

Markov Chain Approximations to Stochastic Differential Equations by Recombination on Lattice Trees

Francesco Cosentino* Harald Oberhauser† Alessandro Abate‡

Abstract

We revisit the classical problem of approximating a stochastic differential equation by a discrete-time and discrete-space Markov chain. Our construction iterates Carathéodory’s theorem over time to match the moments of the increments locally. This allows to construct a Markov chain with a sparse transition matrix where the number of attainable states grows at most polynomially as time increases. Moreover, the MC evolves on a tree whose nodes lie on a “universal lattice” in the sense that an arbitrary number of different SDEs can be approximated on the same tree. The construction is not tailored to specific models, we discuss both the case of uni-variate and multi-variate case SDEs, and provide an implementation and numerical experiments.

1 Introduction

Consider a stochastic differential equation (SDE) with generic vector fields μ and σ

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x \in \mathbb{R}^d \quad (1)$$

and driven by a multi-dimensional Brownian motion W . The aim of this article is to construct a map

$$\text{DISCRETIZE}_n : (\mu, \sigma, x) \mapsto (M_n, x)$$

parametrized by a discretization parameter $n = 1, 2, \dots$ that takes as inputs the drift μ and the diffusion vector field σ , as well as the starting position x , and which outputs the transition matrix M of a Markov chain¹ (MC), denoted as $X^n = (X_i^n)_{i=0,1,2,\dots}$ with starting value $X_0^n = x$, and such that

$\text{Law}(X^n) \rightarrow \text{Law}(X)$ as $n \rightarrow \infty$ and the state space of the Markov Chain X^n is a sparse tree.

Naive discretization strategies generically lead to an explosion of the state space of the MC.

Example 1. Consider the one-dimensional SDE

$$dX_t = \sigma(X_t) dW_t, \quad X_0 = x \in \mathbb{R}$$

and the Markov chain X^n given (implicitly) as the discretized Euler scheme

$$X_{i+1}^n = X_i^n + n^{-1/2} \sigma(X_i^n) B_i,$$

*Mathematical Institute, University of Oxford & The Alan Turing Institute, name.surname@maths.ox.ac.uk

†Mathematical Institute, University of Oxford, oberhauser@maths.ox.ac.uk

‡Department of Computer Science, University of Oxford & The Alan Turing Institute, name.surname@cs.ox.ac.uk

¹Throughout we refer to a discrete-time and discrete-space Markov process when we refer to a Markov chain (MC).

where B_i denote independent Bernoulli random variables. Then X^n converges in law to X , but for generic volatility σ the cardinality of the state space of X^n grows in time i as $O(2^i)$ because the tree nodes of the random walk “do not recombine”. However, the special case $\sigma(x) = 1$ shows that “small state spaces” are possible since in this case the simple symmetric random walk

$$X_{i+1}^n = X_i^n + n^{-1/2}B_i,$$

converges weakly to X – in this case a *Brownian Motion* – and the cardinality of the state space grows only linearly with the number of steps since X^n evolves on a “recombining tree”.

Although Example 1 is elementary, it motivates the arguably most popular approach to time-space discretization, namely to transform the SDE (2) into an SDE with constant volatility where naive time-space discretization strategies work. The main contribution of this work is to completely avoid such transformations (change of measure, Doss–Sussman, Lamperti, etc.) by using ideas from convex geometry about reducing the support of discrete measures; so called *recombination of measures*. This results in several attractive properties of DISCRETIZE_n compared to classical approaches: in particular, it naturally extends to the multi-variate case and even for one-dimensional diffusion it relaxes assumptions on the vector fields μ and σ . Moreover, the construction is universal in the sense that when applied to many SDEs X, Y, \dots , the resulting MCs X^n, Y^n, \dots evolve in the same state space that is a lattice, in fact on the same recombining tree with different branching probabilities.

Related Work. The problem of approximating an SDE by a MC arises in many areas:

Mathematical Finance. One of the earliest motivations came from American option pricing and resulted in the famous Cox–Ross–Rubinstein [9] model which uses a MC to approximate the Black–Scholes SDE. A big improvement was made in the work of Nelson and Ramaswamy [47], which uses the Lamperti transform to construct a recombining tree so that the number of states of the MC increases at most linearly with the number of steps, thus avoiding the exponential explosion of states in [9]. This approach has been refined and extended to models beyond Black–Scholes, such as multi-variate SDEs [3, 12], such as GBM, CIR and CEV models [2, 14, 16, 17, 36, 46, 55], regime-switching [38, 39, 54], stochastic volatility models [15, 35, 35, 44, 44], and many other model-specific solutions [19, 24, 41]. We also highlight that approaches that do not rely on transformations have been developed such as [49, 50, 51] that use probability density functions approximations such as Edgeworth expansions, model specific approaches [19, 41], or moment approximations [7, 8, 29, 34]. All of these have areas where they are advantage but among the drawbacks are they rely on either genuinely one-dimensional arguments, exploit model-specific properties, can result in negative transition probabilities that arise due to transformations or moment approximations.

Optimal Control. American option pricing can be seen as a special case of an optimal control problem where MC approximations are classic; see Kushner [31] and [32, Chapter 10, Theorem 6.2] for general background. An essential tool in this literature is to construct MCs by matching local moments [31, 32]² which ensures convergence to the SDE. These local consistency conditions are also central to our construction (Definition 3 and Lemma 2)³. Note that in general the topology of weak convergence is too coarse to guarantee that the solution of an optimal control problem for MCs can be transferred to SDEs but under additional assumptions this holds; in particular, we show that this is the case for our lattice-tree model approximation under essentially the same assumption used in the approximation of [32]. We also highlight [19, 41]

²See e.g. [32, Equation (1.3), page 71].

³In particular, for the schemes we build the conditions required by [31, 32] can be proved to be equivalent thanks to the bounds of Theorems 4, 8, 10.

Formal Verification. More recently, the formal verification community has started to investigate how to formally verify SDEs [4, 28, 56, 57]. Software tools, such as PRISM [33] or Storm [11], exploit structural properties of the transition matrix of MCs to allow scalable formal proofs that are applied to check whether probabilistic temporal-logic properties hold or not. These tools can be leveraged for the formal verification of SDEs by using them on MC approximations that are close in law to the given SDE. In this context, it is especially important that firstly, the discretization applies to large classes of SDEs under generic assumptions on drift and volatility vector fields; secondly, it is essential that the approximating MC transition matrix has structure, such as sparsity.

Similarly, the topic of reducing the support of a measure while matching statistics given as integrals against a set specified functions (such as monomials) is well studied:

Recombination. If the measure is discrete then it is a direct consequence of Carathéodory’s Theorem that such a reduction is possible. However, the proof of Carathéodory’s Theorem is not constructive and the design of algorithms that carry out such a measure construction efficiently is still the subject of recent research; e.g. just over the last ten years the articles [6, 20, 21, 22, 37, 43, 53] provide novel algorithms. Recombination has already been used in the context of SDE simulations: in order to make “Cubature on Wiener Space” [42] efficient, a recombination step is applied iteratively over time, similar in spirit to our construction; see [37] for a discussion on how powerful this can be for cubature methods on Wiener space. However, there the focus is to match the marginal distribution of an SDE at a fixed time but the total number of possible states that can be visited over time grows very quickly. This makes this approach too expensive to store the whole model which in turn is required for the above mentioned applications in finance, optimal control, and formal verification. Nevertheless, combining the result of this paper with cubature paths for Brownian motion might be an interesting future research venue.

Contribution and Outline. Our approach is inspired by the above mentioned local consistency conditions. The novelty is that we iterate Carathéodory’s recombination to obtain the algorithm $(M_n, x) := \mathbf{DISCRETIZE}_n(\mu, \sigma, x)$, such that the resulting MC $X^n = (X_i^n)$ with transition matrix M_n satisfies the following points:

1. the MC X_n is always a stochastic process, that is the transition matrix M_n is a stochastic matrix. This is noteworthy, since one of the biggest drawbacks of the landmark paper Nelson and Ramaswamy [47] and many of the above mentioned literature, is that it can result in “negative transition probabilities” between nodes, already in the one-dimensional case.
2. the transition matrix M_n is sparse and as a consequence the support of X_n grows slowly. This allows to store the whole model, which in turn allows the use in optimal stopping or formal verification algorithms.
3. the approach is not tailored to specific models under the standard regularity assumptions on the vector fields, and it extends to the multi-dimensional case. For dimensions $d \geq 3$ some additional assumptions are needed, but notice that there are very few baselines to compare to, since most results are model-specific or only apply in the one-dimensional case.
4. The state space is a universal lattice in the terminology of Chen and Yang [5]: for every n there exists a γ_n such that $\mathbf{DISCRETIZE}_n$ applied to many SDEs X, Y, \dots gives MCs X^n, Y^n, \dots that evolve on the same recombining tree that has as nodes a subset of the lattice $\gamma_n \mathbb{Z}^d$.

The structure of the paper is as follows: Section 2 provides general background and informally describes the algorithm $\mathbf{DISCRETIZE}_n$. Sections 3, 4 and 5 contain the main theoretical results and are divided

based on the dimension treated: Section 3 treats one-dimension state spaces and Section 4 treats two-dimensional state spaces. The general case is treated in 5 and makes stronger assumptions on the vector fields. Section 6 compares to previous work and expands on applications such as optimal stopping. Finally, Section 7 turns the theoretical results of the previous sections into algorithms and presents numerical experiments.

Acknowledgments

The authors want to thank The Alan Turing Institute and the University of Oxford for the financial support given. FC is grateful to Christian Bayer and Terry Lyons for helpful remarks. FC is supported by The Alan Turing Institute, TU/C/000021, under the EPSRC Grant No. EP/N510129/1. HO is supported by the EPSRC grant ‘‘Datisig’’ [EP/S026347/1], The Alan Turing Institute, and the Oxford-Man Institute. AA is supported in part by the HICLASS project (113213) from the ATI, BEIS, and Innovate UK.

2 Background

In this Section, we provide background and outline the basic construction.

Lattice Approximations. Throughout we consider a d -dimensional SDEs in Ito form

$$dX_t^{s,x} = \mu(X_t^{s,x}) dt + \sigma(X_t^{s,x}) dW_t, \quad X_s^{s,x} = x \in \mathbb{R}^d, \quad (2)$$

on a time interval $[0, T]$ where $\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times h}$. When there is no confusion, we denote $X^{s,x}$ as X and we use X^x if $s = 0$. We recall that if μ, σ are Lipschitz continuous and have linear growth, then there exists a pathwise unique solution to Equation (2), see for example N. Ikeda [45] and for weaker conditions see [52]. Our goal is find a sequence (X_n) of MC that converges fast in weak topology to X .

Definition 1. Let $(X^n)_{n \geq 1}$ be a sequence of Markov chains and denote $X^n = (X_i^n)_{i \geq 0}$ to emphasize the time-coordinate i . Further, let $(\gamma_n)_{n \geq 1} \subset (0, \infty)$ a sequence that converges to 0. We say that $(X^n)_{n \geq 1}$

1. is a lattice approximation to X if $X_i^n \in \gamma_n \mathbb{Z}^d$ for all $n > 0, i \geq 0$;
2. converges weakly to X with respect to a class of functions \mathbb{F} , if for every $f \in \mathbb{F}, t \in [0, T]$

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_{[nt/T]}^n)] = \mathbb{E}[f(X_t)];$$
3. converges weakly to X with rate α respect to the class of functions \mathbb{F} if for every $f \in \mathbb{F}, t \in [0, T]$

$$\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_{[nt/T]}^n)] = O(n^{-\alpha}).$$

Two important classes \mathbb{F} of test functions for our approach are the continuous differentiable functions with polynomial growth and, resp. with linear growth: $C_p^l(E, F)$ denotes the space of functions from E to F l times continuously differentiable with polynomial growth, including their derivatives; similarly $C_b^l(E, F)$ denotes the space of functions from E to F l times continuously differentiable with linear growth, with uniformly bounded derivatives.

State Space Growth. An essential requirement is that the number of states that the MC can attain grows slowly as time progresses. Example 1 shows that in naive discretization schemes the support of X_i^n grows exponentially in i , but sometimes the growth can be at most polynomial.

Definition 2. A Markov chain $X^n = (X_i^n)_{i=0,1,2,\dots}$ is sparse if $\text{card}[\text{supp}(X_i^n)] = O(i^d)$ as $i \rightarrow \infty$.

The same definition appears in Nelson and Ramaswamy [47] but there they use instead the term “recombining” for what we call a sparse MC in Definition 2. We prefer the term “sparse” for two reasons: firstly, to avoid confusion with the “recombination of measure” that is essential in our construction; secondly, we want to highlight that, unlike previous work, the emphasis of our approach is not the construction of a recombining tree – instead, we focus directly on controlling the size of the support. If a MC is sparse in the sense of Definition 2, then its transition matrix has sparse row entries (since there are d^i lattice points and only $O(i^d)$ can be reached). That the resulting MC evolves on a recombining lattice tree is a consequence of this approach, rather than its main goal (when n gets large, see Section 7 for examples).

Local Consistency. Denoting with

$$\Sigma(x) = \sigma(x)\sigma(x)^\top$$

and applying Ito’s lemma⁴ shows that there exists a $q \geq 1$ and $\alpha > 0$ and natural number N such that for all $n > N$

$$\begin{aligned} |\mathbb{E}[X_{t+n-1} - X_t | X_t = x] - \mu(x)n^{-1}| &\leq c(1 + |x|^q)n^{-1-\alpha} \\ |\mathbb{E}[(X_{t+n-1} - X_t - n^{-1}\mu(X_t))^{\otimes 2} | X_t = x] - \Sigma(x)n^{-1}| &\leq c(1 + |x|^q)n^{-1-\alpha}, \end{aligned}$$

where the constant $c = c(\mu, \sigma)$ depends only the vector fields. These two estimates characterize the SDE and it is a classic result that MCs that locally approximate these two moments converge weakly.

Definition 3. [Local consistency [30, page 328]] We say that a sequence $(X^n)_{n \geq 1}$ of MCs is locally $\delta(n)$ -consistent with X , if there exists a function $n \mapsto \delta(n)$ such that $\delta(n) \rightarrow 0$ as $n \rightarrow \infty$ and for every $i \leq n$

$$\mathbb{E} \left(\mathbb{E} \left[\frac{X_{i+1}^n - X_i^n}{n^{-1}} \middle| X_i^n \right] - \mu(X_i^n) \right)^2 = O(\delta(n)), \text{ and } \mathbb{E} \left(\mathbb{E} \left[\frac{(X_{i+1}^n - X_i^n)^{\otimes 2}}{n^{-1}} \middle| X_i^n \right] - \Sigma(X_i^n) \right)^2 = O(\delta(n)).$$

Theorem 1. [30, Theorem 9.7.4] Let (X^n) be locally $\delta(n)$ consistent with X and further assume

$$\mathbb{E}[|X_{i+1}^n - X_i^n|^6] \leq \delta(n)n^{-2} \text{ and } \mathbb{E}[\max_i |X_i^n|^{2q}] \leq c(1 + |X_0|^{2q}) \quad \text{for every } q \geq 1, \quad (3)$$

and that the vector fields coefficients μ, σ are in C_b^4 . Then⁵

$$|\mathbb{E}f(X_n^n) - \mathbb{E}f(X_T)| \leq c \cdot \left(\min \left\{ \sqrt{\delta(n)}, n^{-1/2} \right\} \right), \quad f \in C_p^3(\mathbb{R}^d, \mathbb{R}), \quad (4)$$

where the constant c depends only on x, μ, σ, f .

Clearly for SDEs the convergence criterion (1) is equivalent to the convergence criterion of Definition 1.

Remark 1. Our regularity assumptions follow closely [30] from which Theorem 1 is taken. The literature also provides similar results under similar regularity assumptions. For example Ethier and Kurtz [13], Stroock and Varadhan [52] provides general results regarding Markov processes, whereas Kushner [31], Kushner and Dupuis [32] generalises to controlled SDEs (we return to this in Section 6.2) although none of these comes with explicit rates, as in Theorem 1.

⁴ $\mu, \sigma \in C_b^4$ is sufficient; [30, 52] for weaker conditions

⁵The rate of convergence is stated in the proof of [30, Theorem 9.7.4].

