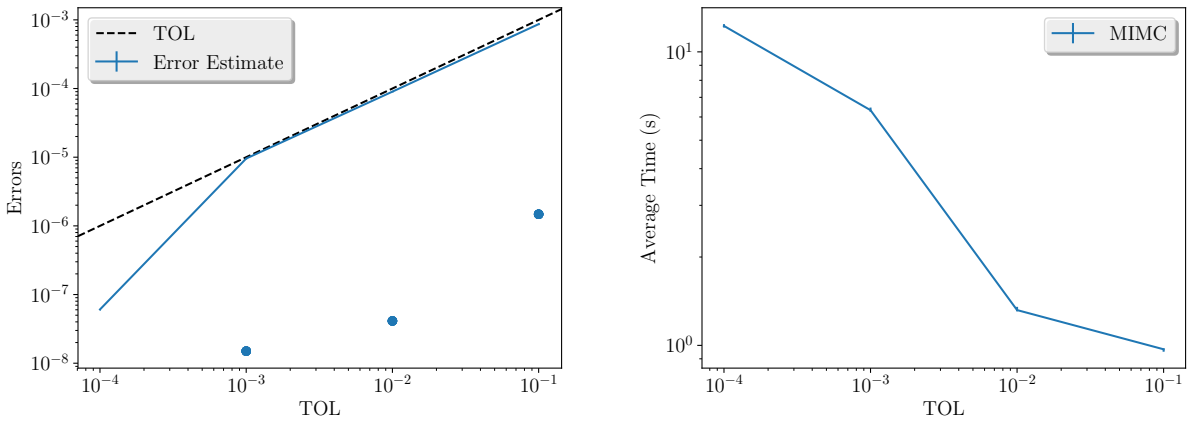
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 37: Convergence and work rates for discretization levels for the $1D$ BS model with the smooth call payoff given by (39).

F.10 Case of smooth payoff (given by (39)) with Richardson extrapolation (level 0: 4 time steps, level 1: 8 time steps)

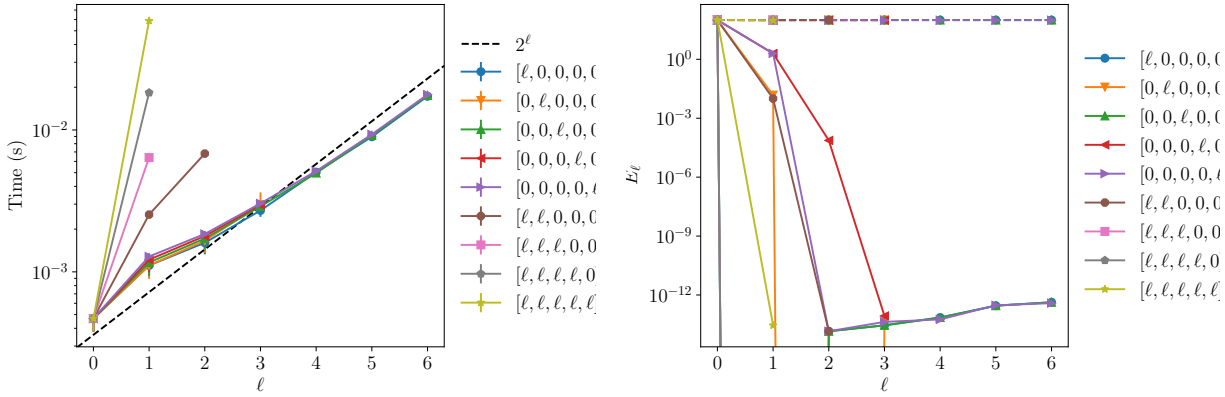
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-4/10$.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 38: Convergence and complexity results for for the $1D$ BS model with the smooth call payoff given by (39) with Richardson extrapolation.



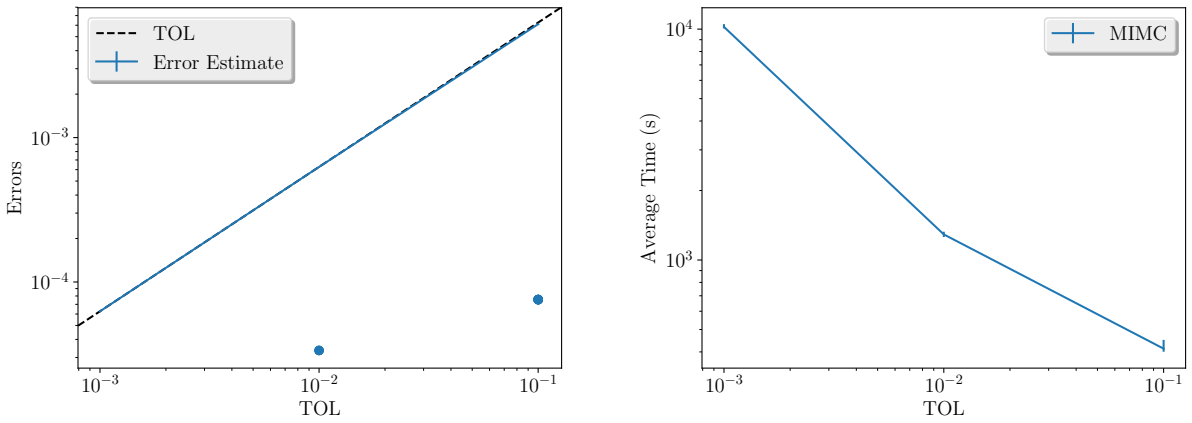
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 39: Convergence and work rates for discretization levels for the $1D$ BS model with the smooth call payoff given by (39) with Richardson extrapolation.

F.11 Case of 16 time steps with non-smooth payoff

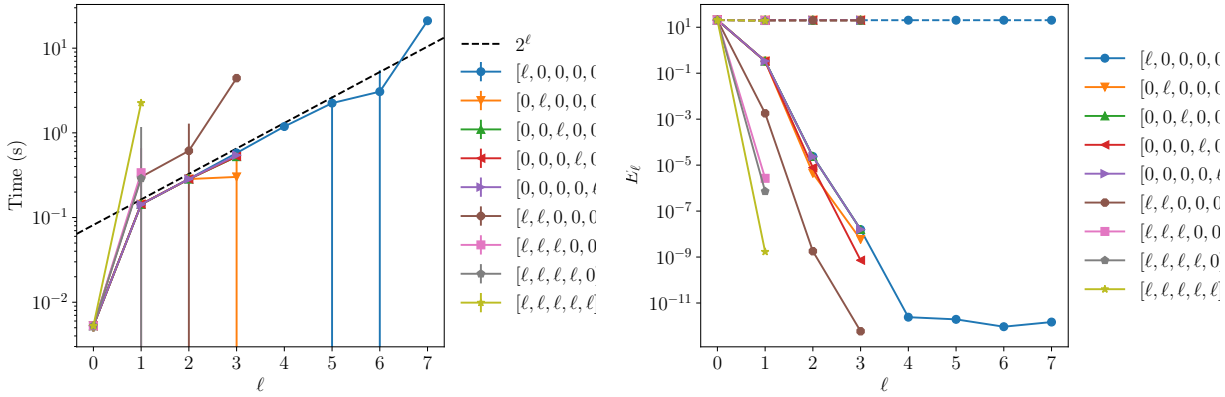
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-7/10$ for 16 time steps.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 40: Convergence and complexity results for for the $1D$ BS model with the non-smooth call payoff.



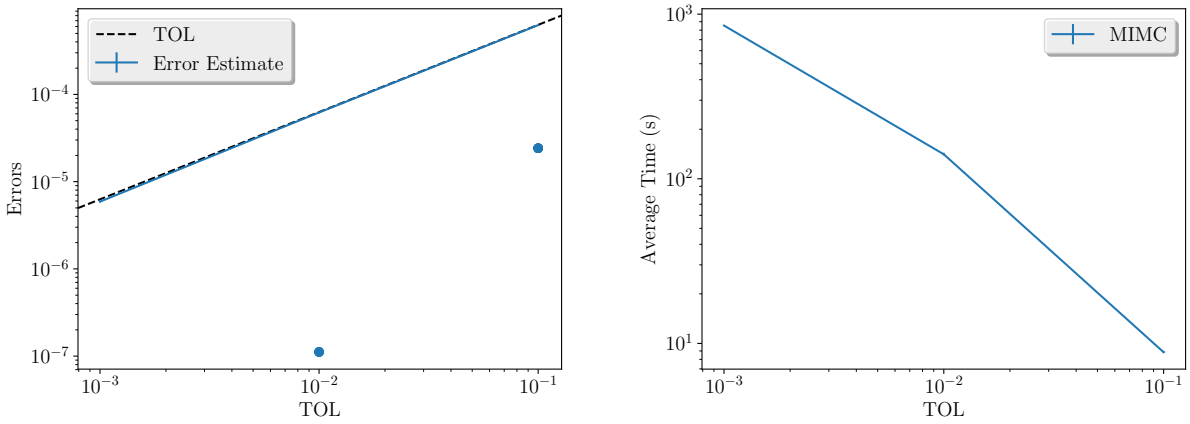
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 41: Convergence and work rates for discretization levels for the $1D$ BS model with the non-smooth call payoff.

