AN EMPIRICAL STUDY OF THE EFFICIENCY OF EA FOR DIFFUSION SIMULATION

STEFANO PELUCHETTI, GARETH O. ROBERTS

ABSTRACT. In this paper we investigate the efficiency of some simulation schemes for the numerical solution of a one dimensional stochastic differential equation (SDE). The schemes considered are: the Exact Algorithm (EA), the Euler, the Predictor-Corrector and the Ozaki-Shoji schemes. The focus of the work is on EA which samples skeletons of SDEs without any approximation. The analysis is carried out via a simulation study using some test SDEs. We also consider efficiency issues arising by the extension of EA to the multi-dimensional setting.

1. INTRODUCTION

In this paper we focus on the efficiency of the Exact Algorithm (EA), introduced by Beskos and Roberts [2005], Beskos et al. [2006a]. The framework that we consider is that of the simulation of a diffusion process, solution of a SDE, whose transition densities are not known. Hence the direct simulation of finite dimensional, or discretised, paths is not feasible. EA is a method that, under suitable conditions, permits the simulation of the discretised diffusion process. The novelty of EA is that we are able to simulate from the true law of the diffusion process, without resorting to any type of approximation.

Numerical schemes for the simulation of diffusion processes have been around for some time, the first contribution probably being that of Maruyama [1955]. However, before the work of Beskos and Roberts [2005], the exact nature of the simulation was confined to a very small class of diffusion processes. The field of numerical schemes for the simulation of diffusion processes is vast and growing rapidly, motivated by the fact that the class of solvable diffusions, that is to say diffusions for which the transition densities have a known tractable from, is quite small. See however some recent results on the topic by Albanese and Kuznetsov [2005]. Subsequently, the importance of having a method that allows for exact simulation is clear, if not for validating purposes. The cost that we have to pay for this achievement is less obvious. Hence our focus on the efficiency of EA.

Some preliminary results on the efficiency of EA, in a Monte Carlo scenario, can be found in Casella [2005]. However this paper gives a much more extensive investigation of EA. We initially consider a class of test models that synthesise a range of one-dimensional diffusive dynamics that are encountered in real world applications. We thus simulate them using three well known discretisation schemes and EA and we compare the results obtained. The

Date: February 2008.

simulation study is replicated in the multi-dimensional settings where only the efficiency of the multi-dimensional EA is examined.

This paper is organised as follows. In Section 2, EA and the three discretisation schemes are briefly introduced. Section 3 consists of the simulation study where the efficiency of the 4 schemes is studied. The main difficulty is comparing a scheme that returns the exact result with schemes that return approximated results. Consequently it is necessary to introduce a comparison criterion that measures a "distance" between the true and the approximated result. We are interested in both the sensitivity of the schemes to the parameters of the test SDEs and the ratio of efficiency between EA and the other schemes. In Section 4 the efficiency of the multi-dimensional extension of EA is investigated, without any comparison with the other discretisation schemes. Section 5 concludes the paper.

2. The simulation schemes

2.1. **The Exact Algorithm.** We begin considering a generic one-dimensional and time homogeneous Stochastic Differential Equation (SDE)

$$\begin{aligned} dY_t &= b(Y_t)\,dt + \sigma(Y_t)\,dB_t \qquad 0 \le t \le T \\ Y_0 &= y \end{aligned} \tag{1}$$

where B is the scalar Brownian Motion (BM) and y is the initial condition. The drift coefficient b and the diffusion coefficient σ are assumed to satisfy the proper conditions for the existence and uniqueness of a strong solution of (1). Let Y be the diffusion process strong solution of (1).

Under the additional requirement that σ is continuously differentiable and strictly positive let

$$\eta(u) := \int^u \sigma^{-1}(z)\,dz \tag{2}$$

be the anti-derivative of σ^{-1}. It follows that $X_t := \eta(Y_t)$ satisfies the unit diffusion coefficient SDE

$$\begin{aligned} dX_t &= \alpha(X_t)\,dt + dB_t \qquad 0 \le t \le T \\ X_0 &= x := \eta(y) \end{aligned} \tag{3}$$

with drift coefficient

$$\alpha(u) := \frac{b\{\eta^{-1}(u)\}}{\sigma\{\eta^{-1}(u)\}} - \frac{\sigma'\{\eta^{-1}(u)\}}{2} \tag{4}$$

SDE (3) is assumed to admit a unique strong solution and we denote with $\mathbb{X}$ the state space of X. The map (2), also known as the Lamperti transform, allows us to consider the simpler problem of simulating from (3) for a vast class of one-dimensional SDEs.

In what follows the laws of stochastic processes are defined on the measurable space of continuous functions $C([0,T],\mathbb{R})$ with its cylinder sigma algebra $\mathcal{C}([0,T],\mathbb{R})$, or on the obvious restrictions of this space. Let $\mathbb{Q}_T^x$ and $\mathbb{W}_T^x$ denote the law of the diffusion X and the

law of a BM respectively on $[0,T]$ both started at x. From now on the following hypotheses are assumed to hold

- $(C1)$ $\forall x \in \mathbb{X}$ $\mathbb{Q}_T^x \ll \mathbb{W}_T^x$ and the Radon-Nikodym derivative is given by Girsanov's formula

$$\frac{d\mathbb{Q}_T^x}{d\mathbb{W}_T^x} = \exp\left\{\int_0^T \alpha(\omega_s)\, dX_s - \frac{1}{2}\int_0^T \alpha^2(\omega_s)\, ds\right\} \tag{5}$$

  where $\omega \in C([0,T],\mathbb{X})$
- $(C2)$ $\alpha \in C^1(\mathbb{X},\mathbb{R})$;
- $(C3)$ $\alpha^2 + \alpha'$ is bounded below on $\mathbb{X}$.

An application of Ito's formula to the function $A(u) = \int_{c\in\mathbb{X}}^u \alpha(z)\, dz$ results in a more tractable form of (5)

$$\frac{d\mathbb{Q}_T^x}{d\mathbb{W}_T^x} = \exp\{A(\omega_T) - A(x)\} \exp\left\{-\int_0^T \frac{\alpha^2+\alpha'}{2}(\omega_s)\, ds\right\} \tag{6}$$

Under the integrability assumption

- $(C4)$ $\forall x \in \mathbb{X}$ $\eta_{x,T} := \mathbb{E}_{\mathbb{W}_T^x}\left[e^{A(\omega_T)}\right] < \infty$

it is possible to get rid of the (possibly unbounded) term $A(\omega_T)$ of (6) introducing a new process Z with law $\mathbb{Z}_T^x$ by the Radon-Nikodym derivative

$$\frac{d\mathbb{Z}_T^x}{d\mathbb{W}_T^x} = e^{A(\omega_T)}/\eta_{x,T} \tag{7}$$

$$\eta_{x,T} = \mathbb{E}_{\mathbb{W}_T^x}\left[e^{A(\omega_T)}\right] \tag{8}$$

We refer to Z as the Biased Brownian Motion (BBM). This process can be alternatively defined as a BM with initial value x conditioned on having its terminal value Z_T distributed according to the density

$$h_{x,T}(u) := \eta_{x,T} \times \exp\left\{A(u) - \frac{(u-x)^2}{2T}\right\} \tag{9}$$

It follows that

$$\frac{d\mathbb{Q}_T^x}{d\mathbb{Z}_T^x}(\omega) = \eta_{x,T} \exp\{-A(x)\} \exp\left\{-\int_0^T \frac{\alpha^2+\alpha'}{2}(\omega_s)\, ds\right\} \tag{10}$$

$$\propto \exp\left\{-\int_0^T \phi(\omega_s)\, ds\right\} \leq 1 \tag{11}$$

where $\phi(u) := \left(\alpha^2(u) + \alpha'(u)\right)/2 - l$ and $l := \inf_{r\in\mathbb{X}}\left(\alpha^2(r) + \alpha'(r)\right)/2 < \infty$. Equation (11) suggests the use of a rejection sampling algorithm to generate realisations from $\mathbb{Q}_T^x$. However it is not possible to generate a sample from Z, being Z an infinite-dimensional variate, and moreover it is not possible to compute analytically the value of the integral in (11).

Let $\mathbb{L}$ denote the law of a unit rate Poisson Point Process (PPP) on $[0,T]\times[0,\infty)$, and let $\Phi=\{\chi,\psi\}$ be distributed according to Φ. We define the event Γ as

$$\Gamma := \bigcap_{j\geq 1} \phi\left(Z_{\chi_j}\right) \leq \psi_j \tag{12}$$

that is the event that all the Poisson points fall into the epigraph of $s\mapsto\phi(Z_s)$. The following theorem is proven in Beskos et al. [2006b]

Theorem 1. ***(Wiener-Poisson factorisation)*** *If* $(Z,\Phi)\sim\mathbb{Z}_T^x\otimes\mathbb{L}\mid\Gamma$ *then* $Z\sim\mathbb{Q}_T^x$

At this stage the result is a purely theoretical, as it is not possible to simulate from the law $\mathbb{L}$. However, in the specific case of ϕ bounded upon by $m<\infty$ it is suffice to consider Φ as a PPP on $[0,T]\times[0,m]$. The reason is that for the determination of the event Γ only the points of Φ below m matter. The algorithm resulting from this restrictive boundedness condition on ϕ is EA1.

It should be noted that this hypothesis can be weakened or even removed, leading to EA2 (Beskos et al. [2006a]) and to EA3 (Beskos et al. [2006b]) respectively. Both extensions involves the simulation of some functional of Z or of an event depending on Z which restrict the range of Z, and by continuity the range of $\phi(Z)$.

We briefly consider the case of EA3. The probability that the BB Z stays in an arbitrary interval can be expressed as an infinite series only. As a consequence the direct simulation of the minimum and the maximum of Z is not feasible. However, we can rearrange the terms of this series so that the sequence of the partial sums s_n satisfies the relations:

$$s_{n-1}\leq l \Rightarrow s_n\geq l \tag{13}$$

$$s_{n-1}\geq l \Rightarrow s_n\leq l \tag{14}$$

where l is the limit value of the serie. As explained in Beskos et al. [2006b] we can consider an increasing collection of nested intervals $\{I_n\}_{n\geq 1}$ which contains the starting and ending values of Z. Due to the behaviour of the partial sums s_n we can simulate the value n so that both the maximum and the minimum of Z are included in a specific I_n and at least one of them is included in $I_n\cap I_{n-1}^C$. Conditional on this event R_n the range of Z is bounded.

It remains to implement an algorithm to sample from $Z\mid R_n$, as we have to compute the value of this process at the time instances given by the PPP Φ. It is not sensible to use Z as a trivial RS proposal, the reason being that the number of proposed paths before the first acceptance has infinite expectation. A better RS algorithm proposes from a mixture of two probability measures with equal weight. One of them is the law of Z conditioned on achieving its minimum in $I_n\cap I_{n-1}^C$. The other one is the law of Z conditioned on achieving its maximum in $I_n\cap I_{n-1}^C$. Crucially, it is possible to sample the constrained minimum (or maximum) m of Z and the time τ at which Z hits this minimum (or maximum). Moreover $Z\mid m,\tau$ gets factorised in the product measure of two 3-dimensional Bessel bridges, whose simulation is trivial. As the Radon-Nikodym derivative of this proposal with respect to $Z\mid R_n$ is available in closed form we are done.