Lattice-Tree models by Recombination. We construct a lattice approximation X^n to X by specifying the transition matrix $P_{x,y}^{(n)} := \mathbb{P}(X_{i+1}^n = y | X_i^n = x)$ inductively, by iterating over i . Let $X_0 = x$ a.s., and define $X_0^n := x$. Now assume that we can construct a lattice-valued random variable Y_x such that

$$\mathbb{E}[Y_x] \approx \mu(x)n^{-1}, \quad \mathbb{E}[Y_x^{\otimes 2}] \approx \mu(x)^{\otimes 2}n^{-2} + \Sigma(x)n^{-1}.$$

We then set $P_{x,y}^{(n)} := \mathbb{P}(Y_x = y - x)$ which determines the row $P_{x,\cdot}^{(n)}$. For each y such that $P_{x,y}^{(n)} \neq 0$ we then repeat the same procedure for the row $P_{y,\cdot}^{(n)}$, see Algorithm **DISCRETIZE**. If additionally each Y_x has a small support and the support is constrained on a lattice $\gamma_n Q \mathbb{Z}^d$, with Q denoting a $d \times d$ -matrix, then it follows that the transition matrix $P^{(n)}$ is sparse and that

$$\mathbb{E}[X_1^n - X_0^n | X_0^n = x] \approx \mu(x)n^{-1}, \quad \mathbb{E}[(X_1^n - X_0^n)^{\otimes 2} | X_0^n = x] \approx \mu(x)^{\otimes 2}n^{-2} + \Sigma(x)n^{-1}, \quad (5)$$

with X^n a MC that has the lattice $Q\gamma\mathbb{Z}^d$ as its state space. So far we have made the strong assumption that the lattice-valued random variable Y_x exists: this is not obvious, since formally this means to solve the non-linear and lattice-constrained system

$$\begin{cases} \sum_{i=1}^r p_i l_i \approx \mu(x)n^{-1} \\ \sum_{i=1}^r p_i l_i^{\otimes 2} \approx \mu(x)^{\otimes 2}n^{-2} + \Sigma(x)n^{-1} \end{cases} \quad (6)$$

subject to : $p_k \geq 0, \quad \sum_{k=1}^r p_k = 1, \quad l_i \in \gamma Q \mathbb{Z}^d,$

where $\sum_i p_i \delta_{l_i}$ denotes the law of the increment Y_i . The main result of the following sections is to show that – perhaps surprisingly – this is in general possible, and it can be done efficiently in terms of rate of convergence as well as size of the state space of X^n . In order to show the existence of such a lattice-valued random variable Y_x that has small support r and that approximately matches the first two moments (2), we make use of a classic theorem of Carathéodory.

Theorem 2 (Carathéodory [40]). *Let $\{x_i\}_{i=1}^N$ be a set of N points in \mathbb{R}^n and $N > n + 1$. Any point z that lies in the convex hull of these N points, z can be expressed as a convex combination of maximum $n + 1$ points, i.e. $z = \sum_{j=1}^{n+1} p_j x_j^*$, where $x_j^* \in \{x_i\}_{i=1}^N$ and $0 \leq p_i \leq 1, \sum_i p_i = 1$.*

Section 3 discusses this for $d = 1$ and Section 4 resp. Section 5 for $d = 2$ resp. $d \geq 3$.

3 Lattice-Tree Models for One-Dimensional Diffusions

We now apply the methodology outlined at the end of Section 2 to one-dimensional SDEs. The first step establishes the existence of a random variable Y that matches (or closely approximates) any given first two moments and is additionally supported on a lattice. The second step uses this random variable to define the MC.

Matching the Moments

For $\gamma > 0, a \in \mathbb{R}$ denote

$$\lfloor a \rfloor_\gamma := \max\{\gamma z, z \in \mathbb{Z} \text{ s.t. } \gamma z \leq a\}, \quad \lceil a \rceil_\gamma := \min\{\gamma z, z \in \mathbb{Z} \text{ s.t. } \gamma z > a\}.$$

If $\gamma = 1$ we suppress the subscript γ and the above reduces to the standard notation for the ceiling and floor functions.

Theorem 3. Let $a, b \in \mathbb{R}$, $b \geq 0$, and $\gamma > 0$. Then there exists a random variable Y such that

1. $\mathbb{E}[Y] = a$ and $|\mathbb{E}[Y^2] - a^2 - b^2| \leq \frac{\gamma^2}{4}$,
2. $\text{supp}(Y) \subset \{y \in \gamma\mathbb{Z} : |y| \leq \sqrt{a^2 + b^2} + \gamma\}$.

Moreover, if

$$a^2 + b^2 \geq (\lceil a \rceil_\gamma^2 - \lfloor a \rfloor_\gamma^2)(a - \lfloor a \rfloor_\gamma)\gamma^{-1} + \lfloor a \rfloor_\gamma^2,$$

then one can additionally assume that $\mathbb{E}[Y^2] = a^2 + b^2$.

Proof. Wlog suppose $a \geq 0$. For a discrete random variable Y denote $p_i := \Pr(Y = y_i)$. If

$$\mathbb{E}[(Y, Y^2)] = \sum_i p_i (y_i, y_i^2) = (a, m) \quad (7)$$

for a m such that $|m - (a^2 + b^2)| \leq \gamma^2/4$ and $y_i \in \gamma\mathbb{Z}$, then the result follows. We distinguish three cases, depending on the position of $(a, a^2 + b^2)$ in relation to the points $(\lceil a \rceil_\gamma, \lceil a \rceil_\gamma^2)$, $(\lfloor a \rfloor_\gamma, \lfloor a \rfloor_\gamma^2)$, as depicted in Figure 1. Formally,

$$\begin{cases} \text{Case 1 : } a^2 \leq a^2 + b^2 \leq \frac{\lceil a \rceil_\gamma^2 - \lfloor a \rfloor_\gamma^2}{\lceil a \rceil_\gamma - \lfloor a \rfloor_\gamma} (a - \lfloor a \rfloor_\gamma) + \lfloor a \rfloor_\gamma^2, \\ \text{Case 2 : } \frac{\lceil a \rceil_\gamma^2 - \lfloor a \rfloor_\gamma^2}{\lceil a \rceil_\gamma - \lfloor a \rfloor_\gamma} (a - \lfloor a \rfloor_\gamma) + \lfloor a \rfloor_\gamma^2 \leq a^2 + b^2 \leq \lceil a \rceil_\gamma^2, \\ \text{Case 3 : } \lceil a \rceil_\gamma^2 \leq a^2 + b^2. \end{cases} \quad (8)$$

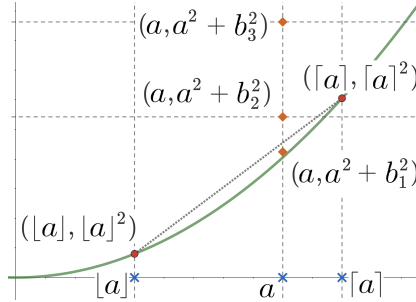


Figure 1: Proof of Theorem 3. b_i represents b under the hypothesis of Case i .

Case 1. We claim that we can choose $y_1 = \lfloor a \rfloor_\gamma$, $y_2 = \lceil a \rceil_\gamma$, $p_1 = \frac{a - \lfloor a \rfloor_\gamma}{\lceil a \rceil_\gamma - \lfloor a \rfloor_\gamma}$ and $p_2 = 1 - p_1$. This gives $\mathbb{E}[Y] = a$ and the bound on $|\mathbb{E}[Y^2] - a^2 - b^2|$ follows from a direct calculation:

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{a - \lfloor a \rfloor_\gamma}{\lceil a \rceil_\gamma - \lfloor a \rfloor_\gamma} \lceil a \rceil_\gamma^2 + \left(1 - \frac{a - \lfloor a \rfloor_\gamma}{\lceil a \rceil_\gamma - \lfloor a \rfloor_\gamma}\right) \lfloor a \rfloor_\gamma^2 = \frac{a - \lfloor a \rfloor_\gamma}{\gamma} (\lceil a \rceil_\gamma^2 - \lfloor a \rfloor_\gamma^2) + \lfloor a \rfloor_\gamma^2 \\ &= \frac{a - \lfloor a \rfloor_\gamma}{\gamma} ((\lfloor a \rfloor_\gamma + \gamma)^2 - \lfloor a \rfloor_\gamma^2) + \lfloor a \rfloor_\gamma^2 = (a - \lfloor a \rfloor_\gamma)(\gamma + 2\lfloor a \rfloor_\gamma) + \lfloor a \rfloor_\gamma^2. \end{aligned}$$

Assuming that b satisfies the inequalities (3)-Case 1, results in

$$\mathbb{E}[Y^2] - a^2 - b^2 = (a - \lfloor a \rfloor_\gamma)(\gamma + 2\lfloor a \rfloor_\gamma) + \lfloor a \rfloor_\gamma^2 - a^2 - b^2$$

$$\begin{aligned}
&\geq (a - \lfloor a \rfloor_\gamma)(\gamma + 2 \lfloor a \rfloor_\gamma) - \frac{\lceil a \rceil_\gamma^2 - \lfloor a \rfloor_\gamma^2}{\gamma}(a - \lfloor a \rfloor_\gamma) \\
&\geq (a - \lfloor a \rfloor_\gamma)(\gamma + 2 \lfloor a \rfloor_\gamma) - (\gamma + 2 \lfloor a \rfloor_\gamma)(a - \lfloor a \rfloor_\gamma) \geq 0.
\end{aligned}$$

Denoting $\theta := (a - \lfloor a \rfloor_\gamma)/\gamma \in [0, 1]$ we write

$$\begin{aligned}
\mathbb{E}[Y^2] - a^2 - b^2 &= (a - \lfloor a \rfloor_\gamma)(\gamma + 2 \lfloor a \rfloor_\gamma) + \lfloor a \rfloor_\gamma^2 - a^2 - b^2 \\
&\leq (a - \lfloor a \rfloor_\gamma)(\gamma + 2 \lfloor a \rfloor_\gamma) + \lfloor a \rfloor_\gamma^2 - a^2 \\
&\leq \theta\gamma(\gamma + 2 \lfloor a \rfloor_\gamma) + \lfloor a \rfloor_\gamma^2 - (\lfloor a \rfloor_\gamma + \theta\gamma)^2 \leq \theta\gamma^2(1 - \theta) \leq \frac{\gamma^2}{4}.
\end{aligned}$$

Case 2. The point $(a, a^2 + b^2)$ is contained in the convex hull spanned by the four points⁶

$$\{(y, y^2) : y \in \{\pm \lceil a \rceil_\gamma, \pm \lfloor a \rfloor_\gamma\}\}. \quad (9)$$

By Carathéodory's Theorem 2, $(a, a^2 + b^2)$ is a convex combination of three points of the set (3). Hence, (3) holds with $m = a^2 + b^2$.

Case 3. Since $a^2 + b^2 \leq \left\lceil \sqrt{a^2 + b^2} \right\rceil_\gamma^2$ the point $(a, a^2 + b^2)$ is contained in the convex hull spanned by the four points $\{(y, y^2) : y \in \{\pm \lceil a \rceil_\gamma, \pm \lceil \sqrt{a^2 + b^2} \rceil_\gamma\}\}$, and by Carathéodory's Theorem 2 we can conclude.

To finish the proof, we need the bounds on the support of the r.v. Y , however it is sufficient to note that this is true by construction. \square

For special cases of a, b one gets stronger results (with simpler proofs).

Corollary 1. Let $b \geq 0$, and $\gamma > 0$. Then there exists a random variable Y with

1. $\mathbb{E}[Y] = 0$ and $\mathbb{E}[Y^2] = b^2$,
2. $\text{supp}(Y) \subset \{y \in \gamma\mathbb{Z} : |y| \leq b + \gamma\}$.

Proof. Let us call $a_{++} = \min\{\gamma z > 0, z \in \mathbb{Z} \text{ s.t. } \gamma^2 z^2 > b^2\}$, then one possible solution (but not the only one) is the r.v. Y with support on $\pm a_{++}$ with probability $b^2/(2a_{++}^2)$ and 0 with probability $1 - b^2/a_{++}^2$. \square

Corollary 2. Let $a, b \in \mathbb{R}$, $b \geq 0$, and $\gamma > 0$. Then there exists a r.v. Y with

1. $\mathbb{E}[Y] = a$ and $\mathbb{E}[Y^2] = a^2 + b^2$;
2. $\text{supp}(Y) \subset \{y \in \gamma\mathbb{Z} \cup \{a\} : |y| \leq \sqrt{a^2 + b^2} + \gamma\}$.

Proof. It should be clear from Figure 1 that adding a to the lattice $\gamma\mathbb{Z}$, the convex hull of the 6 points $(\pm \lfloor a \rfloor_\gamma, \lfloor a \rfloor_\gamma^2), (\pm a, a^2), (\pm \lceil a \rceil_\gamma, \lceil a \rceil_\gamma^2), (\pm \lceil \sqrt{a^2 + b^2} \rceil_\gamma, \lceil \sqrt{a^2 + b^2} \rceil_\gamma^2)$ contains the point $(a, a^2 + b^2)$ and the Carathéodory's Theorem 2 concludes the proof. \square

Remark 2. In some applications, it can be required to constrain the support of the r.v. on specific intervals, e.g. in finance the price quantity should be non-negative. In this case, from Figure 1, it should be clear that one could proceed looking for the smallest point γz_* such that the point $(a, a^2 + b^2)$ lays in the convex hull of the points $(\gamma, \gamma^2), (\lfloor a \rfloor_\gamma, \lfloor a \rfloor_\gamma^2), (\lceil a \rceil_\gamma, \lceil a \rceil_\gamma^2)$ and $(\gamma z_*, \gamma^2 z_*^2)$.

⁶Other solutions can exist.

Constructing the Lattice-Tree Model

Now we are ready to build the Markov chain X^n which approximates X and such that the state space recombines if the SDE coefficients are bounded for the case $d = 1$. We extend to models with dimension greater than 1 in the next Sections.

Theorem 4. *Let $d = 1$ and $\mu, \sigma \in C_b^4$, $\beta > 0$ and $\sigma_{\min}^2 := \inf_x \sigma^2(x)$. Set*

$$\gamma_n := \begin{cases} 2n^{-1/2}\sigma_{\min} & , \text{ if } \sigma_{\min} > 0, \\ \sqrt{n^{-1-\beta}} & , \text{ otherwise.} \end{cases}$$

Then there exists a lattice approximation $(X^n)_{n \geq 1}$ to X on $X_0 \cup \gamma_n \mathbb{Z}$ respect to the class of functions $f \in C_p^3$ with rate $1/2$ if $\sigma_{\min} > 0$, or $\min(\beta, \frac{1}{2})$ otherwise. Moreover

$$\begin{cases} \text{card}[\text{supp}(X_i^n)] \leq \frac{icn^{-1/2}}{\gamma_n} + ic & , \text{ if } \mu, \sigma \text{ are bounded,} \\ \text{card}[\text{supp}(X_i^n)] \leq \frac{i(cn^{-1/2} + \gamma_n)\exp(cin^{-1/2})}{\gamma_n} & , \text{ if } \mu, \sigma \text{ have linear growth.} \end{cases}$$

Note that, if μ, σ are bounded and $\sigma_{\min} > 0$, then $\text{card}[\text{supp}(X_i^n)] \leq ic$, i.e. X^n recombines.

The proof below allows to state the same result but with the lattice approximation supported on the ‘‘shifted lattice’’ $X_0 + \gamma_n \mathbb{Z}$ instead of $X_0 \cup \gamma_n \mathbb{Z}$. However, an advantage of the above formulation is that several diffusions can be supported on the same lattice, namely $\gamma_n \mathbb{Z}$ and the union of their starting points, whereas shifting the lattice would require to build a complete new lattice every time a diffusion is added. We prepare the proof of Theorem 4 with the following Lemma.

Lemma 1. *If there exists $\alpha > 0$ and a $q \geq 1$ such that for $m = 1, 2$*

1. $\left| \mathbb{E}[(X_{i+1} - X_i)^{\otimes m} | X_i = X_i^n] - \mathbb{E}[(X_{i+1}^n - X_i^n)^{\otimes m} | X_i^n = X_i^n] \right| \leq c(1 + |X_i^n|^q)n^{-1-\alpha}$,
2. $\mathbb{E}[\max_i |X_i^n|^q] \leq c(1 + |X_0|^q)$,
3. $\mathbb{E}[|X_{i+1}^n - X_i^n|^6 | X_i^n] \leq c(1 + |X_i^n|^q)n^{-2-2\alpha}$,

then X^n converges weakly with rate $\min\{\alpha, \frac{1}{2}\}$ to X for $f \in C_p^3$.