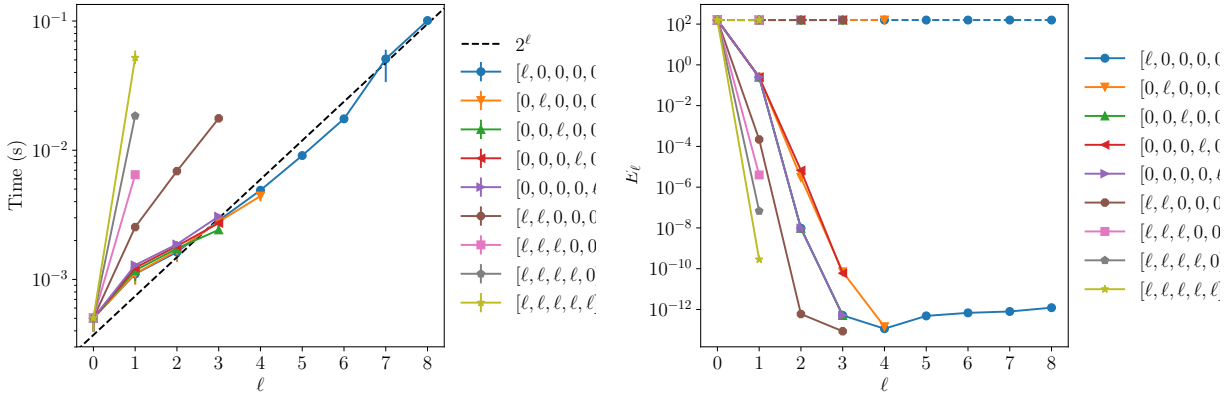
F.12 Case of 16 time steps with smooth payoff (given by (39))

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order -1 for 8 time steps.



(a) Error estimate (b)

Figure 42: Convergence and complexity results for for the $1D$ BS model with the smooth call payoff given by (39).

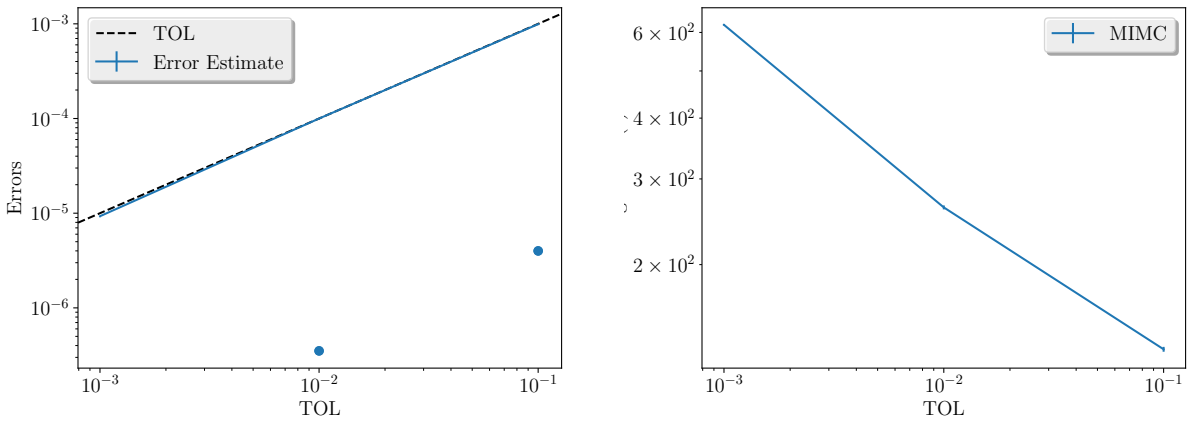


(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 43: Convergence and work rates for discretization levels for the $1D$ BS model with the smooth call payoff given by (39).

F.13 Case of smooth payoff (given by (39)) with Richardson extrapolation (level 0: 8 time steps, level 1: 16 time steps)

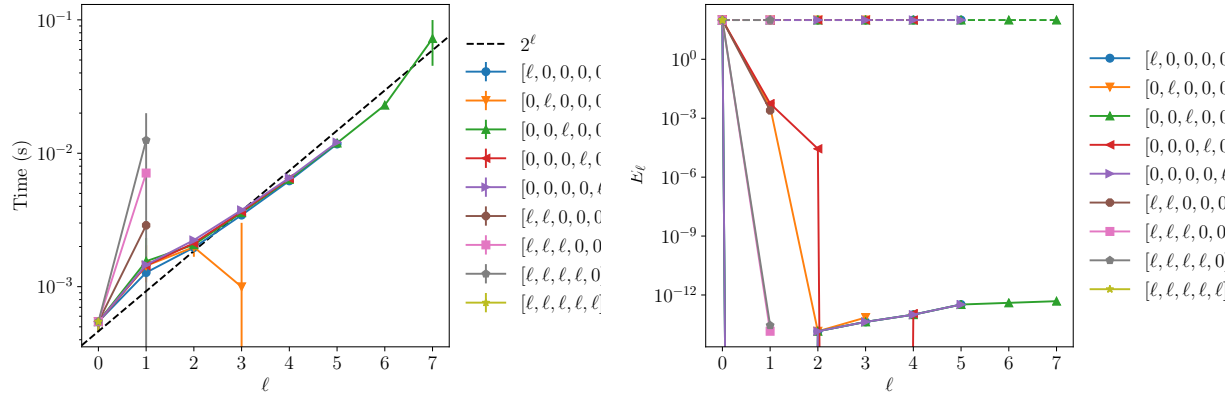
From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-9/25$.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 44: Convergence and complexity results for for the $1D$ BS model with the smooth call payoff given by (39) with Richardson extrapolation.



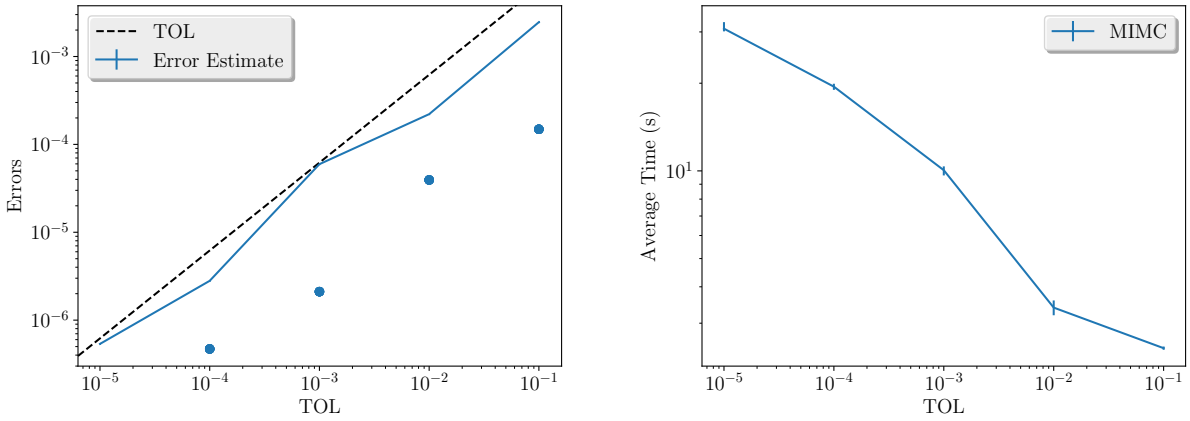
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 45: Convergence and work rates for discretization levels for the $1D$ BS model with the smooth call payoff given by (39) with Richardson extrapolation.

G Checking the bug of richardson extrapolation

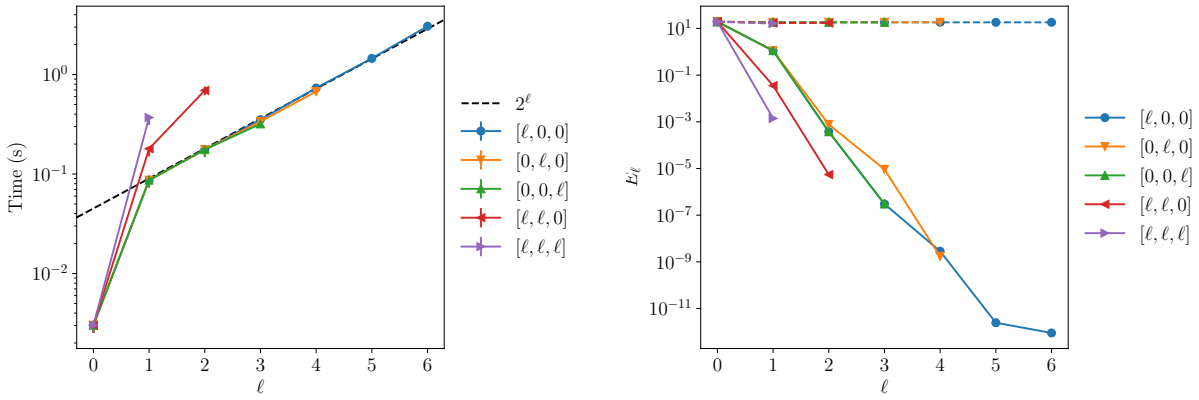
In this section, I check then if we have some bug in the Richardson extrapolation implementation, I present the results for $N = 4$ for 3 cases: non-smooth case without Richardson,non-smooth case with Richardson with just finer level, and non-smooth case with Richardson with just coarser level



(a) Error estimate

(b) Average running time as a function of TOL

Figure 46: Convergence and complexity results for for the $1D$ BS model with the non-smooth call payoff .