2.2. **Optimisation of EA.** From a practical point of view, every version of EA require the simulation from the density (9). This is not a trivial problem as the functional form of (9) depends on the drift coefficient α in (3). Moreover, theoretical results (see Beskos et al. [2006a]) suggest that the acceptance rate of EA typically decreases exponentially with T. It turns out that it is usually more efficient to partition the time interval $[0, T]$ into smaller sub-intervals of length t and apply EA sequentially. This in turn implies that we have to sample from a parametric family of densities $\{h_{x,t}(u)\}_{x\in\mathbb{X}}$, as the starting value x is different on every sub-interval.

Furthermore the time spent in the simulation from $\{h_{x,t}(u)\}_{x\in\mathbb{X}}$ is not negligible in EA. In the particular case of EA1 roughly half of the time is spent in the simulation from $\{h_{x,t}(u)\}_{x\in\mathbb{X}}$. Thus an efficient sampler results in a significantly lower computational cost for the EA. We briefly introduce two adaptive accept-reject samplers that we have developed to sample efficiently from $\{h_{x,t}(u)\}_{x\in\mathbb{X}}$ and we refer to Peluchetti [2007] for a more detailed exposition.

We begin considering the case of a single $h_{x,t}$ for a fixed $x \in \mathbb{X}$ (t is always fixed). The first sampler, ARS1 from now on, requires the following semi sub-linear condition to hold

- $(E1)$ $\exists n^+, N^+, m^-, M^-, \in \mathbb{R},\ c \in \mathbb{X}$:

$$\alpha(u) \leq n^+ + N^+ u \qquad c \leq u \tag{15}$$

$$m^- + M^- u \leq \alpha(u) \qquad u < c \tag{16}$$

The monotonicity of the integral and of the exponential function thus implies the following bounds on $h_{x,t}$

$$h_{x,t}(u) \leq q_+^{u_0}(u) := e^{-\frac{(u-x)^2}{2t} + A(u_0) + n^+(u-u_0) + \frac{N^+}{2}\left(u^2-u_0^2\right)} \qquad c \leq u_0 < u \tag{17}$$

$$h_{x,t}(u) \leq q_-^{u_0}(u) := e^{-\frac{(u-x)^2}{2t} + A(u_0) + m^-(u-u_0) + \frac{M^-}{2}\left(u^2-u_0^2\right)} \qquad u < u_0 < c \tag{18}$$

To construct the envelope, we start by considering the point $u_0 = c$ (c is required to be a point of the envelope in this algorithm). Then, the initial envelope is given by

$$q(u) = q_-^c(u)\, 1_{[u<c]} + q_+^c(u)\, 1_{[c\leq u]} \tag{19}$$

We have successfully bounded $h_{x,t}$ from above with a piece-wise function formed by the kernels of a Gaussian density times finite constants. Using the bounds (17) and (18) it is possible to refine $q(u)$ by adding more points to it too. We illustrate the results of this procedure in Figure 1. If α is sub-linear, a different construction of q results in a tighter envelope for the same number of points.

Considerable attention has been put in the implementation of an efficient algorithm to sample from ARS1:

(1) a binary search is performed (instead of a sequential one) to sample the interval of the piece-wise proposal q;
(2) the same uniform variate used to sample the interval is used to sample from the proper truncated Gaussian distribution by the cdf inversion method;
(3) all the values relevant to the algorithm are cached for re-use.

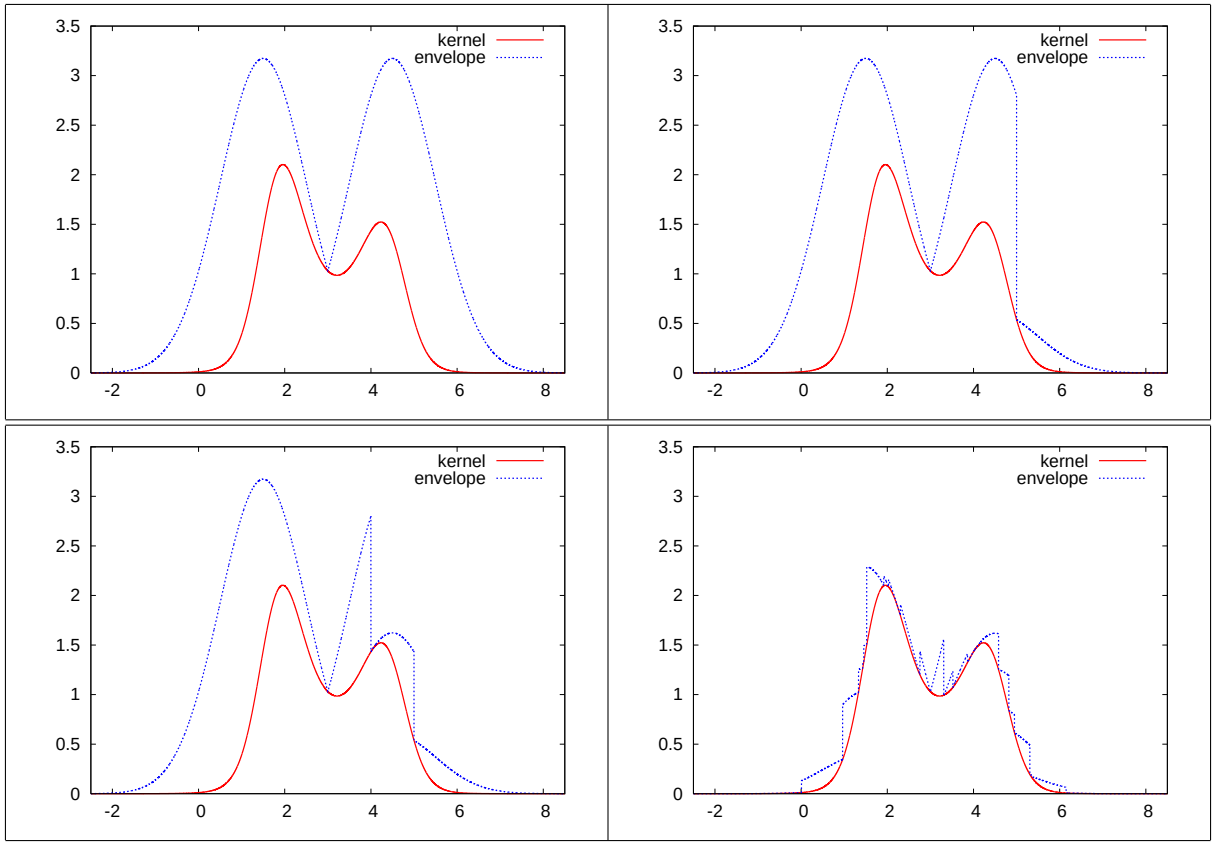

FIGURE 1. The test kernel $h_{x,t}$ and the proposal q constructed from condition $(E1)$ for a test function $h_{x,t}$. Starting from quadrant IV going clockwise we have the envelope constructed from 1, 2, 3 points and the envelope that satisfies an acceptance rate of 95%

The second sampler, ARS2 from now on, has much weaker requirements of ARS1 and is of interest on its own. We basically require the function $h_{x,t}$ to be piece-wise twice differentiable and to exhibit an exponential decay in the tails. This sampler is a generalisation of the adaptive accept-reject sampler introduced in Gilks and Wild [1992], Gilks [1992]. We partition the state space $\mathbb{X}$ into intervals where $h_{x,t}$ is convex/concave and use the geometric interpretation of convexity to construct linear bounds above and below $h_{x,t}$. We illustrate the results of this procedure in Figures 2 and 3.

Similarly to the case of ARS1, considerable attention has been put in the implementation of an efficient algorithm to sample from ARS2. A brief simulation study in Peluchetti [2007] reveals that the efficiency of ARS2 is comparable to that of the Gnu Scientific Library's ad-hoc samplers. ARS1, while somewhat less efficient, is a more robust sampler as it targets a very specific family of densities.

We now consider the more general problem of sampling from $\{h_{x,t}(u)\}_{x\in\mathbb{X}}$. Our idea is to slice the subset $\mathbb{D} \subseteq \mathbb{X}$ where the diffusion X is most likely to stay, to be found by a

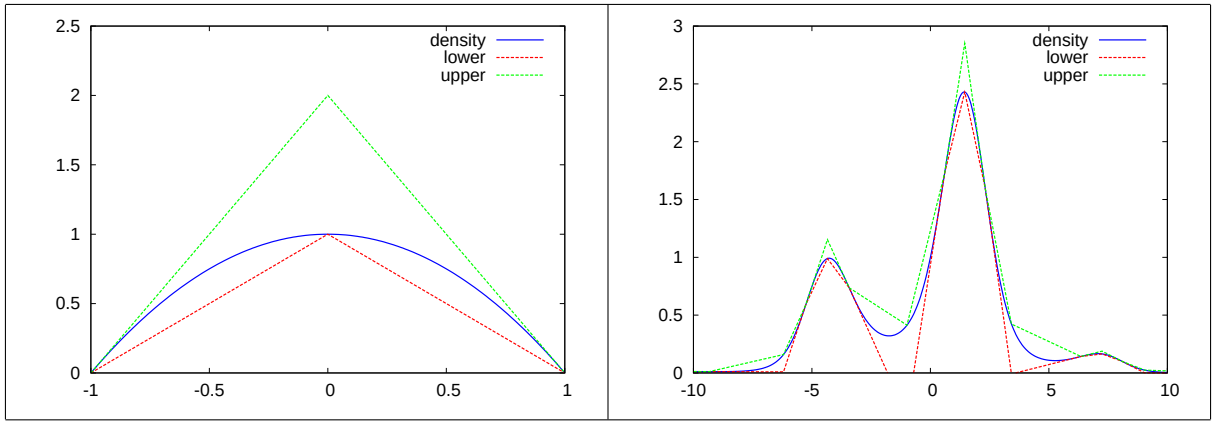

FIGURE 2. The initial construction of the ARS2 on a single interval (left) and on the test density (right)

