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

1 Problem Setting:

In this section, we recall the discussion on May 23.

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

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

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

$$W(t) = W(T)\frac{t}{T} + B(t) \tag{2}$$

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

$$\begin{aligned} X(t) &= X(0) + \int_0^t b(X(s))dB(s) + \frac{W(t)}{t}\int_0^t b(X(s))ds \\ &= X(0) + \int_0^t b(X(s))dB(s) + \frac{Y}{\sqrt{t}}\int_0^t b(X(s))ds, \end{aligned} \tag{3}$$

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

As a consequence,

$$\begin{aligned} E[g(X(T))] &= E^B[E^Y[g(X(T)) \mid B]] \\ &= \frac{1}{\sqrt{2\pi}}E^B[H(B)], \end{aligned} \tag{4}$$

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.

1.1 Smoothing the payoff in the continuous case

1.1.1 1^{st} case: $g(x) = \delta(x-K)$

It is easy to show that

$$
\begin{aligned}
H(B) &= \int \delta(X(T;y,B)-K)\exp(-y^2/2)dy \\
&= \exp(-y_*^2(K)/2)\frac{dy_*}{dx}(K), \qquad (5)
\end{aligned}
$$

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

$$X(T; y_*(x), B) = x \qquad (6)$$

1.1.2 2^{nd} case: $g(x) = (x-K)^+$

$$
\begin{aligned}
H(B) &= \int (X(T;y,B)-K)^+ \exp(-y^2/2)dy \\
&= \int \mathbf{1}_{X(T;y,B)>K} \exp(-y^2/2)dy \\
&= \sqrt{2\pi}P(Y > y_*(K))\frac{dy_*}{dx}(K), \qquad (7)
\end{aligned}
$$

1.2 Numerical Approaches

1.2.1 First approach

- Use sparse grid $\mathcal{D}$ for $\Delta B_0, \dots, \Delta B_{N-1}$.
- Given $(\mathbf{X}^0, \dots, \mathbf{X}^{N-1}) := \mathcal{X} \in \mathcal{D}$ with weights $(\boldsymbol{\omega}^0, \dots, \boldsymbol{\omega}^{N-1})$, add grid points $(y_1(\mathcal{X}), \dots, y_K(\mathcal{X})) = \mathbf{y}$ with weights $(W_1, \dots, W_K)$ such that the mapping $y \to g(\Phi(\Delta t, y, X^0, \dots, X^{N-1}))$ 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.

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

1.2.2 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.

1.2.3 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.

1.3 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.

2 Plan of work and miscancellous observations

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

Given we want to compute

$$E\left[g\left(\Phi(Z_1,\dots,Z_N)\right]\right. \tag{8}$$

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 $h(Z_{-1}) \coloneqq E\left[g\left(\Phi(Z_1,\dots,Z_N) \mid Z_{-1}\right]\right.$ is as smooth as possible, where $Z_{-1} \coloneqq (Z_2,\dots,Z_N)$.

2.1 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.)

2.2 Smoothing

A crucial element of the smoothing property is that the "location of irregularity" $y = y(z_{-1})$ such that g is not smooth at the point $\Phi(y, z_{-1})$. Generally, there might be (for given z_{-1}

- no solution, i.e., the integrand in the definition of $h(z_{-1})$ 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

$$z_{-1} \mapsto h(z_{-1}) \text{ and } z_{-1} \mapsto \widehat{h}(z_{-1})$$

are smooth, where $\widehat{h}$ denotes the numerical approximation of h based on a grid containing $y(z_{-1})$. In particular, y itself should be smooth in 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.

In many situations, case 2 (which is thought to include case 1) can be guaranteed by monotonicity. For instance, in the case of one-dimensional SDEs with Z_1 representing the terminal value of the underlying Brownian motion (and 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.

[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.

2.3 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

$$\widehat{h}(z_{-1}) = \sum_{k=-K}^{K} \eta_k g\left(\Phi\left(\zeta_k(y(z_{-1})), z_{-1}\right)\right),$$

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 z_{-1} becomes

$$\overline{h}(z_{-1}) \coloneqq \sum_{k=-K}^{K} \eta_k g\left(\Phi\left(\zeta_k(\overline{y}(z_{-1})), z_{-1}\right)\right),$$

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 z_{-1}, then *eventually* we will have

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

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 2.1. 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 2.2. In some important cases, g 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.

2.4 Organization

The project should lead to two papers: one application paper and one theoretical paper. The application paper would be finished first. We should consider the following examples:

- Uni- and multi-variate Black-Scholes;
- A relatively simple interest rate model or stochastic volatility model a la CIR or Heston;
- rough Bergomi: Here, no smoothing is needed, but there is great potential for the efficient numerical integration. This example is highly relevant for practice.

The theoretical paper should concentrate on:

- Numerical analysis of the scheme, including all the components such as the Newton iteration. This might require strict conditions, especially in the multi-variate setting.
- Examples for computation of c.d.f.s and densities, possibly in the context of rare events.

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