(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 47: Convergence and work rates for discretization levels for the $1D$ BS model with the non-smooth call payoff.

non-smooth case without Richardson

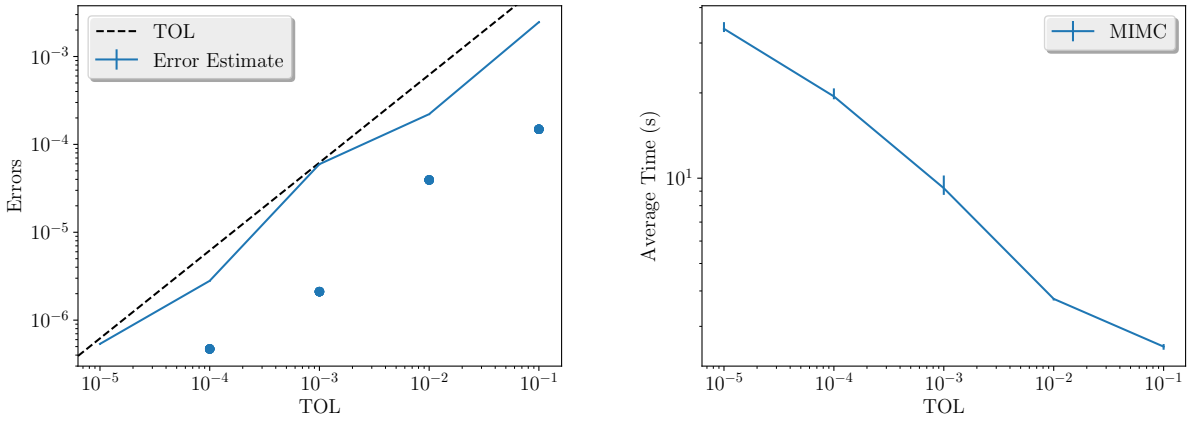
non-smooth case with Richardson (just finer level)

non-smooth case with Richardson (just coarser level)

H Verifying bounds on work and error contributions for the Call option payoff

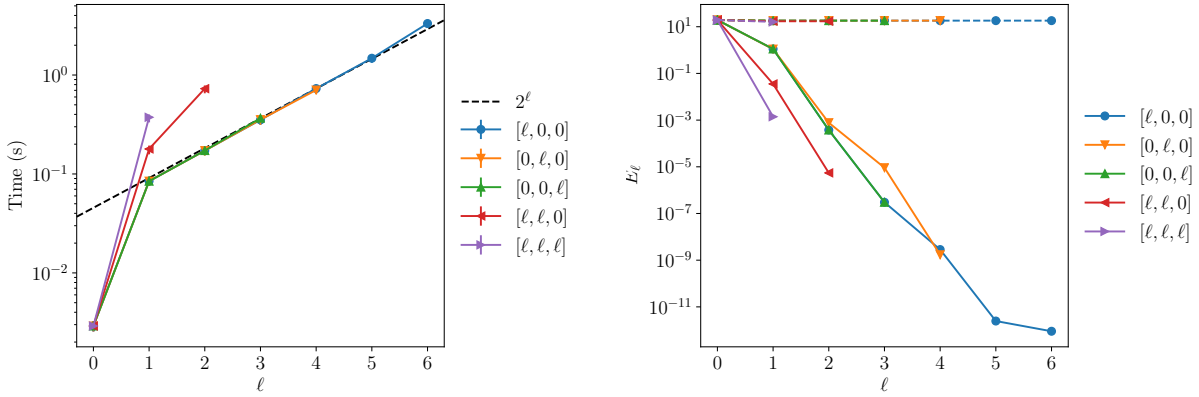
Without Richardson

Below, we show in figures ((52),(53),(54)) the rate of convergence of first and mixed differences for different orders and different number of time steps N. We check that the error is decaying



(a) Error estimate (b) Average running time as a function of TOL

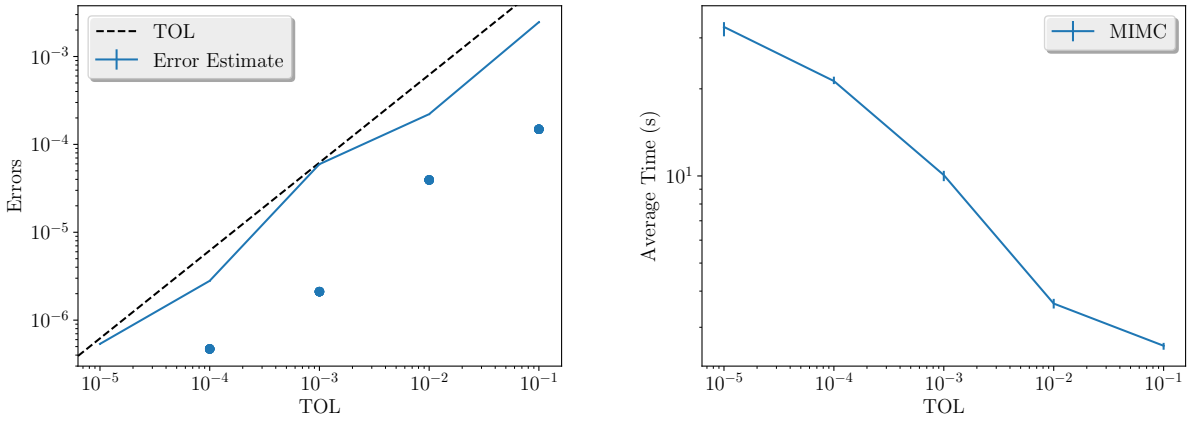
Figure 48: Convergence and complexity results for for the $1D$ BS model with the non-smooth call payoff .



(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 49: Convergence and work rates for discretization levels for the $1D$ BS model with the non-smooth call payoff.

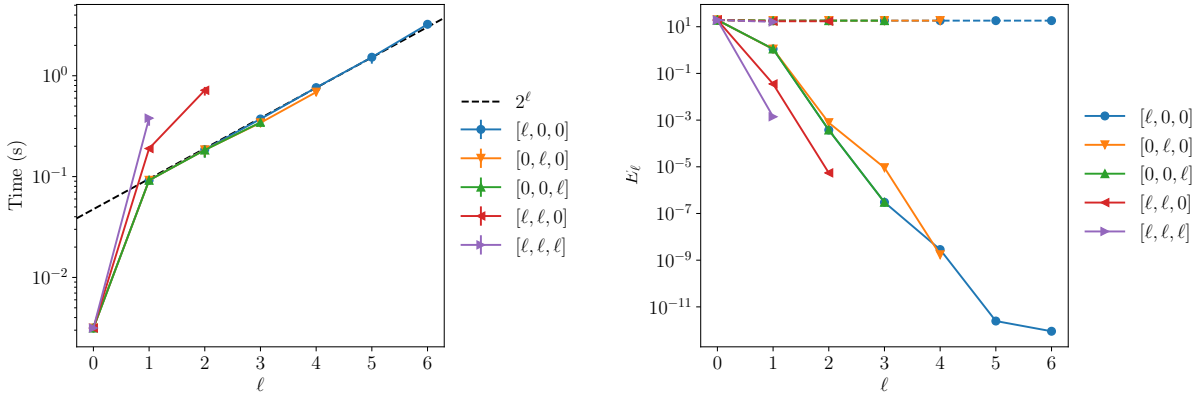
exponentially wrt number of points used in the quadrature.



(a) Error estimate

(b) Average running time as a function of TOL

Figure 50: Convergence and complexity results for for the $1D$ BS model with the non-smooth call payoff .



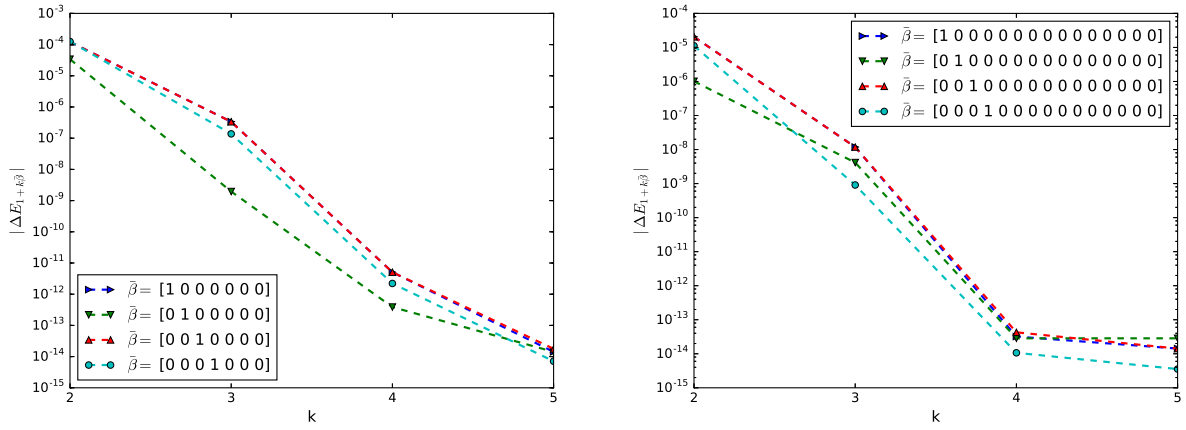
(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 51: Convergence and work rates for discretization levels for the $1D$ BS model with the non-smooth call payoff.