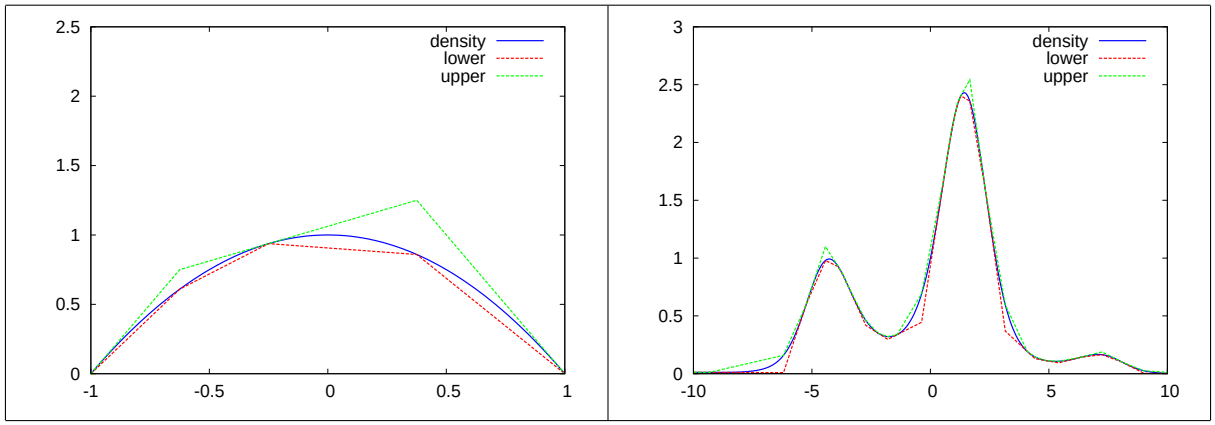

FIGURE 3. The refined construction of the ARS2 on a single interval (left) and on the test density (right)

preliminary simulation, into a finite number of equi-spaced intervals. For each interval, we construct an envelope that uniformly bounds all the $h_{x,t}$ whose x is a point of this interval. To find this uniform bound we notice that for $l < r \in \mathbb{X}$

$$\sup_{l\le x\le r} h_{x,t} = \sup_{l\le x\le r} e^{A(u)-\frac{(u-x)^2}{2t}} \left\{1_{[u<l]} + 1_{[l\le u\le r]} + 1_{[r<u]}\right\} \tag{20}$$

$$\le e^{A(u)-\frac{(u-l)^2}{2t}} 1_{[u<l]} + e^{A(u)} 1_{[l\le u\le r]} + e^{A(u)-\frac{(u-r)^2}{2t}} 1_{[r<u]} \tag{21}$$

$$\le e^{A(u)-\frac{(u-l)^2}{2t}} 1_{[u<l]} + e^{Amax} 1_{[l\le u\le r]} + e^{A(u)-\frac{(u-r)^2}{2t}} 1_{[r<u]} \tag{22}$$

where $Amax = \sup_{l\le u\le r} A(u) < \infty$ as A is a continuous function on a bounded interval, hence A is bounded. The first and the last term of (22) can be easily bounded by envelopes resulting from ARS1 or ARS2. Regarding the central term of (22) we propose the trivial

accept-reject sampling algorithm whose acceptance rate is high if the length of the intervals is reasonably short. We thus pre-compute and cache all these uniform envelopes, one for each intervals in which we split $\mathbb{D}$. During the simulation according to EA, if $x \in \mathbb{D}$ we select the right envelope, otherwise (an event whose probability can be arbitrarily small increasing $\mathbb{D}$) we create an envelope accordingly. As the intervals are equi-spaced there is virtually no efficiency penalty in searching for the right envelope.

2.3. **The discretisation schemes.** We now shortly introduce the three discretisation schemes (DS) whose efficiency, with that of EA, is investigated in the simulation study. All the DSs are assumed to have an equi-spaced discretisation interval of length $\Delta = T/n$, where n is the number of steps and Y^{Δ} denotes a corresponding generic DS. In the following $i = 1, \cdots, n$ and $Y_0 = x$ implicitly.

The Euler scheme is the simplest DS that can be used to approximate the solution of (1). It can be defined by the recursion

$$W_{\Delta}^{i} \overset{iid}{\sim} \mathcal{N}(0, \Delta) \tag{23}$$

$$Y_{i\Delta} = Y_{(i-1)\Delta} + b\left(Y_{(i-1)\Delta}\right)\Delta + \sigma\left(Y_{(i-1)\Delta}\right) W_{\Delta}^{i} \tag{24}$$

The Predictor-Corrector scheme is defined by

$$W_{\Delta}^{i} \overset{iid}{\sim} \mathcal{N}(0, \Delta) \tag{25}$$

$$\overline{Y}_{i\Delta} = Y_{(i-1)\Delta} + b\left(Y_{(i-1)\Delta}\right)\Delta + \sigma\left(Y_{(i-1)\Delta}\right) W_{\Delta}^{i} \tag{26}$$

$$Y_{i\Delta} = Y_{(i-1)\Delta} + \frac{1}{2}\left\{b\left(Y_{(i-1)\Delta}\right) + b\left(\overline{Y}_{i\Delta}\right)\right\}\Delta + \sigma\left(Y_{(i-1)\Delta}\right) W_{\Delta}^{i} \tag{27}$$

The idea behind this DS is to make a Euler prediction $\bar{Y}_{i\Delta}$ by using (26) and adjust $\bar{Y}_{i\Delta}$ by computing an average of the drift's value over the time step $((i-1)\Delta, i\Delta]$ using the trapezoid quadrature formula. This approach results in the correction (27). It is fundamental to_use the same W_{Δ}^{i} in (26) and (27). For more details about the Euler and the Predictor-Corrector schemes see Kloeden and Platen [1992].

Finally we introduce the Ozaki-Shoji scheme. This DS uses a completely different approach that is only applicable to diffusion process with constant diffusion coefficient and, without loss of generality, to (3). This DS belongs to the family of "linearisation schemes" which approximates the drift α of (3) by some sort of linear approximation. The specific version here presented it the one of Shoji and Ozaki [1998]. The idea behind this DS is to approximate the behaviour of $\alpha(X_t)$ in a neighbourhood of X_t using Ito's Lemma

$$d\alpha(X_t) = \alpha'(X_t)\, dX_t + \frac{1}{2}\alpha''(X_t)\, dt \tag{28}$$

$$\alpha(X_{t+h}) \approx \alpha(X_t) + \alpha'(X_t)(X_{t+h} - X_t) + \frac{1}{2}\alpha''(X_t)\, h \tag{29}$$

The law of the Ozaki-Shoji scheme on the time interval $(0, \Delta]$ is given by the solution of the linear SDE

$$dX_t = \left\{ \alpha(x) + \alpha'(x)(X_t - x) + \frac{1}{2}\alpha''(x)\, t \right\} dt + dB_t \tag{30}$$

i.e. a Gaussian process. By the time-homogeneity this DS is defined by the iterative formulae

$$\tilde{W}_\Delta^i \overset{iid}{\sim} \mathcal{N}\left(0, \frac{\exp\left\{2\alpha'\left(Y_{(i-1)\Delta}\right)\Delta\right\} - 1}{2\alpha'\left(Y_{(i-1)\Delta}\right)}\right) \tag{31}$$

$$Y_{i\Delta} = Y_{(i-1)\Delta} + \frac{\alpha\left(Y_{(i-1)\Delta}\right)}{\alpha'\left(Y_{(i-1)\Delta}\right)}\left(\exp\left\{\alpha'\left(Y_{(i-1)\Delta}\right)\Delta\right\} - 1\right) \tag{32}$$

$$+ \frac{\alpha''\left(Y_{(i-1)\Delta}\right)}{2\left(\alpha'\left(Y_{(i-1)\Delta}\right)\right)^2}\left\{\exp\left\{\alpha'\left(Y_{(i-1)\Delta}\right)\Delta\right\} - 1 - \alpha'\left(Y_{(i-1)\Delta}\right)\Delta\right\} + \tilde{W}_\Delta^i \tag{33}$$

3. A simulation study

A standard way to compare DSs is related to the concepts of weak and strong convergence. Y^Δ is said to be a strong approximation of (1) if $\exists \Delta^*, k, \mathcal{S} > 0 : \forall \Delta \leq \Delta^*$

$$\mathbb{E}\left|X_T - Y_T^\Delta\right| \leq k\Delta^{\mathcal{S}} \tag{34}$$

where $\mathcal{S}$ is the rate of convergence. This strong converge criterion basically states the L_1 convergence of the last simulated point Y_T^Δ to X_T. As such, the rate $\mathcal{S}$ is an indicator of how well Y^Δ approximates the paths of X (for a fixed ω). The convergence is not uniform on the time interval $[0, T]$ and the leading order constant k depends on (1).

Y^Δ is said to be a weak approximation of (1) if $\exists \Delta^*, k, \mathcal{W} > 0 : \forall \Delta \leq \Delta^*, g \in \mathcal{G}$

$$\left|\mathbb{E}\left[g\left(X_T\right)\right] - \mathbb{E}\left[g\left(Y_T^\Delta\right)\right]\right| \leq k\Delta^{\mathcal{W}} \tag{35}$$

where $\mathcal{W}$ is the rate of weak convergence and $\mathcal{G}$ is a class of test functions. Here the rate $\mathcal{W}$ is an indicator of how accurately the distribution of Y^Δ approximates the distribution of X. Hence this convergence criterion is more indicated if we are interested in Monte Carlo simulations based on Y^Δ. Similarly to (34) the convergence in not uniform on $[0, T]$ and the constant k of (35) depends on the SDE (1), limiting the practical appeal of these criteria. Our empirical results shows that DSs with the same $\mathcal{W}$ can perform very differently.

The framework of the simulation study is very simple: we consider a unit diffusion coefficient SDE X (3) and a functional F, possibly path-dependent, of interest. In this framework we compare the efficiency of EA and the three DSs previously introduced.

As EA does not clearly involves any discretisation error, its efficiency is inversely proportional to the average computational cost required to sample a single realisation of the functional $F(X)$.

For a given Y^Δ, the smallest computational cost, i.e. the biggest Δ, required for $F\left(Y^\Delta\right)$ to be an accurate approximation of $F(X)$ is then computed. More precisely, we are interested in how similar the distribution of $F\left(Y^\Delta\right)$ is to the distribution of $F(X)$. Our test of choice is the two-sided two-samples Kolmogorov-Smirnov (KS) test. EA is used to sample $F(X)$

exactly. Let $\alpha \in (0,1)$ be a fixed threshold and Δ^* be the biggest value of Δ such that the p-value of the KS test of $\left\{F(X), F\left(Y^{\Delta}\right)\right\}$ is higher then the threshold α. The efficiency of Y^{Δ} is then defined as inversely proportional to the computational cost required for the simulation of a single realisation of the functional $F\left(Y^{\Delta^*}\right)$.

To compute the KS test of $\left\{F(X), F\left(Y^{\Delta}\right)\right\}$ we choose to sample $N \in \mathbb{N}$ skeletons from X using EA and N discretisation using Y^{Δ}. For each one of these samples the value of the functional F is computed resulting in $2N$ samples: N exact and N approximated observations. Finally the p-value of the KS statistic calculated over these $2N$ samples. Moreover to decrease the variance of the KS test (that in this framework is just stochastic noise) we average its value over $M \in \mathbb{N}$ repetitions. All these simulations needs to be repeated until we find the right Δ^* for each of the three DSs considered in the comparison, i.e. the smallest Δ so that we accept the null hypothesis according to the KS test. Finally we repeat all these steps for a reasonable number of combinations of the parameters of the SDE, to obtain computational cost surfaces (as a function of the parameters) for EA and the DSs.

In our simulation study the following arbitrary values are considered: $\alpha = 0.05, N = 10^5, M = 10^3$. The choice of the KS test is arbitrary too, but there are a number of reasons why we opted for the this test. First of all, it has an intuitive meaning. More importantly, it is possible to obtain the limiting distribution of the KS statistic under the null hypothesis. Lastly we want to be cautious about our conclusions. The use of a more powerful goodness of fit test would pose questions about the robustness of our results to the choice of the test statistic considered. This would be especially true for tests that give more importance to the tails of the distribution, as preliminary examination of the histograms of the densities involved reveals that the biggest differences are usually in the tails.