Proof. The first two items imply the local consistency (1) for $\delta(n) = n^{-2\alpha}$, see Appendix, Lemma 2. Similarly, the second and third items imply the conditions required by Theorem 1, in particular the third item, thanks to item 2, implies $\mathbb{E}[|X_{i+1}^n - X_i^n|^6] \leq \delta(n)n^{-1}$ with $\delta(n) = n^{-2\alpha}$. \square

Proof of Theorem 4. We first deal with the strictly elliptic case when $\sigma_{\min}^2 := \inf_x \sigma^2(x) > 0$. Theorem 3 applied with $a = \mu(x)n^{-1}$ and $b = \sigma(x)n^{-1/2}$ guarantees the existence of a random variable Y_x such that

$$\mathbb{E}[Y_x] = \mu(x)n^{-1}, \quad \mathbb{E}[Y_x^2] = \mu^2(x)n^{-2} + \sigma^2(x)n^{-1} \quad (10)$$

and such that

$$\text{supp}(Y_x) = \left\{ z\gamma_n, z \in \mathbb{Z} \text{ s.t. } |z|\gamma_n \leq \sqrt{\mu(x)^2n^{-2} + \sigma(x)^2n^{-1}} + \gamma_n \right\}, \quad (11)$$

if we can show that

$$a^2 + b^2 \geq ([a]_{\gamma_n}^2 - [a]_{\gamma_n}^2)(a - [a]_{\gamma_n})\gamma_n^{-1} + [a]_{\gamma_n}^2. \quad (12)$$

Substituting (3) yields

$$\frac{\sigma(x)^2}{n} \geq \frac{\lceil \frac{\mu(x)}{n} \rceil_{\gamma_n}^2 - \lfloor \frac{\mu(x)}{n} \rfloor_{\gamma_n}^2}{\gamma_n} \left(\frac{\mu(x)}{n} - \lfloor \frac{\mu(x)}{n} \rfloor_{\gamma_n} \right) + \lfloor \frac{\mu(x)}{n} \rfloor_{\gamma_n}^2 - \frac{\mu^2(x)}{n^2} = \theta \gamma_n^2 (1 - \theta), \quad (13)$$

where θ in the last equality is defined as $\theta := (\frac{\mu(x)}{n} - \lfloor \frac{\mu(x)}{n} \rfloor_{\gamma_n}) \gamma_n^{-1}$. Since $\theta \in [0, 1]$, choosing $\gamma_n \leq 2n^{-1/2} \sigma_{\min}$ implies (3) because $\theta(1 - \theta) \in [0, 1/4]$ if $\theta \in [0, 1]$. Hence, the existence of the random variable Y_x matching the moments (3) and with support on the grid (3) follows. By applying this construction to different x we can define a Markov chain X^n as described in Algorithm **DISCRETIZE** by inductively filling out the transition matrix $P^{(n)}$.

If X_0^n is not in the lattice $\gamma_n \mathbb{Z}$ some attention is needed: to have X_1^n in $\gamma_n \mathbb{Z}$ it is necessary to build $Y_{X_0^n}$ on the lattice $\gamma_n \mathbb{Z} - X_0^n$, indeed $X_1^n = X_0^n + Y_{X_0^n}$ would be in $\gamma_n \mathbb{Z}$. This is also equivalent to build the r.v. $\tilde{Y}_{X_0^n}$ on $\gamma_n \mathbb{Z}$ s.t. $\mathbb{E} \tilde{Y}_{X_0^n} = \mu(X_0^n) + X_0^n$ and $\mathbb{E} \tilde{Y}_{X_0^n}^2 = \sigma(X_0^n)^2 + (\mu(X_0^n) + X_0^n)^2$ and define $Y_{X_0^n} := \tilde{Y}_{X_0^n} - X_0^n$ with support on $\gamma_n \mathbb{Z} - X_0^n$ s.t. $\mathbb{E} Y_{X_0^n} = \mu(X_0^n)$ and $\mathbb{V} Y_{X_0^n} = \sigma(X_0^n)^2$.

The resulting Markov chain X^n evolves on the state space $\gamma_n \mathbb{Z}$ and it only remains to show the claimed convergence rate and growth of the support as the time progresses. We do this by showing that the assumptions of Lemma 1 apply. Lemma 1-Item (1) holds by (3). Lemma 1-Item (2) holds in great generality for Markov chains following this construction, see Lemma 3. For the remaining Lemma 1-Item (3) note that for some $z_j \in \mathbb{Z}$

$$\mathbb{E} |Y_x|^6 = \sum_{j=1}^3 p_j |2n^{-1/2} \sigma_{\min} z_j|^6 \leq c \left| n^{-6} \mu(x)^6 + n^{-3} \sigma(x)^6 + 2^6 n^{-3} \sigma_{\min}^3 \right| \leq c(1 + |x|^6) n^{-3},$$

which we can apply to $x = X_i^n$ and shows that the assumptions of Lemma 1 are satisfied which in turn finishes the proof of the convergence rate.

When μ, σ have linear growth, in order to bound the growth of the support we use the discrete Grönwall's Inequality, cf. Lemma 4,

$$|X_i^n| \leq \sum_{j=0}^{i-1} |Y_{X_j^n}| \leq \sum_{j=0}^{i-1} n^{-1/2} c(1 + |X_j^n|) + \gamma_n \leq (cn^{-1/2} i + i\gamma_n) \exp(cin^{-1/2}).$$

Since the points $l_k \in \gamma_n \mathbb{Z}$ such that $|l_k| \leq c$ are c/γ_n we have that

$$\text{card}[\text{supp}(X_i^n)] \leq \frac{i(cn^{-1/2} + \gamma_n) \exp(cin^{-1/2})}{\gamma_n}.$$

If μ, σ are bounded, it is easy to obtain through similar reasoning that

$$|X_i^n| \leq \sum_{j=0}^{i-1} |Y_{X_j^n}| \leq icn^{-1/2} + i\gamma_n,$$

which implies that

$$\text{card}[\text{supp}(X_i^n)] \leq \frac{icn^{-1/2}}{\gamma_n} + ic.$$

This finishes the proof of the elliptic case $\sigma_{\min}^2 > 0$.

The non-elliptic case, $\inf \sigma^2(x) = 0$, follows analogously: as before we rely on Theorem 3 that ensures the existence of a random variable Y_x such that

$$\mathbb{E} [Y_x] = n^{-1} \mu(x), \quad \mathbb{E} [Y_x^2 - n^{-2} \mu^2(x) - n^{-1} \sigma^2(x)] \leq \frac{\gamma_n^2}{4}$$

and that is supported on $\gamma_n \mathbb{Z}$. Note that, unlike in the elliptic case, this time we do not match the second moment exactly, and the second moment condition shows that for any $\beta > 0$ we can take $\gamma_n \leq n^{-\frac{1-\beta}{2}}$.

For the rates of convergence it is enough to observe that $\sqrt{\delta(n)}$ of Theorem 1 is equal to β due to approximation of the second moment, see Lemma 1-Item (1). Moreover, remember that the rate of convergence in Theorem 1 is bounded by $1/2$.

Note that both the choices of γ_n satisfy $n\gamma_n^2 < c$ for some $c > 0$, as required by Lemma 3. \square

For the special case $\mu(x) = 0$ we can prove the following Corollary.

Corollary 3. *Let $d = 1$, $\mu = 0$, $\sigma \in C_b^4$ and for any $c > 0$ set $\gamma_n = cn^{-1/2}$. If $X_0^n \in \gamma_n \mathbb{Z}$, then there exists a lattice approximation $(X^n)_{n \geq 1}$ to X on $\gamma_n \mathbb{Z}$ respect to the class of functions $f \in C_p^3$ with rate $1/2$. Moreover, the same bounds on the cardinality of the support of Theorem 4 hold true.*

Proof. If $\mu(x) = 0$ we can use Corollary 1 and then follow the same reasoning of the previous proof. Since the first two moments can be approximated exactly independently of the lattice considered, we can choose $\gamma_n = cn^{-1/2}$ for any $c > 0$. The dependence of γ_n on n is chosen because required by Lemma 3: γ_n must satisfy $n\gamma_n^2 < c$ for some $c > 0$. \square

4 Lattice-Tree Models for Two-Dimensional Diffusions

We now carry out the same procedure in the two-dimensional case. The key difference to the one-dimensional case is in the first step: the existence of a random variable that matches the moments closely and is supported on a lattice is much more involved to be shown. Nevertheless, we show that by using the eigenvalues of the matrix B the system in (2) can be solved.

Matching the Moments

Throughout this Section we fix $a \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d \times d}$. We assume that B is symmetric and positive definite and denote the eigenvalues of B with $(\lambda_i)_i$ ordered from biggest to smallest, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{\min}$ with λ_{\min} denoting the smallest eigenvalue. We need the following two theorems to prove the main statements of this Section.

Theorem 5. [Hoffman and Wielandt [25]] *If A and B are normal matrices with eigenvalues $\lambda_i[A]$ and $\lambda_j[B]$, then there exists a suitable numbering of the eigenvalues s.t.*

$$\sum_i |\lambda_i[A] - \lambda_i[B]|^2 \leq |A - B|^2.$$

Theorem 6. [Gerschgorin Circle Theorem [18]] *Let $B \in \mathbb{R}^{d \times d}$ with entries b_{ij} , then every eigenvalues λ_i of B lies in at least one of the Gershgorin circles*

$$\text{Circle}(b_{ii}) := \{c \in \mathbb{C} : |c - b_{ii}| \leq \sum_{j \neq i} |b_{ij}|\}.$$

Now we are ready to prove the main statement.

Theorem 7. *Let $d = 2$, $a \in \mathbb{R}^d$, $B \in \mathbb{R}^{d \times d}$ symmetric and positive definite, $\lambda_{\min} > 0$, then for every γ such that $0 < \gamma \leq \sqrt{\lambda_{\min}/3}$ there exists a random variable Y such that*

1. $\mathbb{E}[Y] = a$ and $\mathbb{E}[Y^{\otimes 2}] = a^{\otimes 2} + B$;
2. $\text{supp}(Y) \subset \{y \in \gamma \mathbb{Z}^2 : |y|_\infty \leq |a|_\infty + \sqrt{2\lambda_1} + \sqrt{2\lambda_2} + 6\gamma\}$.

Moreover, if $a = 0$, the above applies for every $0 < \gamma \leq \sqrt{\lambda_{\min}}$.

Proof. We first consider the case when $a = 0$. In this case, note that if \bar{Y} is a discrete random variable that matches the second moment, $\mathbb{E}\bar{Y}^{\otimes 2} = B$, and Z is a Bernoulli random variable independent of \bar{Y} , such that $\mathbb{P}(Z = \pm 1) = \frac{1}{2}$, then the random variable $Y := Z \cdot \bar{Y}$ has mean $\mathbb{E}[Y] = a = 0$ and $\mathbb{E}[Y^{\otimes 2}] = B$. Hence, solving the system (2) reduces to find $l_1, \dots, l_r \in \gamma\mathbb{Z}^2$ such that

$$\sum_{i=1}^r p_i l_i l_i^T = B, \quad p_i \geq 0, \quad \sum_{i=1}^r p_i = 1, \quad (14)$$

for some $r \in \mathbb{N}_+$. Moreover, since B is symmetric, there exists $Q \in \mathbb{R}^{d \times d}$ orthonormal whose columns are the eigenvectors of B , such that $B = Q\Lambda Q^T$ and Λ is diagonal with elements the eigenvalues of B , which are strictly positive by assumption. Applying this, allows to rewrite (4) as

$$\sum_{i=1}^r p_i Q^T l_i l_i^T Q = \Lambda, \quad p_i \geq 0, \quad \sum_{i=1}^r p_i = 1. \quad (15)$$

Henceforth, we denote with $q_1 = (q_{11}, q_{12})$ and $q_2 = (q_{21}, q_{22})$ the two eigenvectors of B . It follows that if q_1 is an eigenvector for Q , then $-q_1$ is also an eigenvector for the same eigenvalue and $-q_1$ still forms an orthonormal basis with $\pm q_2$. We use these properties of eigenvalues in the following without loss of generality.

Case $q_{12} = 0$ or $q_{21} = 0$. Assume $q_{12} = 0$ (the case $q_{21} = 0$ follows analogously), then $q_1 = e_1$ and $q_2 = e_2$. Hence, $Q = I$ and we can refer to Theorem 9, which applies to d greater than 2 and B semi-positive definite.

Case $q_{11}, q_{21} > 0$. We first note that

$$\begin{cases} \langle q_1, q_1 \rangle^2 = q_{11}^2 + q_{12}^2 = 1 \\ \langle q_2, q_2 \rangle^2 = q_{21}^2 + q_{22}^2 = 1 \\ \langle q_1, q_2 \rangle = q_{11}q_{21} + q_{12}q_{22} = 0, \end{cases} \quad \text{if and only if} \quad \begin{cases} q_{11} = \text{sign}(q_{11})\sqrt{1 - q_{12}^2} \\ q_{21} = \text{sign}(q_{21})\sqrt{1 - q_{22}^2} \\ \frac{q_{11}}{q_{12}} = -\frac{q_{22}}{q_{21}}. \end{cases}$$

Using the last equation on the system on the right hand side, we obtain

$$\text{sign}(q_{11})\text{sign}(q_{12})\frac{\sqrt{1 - q_{12}^2}}{|q_{12}|} = -\text{sign}(q_{21})\text{sign}(q_{22})\frac{|q_{22}|}{\sqrt{1 - q_{22}^2}}.$$

This in turn implies that

$$\text{sign}(q_{11})\text{sign}(q_{12}) = -\text{sign}(q_{21})\text{sign}(q_{22}).$$

We now claim that without loss of generality, we can assume that $q_{11} > 0$, $q_{12} > 0$ and $q_{21} > 0, q_{22} < 0$: indeed if we flip the sign of the eigenvectors they are still eigenvectors and we can re-label q_1, q_2 . Moreover, note that

$$\begin{aligned} \frac{q_{11}}{q_{12}} = -\frac{q_{22}}{q_{21}} &\iff q_{22}^2 = \left(\frac{q_{11}}{q_{12}}\right)^2 (1 - q_{22}^2) \iff q_{22}^2 \left(1 + \left(\frac{q_{11}}{q_{12}}\right)^2\right) = \left(\frac{q_{11}}{q_{12}}\right)^2 \\ &\iff q_{22} = -\frac{\frac{q_{11}}{\sqrt{1 - q_{11}^2}}}{\sqrt{1 + \frac{q_{11}^2}{1 - q_{11}^2}}} = -q_{11}, \end{aligned}$$

which implies that

$$q_1 = \left(q_{11}, \sqrt{1 - q_{11}^2} \right) \text{ and } q_2 = \left(\sqrt{1 - q_{11}^2}, -q_{11} \right). \quad (16)$$

In the following, we assume that $1 > q_{11} \geq 1/\sqrt{2}$ - otherwise we can state q_1, q_2 as functions of q_{12} and proceed similarly.⁷ Now note that if there exist r points $l_1, \dots, l_r \in \gamma\mathbb{Z}^2$ s.t.

$$\Lambda \in \text{ConvexHull}\{(Q^T l_1)^{\otimes 2}, \dots, (Q^T l_r)^{\otimes 2}\}, \quad (17)$$

then Equation (4) hold true for some p_i to be computed. It only remains to show that (4) holds. This is a hard problem, but in our case a simple argument works. We make use of the simple observation that if $A = \{a_1, \dots, a_r\} \subset \mathbb{R}^h$ and $b \in \mathbb{R}^h$ s.t.

$$\forall \sim \in \{\leq, \geq\}^h \quad \exists a \in A \text{ such that } a \sim b \text{ then } b \in \text{ConvexHull}(A). \quad (18)$$

We now apply (4) with $b = \Lambda$ and $A = \{(Q^T l_1)^{\otimes 2}, \dots, (Q^T l_r)^{\otimes 2}\}$. Since Λ and $(Q^T l_i)^{\otimes 2}$ are symmetric we work in dimension $h = 3$ instead of $h = 4$. The premise in Equation (4) reduces to show that for each of the eight possible choices of $\sim = (\sim_2, \sim_1, \sim_1, \sim_3) \in \{\geq, \leq\}^3$, there exists a point $l \in \gamma\mathbb{Z}^2$ such that

$$\begin{pmatrix} \langle q_1, l \rangle^2 & \langle q_1, l \rangle \langle q_2, l \rangle \\ \langle q_1, l \rangle \langle q_2, l \rangle & \langle q_2, l \rangle^2 \end{pmatrix} \sim \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Spelled out in coordinates the above reads as

$$\begin{cases} \langle q_1, l \rangle \langle q_2, l \rangle \sim_1 0 \\ \langle q_1, l \rangle^2 \sim_2 \lambda_1 \\ \langle q_2, l \rangle^2 \sim_3 \lambda_2 \end{cases} \xleftrightarrow{\text{Using Equation (4)}} \begin{cases} \left(q_{11} l_{i1} + \sqrt{1 - q_{11}^2} l_{i2} \right) \left(\sqrt{1 - q_{11}^2} l_{i1} - q_{11} l_{i2} \right) \sim_1 0 \\ \left(q_{11} l_{i1} + \sqrt{1 - q_{11}^2} l_{i2} \right)^2 \sim_2 \lambda_1 \\ \left(\sqrt{1 - q_{11}^2} l_{i1} - q_{11} l_{i2} \right)^2 \sim_3 \lambda_2. \end{cases}$$

We denote $\varphi := \frac{\sqrt{1 - q_{11}^2}}{q_{11}}$, therefore the system we study is

$$\begin{cases} (l_{i1} + \varphi l_{i2}) (\varphi l_{i1} - l_{i2}) \sim_1 0 \\ (l_{i1} + \varphi l_{i2})^2 \sim_2 \frac{\lambda_1}{q_{11}^2} \\ (\varphi l_{i1} - l_{i2})^2 \sim_3 \frac{\lambda_2}{q_{11}^2}, \end{cases} \quad (19)$$

where $l_i = (l_{i1}, l_{i2})$, $l_{i1} \in \gamma\mathbb{Z}$ and $l_{i2} \in \gamma\mathbb{Z}$. We now claim that for every of the $r = 8$ possible choices of $\sim_1, \sim_2, \sim_3 \in \{\leq, \geq\}^3$ one can find a $l_i \in \gamma\mathbb{Z}^2$ such that Equation (4) holds. All the possible combinations of \sim_1, \sim_2, \sim_3 are $\{\geq\geq\geq, \geq\geq\leq, \geq\leq\geq, \geq\leq\leq, \leq\geq\geq, \leq\geq\leq, \leq\leq\geq, \leq\leq\leq\}$ and we study them in order. We suppress the subscript i for $l_i = (l_{i1}, l_{i2})$. Let us also resume some bounds

$$\frac{1}{\sqrt{2}} \leq q_{11} < 1, \quad 1 \leq \frac{1}{q_{11}} < \sqrt{2}, \quad 0 \leq \varphi \leq 1, \quad 1 \leq \frac{1}{\varphi} < \infty. \quad (20)$$

⁷The implications regarding q_1, q_2 at first glance can seem hostile, however geometrically they are more intuitive. They depend on the fact that q_1, q_2 form an orthonormal basis and changing the signs they are still eigenvectors and they still form an orthonormal basis.