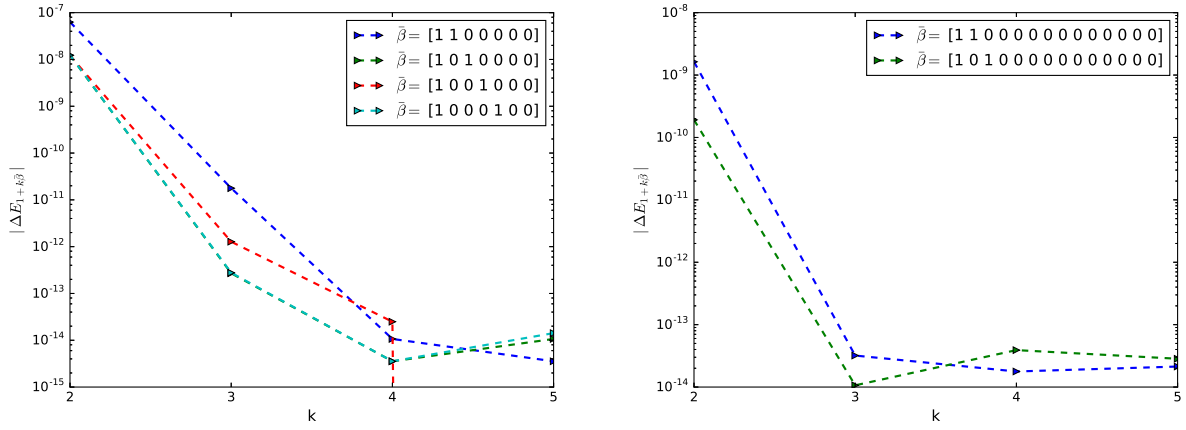
With Richardson

Below, we show in figures ((55),(56)) the rate of convergence of first and mixed differences for different orders and different number of time steps N. We check that the error is decaying exponentially wrt number of points used in the quadrature.



(a) (b)

Figure 52: The rate of convergence of first order differences of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$): a) $N = 8$. b) $N = 16$.



(a) (b)

Figure 53: The rate of convergence of second order mixed differences of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$)): a) $N = 8$ and b) $N = 16$.

I MISC results for the binary option example

I.1 Case of 4 time steps

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-1/3$ for 4 time steps.

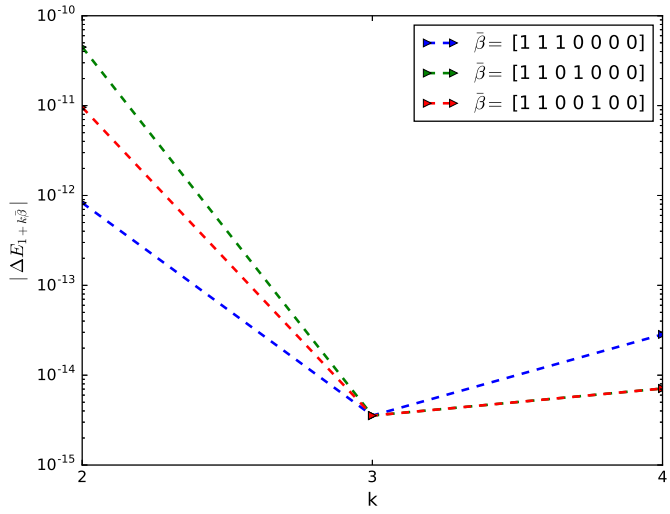
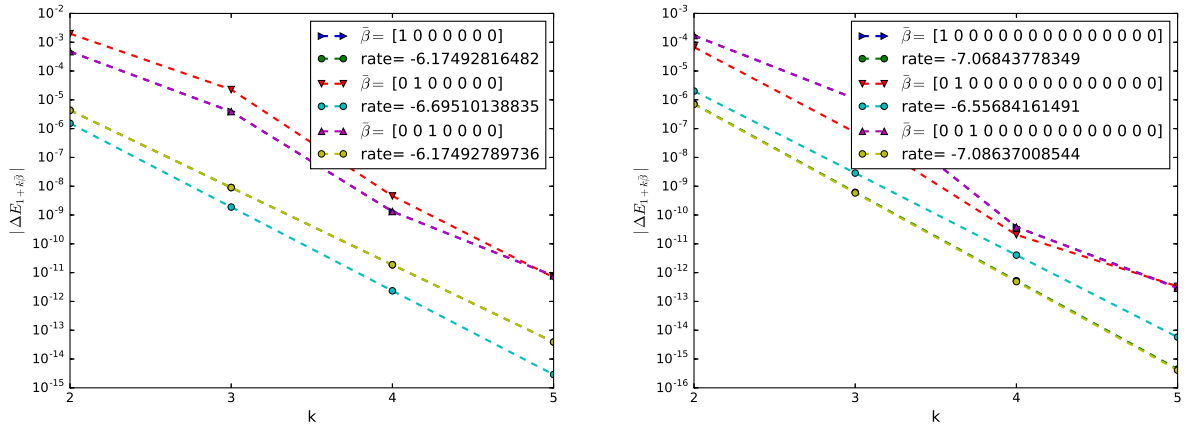


Figure 54: The rate of convergence of third order mixed differences of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$)): for $N = 8$.

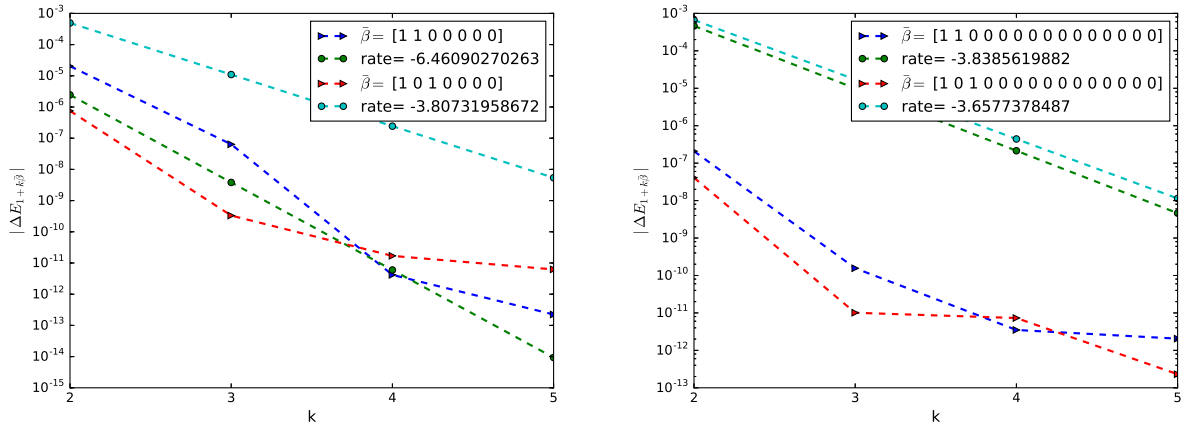


(a) (b)

Figure 55: The rate of convergence of first order differences of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$): a) $N = 8$. b) $N = 16$.

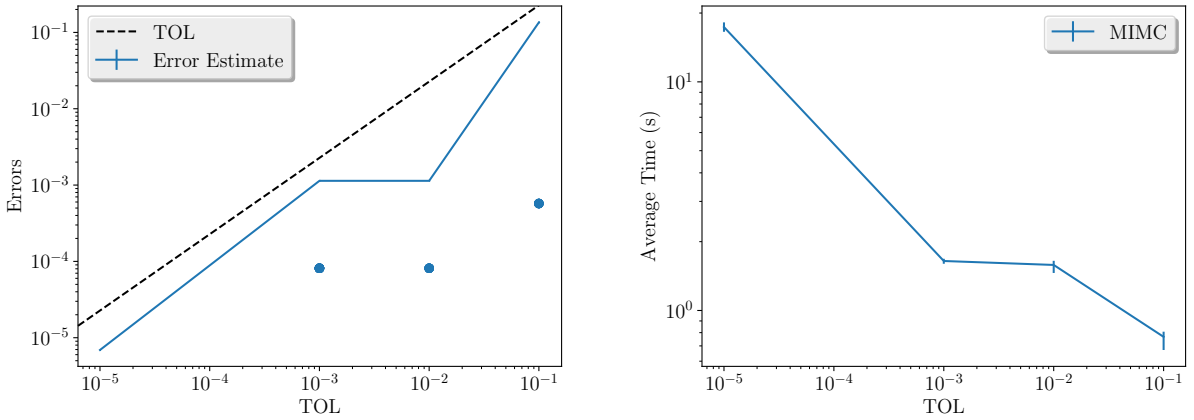
I.2 Case of non smooth payoff with Richardson extrapolation (level 0: 2 time steps, level 1: 4 time steps)

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-4/10$.