The aim of this simulation study is to obtain useful indication about the efficiency of EA and the three DSs. The choice of the diffusion models that we take into account reflects this objective, they are "toy examples".

3.1. **The case of EA1.** The class of parametric diffusion models that can be considered is limited by the assumptions of EA1. We focus on the following three models:

- The PSINE SDE

$$dX_t = \theta \sin(\gamma X_t)\, dt + dB_t \qquad \theta > 0, \gamma > 0 \tag{36}$$

- The NSINE SDE

$$dX_t = \theta \sin(\gamma X_t)\, dt + dB_t \qquad \theta < 0, \gamma > 0 \tag{37}$$

- The PTANH SDE

$$dX_t = \theta \tanh(\gamma X_t)\, dt + dB_t \qquad \theta > 0, \gamma > 0 \tag{38}$$

- The NTANH SDE

$$dX_t = \theta \tanh(\gamma X_t)\, dt + dB_t \qquad \theta < 0, \gamma > 0 \tag{39}$$

We take into account these models because they summarise a good range of diffusion dynamics. In every model the starting point x and the terminal time T are fixed to 0 and 1 respectively.

The functionals considered are the last point $L(X) = X_T$ and the maximum of the path $M(X) = \sup_{0 \leq s \leq T} X_s$. For $M(X)$ we simulate the maximum of a BB between each discretized value even when dealing with DSs.

In Figures 4 to 11 the four plots on the top of each figure represents on the Z-axis the computational time required by EA and by the three DSs to complete the simulation (with the required level of accuracy) as function of the values of the SDE's parameters.

In the remaining 3 plots of each figure, the ratio of the computational time of a DS over the computational time of EA is represented on the Z-axis, again as a function of the SDE's parameters. Whenever possible, the white colour represents a unitary ratio, the red colour a ratio lower than 1 and the blue colour a ratio higher than 1. We remark that these ratios are the results of our arbitrary choices. For example comparing a higher number of observations would increase the power of the test and this would result in a lower efficiency of the DSs.

Moreover the shape of these surfaces is of interest on its own, as it says how the DSs behave with respect to parametric classes of drift and diffusion coefficients. From this point of view EA is a valuable validation tool.

The two main goals of this simulation are: commenting the efficiency of EA with respect to other DSs and study the behaviour of EA and of the DSs with respect to qualitative characteristics of the diffusion model X. Regarding the first of these points, we note that:

(1) EA1 has a computational cost that is comparable to that of good DSs such as the Predictor-Corrector scheme. This means that there is generally not a huge difference between simulating from the approximated or the exact law of the process.
(2) EA1 is favoured when we consider the functional $M(X)$. One possible explanation for this is that while simulating $L(X)$ the discretisation errors of every step are likely to cancel, but when simulating $M(X)$ the errors are likely to accumulate. Moreover, we are using two levels of approximations: we approximate the discretized path and also the maximum of the path conditionally on the discretisation.
(3) While all DSs share a very good performance when γ is very low, independently of the value of θ, this is not the case with EA1. While the computational cost in EA1 remains very contained it increases with $|\theta|$ more rapidly. Conversely, EA1 has a better efficiency than DSs when $|\theta|$ is low.
(4) There are situations where EA1 performs much better, and this is the case of the PTANH model. This happens because if $\alpha^2 = \alpha'$ in (3)_it follows that EA always accept the proposed skeleton. In this case we actually know the transition density of X. This is the case when $\gamma = \theta$ in the PTANH. When we move away from the diagonal the range of $\alpha^2 + \alpha'$ increases and so does the rejection rate.

Concerning the second of these points, we note that:

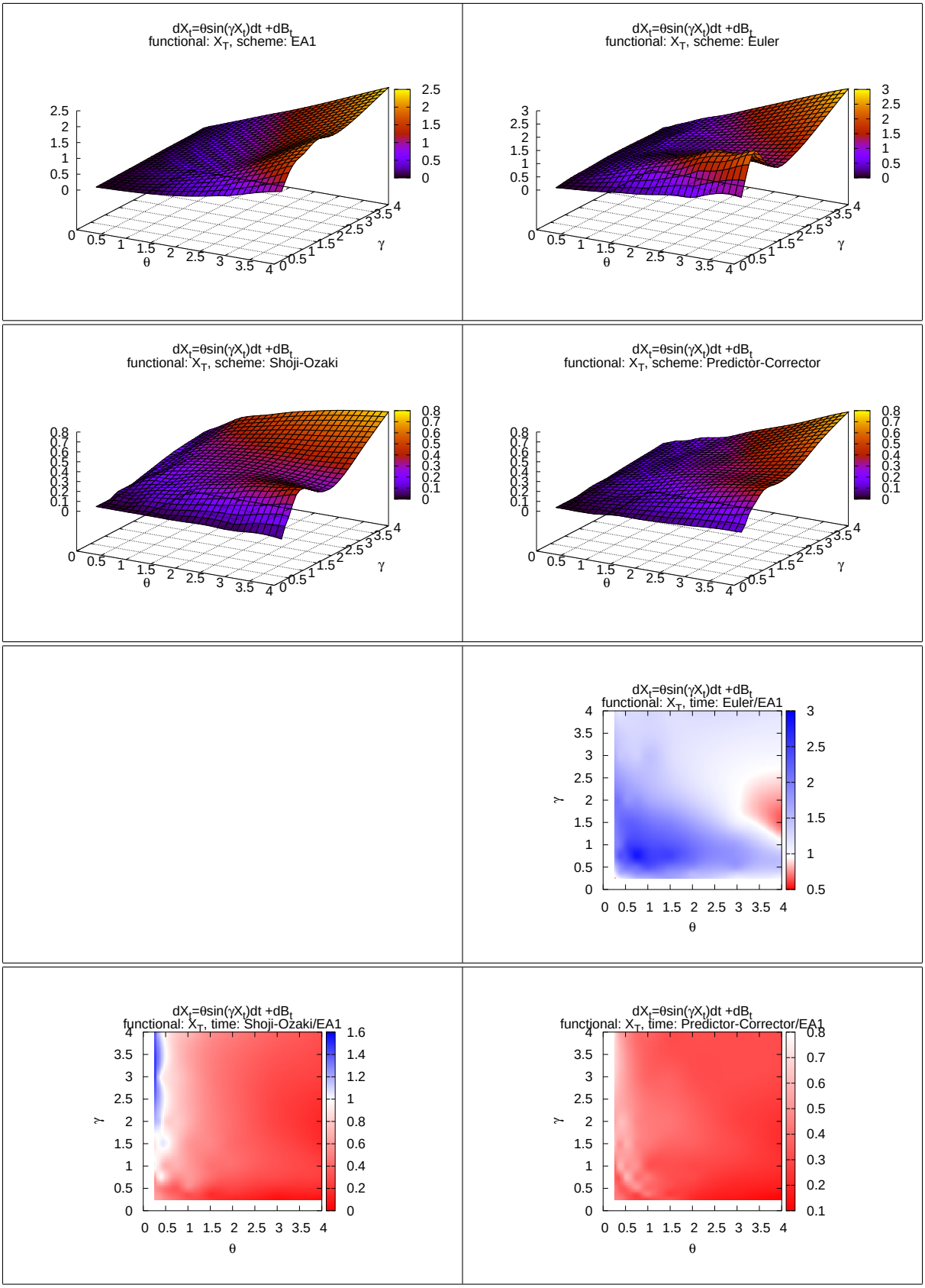

FIGURE 4. model: PSINE, functional: $L(X)$

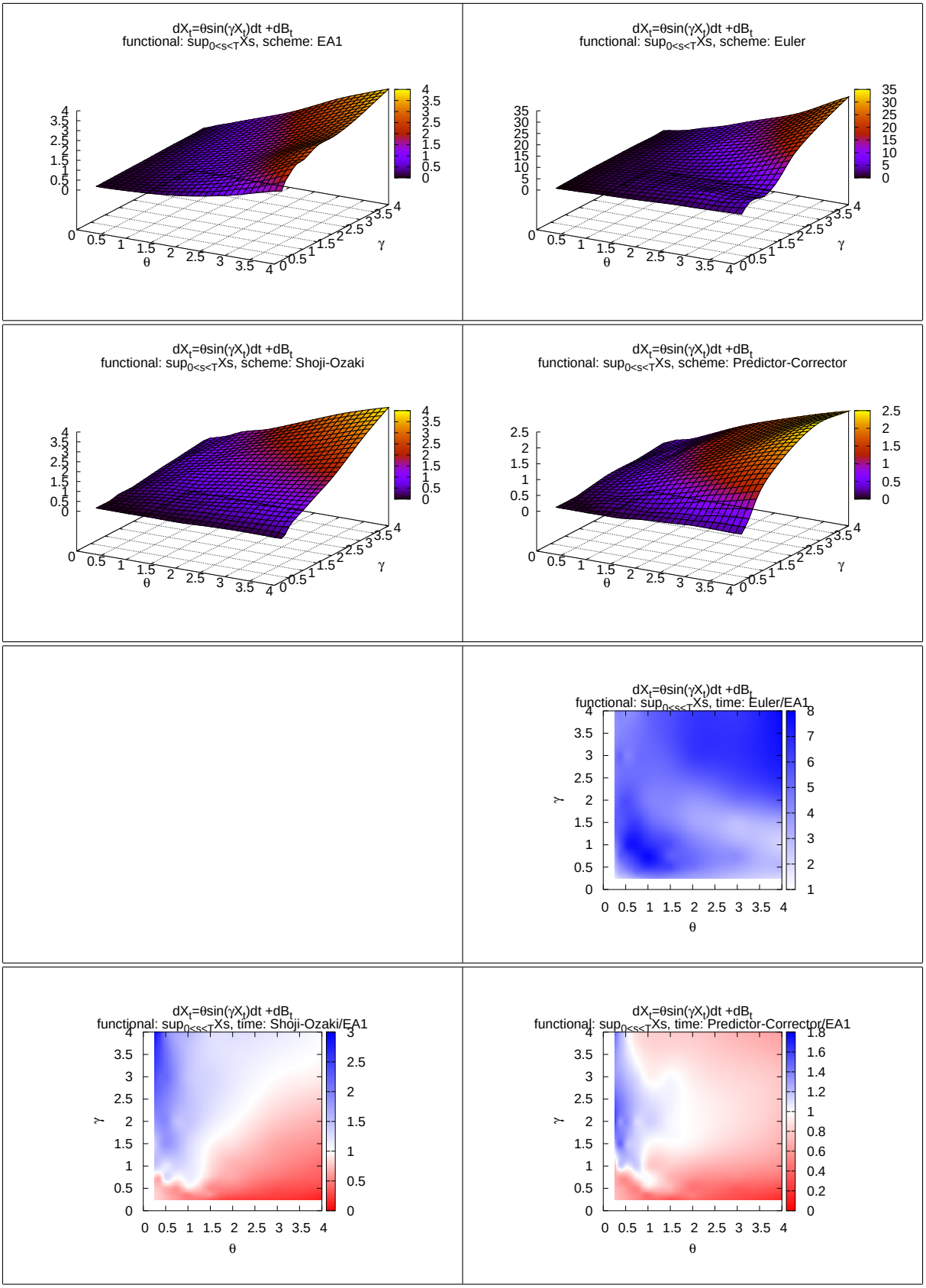

FIGURE 5. model: PSINE, functional: $M(X)$

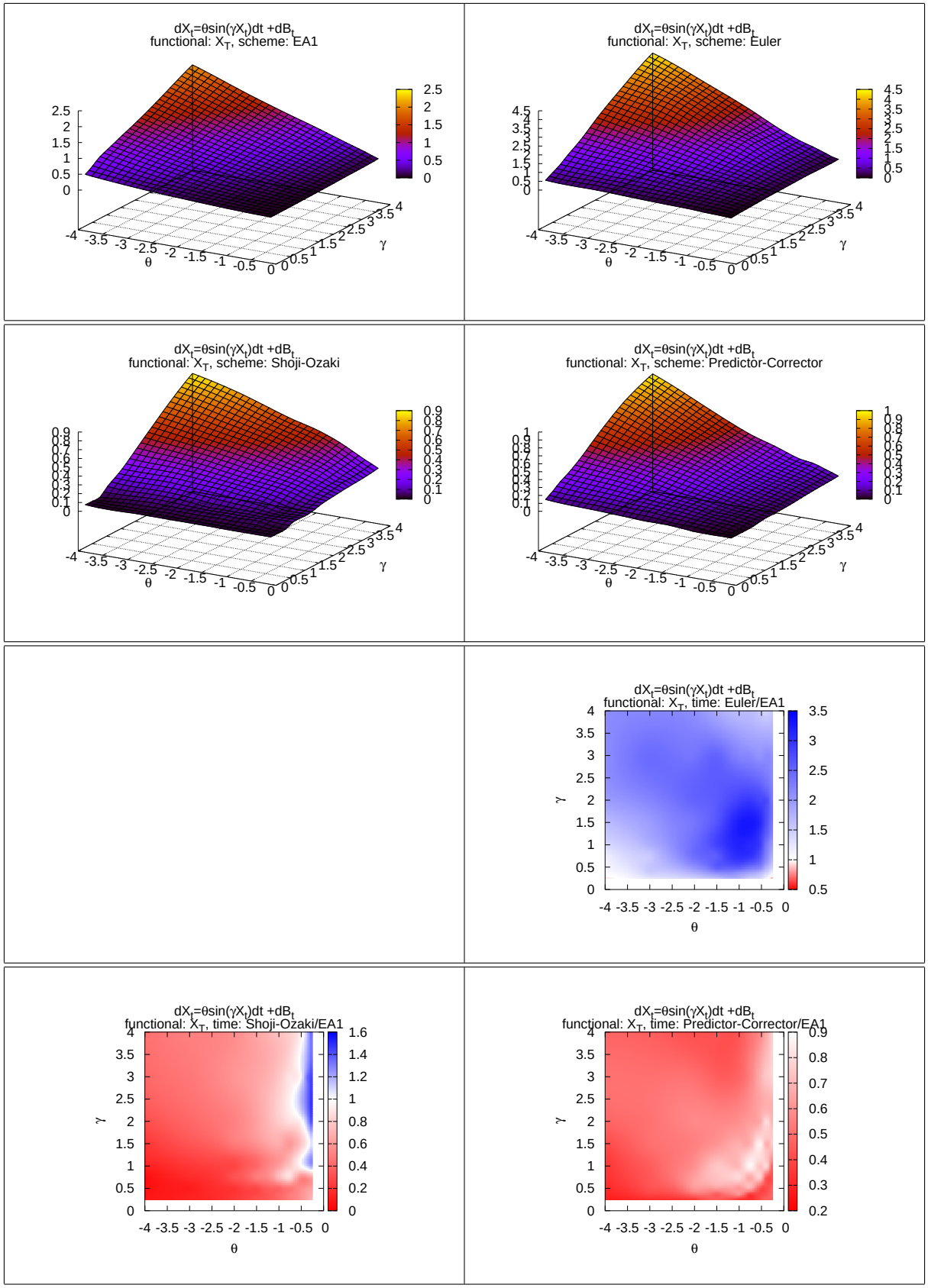

FIGURE 6. model: NSINE, functional: $L(X)$

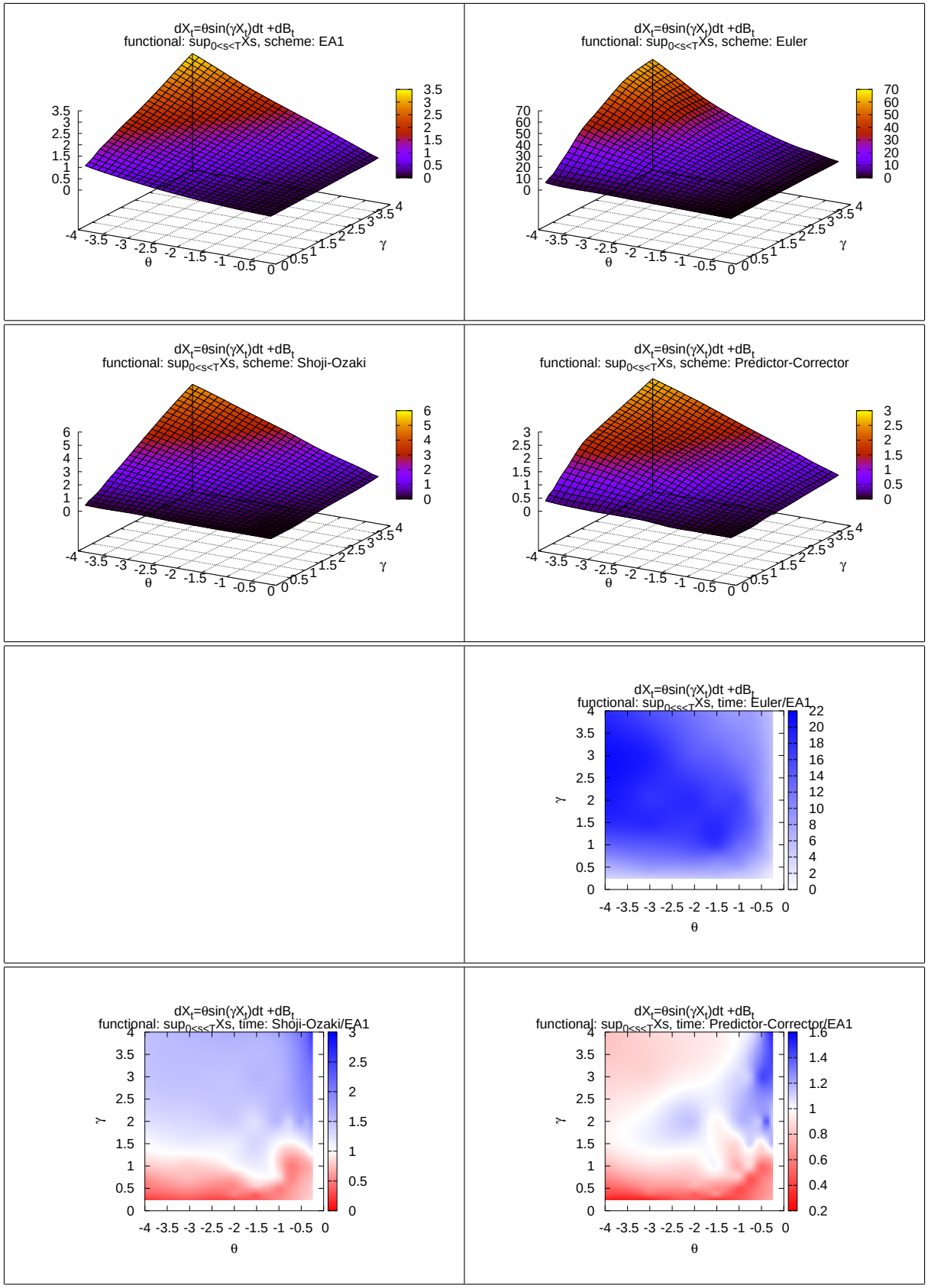

FIGURE 7. model: NSINE, functional: $M(X)$

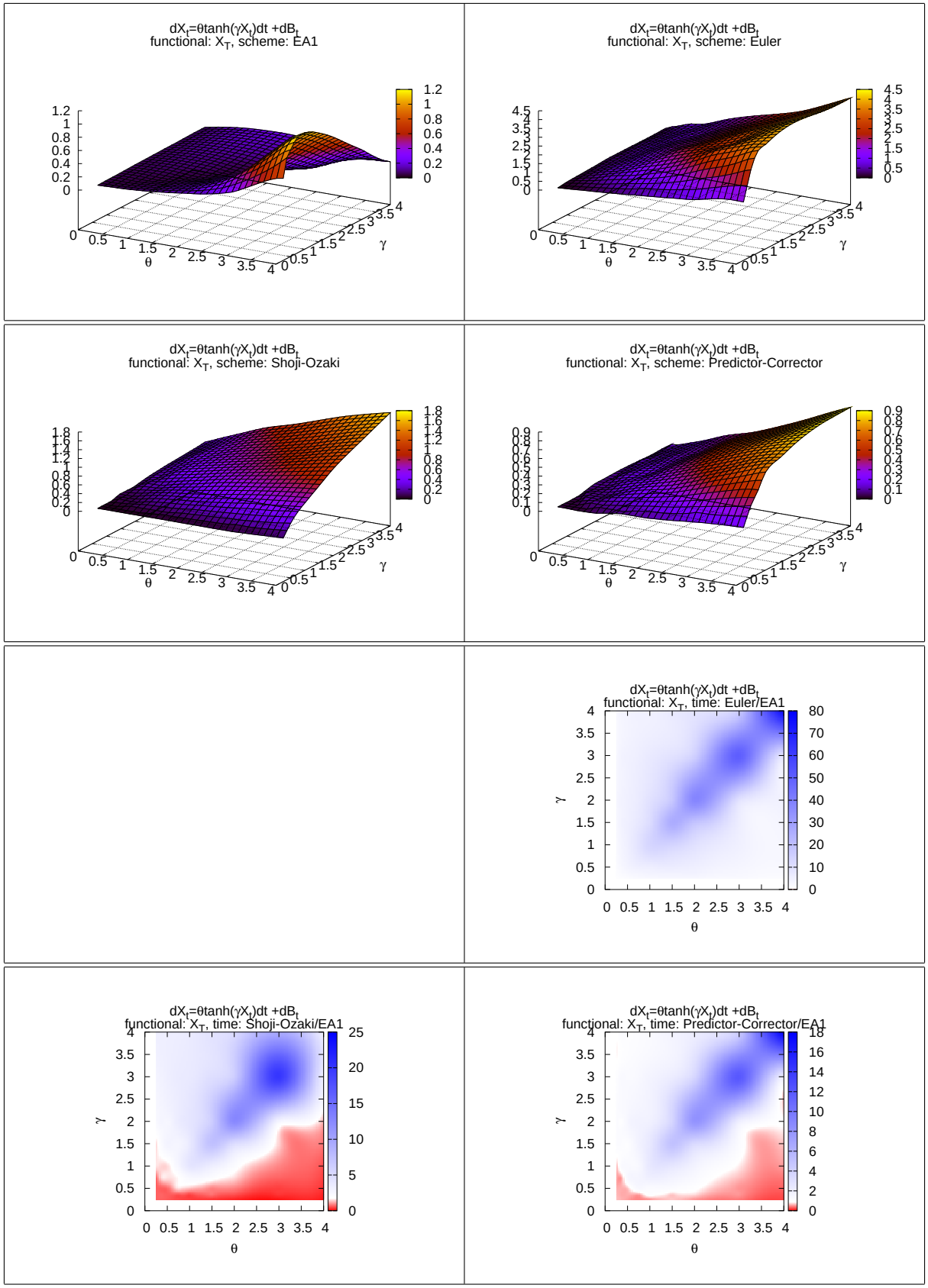

FIGURE 8. model: PTANH, functional $L(X)$. Ozaki-Shoji scheme does not converge if $-\theta = \gamma = 4$