- ($\geq \geq$) We have the following system, whose solutions are on the right

$$\begin{cases} (l_1 + \varphi l_2)(\varphi l_1 - l_2) \geq 0 \\ (l_1 + \varphi l_2)^2 \geq \frac{\lambda_1}{q_{11}^2} \\ (\varphi l_1 - l_2)^2 \geq \frac{\lambda_2}{q_{11}^2} \end{cases} \iff \begin{cases} \text{Solution 1 } \geq \geq & \text{Solution 2 } \leq \leq \\ l_1 \geq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 & \text{or } l_1 \leq -\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \\ l_1 \geq \frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 & l_2 \leq -\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2. \end{cases}$$

“Solution 1 $\geq \geq$ ” means that we consider the case where $l_1 + \varphi l_2 \geq 0$, $\varphi l_1 - l_2 \geq 0$ and similarly for the other case “Solution 2 $\leq \leq$ ”. In this and the successive cases we focus on “Solution 1 $\geq \geq$ ”. In this case the solution can be obtained choosing e.g. $l_2 = 0$ and l_1 big enough, which however would result in an unbounded solution given that $1/\varphi \geq 1$ is not bounded from above. A bounded solution is⁸

$$l_2 = \left[-\frac{\sqrt{\lambda_2}}{q_{11}} \right]_{\gamma} \in [-\sqrt{2\lambda_2} - \gamma, -\sqrt{\lambda_2}]$$

$$l_1 = \left[\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi \left[-\frac{\sqrt{\lambda_2}}{q_{11}} \right]_{\gamma} \right]_{\gamma} \in [\sqrt{\lambda_1}, \sqrt{2\lambda_1} + \sqrt{2\lambda_2} + 2\gamma],$$

which is bounded. This is obtained choosing l_2 so that $\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \approx 0$ and then using it in the first equation of “Solution 1”.

- ($\geq \leq$) We have the following system whose, solutions are on the right

$$\begin{cases} (l_1 + \varphi l_2)(\varphi l_1 - l_2) \geq 0 \\ (l_1 + \varphi l_2)^2 \geq \frac{\lambda_1}{q_{11}^2} \\ (\varphi l_1 - l_2)^2 \leq \frac{\lambda_2}{q_{11}^2} \end{cases} \iff \begin{cases} \text{Solution 1 } \geq \geq & \text{Solution 2 } \leq \leq \\ l_1 \geq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 & \text{or } l_1 \leq -\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \\ \frac{1}{\varphi} l_2 \leq l_1 \leq \frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 & -\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \leq l_1 \leq \frac{1}{\varphi} l_2. \end{cases}$$

This case is more tedious. We choose $l_1 = \gamma n$, $l_2 = \gamma \lfloor n\varphi \rfloor$, where n is the smallest integer satisfying $\gamma n \geq \sqrt{2\lambda_1}$. This implies that the first equation of “Solution 1” is satisfied: using Equation (4)

$$\gamma n \geq \sqrt{2\lambda_1} \geq \frac{\sqrt{\lambda_1}}{q_{11}} \geq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi \gamma \lfloor n\varphi \rfloor.$$

We choose therefore $l_1 = \gamma n = \lceil \sqrt{2\lambda_1} \rceil_{\gamma} \in [0, \sqrt{2\lambda_1} + \gamma]$. The same bound for l_2 holds true since $l_2 = \gamma \lfloor n\varphi \rfloor \leq l_1$, again using Equation (4).

It remains to prove that the second equation of “Solution 1” is satisfied. Dividing the second equation of “Solution 1” by l_2 we prove the left inequality

$$\frac{l_1}{l_2} = \frac{\gamma n}{\gamma \lfloor n\varphi \rfloor} \geq \frac{n}{n\varphi} = \frac{1}{\varphi},$$

and, since by assumption $\gamma \leq \sqrt{\lambda_2}$ we conclude that

$$\frac{l_1}{l_2} - \frac{1}{\varphi} = \frac{\gamma n\varphi - \gamma \lfloor n\varphi \rfloor}{\gamma \lfloor n\varphi \rfloor \varphi} \leq \frac{\gamma}{\gamma \lfloor n\varphi \rfloor \varphi} = \frac{\gamma}{l_2 \varphi} \implies \frac{l_1}{l_2} \leq \frac{1}{\varphi} \left(1 + \frac{\gamma}{l_2} \right) \leq \frac{1}{\varphi} \left(1 + \frac{1}{l_2} \frac{\sqrt{\lambda_2}}{q_{11}} \right)$$

because $1 \leq \frac{1}{q_{11}} \leq \sqrt{2}$, which is equivalent to the second equation of “Solution 1”.

⁸In this and the successive cases, we have found really useful plotting the lines (l_1 as function of l_2) represented by the two equations in “Solution 1”.

- ($\geq \leq \geq$) We have the following system, whose solutions are on the right

$$\begin{cases} (l_1 + \varphi l_2)(\varphi l_1 - l_2) \geq 0 \\ (l_1 + \varphi l_2)^2 \leq \frac{\lambda_1}{q_{11}^2} \\ (\varphi l_1 - l_2)^2 \geq \frac{\lambda_2}{q_{11}^2} \end{cases} \iff \begin{cases} \text{Solution 1 } \geq \geq \\ -\varphi l_2 \leq l_1 \leq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \\ l_1 \geq \frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \end{cases} \quad \text{or} \quad \begin{cases} \text{Solution 2 } \leq \leq \\ -\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \leq l_1 \leq -\varphi l_2 \\ l_1 \leq -\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2. \end{cases}$$

In this case we can take

$$l_2 = \left[-\frac{\sqrt{\lambda_2}}{q_{11}} \right]_{\gamma} \in [-\sqrt{2\lambda_2} - \gamma, -\sqrt{\lambda_2}], \quad l_1 = \left[-\varphi \left[-\frac{\sqrt{\lambda_2}}{q_{11}} \right]_{\gamma} \right]_{\gamma} \in [0, \sqrt{2\lambda_2} + 2\gamma].$$

This solution has been found choosing l_2 so that $\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \approx 0$ and then using it in the first equation of ‘‘Solution 1’’. The second equation of ‘‘Solution 1’’ is satisfied since l_1 is positive. For the first one, the left inequality is satisfied by definition of l_1 , whilst the right inequality because $\gamma \leq \sqrt{\lambda_1} \leq \sqrt{\lambda_1}/q_{11}$.

- ($\geq \leq \leq$) We have the following system, whose solutions are on the right

$$\begin{cases} (l_1 + \varphi l_2)(\varphi l_1 - l_2) \geq 0 \\ (l_1 + \varphi l_2)^2 \leq \frac{\lambda_1}{q_{11}^2} \\ (\varphi l_1 - l_2)^2 \leq \frac{\lambda_2}{q_{11}^2} \end{cases} \iff \begin{cases} \text{Solution 1 } \geq \geq \\ -\varphi l_2 \leq l_1 \leq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \\ \frac{1}{\varphi} l_2 \leq l_1 \leq \frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \end{cases} \quad \text{or} \quad \begin{cases} \text{Solution 2 } \leq \leq \\ -\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \leq l_1 \leq -\varphi l_2 \\ -\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \leq l_1 \leq \frac{1}{\varphi} l_2. \end{cases}$$

In this case two possible solutions for the case ‘‘Solution 1’’ are $(l_1, l_2) = (0, 0)$ and $(l_1, l_2) = (\gamma, 0)$, since by assumption $\gamma \leq \min\{\sqrt{\lambda_1}, \sqrt{\lambda_2}\}$.

- ($\leq \geq \geq$) We have the following system, whose solutions are on the right

$$\begin{cases} (l_1 + \varphi l_2)(\varphi l_1 - l_2) \leq 0 \\ (l_1 + \varphi l_2)^2 \geq \frac{\lambda_1}{q_{11}^2} \\ (\varphi l_1 - l_2)^2 \geq \frac{\lambda_2}{q_{11}^2} \end{cases} \iff \begin{cases} \text{Solution 1 } \geq \leq \\ l_1 \geq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \\ l_1 \leq -\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \end{cases} \quad \text{or} \quad \begin{cases} \text{Solution 2 } \leq \geq \\ l_1 \leq -\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \\ l_1 \geq \frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2. \end{cases}$$

We could take simply $l_1 = 0$ and l_2 big enough, but this solution would be unbounded (depending on $1/\varphi$). Therefore, we reason similarly to the case ($\geq \geq \geq$): let us rewrite the equations for ‘‘Solution 1’’ respect to l_2

$$l_2 \geq \frac{1}{\varphi} \frac{\sqrt{\lambda_1}}{q_{11}} - \frac{1}{\varphi} l_1, \quad l_2 \geq \frac{\sqrt{\lambda_2}}{q_{11}} + \varphi l_1,$$

and we can take as a solution

$$l_1 = \left[\frac{\sqrt{\lambda_1}}{q_{11}} \right]_{\gamma} \in [\sqrt{\lambda_1} - \gamma, \sqrt{2\lambda_1}]$$

$$l_2 = \left[\frac{\sqrt{\lambda_2}}{q_{11}} + \varphi \left[\frac{\sqrt{\lambda_1}}{q_{11}} \right]_{\gamma} \right]_{\gamma} \in [\sqrt{\lambda_2}, \sqrt{2\lambda_2} + \sqrt{2\lambda_1} + 2\gamma].$$

This solution has been found choosing l_1 so that $\frac{1}{\varphi} \frac{\sqrt{\lambda_1}}{q_{11}} - \frac{1}{\varphi} l_1 \approx 0$ and then using it in the other equation.

- ($\leq \geq \leq$) We have the following system, whose solutions are on the right

$$\begin{cases} (l_1 + \varphi l_2)(\varphi l_1 - l_2) \leq 0 \\ (l_1 + \varphi l_2)^2 \geq \frac{\lambda_1}{q_{11}} \\ (\varphi l_1 - l_2)^2 \leq \frac{\lambda_2}{q_{11}} \end{cases} \iff \begin{cases} \text{Solution 1 } \geq \leq & \text{Solution 2 } \leq \geq \\ l_1 \geq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 & \text{or } l_1 \leq -\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \\ -\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \leq l_1 \leq \frac{1}{\varphi} l_2 & \frac{1}{\varphi} l_2 \leq l_1 \leq \frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2. \end{cases}$$

We choose $l_1 = \gamma n$, $l_2 = \gamma \lceil n\varphi \rceil$, where n is the smallest integer satisfying $\gamma n \geq \sqrt{2\lambda_1}$. This implies that the first equation of “Solution 1” is satisfied: using Equation (4)

$$\gamma n \geq \sqrt{2\lambda_1} \geq \frac{\sqrt{\lambda_1}}{q_{11}} \geq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi \gamma \lceil n\varphi \rceil.$$

We choose therefore $l_1 = \gamma n = \lceil \sqrt{2\lambda_1} \rceil \gamma \in [0, \sqrt{2\lambda_1} + \gamma]$. The same bound for l_2 holds true since $l_2 = \gamma \lceil n\varphi \rceil \leq l_1$, again using Equation (4).

It remains to prove that the second equation of “Solution 1” is satisfied. Dividing the second equation of “Solution 1” by l_2 we prove the left inequality

$$\frac{l_1}{l_2} = \frac{\gamma n}{\gamma \lceil n\varphi \rceil} \leq \frac{n}{n\varphi} = \frac{1}{\varphi},$$

and, since by assumption $\gamma \leq \sqrt{\lambda_2}$ we conclude that

$$\frac{l_1}{l_2} - \frac{1}{\varphi} = \frac{\gamma n\varphi - \gamma \lceil n\varphi \rceil}{\gamma \lceil n\varphi \rceil \varphi} \geq \frac{-\gamma}{\gamma \lceil n\varphi \rceil \varphi} = \frac{-\gamma}{l_2 \varphi} \implies \frac{l_1}{l_2} \geq \frac{1}{\varphi} \left(1 - \frac{\gamma}{l_2}\right) \geq \frac{1}{\varphi} \left(1 - \frac{1}{l_2} \frac{\sqrt{\lambda_2}}{q_{11}}\right)$$

because $-1 \geq -\frac{1}{q_{11}} \geq -\sqrt{2}$, which is equivalent to the second equation of Solution 1.

- ($\leq \leq \geq$) We have the following system, whose solutions are on the right

$$\begin{cases} (l_1 + \varphi l_2)(\varphi l_1 - l_2) \leq 0 \\ (l_1 + \varphi l_2)^2 \leq \frac{\lambda_1}{q_{11}} \\ (\varphi l_1 - l_2)^2 \geq \frac{\lambda_2}{q_{11}} \end{cases} \iff \begin{cases} \text{Solution 1 } \geq \leq & \text{Solution 2 } \leq \geq \\ -\varphi l_2 \leq l_1 \leq \frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 & \text{or } -\frac{\sqrt{\lambda_1}}{q_{11}} - \varphi l_2 \leq l_1 \leq -\varphi l_2 \\ l_1 \leq -\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 & l_1 \geq \frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2. \end{cases}$$

We can proceed as in the case $\geq \leq \geq$: we can take

$$l_2 = \left\lceil \frac{\sqrt{\lambda_2}}{q_{11}} \right\rceil_{\gamma} \in \left[\sqrt{\lambda_2}, \sqrt{2\lambda_2} + \gamma \right], \quad l_1 = \left\lceil -\varphi \left\lceil \frac{\sqrt{\lambda_2}}{q_{11}} \right\rceil_{\gamma} \right\rceil_{\gamma} \in \left[-\sqrt{2\lambda_2} - \gamma, 0 \right].$$

This solution has been found choosing l_2 so that $\frac{1}{\varphi} \frac{\sqrt{\lambda_2}}{q_{11}} + \frac{1}{\varphi} l_2 \approx 0$ and then using it in the first equation of “Solution 1”. The second equation of “Solution 1” is satisfied since l_1 is negative. For the first one, the left inequality is satisfied by definition of l_1 , whilst the right inequality because $\gamma \leq \sqrt{\lambda_1} \leq \sqrt{\lambda_1}/q_{11}$.

- ($\leq \leq \leq$) In this case it is easy to see that we can take the origin.

In all the cases we have studied the “Solution 1”, however, for the “Solution 2” we can take $-(l_1, l_2)$. This concludes the proof for the case $a = 0$.