(a) (b)

Figure 56: The rate of convergence of second order mixed differences of $|\Delta E_\beta|$ ($\beta = \mathbf{1} + k\overline{\beta}$)): a) $N = 8$ and b) $N = 16$.

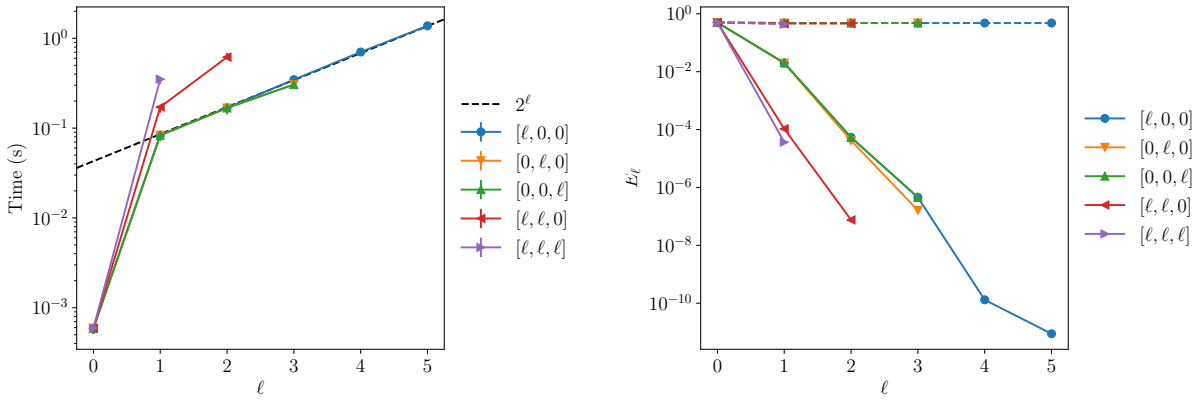


(a) Error estimate (b) Average running time as a function of TOL

Figure 57: Convergence and complexity results for the $1D$ BS model with binary option payoff.

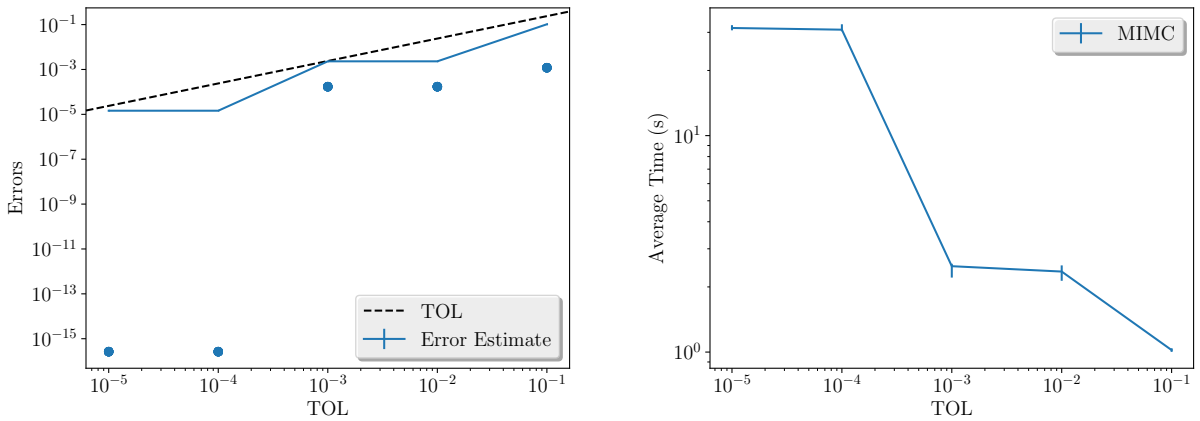
I.3 Case of 8 time steps

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-4/10$ for 8 time steps.



(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 58: Convergence and work rates for discretization levels for the $1D$ BS model with binary option payoff.

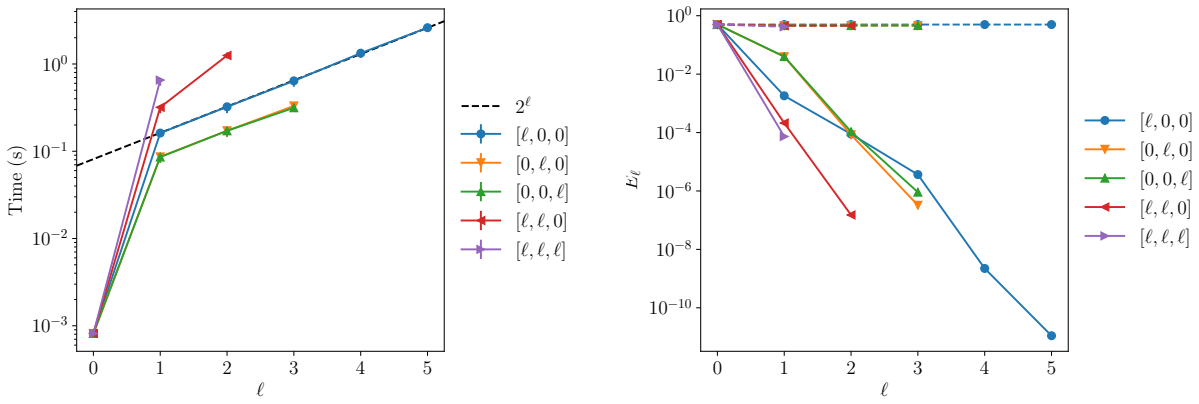


(a) Error estimate (b) Average running time as a function of TOL

Figure 59: Convergence and complexity results for the $1D$ BS model with binary option payoff with Richardson extrapolation.

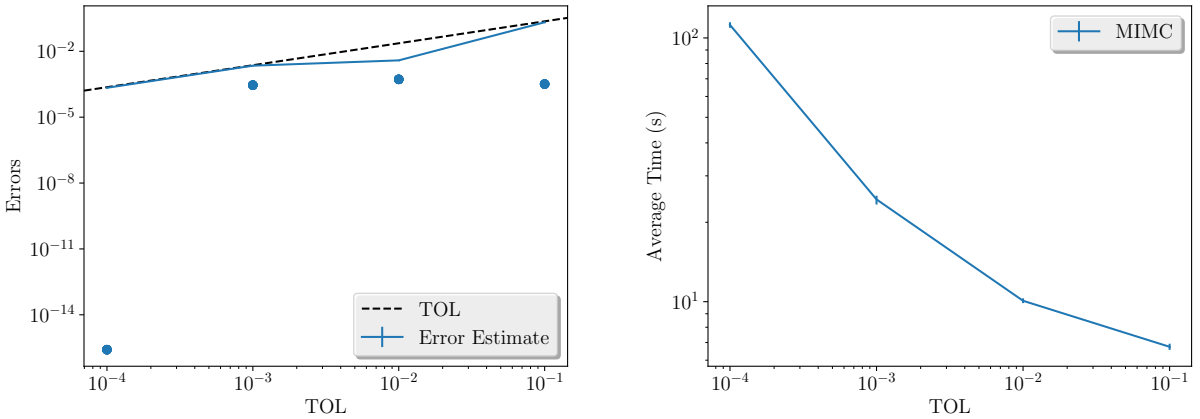
I.4 Case of non smooth payoff with Richardson extrapolation (level 0: 4 time steps, level 1: 8 time steps)

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-6/10$.



(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 60: Convergence and work rates for discretization levels for the $1D$ BS model with binary option payoff with Richardson extrapolation.

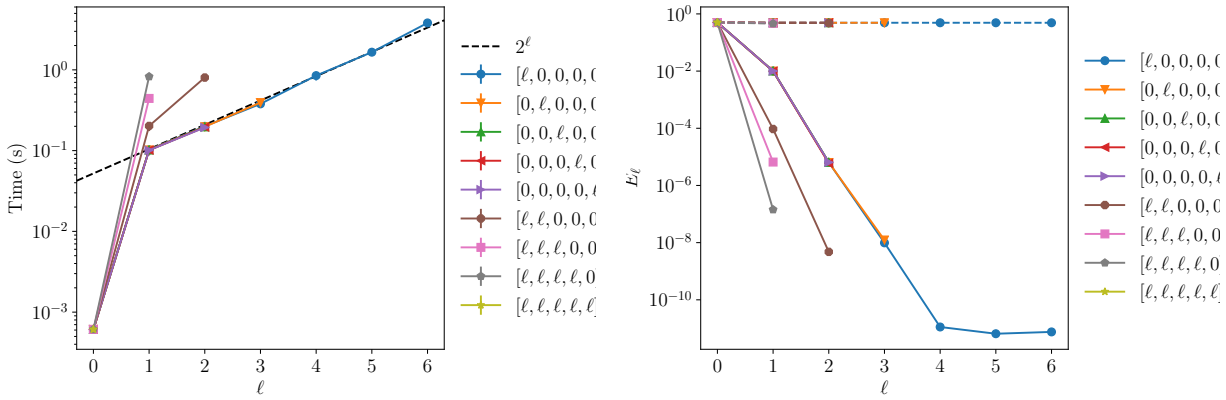


(a) Error estimate (b) Average running time as a function of TOL

Figure 61: Convergence and complexity results for the $1D$ BS model with binary option payoff.