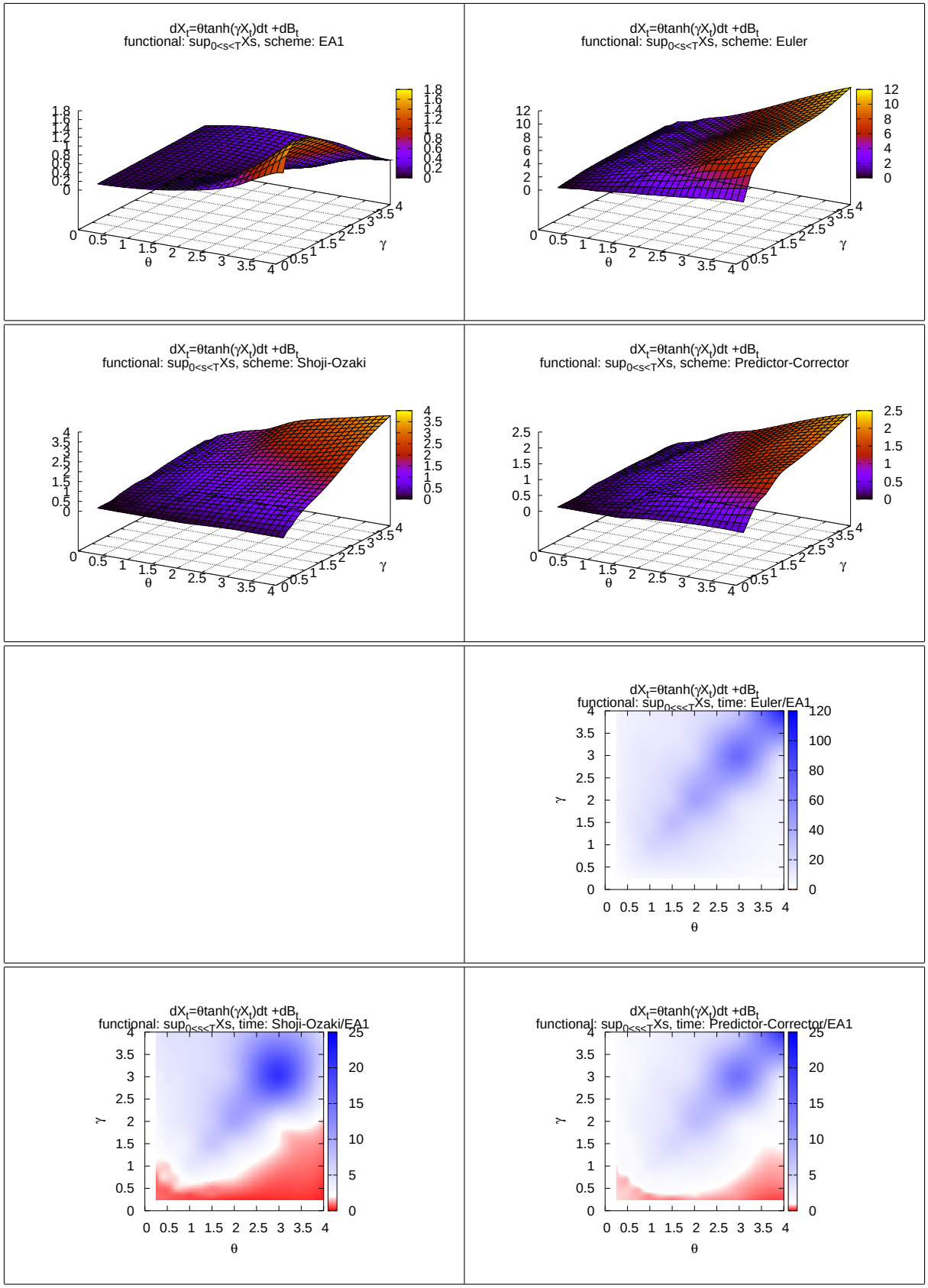

FIGURE 9. model: PTANH, functional: $M(X)$. Ozaki-Shoji scheme does not converge if $\theta = \gamma = 4$