For a general $a \in \mathbb{R}^d$ we can build for any $\gamma > 0$ a random variable \tilde{Y} with support on $\gamma \mathbb{Z}^d$ such that $\mathbb{E}\tilde{Y} = a$. This follows from Carathéodory’s Theorem 2, considering the smallest hypercube with vertices

in $\gamma\mathbb{Z}^d$ that contains a , that is of the 2^d possible choices of $\sim \in \{\leq, \geq\}^d$ determines a point y_\sim on the hypercube around a s.t.

$$y_\sim \in \gamma\mathbb{Z}^d \text{ and } y_\sim \sim a.$$

We denote the resulting 2^d points as y_1, \dots, y_{2^d} and now estimate the variance $\mathbb{V}[\tilde{Y}]$,

$$|\mathbb{V}[\tilde{Y}]| = \left| \mathbb{E} \left[(\tilde{Y} - a)^{\otimes 2} \right] \right| = \left| \sum_i \tilde{p}_i (y_i - a)^{\otimes 2} \right| \leq \sum_i \tilde{p}_i |y_i - a|^2 \leq d\gamma^2.$$

We now build a random variable Y_\star supported on $\gamma\mathbb{Z}^2$ ($d = 2$) that is independent of Y and has mean and variance $\mathbb{E}Y_\star = 0$ and $\mathbb{V}Y_\star = \mathbb{E}Y_\star^{\otimes 2} = B - \mathbb{V}\tilde{Y}$. It follows that $Y := \tilde{Y} + Y_\star$ has mean and variance $\mathbb{E}Y = a$ and $\mathbb{V}Y = B$. To build Y_\star we use the first part of this proof. Therefore we need to verify that the eigenvalues of $B - \mathbb{V}\tilde{Y}$ are strictly positive. Since $\mathbb{V}\tilde{Y}$ and B are symmetric it follows from the Hoffman–Wielandt Theorem, Theorem 5, that there exists a permutation π such that for any k

$$-\mathbb{V}\tilde{Y} \leq \lambda_k [B - \mathbb{V}\tilde{Y}] - \lambda_{\pi(k)} [B] \leq |\mathbb{V}\tilde{Y}|,$$

where $\lambda_k [B - \mathbb{V}\tilde{Y}]$ indicates the k -th eigenvalue of $B - \mathbb{V}\tilde{Y}$. Therefore,

$$\lambda_{\min} [B - \mathbb{V}\tilde{Y}] \geq \lambda_{\pi(\min)} [B] - |\mathbb{V}\tilde{Y}| \geq \lambda_{\min} - 2\gamma^2.$$

Hence, choosing $\gamma \leq \sqrt{\lambda_{\min}/3}$ guarantees $\gamma \leq \sqrt{\lambda_{\min} [B - \mathbb{V}\tilde{Y}]}$ since

$$\lambda_{\min} - 2\gamma^2 \geq \gamma^2 \iff \gamma \leq \sqrt{\lambda_{\min}/3}.$$

Note that the cardinality of the support of $Y = \tilde{Y} + Y_\star$, eventually, can be reduced using the Caratheodory's Theorem 2. We have proved that, by construction, the support of the r.v. Y_\star is included in

$$\text{supp}(Y_\star) \subset \{y \in \gamma\mathbb{Z}^2 : |y|_\infty \leq \sqrt{2\lambda_1 [B - \mathbb{V}\tilde{Y}]} + \sqrt{2\lambda_2 [B - \mathbb{V}\tilde{Y}]} + 2\gamma\}.$$

Thanks again to the Hoffman–Wielandt Theorem, Theorem 5 we have that

$$\begin{aligned} \sqrt{2\lambda_1 [B - \mathbb{V}\tilde{Y}]} + \sqrt{2\lambda_2 [B - \mathbb{V}\tilde{Y}]} &\leq \sqrt{2 \left(|\mathbb{V}\tilde{Y}| + \lambda_{\pi(1)} [B] \right)} + \sqrt{2 \left(|\mathbb{V}\tilde{Y}| + \lambda_{\pi(2)} [B] \right)} \\ &\leq \sqrt{2\lambda_{\pi(1)} [B]} + \sqrt{2\lambda_{\pi(2)} [B]} + 2\sqrt{2|\mathbb{V}\tilde{Y}|} \\ &\leq \sqrt{2\lambda_1 [B]} + \sqrt{2\lambda_2 [B]} + 2\gamma\sqrt{2d}. \end{aligned}$$

This implies that the support of $Y = \tilde{Y} + Y_\star$ is included in

$$\text{supp}(Y) \subset \{y \in \gamma\mathbb{Z}^2 : |y|_\infty \leq |a|_\infty + \sqrt{2\lambda_1 [B]} + \sqrt{2\lambda_2 [B]} + 6\gamma\}.$$

□

Remark 3. We conjecture that Theorem 7 generalizes to dimensions higher than $d = 2$, however the above brute-force proof strategy is not helpful, since it is infeasible to solve all the inequalities one by one for the equivalent of System (4) in higher dimensions. Concretely, we believe that for $a \in \mathbb{R}^d$, $B \in \mathbb{R}^{d \times d}$ symmetric and positive definite, $\lambda_{\min} > 0$, then for every γ such that $0 < \gamma \leq \sqrt{\lambda_{\min}/(d+1)}$ there exists a random variable Y such that

1. $\mathbb{E}[Y] = a$ and $\mathbb{E}[Y^{\otimes 2}] = a^{\otimes 2} + B$;
2. $\text{supp}(Y) \subset \{y \in \gamma\mathbb{Z}^2 : |y|_\infty \leq |a|_\infty + \sum_{i=1}^d \sqrt{2\lambda_i} + c_d\gamma\}$.

Note that it only remains to show the existence of a random variable Y_\star , such that $\mathbb{E}[Y_\star] = 0$ and $\mathbb{E}[Y_\star^{\otimes 2}] = B - \mathbb{V}\tilde{Y}$ for any positive definite matrix B , since this allows to simply proceed as in the second part of the proof of Theorem 7.

Constructing the Lattice-Tree Model

Now we are ready to build the Markov chain X^n , which approximates X for the 2-dimensional case and such that the state space recombines if the SDE coefficients are bounded. We denote $\Sigma = \sigma(x)\sigma(x)^T$ and with $\lambda_i(x)$ its eigenvalues.

Theorem 8. *Let $\mu, \sigma \in C_b^4$, $d = 2$, $\varepsilon := \inf_x \lambda_{\min}(x) > 0$ and $\gamma_n = n^{-1/2} \sqrt{\varepsilon/3}$. There exists a lattice approximation $(X^n)_{n \geq 1}$ to X on $X_0^n \cup \gamma_n \mathbb{Z}^2$ with respect to the class of functions $f \in C_p^3$ with rate $1/2$. Moreover,*

$$\begin{cases} \text{card}[\text{supp}(X_i^n)] \leq ci^2 & , \text{ if } \mu, \sigma \text{ are bounded,} \\ \text{card}[\text{supp}(X_i^n)] \leq ci^2 \exp(2cin^{-1/2}) & , \text{ if } \mu, \sigma \text{ have linear growth.} \end{cases}$$

Proof. We argue as in Theorem 4, but using Theorem 7 above. In contrast to the one-dimensional case, the support of the random variable Y_x that gives the increment $X_{i+1}^n - X_i^n$ when $X_i^n = x$ now depends also on the eigenvalues of Σ . However, the Gerschgorin's Circle Theorem [18], Theorem 6, bounds the eigenvalues of Σ . We conclude by noting that the number of points of the lattice $\gamma_n \mathbb{Z}^2$ in a square of side c are $(c/\gamma_n)^2$, which yields the square factor in the cardinality of the support.

If X_0^n is not in the lattice $\gamma_n \mathbb{Z}^2$ more care is needed: for X_1^n to be supported on $\gamma_n \mathbb{Z}^2$ it is necessary to build the random variable $Y_{X_0^n}$ with support on the lattice $\gamma_n \mathbb{Z}^2 - X_0^n$, indeed $X_1^n = X_0^n + Y_{X_0^n}$ would be in $\gamma_n \mathbb{Z}^2$. This is equivalent to build the random variable $\tilde{Y}_{X_0^n}$ on $\gamma_n \mathbb{Z}^2$ such that $\mathbb{E}\tilde{Y}_{X_0^n} = \mu(X_0^n) + X_0^n$ and $\mathbb{E}\tilde{Y}_{X_0^n}^2 = \Sigma(X_0^n) + (\mu(X_0^n) + X_0^n)^{\otimes 2}$. We define $Y_{X_0^n} := \tilde{Y}_{X_0^n} - X_0^n$ with support on $\gamma_n \mathbb{Z}^2 - X_0^n$ s.t. $\mathbb{E}Y_{X_0^n} = \mu(X_0^n)$ and $\mathbb{V}Y_{X_0^n} = \Sigma(X_0^n)$. To build $\tilde{Y}_{X_0^n}$ using Theorem 7, we need to bound the eigenvalues of $\Sigma(X_0^n) + (\mu(X_0^n) + X_0^n)^{\otimes 2}$ using that

$$\begin{aligned} \lambda_{\min}[\Sigma(X_0^n) + (\mu(X_0^n) + X_0^n)^{\otimes 2}] &= \inf_{d \neq 0} \frac{d^T [\Sigma(X_0^n) + (\mu(X_0^n) + X_0^n)^{\otimes 2}] d}{d^T d} \\ &\geq \inf_{d \neq 0} \frac{d^T \Sigma(X_0^n) d}{d^T d} + \inf_{e \neq 0} \frac{e^T (\mu(X_0^n) + X_0^n)^{\otimes 2} e}{e^T e} \\ &\geq \lambda_{\min}[\Sigma(X_0^n)] + \lambda_{\min}[(\mu(X_0^n) + X_0^n)^{\otimes 2}] \\ &\geq \lambda_{\min}[\Sigma(X_0^n)], \end{aligned}$$

noting that $\lambda_{\min}[(\mu(X_0^n) + X_0^n)^{\otimes 2}] \geq 0$, which concludes the proof. \square

5 Lattice-Tree Models for Multi-Dimensional Diffusions

We now carry out the same procedure in the multi-dimensional case. In this case, our assumptions will be more stringent and we will consider the lattice to depend on the structure of the matrix B .

Matching the Moments

In this Subsection we solve System (2) for a general dimension d , but on pre-specified lattice. We call $Q \in O(d)$ the orthonormal matrix such that $B = Q\Lambda Q^T$, with Λ the diagonal matrix consisting of the eigenvalues of B .

Theorem 9. *Let $B \in \mathbb{R}^{d \times d}$ symmetric and positive semi-definite, $\gamma > 0$, then there exists a random variable Y such that*

1. $\mathbb{E}[Y] = 0$ and $\mathbb{E}[Y^{\otimes 2}] = B$;

$$2. \text{supp}(Y) \subset \{y \in \gamma Q \mathbb{Z}^d : |y|_\infty \leq \sqrt{\sum_i \lambda_i} \cdot \max_i |Qe_i| + \gamma\}.$$

Proof. If all the eigenvalues λ_i of B are 0, then $Y = 0$ a.s. is a solution, so let us suppose that at least one eigenvalue is strictly positive.

Recall that, from the proof of Theorem 7, if \bar{Y} is a discrete random variable that matches the second moment, $\mathbb{E}\bar{Y}^{\otimes 2} = B$ and Z is a Bernoulli random variable independent of \bar{Y} , $\mathbb{P}(Z = \pm 1) = \frac{1}{2}$, then the random variable $Y := Z \cdot \bar{Y}$ has mean $\mathbb{E}[Y] = 0$ and $\mathbb{E}[Y^{\otimes 2}] = B$. Hence, solving the system (2) reduces to find $\bar{l}_i \in \gamma Q \mathbb{Z}^d$ s.t.

$$\sum_i \bar{p}_i \bar{l}_i \bar{l}_i^T = B, \quad \bar{p}_i > 0, \quad \sum_i \bar{p}_i = 1,$$

then the measure on the points $\bar{l}_i, -\bar{l}_i$ with weights respectively $p_i = \bar{p}_i/2$ represents the r.v. Y we are looking for. Since we suppose B to be positive semi-definite and symmetric, there exists an orthonormal matrix Q such that $B = Q\Lambda Q^T$, whose columns are the eigenvectors of B , and Λ is a diagonal matrix whose values are the non-negative eigenvalues. If we indicate $\bar{l}_i = \gamma Q \bar{z}_i$, $\bar{z}_i \in \mathbb{Z}^d$, we have that

$$\sum_i^r \bar{p}_i \bar{l}_i \bar{l}_i^T = B = Q\Lambda Q^T \Leftrightarrow \sum_i^r \bar{p}_i \gamma^2 Q \bar{z}_i \bar{z}_i^T Q^T = Q\Lambda Q^T \Leftrightarrow \sum_i^r \bar{p}_i \gamma^2 \bar{z}_i \bar{z}_i^T = \Lambda.$$

Therefore, rearranging the last equation we obtain

$$\sum_i^r p_i \gamma^2 \bar{z}_i \bar{z}_i^T = \sum_{j=1}^d \frac{\lambda_j}{\sum_{k=1}^d \lambda_k} \left(\sqrt{\sum_{k=1}^d \lambda_k} \cdot e_j \right)^{\otimes 2}.$$

Since r is a free parameter for us, we can take $r = d^* + 1$, where d^* is the number of different eigenvalues not equal to 0, and we can solve the problem independently for any eigenvalue, then we merge together all the solutions. Thus, we obtain

$$\begin{aligned} \bar{l}_{d^*+1} &= 0, & \bar{p}_{d^*+1} &= 1 - \left\{ \frac{\sum_{j=1}^d \lambda_j}{\left[\sqrt{\sum_{j=1}^d \lambda_j} \right]_\gamma^2} \right\}, \\ \bar{l}_i &= Q \left[\sqrt{\sum_{j=1}^d \lambda_j} \right]_\gamma e_i, & \bar{p}_i &= \frac{m(\lambda_i) \lambda_i}{\sum_{j=1}^d \lambda_j} \left\{ \frac{\sum_{j=1}^d \lambda_j}{\left[\sqrt{\sum_{j=1}^d \lambda_j} \right]_\gamma^2} \right\}, \end{aligned}$$

where $m(\lambda_i)$ is the algebraic multiplicity of the eigenvalue λ_i . At this point we have a probability measure comprising the $d^* + 1$ weights \bar{p}_i and the points $\bar{l}_1, \dots, \bar{l}_{d^*}, 0 \in \gamma Q \mathbb{Z}^d$ and we refer to this random variable by \bar{Y} . We can now build the r.v. Y taking the $2d^* + 1$ points $\{\bar{l}_i, -\bar{l}_i\}$ with probability $p_i = \bar{p}_i/2$, such that $\mathbb{E}Y = 0$ and $\mathbb{E}Y^{\otimes 2} = B$. \square

Constructing the Lattice-Tree Model

Although we have not proved Theorem 8 for general dimension d , we can still prove one particular case. We are ready to build the Markov chain X^n which approximates X in the multi-dimensional case under more stringent conditions than before.

Theorem 10. Let $\mu(x) = 0, \sigma \in C_b^4, d \geq 2, \Sigma(x) = Q\Lambda(x)Q^T, \Lambda(x) \in \mathbb{R}^{d \times d}$ a diagonal matrix and Q an orthogonal matrix. For any $c > 0$, set $\gamma_n = cn^{-1/2}$ and let us suppose that $X_0^n \in \gamma_n Q\mathbb{Z}^d$. There exists a lattice approximation $(X^n)_{n \geq 1}$ to X on $\gamma_n Q\mathbb{Z}^d$ respect to the class of functions $f \in C_p^3$ with rate $1/2$. Moreover,

$$\begin{cases} \text{card}[\text{supp}(X_i^n)] \leq i^d c. & , \text{ if } \mu, \sigma \text{ are bounded,} \\ \text{card}[\text{supp}(X_i^n)] \leq ci^d \exp(dci n^{-1/2}) & , \text{ if } \mu, \sigma \text{ have linear growth.} \end{cases}$$

Proof. The proof follows the same reasoning used in Theorem 4 and 8, but using Theorem 9. The bounds on the cardinality of the support can be obtained as explained in the proof of Theorem 8 \square

It is relevant to recall that to obtain $\mu(x) = 0$ the Girsanov Theorem can be helpful.

6 Variations and Extensions

We briefly discuss variations and possible extensions.

6.1 Combining State-Space Transformations and Recombination

The classical transformation approach and the recombination approach we present here are not mutually exclusive and can be combined in way that leverages their strengths. To demonstrate this, we revisit the problem of approximating an SDE with drift and volatility that are not bounded but have linear growth. Our Theorem 4 guarantees under ellipticity assumptions on σ , that

$$\begin{cases} \text{card}[\text{support}(X_i^n)] \leq c \left(i \times e^{n^{-1/2}i} \right), & \text{if } \mu \text{ and } \sigma^2 \text{ have linear growth,} \\ \text{card}[\text{support}(X_i^n)] \leq ci, & \text{if } \mu \text{ and } \sigma^2 \text{ are bounded,} \end{cases}$$

which in the case of linear growth leads to exponential growth in time. However, inspired by Nelson and Ramaswamy [47], we can look for a state-space transformation to first turn the SDE into an SDE with bounded vector fields and subsequently apply Theorem 4, but now for the case of bounded vector fields. Formally, this means that we look for a state-space transformation given by smooth function f such that $Y_t = f(X_t)$, where X_t is the solution of the SDE (2), with bounded μ, σ . Thus, using the Ito's lemma, Y solves

$$dY_t = \left[\mu(X_t) \partial f(X_t) + \frac{\sigma(X_t)^2}{2} \partial^2 f(X_t) \right] dt + \sigma(X_t) \partial f(X_t) dW_t. \quad (21)$$

Going back to Equation (6.1) we can see that⁹

$$\begin{aligned} \mathbb{E}_t[Y_{t+n-1} - Y_t] &= \left[\mu(X_t) \partial f(X_t) + \frac{\sigma(X_t)^2}{2} \partial^2 f(X_t) \right] n^{-1} + O(n^{-2}), \\ \mathbb{E}_t[Y_{t+n-1} - Y_t]^2 &= [\sigma(X_t) \partial f(X_t)]^2 n^{-1} + O(n^{-2}). \end{aligned}$$

Now, if we suppose to have built the Markov chain approximation X_i^n as in Theorem 4, then we want to see if $Y_i^n = f(X_i^n)$ satisfy Lemma 1-Item (1), respect to Y . Using a Taylor expansion we get

$$\mathbb{E}_i Y_{i+1}^n - Y_i^n = \mathbb{E}_i [f(X_{i+1}^n) - f(X_i^n)]$$

⁹We have to suppose $\mathbb{E}_t \int_t^{t+n-1} \sigma(X_s) \partial f(X_s) dW_s = 0$, which is for example satisfied when the integrand is (locally) square integrable.