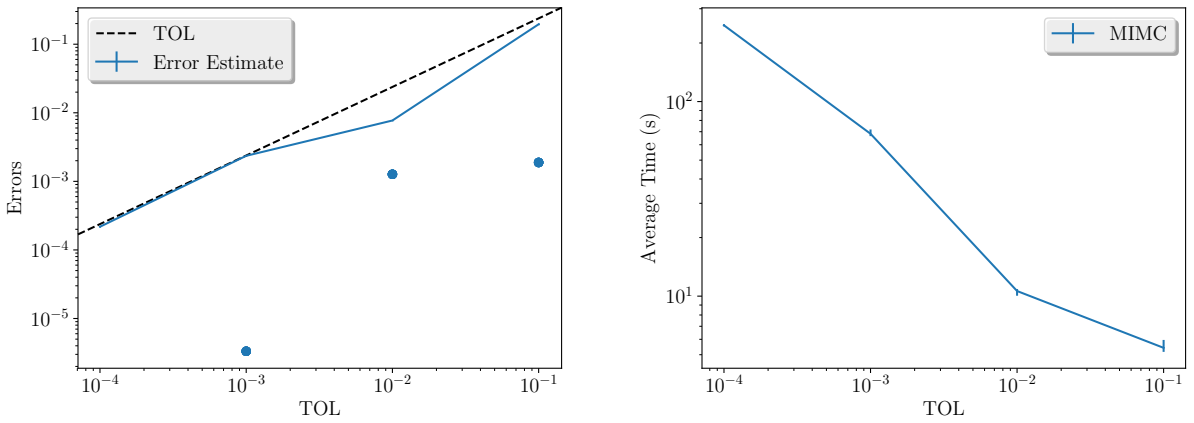
I.5 Case of 16 time steps

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-17/25$ for 16 time steps.



(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 62: Convergence and work rates for discretization levels for the $1D$ BS model with binary option payoff.

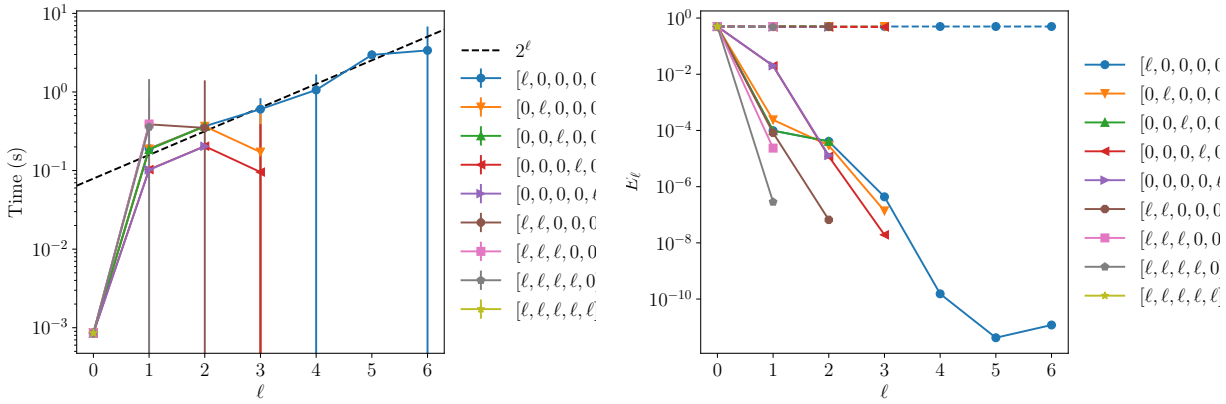


(a) Error estimate (b) Average running time as a function of TOL

Figure 63: Convergence and complexity results for the $1D$ BS model with binary option payoff with Richardson extrapolation.

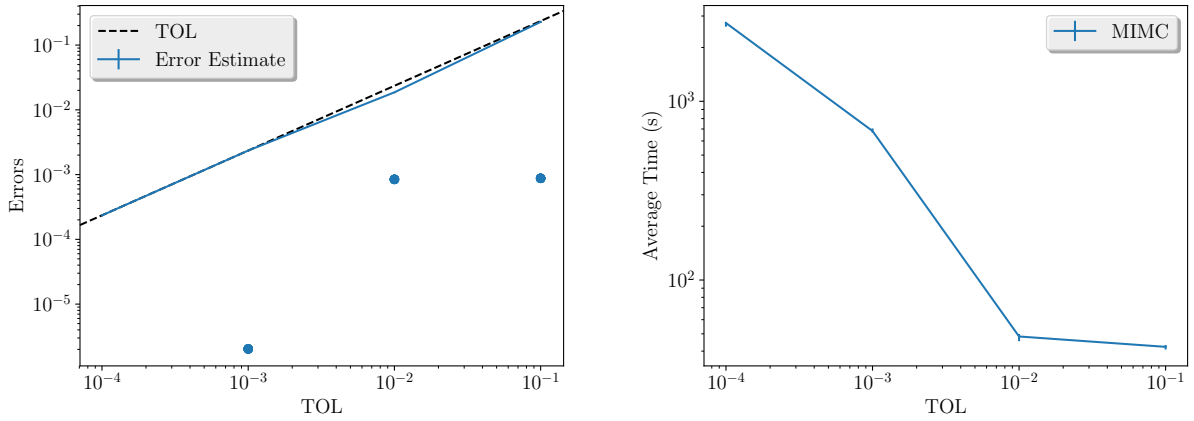
I.6 Case of non smooth payoff with Richardson extrapolation (level 0: 8 time steps, level 1: 16 time steps)

From the following plots, we confirm the exponential rate of convergence for the stochastic parameters, we achieve the prescribed tolerance and we get a rate of complexity (in terms of average running time) approximately of order $-8/10$.



(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 64: Convergence and work rates for discretization levels for the $1D$ BS model with binary option payoff with Richardson extrapolation.



(a) Error estimate (b) Average running time as a function of TOL

Figure 65: Convergence and complexity results for the $1D$ BS model with binary option payoff.

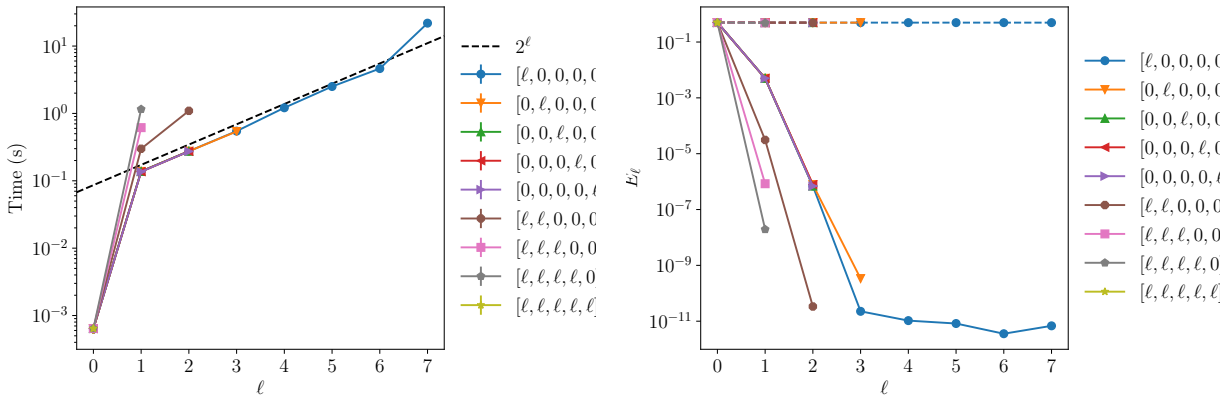
J Problem Setting:

We aim at approximating $E[g(X(t))]$ given $g : \mathbb{R}^d \to \mathbb{R}$, where $X \in \mathbb{R}^d$ solves

$$X(t) = X(0) + \int_0^t a(s, X(s))ds + \sum_{\ell=1}^{\ell_0} \int_0^t b^\ell(s, X(s))dW^\ell(s) \tag{40}$$