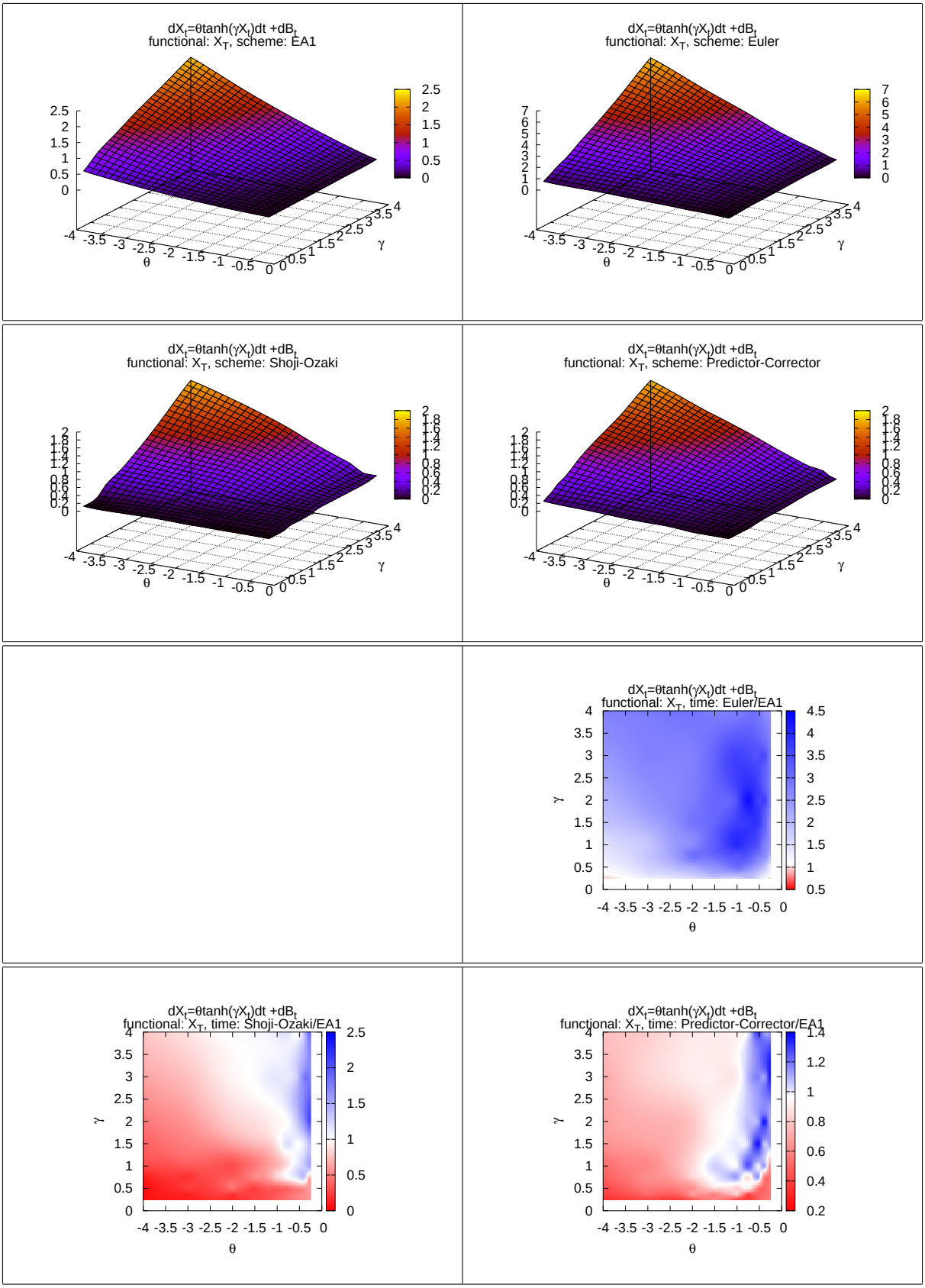

FIGURE 10. model: NTANH, functional: $L(X)$

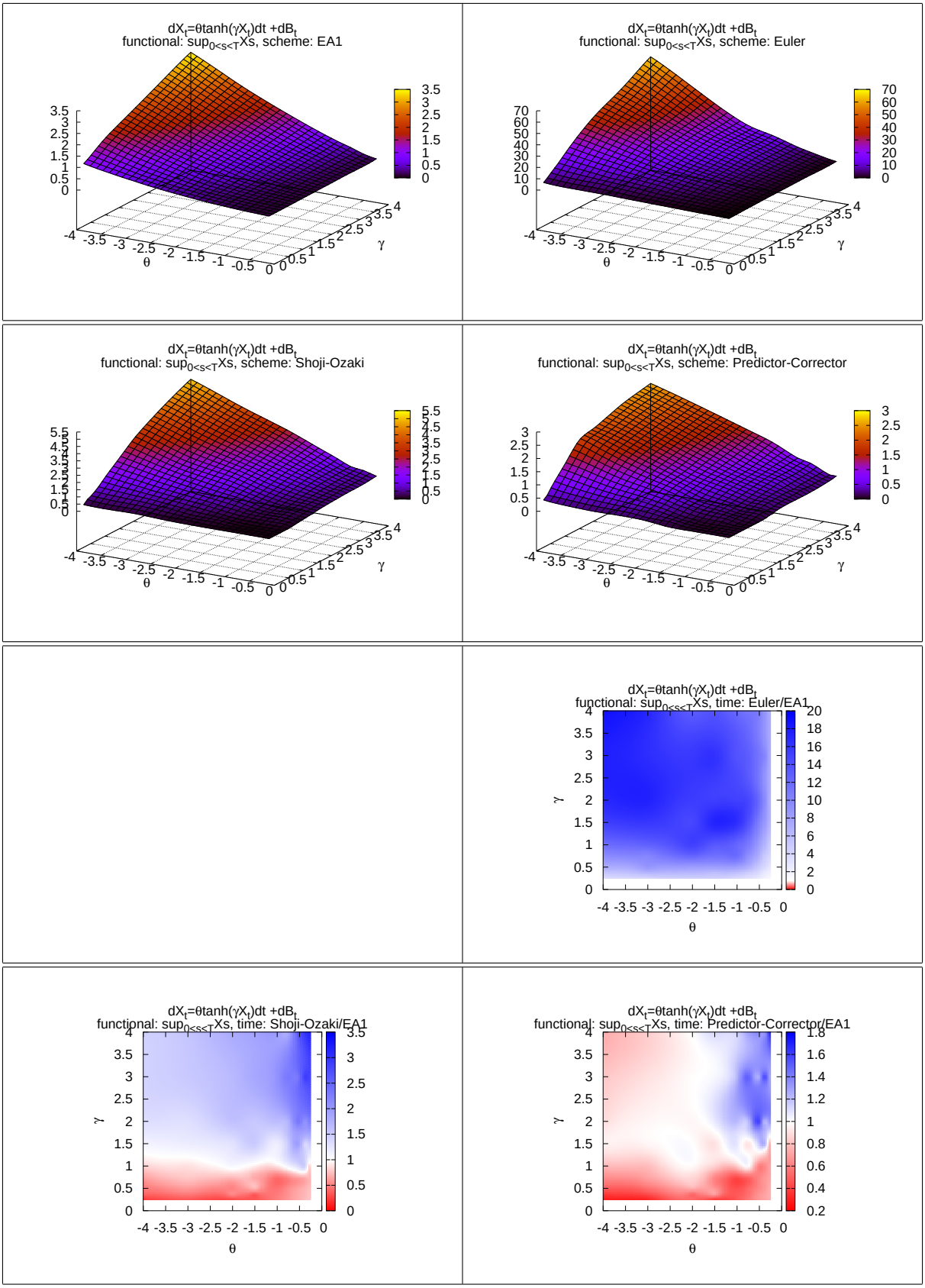

FIGURE 11. model: NTANH, functional: $M(X)$

(1) Euler scheme is clearly the least efficient DS. In some situation it can be 20 times more inefficient than the other two DSs. Moreover the implementation difficulty off all these DSs is comparable.
(2) Predictor-Corrector and Ozaki-Shoji scheme shares more or less the same efficiency, even if in the same situations the former can be two times more efficient than the latter. Furthermore, the Ozaki-Shoji scheme exhibits numerical instabilities every time $\alpha'\left(Y_{(i-1)\Delta}\right) \approx 0$. Hence it is necessary to introduce an extra check for the algorithm that would slow down the simulation even more. All this suggests that the Predictor-Corrector scheme should be the first choice in most situations.
(3) As already stated, the weak convergence criterion is not very useful from a practitioner point of view. In fact both the Euler DS and the Predictor-Corrector DS share the same unit-order of weak convergence.
(4) It is very difficult with this limited amount of information to infer a rule of thumbs that links the efficiency of the DSs to the qualitative behaviour of the target diffusion model X. We just notice that the computational time surface has more or less the same shape in all the DSs. The difference is in the multiplicative factor.

3.2. **The case of EA3.** We consider the following diffusion models

- the LANG SDE

$$dX_t = -k\text{sign}\left(X_t\right)\left|X_t\right|^{\beta} dt + dB_t \qquad k > 0, \beta \in \mathbb{N} \tag{40}$$

- the XXCUBE SDE

$$dX_t = \left\{-\alpha X_t^3 + \beta X_t\right\} dt + dB_t \qquad \alpha > 0, \beta > 0 \tag{41}$$

In the case of EA3, we can no longer easily and exactly simulate from the law of $M\left(X\right)$, hence the comparison is only limited to the $L\left(X\right)$ functional. As the results of Section 3.1 suggests that Shoji-Ozaki scheme does not offer any clear advantage against Predictor-Corrector scheme, while showing numerical instabilities, we decide to include the Euler DS and the Predictor-Corrector DS in the comparison only.

Regarding the efficiency of EA3 with respect to Predictor-Corrector scheme we notice that the former is always less efficient then the the latter. The most obvious reason is that EA3 is much more complicated from an algorithmic point of view than EA1, and this results in a higher computational time. However, everything is relative to the choice of the specific comparison criterion considered. As a rule of thumb we can say that EA3 is a factor of 10 slower than EA1.

Given these results, there is no obvious link between qualitative behaviour of the diffusion model X and the expected efficiency of the DSs. The relative efficiency of Euler with respect to Predictor-Corrector is confirmed. But for the first time we observe a difference in the shape of the computational time surfaces of the Euler and the Predictor-Corrector schemes. This is the case of the LANG model. More investigation is needed to find the reasons of this result.

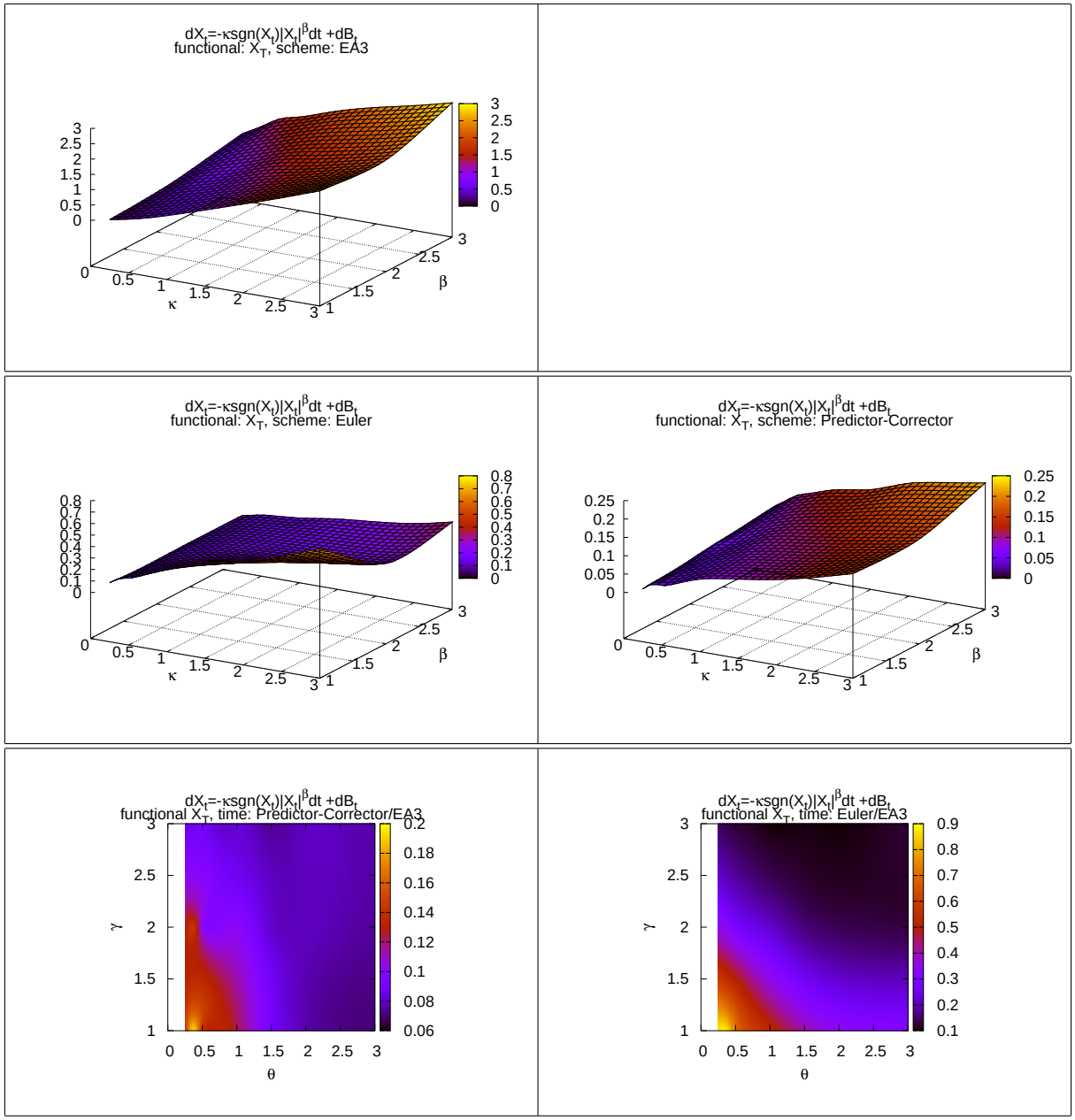

FIGURE 12. model: LANG, functional: $L(X)$

4. THE MULTI-DIMENSIONAL SETTING

We now concentrate on the unit-diffusion d-dimensional SDE

$$d\mathbf{X}_t = \alpha(\mathbf{X}_t)\,dt + d\mathbf{B}_t \qquad t \in [0,T] \tag{42}$$
$$\mathbf{X}_0 = \mathbf{x}$$