$$\begin{aligned}
&= \mathbb{E}_i \left[\partial f(X_i^n) \Delta X_i^n + \frac{1}{2} \partial^2 f(X_i^n) (\Delta X_i^n)^2 + O(|\Delta X_i^n|^3) \right] \\
&= \partial f(X_i^n) \mathbb{E}_i [\Delta X_i^n] + \frac{1}{2} \partial^2 f(X_i^n) \mathbb{E}_i [(\Delta X_i^n)^2] + O(n^{-3/2}),
\end{aligned}$$

where $\Delta X_i^n = X_{i+1}^n - X_i^n$ and by construction $\mathbb{E}_i O(|\Delta X_i^n|^3) \leq c \sum_j p_j \left| n^{-1/2} z_j \right|^3 \leq cn^{-3/2}$, $z_j \in \mathbb{Z}$. Given that X_i^n satisfies Lemma 1-Item (1) respect to X , we have that $\mathbb{E}_i [Y_{i+1}^n - Y_i^n] = \mathbb{E}_i [Y_{i+n-1} - Y_i] + O(n^{-3/2})$. For the second moment we proceed similarly

$$\begin{aligned}
\mathbb{E}_i [Y_{i+1}^n - Y_i^n]^2 &= \mathbb{E}_i [f(X_{i+1}^n) - f(X_i^n)]^2 \\
&= \mathbb{E}_i \left[\partial f(X_i^n) \Delta X_i^n + O(|\Delta X_i^n|^2) \right]^2 \\
&= \partial f(X_i^n)^2 \mathbb{E}_i [(\Delta X_i^n)^2] + 2\partial f(X_i^n) \mathbb{E}_i [\Delta X_i^n \times O(|\Delta X_i^n|^2)] + \mathbb{E}_i O(|\Delta X_i^n|^4),
\end{aligned}$$

note that $|\mathbb{E}_i [\Delta X_i^n \times O(|\Delta X_i^n|^2)]| \leq cn^{-3/2}$. Since

$$\mathbb{E}_i [(\Delta X_i^n)^2] = \mu (X_i^n)^2 n^{-2} + \sigma (X_i^n)^2 n^{-1} + O(n^{-3/2}) = \sigma (X_i^n)^2 n^{-1} + O(n^{-2}),$$

we can conclude that $\mathbb{E}_i [Y_{i+1}^n - Y_i^n]^2 = \mathbb{E}_i [Y_{i+n-1} - Y_i]^2 + O(n^{-3/2})$. In the same way we obtain $\mathbb{E}_i |Y_{i+1}^n - Y_i^n|^3 = O(n^{-3/2})$, therefore we can conclude that Y^n satisfies Lemma 1-Item (1) respect to Y for $\alpha = 1/2$. Already the example of Y a Geometric Brownian Motion is interesting where $f(x) = e^x$ yields that X is a Brownian Motion with a bounded mean and variance, i.e.

$$\frac{dY_t}{Y_t} = \mu dt + \sigma dW_t, \quad dX_t = \left[\mu - \frac{\sigma^2}{2} \right] dt + \sigma dW_t.$$

The state space is not any more a lattice a priori, indeed it is a non linear transformation of a lattice. To sum up, this simple variation allows to treat linear growth vector fields by a MC.

6.2 Optimal Control and Optimal Stopping

In Section 6.2 we cited optimal stopping as one of the motivations to consider MC approximations. However, in general the topology of weak convergence is too coarse to provide robustness of the discretization procedures; for example, the solution of the optimal stopping problem for the MC does in general not converge to the solution of the stopping problem for the SDE; see [27]. Nevertheless, we can mimic the arguments in [31, 32] to show that solving the optimal stopping for MC produces by **DISCRETIZE_n** yields solutions that are close to the continuous time optimal stopping problem.

Theorem 11. ([32, Chapter 10, Theorem 6.2]) *Let X be the solution of the SDE (2) and X^n be one of the schemes built in Theorems 4, 8, 10. Denote with \mathcal{T} and \mathcal{T}^n the set of almost sure finite stopping times with respect to the natural filtrations of the SDE (2) resp. the filtration generated by X^n . Define the value functions*

$$V(x) := \min_{\tau \in \mathcal{T}} \mathbb{E} \left[\int_0^{\tau \wedge 1} k(X_s) ds + g(X_{\tau \wedge 1}) \mid X_0 = x \right] \text{ and } V^n(x) := \min_{\tau^n \in \mathcal{T}^n} \mathbb{E} \left[\frac{1}{n} \sum_{i=0}^{\tau^n \wedge n} k(X_i^n) + g(X_{\tau^n \wedge n}) \mid X_0^n = x \right].$$

If μ, σ, g, k are bounded Lipschitz continuous¹⁰, then $V^n(x) \rightarrow V(x)$, as $n \rightarrow \infty$.

¹⁰Other rather general and technical assumptions on V are needed.

Proof. Thanks to [32, Chapter 10, Theorem 6.2], it is enough to check that the conditions in [32, Equation (1.3), page 71] are satisfied for the approximation schemes of Theorems 4, 8, 10. [32, Equation (1.3), page 71] has two different types of conditions: the first type of conditions is equivalent to Lemma 1-(1); the second type of conditions requires that a.s. $\sup_i |X_{i+1}^n - X_i^n| \rightarrow 0$, as $n \rightarrow \infty$, but this is a consequence of the fact that the r.v. $Y_{(\cdot)}$ built in the proof of Theorems 4, 8, 10 are bounded by the bounds of Theorems 3, 7, 9. \square

For brevity we do not pursue this further here, but note that the same logic extends to more general control problems, at least in principle: one needs to introduce controls in the coefficients μ and σ and define the related value functions V, V^n ; see [31, 32]. In this case, one would need to construct the random variable $Y_{(\cdot)}$ for the different controls to allow for effective computation.

6.3 Extensions.

Hypo-elliptic SDEs. If we do not know a priori if Σ is elliptic, we can take a $\gamma > 0$ reasonably small and then start building the random variables $Y_{(\cdot)}$ for the lattice $\gamma\mathbb{Z}^d$. If we arrive in a point x where $\sqrt{\lambda_{\min}[\Sigma(x)]} \leq \gamma$, then it is enough to add to the lattice $\gamma\mathbb{Z}^d$ some points in the lattice $\mu(x) + \gamma Q\mathbb{Z}^d$, where Q depends as usual from $\Sigma(x)$.

Switching SDEs Switching diffusions can be treated in a straightforward manner. Suppose that the functions Σ and μ depend on one (or more) independent dynamic process $\theta(t)$, the “switch”, in a Markovian way. For simplicity, let us suppose that $\theta(t) \in \{0, 1\}$. We consider four functions Σ_0, μ_0 and Σ_1, μ_1 :

$$\mathbb{E}_t[X_{t+n-1} - X_t] \approx \begin{cases} \mu_1(X_t) & \text{if } \theta_t = 1 \\ \mu_0(X_t) & \text{if } \theta_t = 0 \end{cases}, \quad \mathbb{V}_t[X_{t+n-1} - X_t] \approx \begin{cases} \Sigma_1(X_t) & \text{if } \theta_t = 1 \\ \Sigma_0(X_t) & \text{if } \theta_t = 0 \end{cases}.$$

Thus, we can approximate two different processes defined by Σ_0, μ_0 and Σ_1, μ_1 independently, assuming to know or to be able to build a “good” discrete time approximation θ^n of $\theta(t)$.

Time-inhomogeneous SDEs. We have presented the case of time-homogeneous SDEs, but the convergence results in [30] apply to time-inhomogeneous SDEs, hence the same approach extends to time-inhomogeneous SDEs. Nevertheless, from a computational perspective, time-inhomogeneous SDEs are more challenging since if lattice points are re-visited then the recombination computation has to be redone because time has increased since the last visit. In principle, it is also possible to go beyond SDEs, but the bottleneck there is to find suitable conditions to replace those given in Definition 3 to guarantee the weak convergence.

7 Algorithm and Experiments

First, let us state the pseudo-code of the algorithm for the construction of the Transition Matrix. We suppose for simplicity $X_0 = 0$, so that $X_0^n \in \gamma_n Q\mathbb{Z}^d$ for any n, Q . Remember that, if X_0^n is not in the lattice, X_1^n must be built directly to be in the considered lattice, see the proof of Theorem 4.

Step 7: Solving a non-linear Moment System. Step 7 is the computationally most challenging step of Algorithm **DISCRETIZE** since it requires to solve the non-linear lattice-constrained system (2). We now show that previous work on recombination allows to do this step efficiently. Since our experiments are in dimension $d = 2$, we provide the details for the two-dimensional case, but the same approach works for general dimension d under minor modifications. Assuming the assumptions of Theorem 7, 8 are met, then one can replace both \approx signs in (2) by $=$ and solve the resulting system by either

Algorithm DISCRETIZE Lattice-Tree Markov Chain Approximation

Inputs: μ, σ, x_0, n

```

1: Choose  $\gamma_n$  and  $Q$  as prescribed by Theorem 4, 8, 10
2: queue_todo.append[ $x_0$ ]
3:  $i \leftarrow 1$ 
4: while  $i \leq n$  do
5:   while queue_todo is not empty do
6:      $x \leftarrow$  queue_todo[0]
7:     Solve the non-linear moment matching system (2) to get  $(p_j, l_j)_j \subset \mathbb{R}^+ \times \gamma_n Q \mathbb{Z}^d$ 
8:     Transition Matrix  $M_n \leftarrow \{x \text{ can go to } \{x + l_j\} \text{ with probabilities } \{p_j\}\}$ 
9:     queue_todo_nextstep.append[ for any  $j$ ,  $x + l_j$  such that  $x + l_j \notin$  queue_done  $\cup$ 
       queue_todo_nextstep ]
10:    queue_todo.remove[ $x$ ]
11:    queue_done.append[ $x$ ]
12:  end while
13:  queue_todo  $\leftarrow$  queue_todo_nextstep
14:  queue_todo_nextstep.remove[all elements]
15:   $i \leftarrow i + 1$ 
16: end while
17: Return Transition Matrix  $M_n$  and starting value  $x_n$ 

```

(i) following the proof of Theorem 7, or

(ii) by using the randomized recombination Algorithm 2 in Cosentino et al. [6].

The approach (i) has two disadvantages: firstly the number r of weight-point tuples $(w_j, l_j)_{j=1, \dots, r}$ it returns is in general larger than needed whereas (ii) returns the smallest number of tuples that is guaranteed by Carathéodory's theorem; secondly, and more importantly in practice, is that approach (i) is slower than the optimized recombination Algorithm 2 used in approach (ii).

Remark 4. *There are other recombination algorithms [20, 21, 22, 37, 43, 53] besides Algorithm 2 from [6]. However, what makes Algorithm 2 especially attractive in the current situation is that it can take as input a sequence of points – in our case, points sampled uniformly on the lattice – and the moments that need to be matched. This is in contrast to all the other mentioned algorithms, which need as input a discrete measure, that is a collection of points and weights. Although subtle, this is an important difference that adds to the efficiency of the resulting algorithm.*

If the assumptions of Theorem 8 are not met, one could still follow the approach “in spirit”, by trying to construct a discrete probability measure with a small support that approximates the first and second moments. Formally, this means we

(iii) try to solve the constrained minimization problem

$$\begin{aligned}
 (w_j, l_j)_j := \arg \min & \left| \sum_j \begin{pmatrix} z_j \cdot w_j \\ z_j^{\otimes 2} \cdot w_j \end{pmatrix} - \begin{pmatrix} n^{-1} \mu(x) \\ n^{-2} \mu(x)^{\otimes 2} + n^{-1} \Sigma(x) \end{pmatrix} \right|^2 \\
 \text{s.t. } & l_j \in \gamma \mathbb{Z}^2, \quad |l_j| \leq c \\
 & 0 \leq w_j \leq 1, \quad \sum_j w_j = 1,
 \end{aligned} \tag{22}$$

and then reduce the discrete probability measure obtained by (iii) further by using any of the recombination algorithms in [20, 21, 22, 37, 43, 53].

We re-emphasize that in the case of (iii) none of the theoretical guarantees of the previous sections apply; in fact, it is not even guaranteed that the minimum of the optimization problem (iii) is small enough. Nevertheless, highly optimized optimization software exists for such constrained problems and the intuition remains that the resulting MC is a good approximation in view of its construction. In fact, our numerical example below suggests that this procedure is quite robust.

We have implemented the above algorithms and below we apply it to two concrete problems: Section 7.1 applies **DISCRETIZE** to a two-dimensional SDE and Section 7.2 studies the robustness of **DISCRETIZE** by applying it to a Heston model where the assumptions of Theorem 8 are not met and we use case (iii) above to solve a constrained minimization problem. Python code to replicate the two experiments in the next Subsections is publicly available¹¹; see also Section B.

7.1 Example: toy model

We consider the following toy model

$$d \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \sin(X) \\ \cos(Y) \end{pmatrix} dt + \begin{pmatrix} \cos(Y) + 2 & 0 \\ 0 & \sin(X) + 2 \end{pmatrix} dW_t,$$

$$X_0 = Y_0 = 0.$$

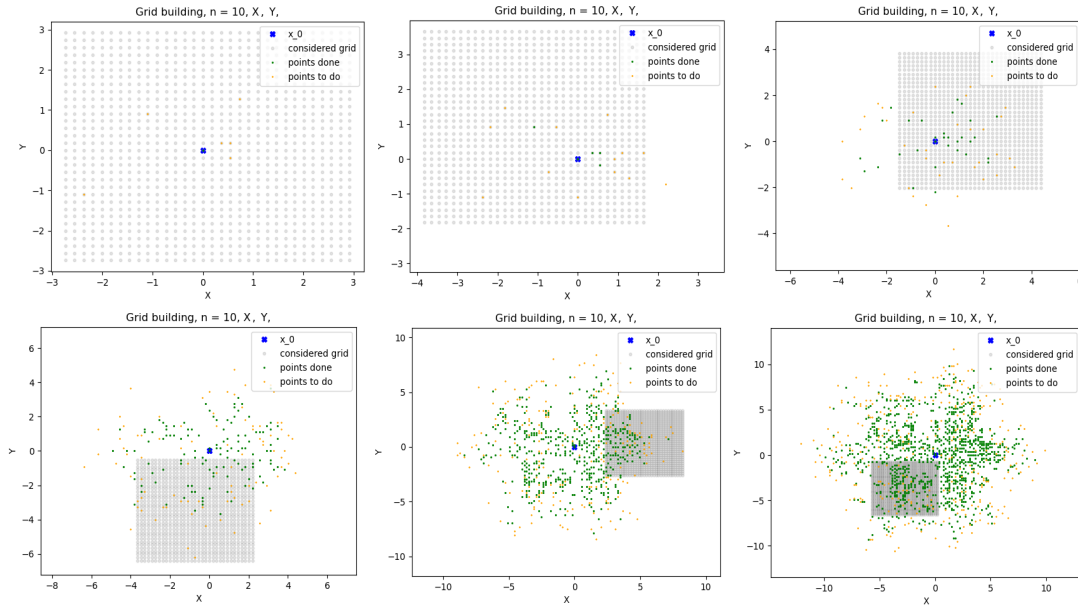


Figure 2: Construction of the transition matrix. Starting from x_0 we consider to build the random variable X_1 , such that it has support on the grey points and the first two moments are matched exactly. In the top-left plot, X_1 has support on the yellow points. Then this procedure is iterated for all the possible points in the support of the measure, following Algorithm **DISCRETIZE**.

¹¹<https://github.com/FraCose/Tree-Recombination>

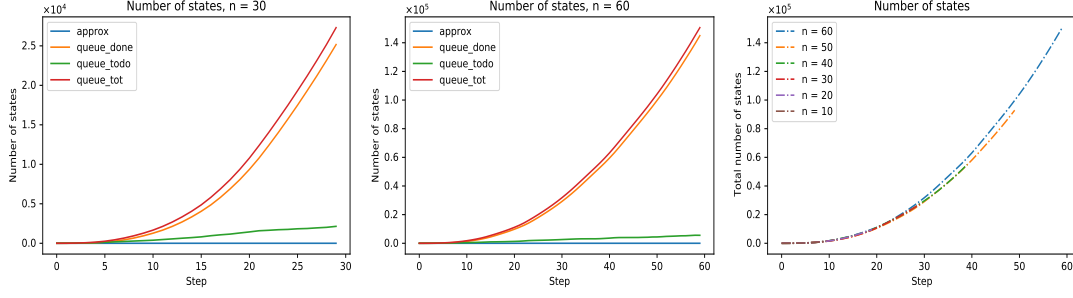


Figure 3: The left and middle plots show the growth of the cardinalities of the queues created by Algorithm **DISCRETIZE** for fixed n as i increases. The right plot shows a comparison of the growth of the number of states for different n as i increases.