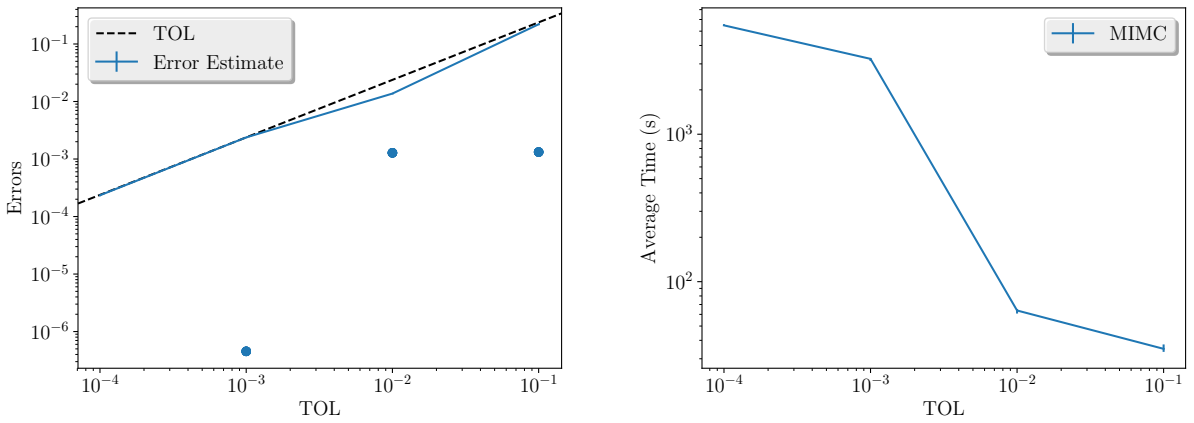
Let us decompose the Wiener process in the interval $[0, T]$ as

$$W(t) = W(T)\frac{t}{T} + B(t) \tag{41}$$



(a) Average Computational time per level (b) The convergence rate of first differences per level

Figure 66: Convergence and work rates for discretization levels for the $1D$ BS model with binary option payoff.



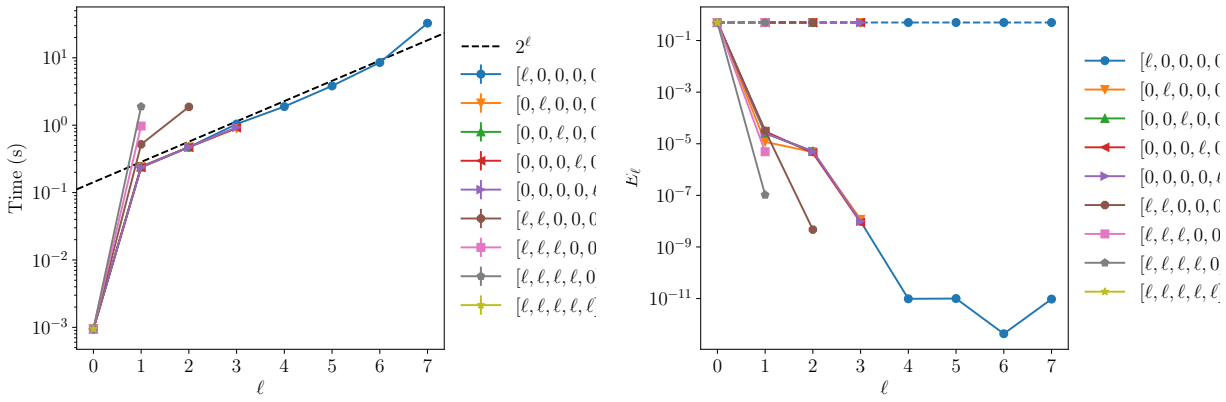
(a) Error estimate (b) Average running time as a function of TOL

Figure 67: Convergence and complexity results for the $1D$ BS model with binary option payoff with Richardson extrapolation.

with $B(t)$ a Brownian bridge with zero end value. Then, for each $t \in [0, T]$ we have

$$\begin{aligned} X(t) &= X(0) + \int_0^t b(X(s))dB(s) + \frac{W(t)}{t}\int_0^t b(X(s))ds \\ &= X(0) + \int_0^t b(X(s))dB(s) + \frac{Y}{\sqrt{t}}\int_0^t b(X(s))ds, \end{aligned} \tag{42}$$

Where $Y \sim \mathcal{N}(0,1)$ and B and Y are independent.



(a) Average Computational time per level

(b) The convergence rate of first differences per level

Figure 68: Convergence and work rates for discretization levels for the $1D$ BS model with binary option payoff with Richardson extrapolation.

As a consequence,

$$\begin{aligned} E[g(X(T))] &= E^B[E^Y[g(X(T)) \mid B]] \\ &= \frac{1}{\sqrt{2\pi}} E^B[H(B)], \end{aligned} \tag{43}$$

where $H(B) = \int g(X(T; y, B)) \exp(-y^2/2) dy$.

We note that $H(B)$ has for many practical cases, a smooth dependence wrt to X due to the smoothness of the pdf of Y.

For illustration, we have $W_t = \frac{t}{T} W_T + B_t$ and

$$\begin{aligned} \Delta W_i &= (B_{t_{i+1}} - B_{t_i}) + \Delta t \frac{Y}{\sqrt{T}} \\ &= \Delta B_i + \Delta t \frac{Y}{\sqrt{T}}, \end{aligned} \tag{44}$$

implying that the numerical approximation of $X(T)$ satisfies

$$\overline{X}_T = \Phi(\Delta t, W_1(y), \Delta B_0, \ldots, \Delta B_{N-1}), \tag{45}$$

for some path function Φ.

K Numerical Approaches

K.1 First approach

- Use sparse grid $\mathcal{D}$ for $\Delta B_0, \ldots, \Delta B_{N-1}$.

- Given $(\mathbf{X}^0, \ldots, \mathbf{X}^{N-1}) := \mathcal{X} \in \mathcal{D}$ with weights $(\boldsymbol{\omega}^0, \ldots, \boldsymbol{\omega}^{N-1})$, add grid points $(y_1(\mathcal{X}), \ldots, y_K(\mathcal{X})) = \mathbf{y}$ with weights $(W_1, \ldots, W_K)$ such that the mapping $y \to g(\Phi(\Delta t, y, X^0, \ldots, X^{N-1}))$ is smooth outside the kink point. Mainly here we will use the **Newton iteration** to determine the kink point.

- Construct our estimator for $E[g(X(T))]$ by looping over step 1 and 2 such that we choose the optimal indices of sparse grids that achieves a global error of order TOL.

$$E[g(X(T))] = \sum_{n=0}^{N-1} \sum_{j} \sum_{i=1}^{K} W_i g(\Phi(\Delta t, \mathbf{y}, \mathcal{X}))\omega_j^n$$

K.1.1 Some discussion on the complexity and errors

- We expect that the global error of our procedure will be bounded by the weak error which is in our case of order $O(\Delta t)$. In this case, the overall complexity of our procedure will be of order $O(TOL^{-1})$. We note that this rate can be improved up to $O(TOL^{-\frac{1}{2}})$ if we use **Richardson extrapolation**. Another way that can improve the complexity could be based on **Cubature on Wiener Space** (This is left for a future work). The aimed complexity rate illustrates the contribution of our procedure which outperforms Monte Carlo forward Euler (MC-FE) and multi-level MC-FE, having complexity rates of order $O(TOL^{-3})$ and $O(TOL^{-2}log(TOL)^2)$ respectively.

- We need to check the impact of the error caused by the Newton iteration on the integration error. In the worst case, we expect that if the error in the Newton iteration is of order $O(\epsilon)$ than the integration error will be of order $\log(\epsilon)$. But we need to check that too.

K.2 Second approach

An alternative approach could be achieved by tensorizing all the quadrature rules (this is not clear to me how to do it yet). The advantage of this procedure is that the additional cost that we will pay by using fine quadrature in the dimension of y will be rewarded by the ability of using coarser quadratures in the remaining dimensions.

L Plan of work and miscancellous observations

We recall the discussion between Raul and Christian on June 1st.

Given we want to compute

$$E\left[g\left(\Phi(Z_1, \ldots, Z_N)\right]\right.$$

for some non-smooth function g and a Gaussian vector Z. (Here, the discretization of the SDE is in the function Φ.) We assume that Z is already rotated such that $h(Z_{-1}) \coloneqq E\left[g\left(\Phi(Z_1, \ldots, Z_N) \mid | \mid Z_1\right]\right.$ is as smooth as possible, where $Z_{-1} \coloneqq (Z_2, \ldots, Z_N)$.

M Plan of implementation

- The first example should probably be the discretized Black-Scholes model, as we discussed together. There, we could also compare different ways to identify the location of the kink, such as:
  - Exact location of the continuous problem
  - Exact location of the discrete problem found by finding the root of a polynomial in y
  - Newton iteration
- As we also discussed, the impact of the Brownian bridge will disappear in the limit, which may make the effect of the smoothing, but also of the errors in the kink location difficult to identify. For this reason, we suggest to study a more complicated 1-dimensional problem next. We suggest to use a CIR process. To avoid complications at the boundary, we suggest "nice" parameter choices, such that the discretized process is very unlikely to hit the boundary (Feller condition).
- Extension to the multi-dimensional situation. Here, we suggest to return to the Black-Scholes model, but in multi-d. In this case, linearizing the exponential, suggest that a good variable to use for smoothing might be the sum of the final values of the Brownian motion. In general, though, one should probably eventually identify the optimal direction(s) for smoothing via the duals / algorithmic differentiation.