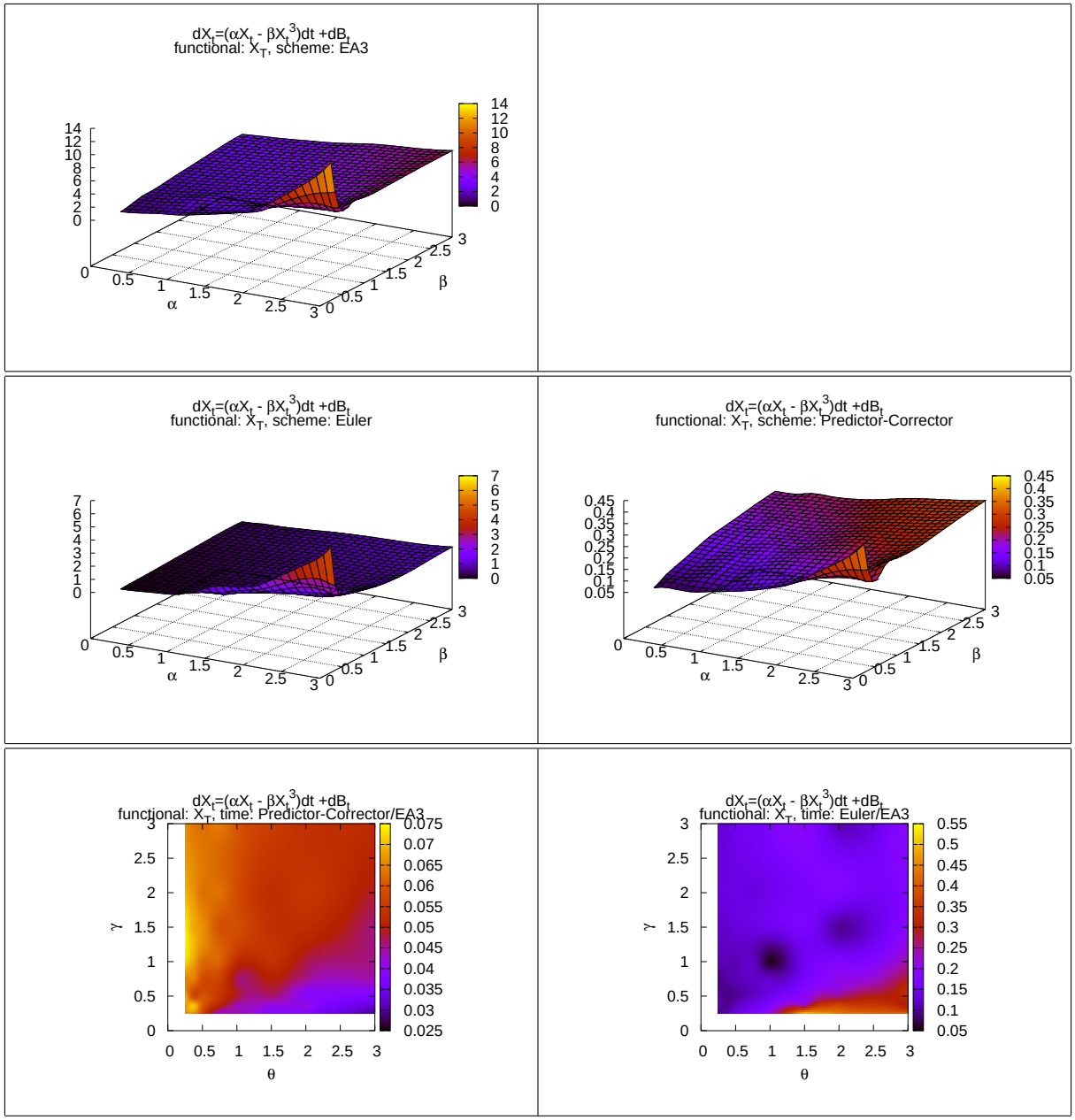

FIGURE 13. model: XXCUBE, functional: $L(X)$

where $\mathbf{B}_t$ is the d-dimensional BM. The drift coefficient α is assumed to satisfy proper conditions that guarantee the existence of a unique non-explosive strong solution of (42). In this section $\mathbb{Q}_T^{\mathbf{x}}$ and $\mathbb{W}_T^{\mathbf{x}}$ represent the law of the diffusion process $\mathbf{X}$ solution of (42) and the d-dimensional Wiener measure for the initial condition $\mathbf{B}_0 = \mathbf{x}$ respectively. Let $\mathbb{X}$ be the state space of $\mathbf{X}$.

It is possible to find equivalent conditions to $(C1)$-$(C4)$ for the d-dimensional framework and we refer to Beskos et al. [2006b] for a formal development of EA in this setting. The main theoretical limitations of EA in the d-dimensional setting are:

(1) the necessary and sufficient condition for the existence of a transformation from a generic d-dimensional SDE to the unit diffusion coefficient SDE (42) is quite demanding (see Ait-Sahalia [2002]);
(2) we require the existence of a potential function $A : \mathbb{R}^d \to \mathbb{R}$ such that $\alpha(\mathbf{u}) = \nabla A(\mathbf{u})$.

EA then generalises to this setting in a simple way. We define the d-dimensional BBM $\mathbf{Z}$ as a d-dimensional BM with initial value $\mathbf{x}$ conditioned on having its final value $\mathbf{Z}_T$ distributed according to $h_{\mathbf{x},t}(\mathbf{u})$ where

$$h_{\mathbf{x},t}(\mathbf{u}) \propto \exp\left\{A(\mathbf{u}) - \frac{\| \mathbf{u}-\mathbf{x} \|^2}{2T}\right\} \tag{43}$$

and denote with $\mathbb{Z}_T^{\mathbf{x}}$ its law. Let $\phi : \mathbb{X} \to \mathbb{R}$, assumed to be bounded below, be defined as $\phi(\mathbf{u}) := (\| \alpha(\mathbf{u}) \|^2 + \mathrm{div}\alpha(\mathbf{u}))/2 - l$ and $l := \inf_{r\in\mathbb{X}} \phi(\mathbf{r}) < \infty$. As before $\mathbb{L}$ denote the law of a unit rate Poisson Point Process (PPP) on $[0,T]\times[0,\infty)$, and let $\Phi = \{\chi, \psi\}$ be distributed according to Φ. We define the event Γ as

$$\Gamma := \bigcap_{j\geq 1} \phi\left(\mathbf{Z}_{\chi_j}\right) \leq \psi_j \tag{44}$$

The following extension of Theorem 1 holds

Theorem 2. ***(Multivariate Wiener-Poisson factorisation)*** *If* $(\mathbf{Z}, \Phi) \sim \mathbb{Z}_T^{\mathbf{x}} \otimes \mathbb{L} \mid \Gamma$ *then* $\mathbf{Z} \sim \mathbb{Q}_T^{\mathbf{x}}$

Proof. see Beskos et al. [2006b] □

Using Theorem 2, the extension of EA1 to the d-dimensional setting is immediate. The only difficulty is finding the global maximum of ϕ over the domain $\mathbb{X}$. The extension of EA3 to the d-dimensional setting is similarly immediate, with the added difficulty that we now have to compute the maximum of ϕ over a bounded d-dimensional hyper-rectangle in $\mathbb{X}$.

As in the case of the one-dimensional EA the simulation of $\mathbf{Z}$ requires to sample from $\{h_{\mathbf{x},T}(\mathbf{u})\}_{\mathbf{x}\in\mathbb{X}}$. Unfortunately the high dimensionality of the problem makes any adaptive approach, such as the ones in Peluchetti [2007], infeasible. However, if we can find a d-dimensional matrix K, a vector $\mathbf{v}$ and a constant k such that

(1) $\forall \mathbf{u} \in \mathbb{X}\, A(\mathbf{u}) \leq (\mathbf{u}-\mathbf{v})' K (\mathbf{u}-\mathbf{v}) + k$
(2) $\int_{\mathbb{X}} \exp\left\{(\mathbf{u}-\mathbf{v})' K (\mathbf{u}-\mathbf{v}) - \frac{\|\mathbf{u}-\mathbf{x}\|^2}{2T}\right\} < \infty$

it is possible to implement a simple accept-reject sampler using a multivariate Gaussian variate as proposal (the LPS from now on). In most diffusion models of interest it is possible to find such $K, \mathbf{v}, k$ that satisfies these conditions (at least for T small enough) indeed.

To see how the computation cost of EA scales as d increases we considered two test d-dimensional SDEs defined by their potential function A:

Dimension	1	2	4	8	16
EA1 comp.cost	0.48	0.92	1.85	5.56	27.51

TABLE 1. The multi-dimensional EA1

Dimension	1	2	3	4	5	6	7	8	9	10	11
LPS acceptance	83.9	71.3	61.3	52.3	44.4	37.4	32.5	27.2	23.3	19.3	17.2
EA3 comp. cost	0.39	0.31	0.40	0.46	0.75	1.21	2.15	5.87	15.9	45.9	129.9

TABLE 2. The multi-dimensional EA3

- the d-dimensional SINE, $A(\mathbf{u}) = -\cos\left(\sum_{i=1}^{d} u_i\right)$
- the d-dimensional LANG, $A(\mathbf{u}) = -\sum_{i=1}^{n} u_i^4$

The initial value $\mathbf{x}$ is the origin of $\mathbb{R}^d$ and $T = 1$. Theoretical consideration suggests that partitioning $[0, T]$ in sub-intervals of length T/d (and applying EA sequentially) would keep the acceptance rate of EA stable as d changes. Our simulation study suggests that this intuition is correct and we adopt this strategy.

In Table 1 we report the computational cost (in seconds) required to sample 1000 observations from the d-dimensional SINE SDE using EA1. We see that, apart from variations due to the implementation, the computational cost increases linearly with d. Due to the bounded nature of this example the acceptance rate of the LPS is stable.

In Table 2 we report the computational cost of the d-dimensional EA3 required to sample 100 observations from the d-dimensional LANG SDE. While the acceptance rate of the LPS decreases with d (as expected) this is not the reason of the explosive behaviour of the d-dimensional EA3's computational cost. The problem is the computation of the maximum of over a bounded d-dimensional hyper-rectangle that requires at least 2^d computations.

5. Conclusions

In this paper we have performed a simulation study of EA's efficiency. We have investigated the computational time required by EA1 and EA3 in different scenarios, both in the one and d-dimensional setting. In the one-dimensional case the results of this simulation are compared with the computational time required by three other numerical schemes too. The results are quite encouraging: EA1 proves to be very competitive with respect to the other DSs as the computational time required for an accurate approximation using traditional DSs is comparable to that necessary for an exact simulation of the SDE. Thus its exact nature makes EA1 the preferred discretization scheme, according to our opinion. Additionally, knowing the true distribution of the path of the process conditioned on the returned skeleton makes the exact simulation of some path-dependent functionals possible.

In the case of EA3, the added complexity of the algorithm has resulted in a less efficient scheme. The choice of the suggested discretization schemes thus depends on the particular application. When a very precise simulation is needed, EA3 still presents a reasonable efficiency, being roughly a factor of 10 slower then EA1.

In the d-dimensional case EA1 scales quadratically with the dimension d, while in most cases EA3 scales exponentially. However, the exact simulation of not too high-dimensional problems remains feasible.

Moreover, the exact nature of EA is of great importance when efficiency is not the first concern. Thanks to EA we have been able to analyse the efficiency of other discretization schemes with a high degree of accuracy. And we did so by considering diffusion models for which the exact solution is not available in a closed form. Another example of an application of EA as a validation tool is that of inference for diffusion model. Some methods rely on Euler-like approximations of the diffusion process. By using the Euler (or similar) discretization scheme again for generating paths to test the method would false the conclusions of the experiment.

Concerning the behaviour of the DSs with respect to the SDEs' parameters, the small amount of information collected is not sufficient to establish any rule of thumb. Additional research is needed to gain more insight on this interesting topic.

References

Y. Ait-Sahalia. Closed-Form Likelihood Expansions for Multivariate Diffusions. 2002.

C. Albanese and A. Kuznetsov. Transformations of Markov processes and classification scheme for solvable driftless diffusions. *preprint, www3. imperial. ac. uk/mathfin/people/calban/papersmathfin*, 2005.

A. Beskos and G.O. Roberts. Exact simulation of diffusions. *Ann. Appl. Probab*, 15:2422–2444, 2005.

A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12:1077–1098, 2006a.

A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. A new factorisation of diffusion measure and finite sample path construction. *Methodology and Computing in Applied Probability, Submitted*, 2006b.

Bruno Casella. *Exact MC simulation for diffusion and jump-diffusion processes with financial applications*. PhD thesis, IMQ - Bocconi University, 2005.

WR Gilks. Derivative-free adaptive rejection sampling for Gibbs sampling. *Bayesian Statistics*, 4(2):641–649, 1992.

WR Gilks and P. Wild. Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics*, 41(2):337–348, 1992.

P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

G. Maruyama. Continuous markov processes and stochastic equations. *Rend. Circ. Mat. Palermo*, 4:48–90, 1955.

Stefano Peluchetti. *An analysis of the efficiency of the Exact Algorithm*. PhD thesis, IMQ - Universitá Commerciale Luigi Bocconi, 2007.

I. Shoji and T. Ozaki. Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, 16(4):733–752, 1998.