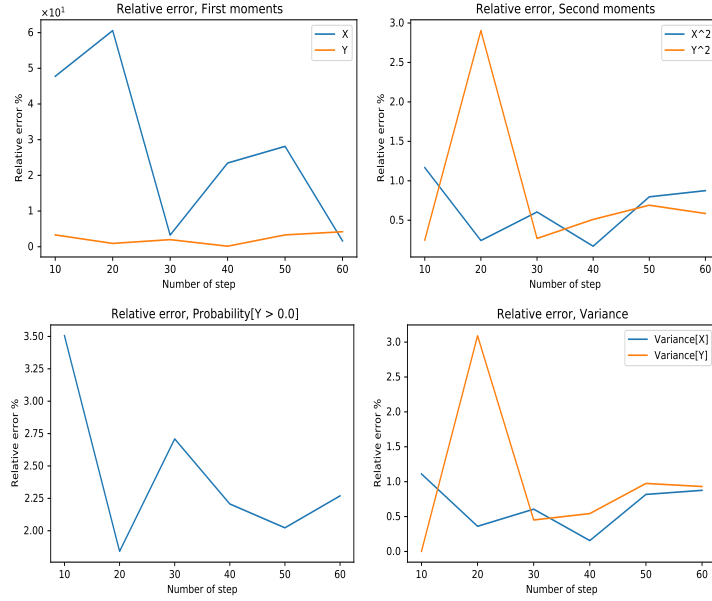


Figure 4: Relative errors for some functionals of X_1, Y_1 . We compute the first and the second moments, the variances and the probability that Y_1 is greater than 0, respectively $\mathbb{E}[X_1], \mathbb{E}[Y_1], \mathbb{E}[X_1^2], \mathbb{E}[Y_1^2], \mathbb{V}[X_1], \mathbb{V}[Y_1]$ and $\mathbb{P}[Y_1 > 0]$. As ground truth we consider an Euler scheme Monte Carlo simulation of the mode with $n = 1000$ steps.

It is easy to check that this model satisfies the assumptions of Theorem 8, indeed the vector fields are bounded and the SDE is strictly elliptic: $\varepsilon = \inf_x \lambda_{\min}(x) = \inf_x \text{eigenvalues}(\sigma \sigma^T(x)) = 1$, thus we can consider $\gamma_n = n^{-1/2} \sqrt{1/3}$.

Results. Figure 2 visualizes the iterative nature of the MC construction: starting from the initial point x_0 the algorithm selects at every time step a few possible attainable states on the two-dimensional lattice, such that the local consistency condition is satisfied. As time progresses, new states get added and many previous states are revisited (leading to a recombining tree). Figure 3 shows the growth of the state space as a function of time for different values of n , and Figure 4 shows the relative error. Finally, Figure 5

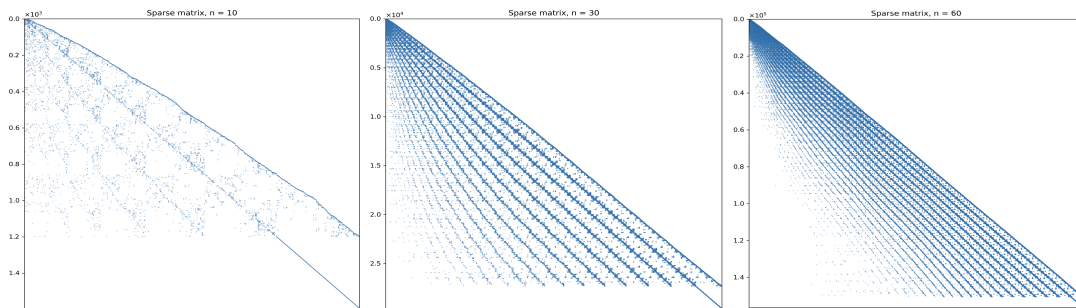


Figure 5: Representation as sparse matrix of the Probability Transition Matrix for different values of n .

shows the resulting sparse transition matrices. Unsurprisingly, the transition matrices are sparse, as seen in Figure 5, since this is optimized by construction, see also Appendix B.

7.2 Example: Mean reverting Heston model

For the second experiment, we consider the Heston model (HM) [23], a popular stochastic volatility model in mathematical finance

$$\begin{aligned} dP_t &= \mu P_t dt + \sqrt{V_t} P_t dW_t^P, \\ dV_t &= \lambda(\theta - V_t) dt + \xi \sqrt{V_t} dW_t^V, \end{aligned} \quad (23)$$

where P_t is the price of an asset and V_t its instantaneous variance; W_t^P, W_t^V are two standard Brownian Motions with correlation parameter ρ and $\lambda, \theta, \mu, \xi$ are all constants. We recall that if $2\lambda\theta > \xi^2$, then V_t is a.s. positive, see for example [10]. Tree approximations that are tailor-made to the HM model in (7.2), e.g. Akyıldırım et al. [1], Zeng and Zhu [58], have been developed but, to the best of our knowledge, these do not apply when V depends on P . The latter is a desirable model property since empirical observation suggest the price and the volatility to be inversely related, and that V has a mean reverting behaviour; for theoretical properties of such an SDE, see Romano and Touzi [48]. A simple example of such price-variance relation that we use for our numerical experiment is

$$\theta_t := \theta(P_t) := c \frac{1}{P_t + 1} + k.$$

Indeed, one can check that in this case V is mean-reverting¹², $\lim_{t \rightarrow \infty} \mathbb{E}V_t = \mathbb{E}\theta_\infty$. To sum up, we consider the system

$$d \begin{pmatrix} P_t \\ V_t \end{pmatrix} = \begin{pmatrix} \mu P_t \\ \lambda \left(\frac{c}{1+P_t} + k - V_t \right) \end{pmatrix} dt + \sqrt{V_t} \begin{pmatrix} P_t & 0 \\ 0 & \xi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} d \begin{pmatrix} W_t^1 \\ W_t^2 \end{pmatrix},$$

where W_t^1, W_t^2 are two independent Brownian Motions. Moreover, in a finance context it is more natural model $\ln P_t$ instead of P_t , and we reformulate the above system as

$$d \begin{pmatrix} \ln P_t \\ V_t \end{pmatrix} = \begin{pmatrix} \mu - \frac{1}{2} V_t \\ \lambda \left(\frac{c}{1+P_t} + k - V_t \right) \end{pmatrix} dt + \sqrt{V_t} \begin{pmatrix} 1 & 0 \\ \xi \rho & \xi \sqrt{1-\rho^2} \end{pmatrix} d \begin{pmatrix} W_t^{(1)} \\ W_t^{(2)} \end{pmatrix}.$$

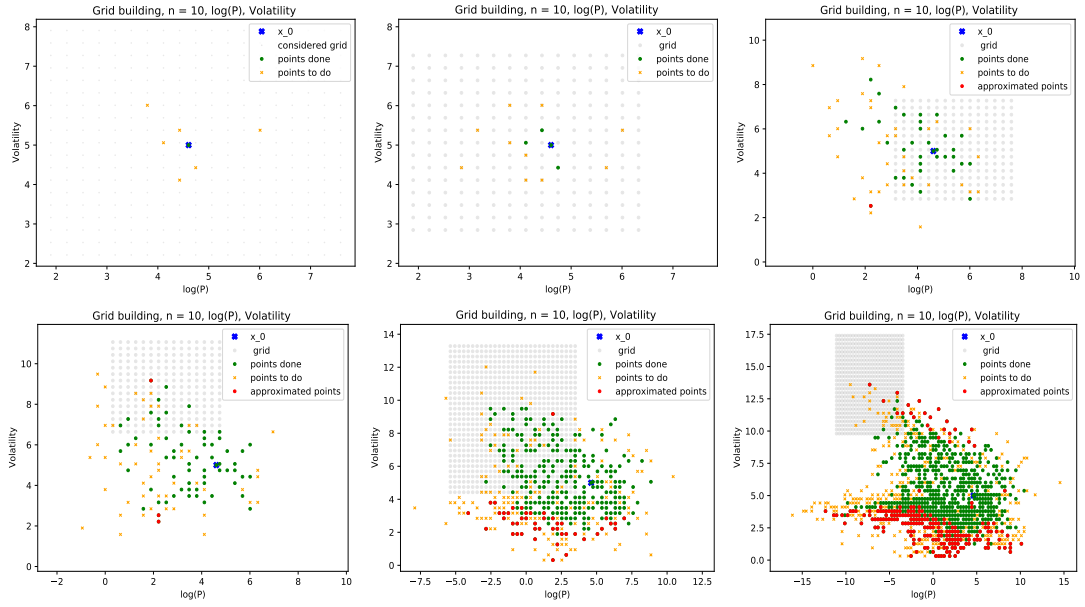


Figure 6: Construction of the transition matrix. Starting from x_0 we consider to build the random variable X_1 such that it has support on the grey points and the first two moments are matched exactly. In the top-left plot, X_1 has support on the yellow points. Then this procedure is iterated for all the possible points in the support of the measure, following Algorithm **DISCRETIZE**. In contrast to Figure 2 of the previous example, now also red points appear since we are in the situation of case (iii) discussed above.

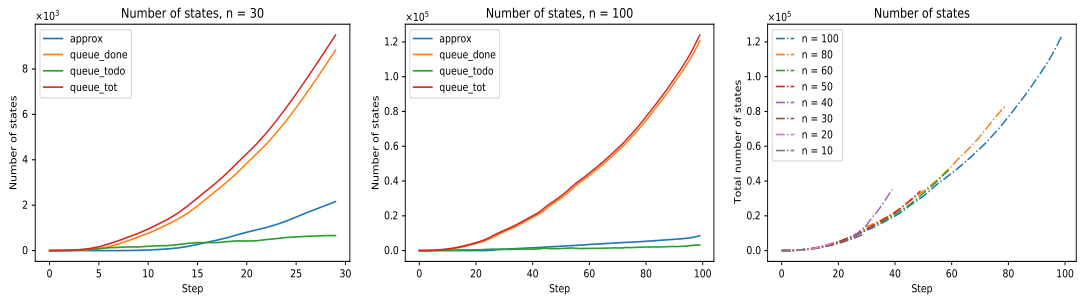


Figure 7: The left and middle plots show the growth of the cardinalities of the queues created by Algorithm **DISCRETIZE** for fixed n as i increases. The right plot shows a comparison of the growth of the number of states for different n as i increases.

The assumptions of Theorem 8 are not met since the eigenvalues of the volatility are not bounded away from 0 due to the presence of V_i . Nevertheless, we can follow the approach outlined at the end of the above section, that is step (iii) which aims to solve the constrained optimization problem (iii).

¹²By using the differential form of $d(e^{t\lambda}V_t)$ and Ito's rule. We have to suppose the existence of $\mathbb{E}\theta_\infty$.

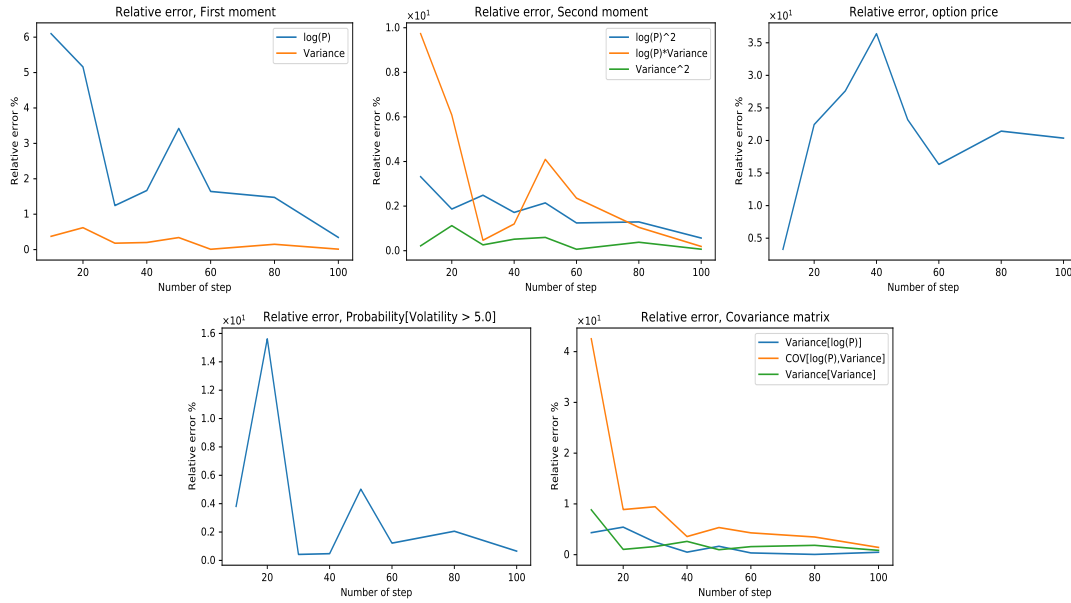


Figure 8: Relative errors for some functionals of P_1, V_1 . For the option, we consider a European Option with strike 100. We compute the first and the second moments, the variances and the probability that Y_1 is greater than 5, respectively $\mathbb{E}[(\log(P_1), V_1)]$, $\mathbb{E}[(\log(P_1), V_1)^{\otimes 2}]$, $\mathbb{V}[(\log(P_1), V_1)]$ and $\mathbb{P}[V_1 > 5]$. The formula for the option price is $\mathbb{E}[\max(P_1 - 100, 0)]$. As ground truth we consider an Euler scheme Monte Carlo simulation of the mode with $n = 1000$ steps.

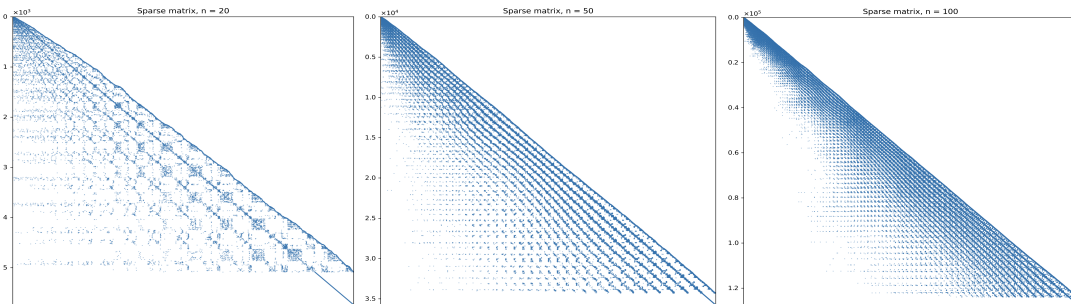


Figure 9: Representation as sparse matrix of the Probability Transition Matrix for different values of n .

Results. For our simulations we set $\log(P_0) = \log(100)$, $V_0 = 5$, $\mu = 0$, $\lambda = 2$, $\rho = 0.2$, $\xi = 1$, $c = 2$, $k = 5$ and $t \in [0, 1]$. The set of experiments is the same as those presented in the previous Subsection. Figure 6 visualizes the iterative nature of the MC construction. It can be noticed that Figure 6 includes the presence of red points, which represent approximated points, i.e. the ones “recombined” numerically using the optimization problem iii, whilst in Figure 2 these are not present since the assumptions of Theorem 8 are satisfied. Figure 7 shows the growth of the state space as a function of time for different values of n , and Figure 8 shows the relative error. Finally, Figure 9 shows the resulting sparse transition matrices. Note that the approximations error, Figure 8, supports the intuition we have used in step (iii): although the assumptions of Theorem 8 are not met, the model error resulting from using the MC is small. Again, it is not surprising that the transition matrices are sparse, Figure 9, since this is optimized by construction.

References

- [1] E. Akyıldırım, Y. Dolinsky, and H. M. Soner. Approximating stochastic volatility by recombinant trees. *The Annals of Applied Probability*, 24(5), oct 2014. doi: 10.1214/13-aap977.
- [2] P. P. Boyle. A lattice framework for option pricing with two state variables. *The Journal of Financial and Quantitative Analysis*, 23(1):1, mar 1988. doi: 10.2307/2331019.
- [3] P. P. Boyle, J. Evnine, and S. Gibbs. Numerical evaluation of multivariate contingent claims. *Review of Financial Studies*, 2(2):241–250, apr 1989. doi: 10.1093/rfs/2.2.241.
- [4] M. Bujorianu and J. Lygeros. Toward a general theory of stochastic hybrid systems. In H. Blom and J. Lygeros, editors, *Stochastic Hybrid Systems*, volume 337 of *Lecture Notes in Control and Information Science*. Springer, 2006.
- [5] R.-R. Chen and T. T. Yang. A universal lattice. *Review of Derivatives Research*, 3(2):115–133, 1999. doi: 10.1023/a:1009646809675.
- [6] F. Cosentino, H. Oberhauser, and A. Abate. A randomized algorithm to reduce the support of discrete measures. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 15100–15110. Curran Associates, Inc., 2020. URL <https://proceedings.neurips.cc/paper/2020/file/ac4395adcb3da3b2af3d3972d7a10221-Paper.pdf>.
- [7] M. Costabile, I. Massabò, and E. Russo. A forward shooting grid method for option pricing with stochastic volatility. *The Journal of Derivatives*, 20(2):67–78, nov 2012. doi: 10.3905/jod.2012.20.2.067.
- [8] M. Costabile, I. Massabò, and E. Russo. A shifted tree model for the efficient evaluation of options with fixed dividends. *IMA Journal of Management Mathematics*, page dpw002, jun 2016. doi: 10.1093/imaman/dpw002.
- [9] J. C. Cox, S. A. Ross, and M. Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, sep 1979. doi: 10.1016/0304-405x(79)90015-1.
- [10] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385, mar 1985. doi: 10.2307/1911242.
- [11] C. Dehnert, S. Junges, J.-P. Katoen, and M. Volk. A storm is coming: A modern probabilistic model checker. In *International Conference on Computer Aided Verification*, pages 592–600. Springer, 2017.
- [12] N. Ekvall. A lattice approach for pricing of multivariate contingent claims. *European Journal of Operational Research*, 91(2):214–228, jun 1996. doi: 10.1016/0377-2217(95)00279-0.
- [13] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.
- [14] T. J. Finucane and M. J. Tomas. American stochastic volatility call option pricing: A lattice based approach. *Review of Derivatives Research*, 1(2):183–201, jun 1996. doi: 10.1007/bf01531598.
- [15] I. Florescu and F. G. Viens. Stochastic volatility: Option pricing using a multinomial recombining tree. *Applied Mathematical Finance*, 15(2):151–181, apr 2008. doi: 10.1080/13504860701596745.