N Smoothing via conditional sampling step

Let us decompose the Wiener process in the interval $[0, T]$ as

$$W(t) = W(T)\frac{t}{T} + B(t) \tag{46}$$

with $B(t)$ a Brownian bridge with zero end value. Then, for each $t \in [0, T]$ we have

$$\begin{aligned} X(t) &= X(0) + \int_0^t b(X(s))dB(s) + \frac{W(t)}{t}\int_0^t b(X(s))ds \\ &= X(0) + \int_0^t b(X(s))dB(s) + \frac{Y}{\sqrt{t}}\int_0^t b(X(s))ds, \end{aligned} \tag{47}$$

Where $Y \sim \mathcal{N}(0, 1)$ and B and Y are independent.

As a consequence,

$$\begin{aligned} E[g(X(T))] &= E^B[E^Y[g(X(T)) \mid B]] \\ &= \frac{1}{\sqrt{2\pi}}E^B[H(B)], \end{aligned} \tag{48}$$

where $H(B) = \int g(X(T; y, B)) \exp(-y^2/2)dy$.

We note that $H(B)$ has for many practical cases, a smooth dependence wrt to X due to the smoothness of the pdf of Y.

N.1 Smoothing the payoff in the continuous case

N.1.1 1^{st} case: $g(x) = \delta(x - K)$

It is easy to show that

$$\begin{aligned} H(B) &= \int \delta(X(T; y, B) - K) \exp(-y^2/2) dy \\ &= \exp(-y_*^2(K)/2) \frac{dy_*}{dx}(K), \end{aligned} \tag{49}$$

where $y_*(x)$, is an invertible function that satisfies

$$X(T; y_*(x), B) = x \tag{50}$$

N.1.2 2^{nd} case: $g(x) = \mathbf{1}_{x>K}$

$$\begin{aligned} H(B) &= \int \mathbf{1}_{X(T;y,B)>K} \exp(-y^2/2) dy \\ &= \sqrt{2\pi} P(Y > y_*(K)) \frac{dy_*}{dx}(K), \end{aligned} \tag{51}$$

O Description of our approach and implementation details

Given we want to compute

$$E\left[g\left(\Phi(Z_1, \ldots, Z_N)\right]\right. \tag{52}$$

for some non-smooth function g and a Gaussian vector $Z = (Z_1, \ldots, Z_N)$. (Here, the discretization of the SDE is in the function Φ.) We assume that Z is already rotated such that $h(Z_{-1}) \coloneqq E\left[g\left(\Phi(Z_1, \ldots, Z_N) \mid Z_{-1}\right]\right.$ is as smooth as possible, where $Z_{-1} \coloneqq (Z_2, \ldots, Z_N)$.

O.1 Choice of functional

We should restrict ourselves to a few possible choices g such as:

- hockey-stick function, i.e., put or call payoff functions;
- indicator functions (both relevant in finance and in other applications of estimation of probabilities of certain events);
- delta-functions for density estimation (and derivatives thereof for estimation of derivatives of the density).

More specifically, g should be the composition of one of the above with a smooth function. (For instance, the basket option payoff as a function of the log-prices of the underlying.)

O.2 Steps of the numerical approach

O.2.1 First approach

1. Determine the location of the kink point(s), $y^{\text{kink}} = y^{\text{kink}}(z_{-1})$ either exactly or approximately, $\overline{y}^{\text{kink}}$, using **Newton iteration method**.
2. Use sparse grid $\mathcal{D}$ for $Z_{-1} = (Z_2, \ldots, Z_N)$.
3. Given $(\mathbf{X}^2, \ldots, \mathbf{X}^N) := \mathcal{X} \in \mathcal{D}$ with weights $(\boldsymbol{\omega}^2, \ldots, \boldsymbol{\omega}^N)$, add grid $\boldsymbol{\zeta} = (\zeta_{-K}(y^{\text{kink}}(\mathcal{X})), \ldots, \zeta_K(y^{\text{kink}}(\mathcal{X})))$ with weights $(\eta_{-K}, \ldots, \eta_K)$ such that $\widehat{h}\ (\overline{h})$ is smooth.
4. Construct the adaptive collocation to compute the integral of $\widehat{h}\ (\overline{h})$ that approximates $E[g(\phi(Z_1, \ldots, Z_N))]$ by choosing the optimal indices of sparse grids that achieves a global error of order TOL.

P Richardson extrapolation

We recall that the Euler (often) scheme has weak order 1 so that

$$\left|\mathrm{E}\left[f(\widehat{X}_T^h)\right] - \mathrm{E}\left[f(X_T)\right]\right| \leq Ch \tag{53}$$

for some constant C, all sufficiently small h and suitably smooth f. It was shown that 53 can be improved to

$$\mathrm{E}\left[f(\widehat{X}_T^h)\right] = \mathrm{E}\left[f(X_T)\right] + ch + \mathcal{O}\left(h^2\right), \tag{54}$$

where c depends on f.
Applying 54 with discretization step $2h$, we obtain

$$\mathrm{E}\left[f(\widehat{X}_T^{2h})\right] = \mathrm{E}\left[f(X_T)\right] + 2ch + \mathcal{O}\left(h^2\right), \tag{55}$$

implying

$$2\mathrm{E}\left[f(\widehat{X}_T^{2h})\right] - \mathrm{E}\left[f(\widehat{X}_T^h)\right] = \mathrm{E}\left[f(X_T)\right] + \mathcal{O}\left(h^2\right), \tag{56}$$

Q Binary option results

In table 13, we summarize the results that we get in terms of complexity rates for different methods using different number of time steps for the binary option case. We clarify that the results for the Richardson extrapolation are given for just one level, where the associated steps in the table corresponds to those of level 1. Detailed plots for each case are given in appendix I. Mainly, from the plots, we checked that we achive the prescribed tolerance using MISC, the convergence rates of mixed differences which is a basic assumption for using MISC (we observe exponential decay of error rates wrt to the number of quadrature points) and finally the complexity rates.

We got no improvement using Richardson extrapolation.

Method \Steps	4	8	16
without Richardson extrapolation	$-1/3$	$-4/10$	$-17/25$
with Richardson extrapolation	$-4/10$	$-6/10$	$-8/10$

Table 13: Complexity rates of the different methods for different number of time steps

R Analysis of the numerical procedure

Actually, I am no more convinced, why we should get an improvement using Richardson extrapolation the way it is used, since the gain in accuracy is in simulating the paths but then we apply kind of non-linear operation (The newton iteration, with the same accuracy), which somehow may deteriorate that gain, because the only operation to compute the integrand in the binary option is solving the kink point (no quadrature for the last dimension of Brownian increment). To have a better understanding, we need to analyze our procedure in terms of the time step size Δt and see if we can explain the observed behavior. We will do this for the case of binary option for simplicity.

Without Richardson extrapolation

We start with the case where we do not use Richardson extrapolation. We are mainly interested in the behavior of the integrand that will be fed to MISC. The first step is identifying the kink point, $y_*(B, \Delta t)$, which is related to solving

$$X(T; y_*, B) = K. \tag{57}$$

using Newton iteration. If we call the Newton iteration operator as $\overline{f}$, then we can write

$$y_*(B, \Delta t) = \overline{f}\left(\Phi(\Delta t, B, y_*), \frac{\partial \Phi(\Delta t, B, y_*)}{\partial y_*}\right), \tag{58}$$

where Φ is given by 29 and $\frac{\partial \Phi(\Delta t, B, y_*)}{\partial y_*}$ is given by 30.

In the second step, we compute $H(y_*(B, \Delta t))$ as given by 31, so at the end, the integrand will be given by

$$\begin{aligned} H(B, \Delta t) &= H(y_*(B, \Delta t)) \\ &= H\left(\overline{f}\left(\Phi(\Delta t, B, y_*), \frac{\partial \Phi(\Delta t, B, y_*)}{\partial y_*}\right)\right) \end{aligned} \tag{59}$$

The question is how exactly $H(B, \Delta t)$ depends on Δt?

With Richardson extrapolation

In this case, in the first step, we will determine two kink points that we denote y_*^c, y_*^f, for coarser and finer grid respectively, and they are given by

$$y_*^c(B^c, \Delta t) = \overline{f}\left(\Phi(\Delta t, B^c, y_*^c), \frac{\partial \Phi(\Delta t, B^c, y_*^c)}{\partial y_*^c}\right)$$

$$y_*^f(B^f, \Delta t/2) = \overline{f}\left(\Phi(\Delta t/2, B^f, y_*^f), \frac{\partial \Phi(\Delta t/2, B^f, y_*^f)}{\partial y_*^f}\right), \tag{60}$$

Then the integrand fed to MISC will be given by

$$H(B^f, \Delta t) = 2H(y_*^f(B^f, \Delta t/2)) - H(y_*^c(B^c, \Delta t)). \tag{61}$$

We need to understand if using 61 guarantees that we have an improvement of the rate with respect to Δt. This is related to the understanding of the previous question!!