- [16] A. Gamba and L. Trigeorgis. An improved binomial lattice method for multi-dimensional options. *Applied Mathematical Finance*, 14(5):453–475, dec 2007. doi: 10.1080/13504860701532237.
- [17] A. Gamba and L. Trigeorgisz. A log-transformed binomial lattice extension for multi-dimensional option problems. In *Real Options*, 2001.
- [18] S. Gerschgorin. Über die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, 6:749–754, 1931.
- [19] T. J. Haahtela. Recombining trinomial tree for real option valuation with changing volatility. *SSRN Electronic Journal*, 2010. doi: 10.2139/ssrn.1932411.
- [20] S. Hayakawa. Monte-carlo cubature construction. *ArXiv*, abs/2001.00843, 2020.
- [21] S. Hayakawa, T. Lyons, and H. Oberhauser. Estimating the probability that a given vector is in the convex hull of a random sample. *arXiv preprint arXiv:2101.04250*, 2021.
- [22] S. Hayakawa, H. Oberhauser, and T. Lyons. Positively weighted kernel quadrature via subsampling. *arXiv preprint arXiv:2107.09597*, 2021.
- [23] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, apr 1993. doi: 10.1093/rfs/6.2.327.
- [24] T.-S. Ho, R. C. Stapleton, and M. G. Subrahmanyam. Multivariate binomial approximations for asset prices with nonstationary variance and covariance characteristics. *Review of Financial Studies*, 8(4): 1125–1152, oct 1995. doi: 10.1093/rfs/8.4.1125.
- [25] A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20(1):37–39, mar 1953. doi: 10.1215/s0012-7094-53-02004-3.
- [26] J. M. Holte. Discrete gronwall lemma and applications. MAA North Central Section Meeting, 2009. URL <http://homepages.gac.edu/~holte/publications/GronwallLemma.pdf>.
- [27] D. N. Hoover and H. J. Keisler. Adapted probability distributions. *Transactions of the American Mathematical Society*, 286(1):159–201, 1984.
- [28] J. Hu and M. Prandini. Aircraft conflict detection: A method for computing the probability of conflict based on markov chain approximation. In *7th European Control Conference, ECC 2003, Cambridge, UK, September 1-4, 2003*, pages 2225–2230. IEEE, 2003.
- [29] D. Ji and B. W. Brorsen. A recombining lattice option pricing model that relaxes the assumption of lognormality. *Review of Derivatives Research*, 14(3):349–367, dec 2010. doi: 10.1007/s11147-010-9060-3.
- [30] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992. ISBN 3-540-54062-8. doi: 10.1007/978-3-662-12616-5.
- [31] H. J. Kushner. Numerical methods for stochastic control problems in continuous time. *SIAM Journal on Control and Optimization*, 28(5):999–1048, 1990.
- [32] H. J. Kushner and P. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer New York, 2001. doi: 10.1007/978-1-4613-0007-6.

- [33] M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic symbolic model checking with PRISM: A hybrid approach. *International Journal on Software Tools for Technology Transfer*, 6(2):128–142, 2004.
- [34] A. Leccadito, P. Toscano, and R. S. Tunaru. Hermite binomial trees: a novel technique for derivatives pricing. *International Journal of Theoretical and Applied Finance*, 15(08):1250058, dec 2012. doi: 10.1142/s0219024912500586.
- [35] D. P. Leisen. A stochastic volatility lattice. *IFAC Proceedings Volumes*, 31(16):75–80, jun 1998. doi: 10.1016/s1474-6670(17)40461-7.
- [36] A. Li, P. Ritchken, and L. Sankarasubramanian. Lattice models for pricing american interest rate claims. *The Journal of Finance*, 50(2):719–737, jun 1995. doi: 10.1111/j.1540-6261.1995.tb04802.x.
- [37] C. Litterer and T. Lyons. High order recombination and an application to cubature on wiener space. *The Annals of Applied Probability*, 22(4), aug 2012. doi: 10.1214/11-aap786.
- [38] R. Liu and J. Zhao. A lattice method for option pricing with two underlying assets in the regime-switching model. *Journal of Computational and Applied Mathematics*, 250:96–106, oct 2013. doi: 10.1016/j.cam.2013.02.012.
- [39] R. H. Liu. Regime-switching recombining tree for option pricing. *International Journal of Theoretical and Applied Finance*, 13(03):479–499, may 2010. doi: 10.1142/s0219024910005863.
- [40] J. A. D. Loera, R. Hemmecke, and M. Köppe. *Algebraic and Geometric Ideas in the Theory of Discrete Optimization*. Society for Industrial and Applied Mathematics, dec 2012. doi: 10.1137/1.9781611972443.
- [41] U. H. Lok and Y.-D. Lyuu. Efficient trinomial trees for local-volatility models in pricing double-barrier options. *Journal of Futures Markets*, 40(4):556–574, dec 2019. doi: 10.1002/fut.22080.
- [42] T. Lyons and N. Victoir. Cubature on wiener space. 460(2041):169–198, jan 2004. doi: 10.1098/rspa.2003.1239.
- [43] A. Maalouf, I. Jubran, and D. Feldman. Fast and accurate least-mean-squares solvers. In *Advances in Neural Information Processing Systems*, pages 8305–8316, 2019.
- [44] E. Moretto, S. Pasquali, and B. Trivellato. Derivative evaluation using recombining trees under stochastic volatility. *Advances and Applications in Statistical Science*, 1:453–480, Aug. 2009.
- [45] S. W. N. Ikeda. *Stochastic Differential Equations and Diffusion Processes*. Elsevier S&T, June 2014. URL https://www.ebook.de/de/product/23481663/n_ikeda_s_watanabe_stochastic_differential_equations_and_diffusion_processes.html.
- [46] S. K. Nawalkha and N. Beliaeva. Efficient trees for CIR and CEV short rate models. *SSRN Electronic Journal*, 2007. doi: 10.2139/ssrn.976819.
- [47] D. B. Nelson and K. Ramaswamy. Simple binomial processes as diffusion approximations in financial models. *Review of Financial Studies*, 3(3):393–430, jul 1990. doi: 10.1093/rfs/3.3.393.
- [48] M. Romano and N. Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412, oct 1997. doi: 10.1111/1467-9965.00038.
- [49] M. Rubinstein. Implied binomial trees. *The Journal of Finance*, 49(3):771–818, jul 1994. doi: 10.1111/j.1540-6261.1994.tb00079.x.

- [50] M. Rubinstein. Edgeworth binomial trees. *The Journal of Derivatives*, 5(3):20–27, feb 1998. doi: 10.3905/jod.1998.407994.
- [51] J.-G. Simonato. Johnson binomial trees. *SSRN Electronic Journal*, 2008. doi: 10.2139/ssrn.1346110.
- [52] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer Berlin Heidelberg, 2006. doi: 10.1007/3-540-28999-2.
- [53] M. Tchernychova. *Caratheodory cubature measures*. PhD thesis, University of Oxford, 2016.
- [54] M. Wahab and C.-G. Lee. A lattice approach to pricing of multivariate contingent claims with regime switching. *The Journal of Derivatives*, 17(1):49–61, aug 2009. doi: 10.3905/jod.2009.17.1.049.
- [55] R. Xu and S. Li. A tree model for pricing credit default swaps with equity, market and default risk. In *2009 International Conference on Management and Service Science*. IEEE, sep 2009. doi: 10.1109/icmss.2009.5301932.
- [56] M. Zamani, P. M. Esfahani, R. Majumdar, A. Abate, and J. Lygeros. Symbolic control of stochastic systems via approximately bisimilar finite abstractions. *IEEE Transactions on Automatic Control*, 59(12):2825–2830, 2014.
- [57] M. Zamani, I. Tkachev, and A. Abate. Towards scalable synthesis of stochastic control systems. *Discrete Event Dynamic Systems*, 27(2):341–369, 2017.
- [58] X.-C. Zeng and S.-P. Zhu. A new simple tree approach for the heston’s stochastic volatility model. *Computers & Mathematics with Applications*, 78(6):1993–2010, sep 2019. doi: 10.1016/j.camwa.2019.03.030.

A Appendix - Lemmas

Lemma 2. *If $\mathbb{E}|X_i^n|^q \leq c(q, T, X_0)$, then Lemma 1-Item (1) implies Equation (1).*

Proof. Let us recall [30] that if the coefficients $\mu, \sigma \in C_b^4$, we have that for any $x \in \mathbb{R}^d$ and for some $q \geq 0$

$$\begin{aligned}\mathbb{E}[X_{i+1} - X_i | X_i = x] - n^{-1}\mu(x) &= O\left((1 + |x|^q)n^{-2}\right), \\ \mathbb{E}[(X_{i+1} - X_i)^{\otimes 2} | X_i = x] - n^{-2}\mu(x)^{\otimes 2} + n^{-1}\Sigma(x) &= O\left((1 + |x|^q)n^{-2}\right),\end{aligned}$$

in particular the second inequality implies that

$$\mathbb{E}[(X_{i+1} - X_i)^{\otimes 2} | X_i = x] - n^{-1}\Sigma(x) = O\left((1 + |x|^q)n^{-2}\right)$$

From Lemma 1-Item (1), dividing by n^{-1} , we obtain that

$$\begin{aligned}\mathbb{E}\left[\frac{X_{i+1}^n - X_i^n}{n^{-1}} \middle| X_i^n\right] - \mu(X_i^n) &= O\left((1 + |X_i^n|^q)n^{-\alpha}\right), \\ \mathbb{E}\left[\frac{(X_{i+1}^n - X_i^n)^{\otimes 2}}{n^{-1}} \middle| X_i^n\right] - \Sigma(X_i^n) &= O\left((1 + |X_i^n|^q)n^{-\alpha}\right).\end{aligned}$$

Taking the square on both sides and then the expectation, using that $\mathbb{E}|X_i^n|^q \leq c(q, T, X_0)$, we obtain the thesis. \square

Lemma 3. *If X^n represents the schemes built in Theorem 4, 8, 10 and $n\gamma_n^2 < c$ for some $c > 0$, then*

$$\mathbb{E}[\max_i |X_i^n|^{2q}] \leq c(1 + |X_0|^{2q}), \quad \text{for } q \geq 1.$$

Proof. Let us first note that

$$X_{i+1}^n - X_i^n | X_i^n = x \sim Y_x$$

In Theorems 4, 8 and 10 we have proved that if the SDE coefficients are bounded the r.v. Y_x , can be built such that $|Y_x| \leq c_{\mu, \sigma}n^{-1/2} + k\gamma_n$, whilst if the coefficients are linearly bounded then the r.v. Y_x , can be built s.t. $|Y_x| \leq c_{\mu, \sigma}(1 + |x|)n^{-1/2} + k\gamma_n$, for some $k > 0$. Since we are interested in the finite time horizon we suppose for simplicity $T = 1$, which means $i \leq n$.

Let us define $\mathcal{Y}_i^n = \sum_{j=0}^{i-1} Y_{X_j^n} - n^{-1}\mu(X_j^n)$. By construction, to satisfy Item Lemma 1-(1), we build the r.v. $Y_{(\cdot)}$ s.t. $\mathbb{E}_j Y_{X_j^n} = n^{-1}\mu(X_j^n)$, which implies that \mathcal{Y}_i^n is a martingale. Thus, we can apply the Burkholder-Davis-Gundy (BDG) Inequality

$$\begin{aligned}\mathbb{E}\left[\max_{i \leq n} |\mathcal{Y}_i^n|^{2q}\right] &\leq c_q \mathbb{E}\langle \mathcal{Y}_n^n \rangle^q \leq c_q \mathbb{E}\left[\sum_{j=0}^{n-1} |Y_{X_j^n} - n^{-1}\mu(X_j^n)|^2\right]^q \\ &\leq c_q n^{q-1} \sum_{j=0}^{n-1} \mathbb{E}\left|Y_{X_j^n} - n^{-1}\mu(X_j^n)\right|^{2q} \\ &\leq c_q n^{q-1} \left[\sum_{j=0}^{n-1} c_{\mu, \sigma}(1 + \mathbb{E}|X_j^n|^{2q})n^{-q} + k\gamma_n^{2q}\right] \\ &\leq c_q n^{q-1} \left[nk\gamma_n^{2q} + c_{\mu, \sigma}n^{-q}n + c_{\mu, \sigma}n^{-q} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q}\right]\end{aligned}$$

$$\begin{aligned}
&\leq c_q (n\gamma_n^2)^q + c_{q,\mu,\sigma} + c_{q,\mu,\sigma} n^{-1} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q} \\
&\leq c_{\mu,\sigma,q} \left(1 + n^{-1} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q} \right),
\end{aligned}$$

where at the 4-th step (3-rd line) we have used the fact that $|Y_{X_j^n}| \leq c_{\mu,\sigma}(1 + |X_j^n|)n^{-1/2} + k\gamma_n$ and $\mathbb{E}Y_{X_j^n} = n^{-1}\mu(X_j^n)$, so $n^{-1}\mu(X_j^n) \in \text{Convex Hull}\{\text{supp}(Y_{X_j^n})\}$; at the last step we have used the assumption that $n\gamma_n^2 < c$, for some $c > 0$.

We can now proceed to bound $\mathbb{E} \left[\max_{i \leq n} |X_i^n|^{2q} \right]$

$$\begin{aligned}
\mathbb{E} \left[\max_{i \leq n} |X_i^n|^{2q} \right] &= \mathbb{E} \left[\max_{i \leq n} \left| X_0 + \sum_{j=0}^{i-1} Y_{X_j^n} \right|^{2q} \right] \\
&= \mathbb{E} \left[\max_{i \leq n} \left| X_0 + \sum_{j=0}^{i-1} n^{-1}\mu(X_j^n) + \sum_{j=0}^{i-1} Y_{X_j^n} - n^{-1}\mu(X_j^n) \right|^{2q} \right] \\
&\leq c_q \left\{ |X_0|^{2q} + \mathbb{E} \left[\max_{i \leq n} \left| n^{-1} \sum_{j=0}^{i-1} \mu(X_j^n) \right|^{2q} \right] + \mathbb{E} \left[\max_{i \leq n} \left| \sum_{j=0}^{i-1} Y_{X_j^n} - n^{-1}\mu(X_j^n) \right|^{2q} \right] \right\} \\
&\leq c_q \left\{ |X_0|^{2q} + n^{-2q} n^{2q-1} \sum_{j=0}^{n-1} \mathbb{E} \left| \mu(X_j^n) \right|^{2q} + \mathbb{E} \left[\max_{i \leq n} |\mathcal{Y}_i^n|^{2q} \right] \right\} \\
&\leq c_q \left\{ |X_0|^{2q} + n^{-1} \sum_{j=0}^{n-1} c_{q,\mu,\sigma} \left(1 + \mathbb{E}|X_j^n|^{2q} \right) + c_{\mu,\sigma,q} + c_{\mu,\sigma,q} n^{q-1} n^{-q} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q} \right\} \\
&\leq c_q \left\{ |X_0|^{2q} + c_{q,\mu,\sigma} n^{-1} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q} + 2c_{q,\mu,\sigma} + c_{q,\mu,\sigma} n^{-1} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q} \right\} \\
&\leq c_{q,\mu,\sigma} \left\{ |X_0|^{2q} + 1 + n^{-1} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q} \right\}.
\end{aligned}$$

It remains now to bound $\mathbb{E}|X_i^n|^{2q}$: applying the discrete Grönwall's Lemma 4, from the previous equation we know that

$$\begin{aligned}
\mathbb{E}|X_i^n|^{2q} &\leq \mathbb{E} \left[\max_{1 \leq i \leq n} |X_i^n|^{2q} \right] \\
&\leq c_{q,\mu,\sigma} \left\{ |X_0|^{2q} + 1 + n^{-1} \sum_{j=0}^{n-1} \mathbb{E}|X_j^n|^{2q} \right\} \\
&\leq c_{q,\mu,\sigma} \left\{ |X_0|^{2q} + 1 \right\} + c_{q,\mu,\sigma} \left\{ |X_0|^{2q} + 1 \right\} \sum_{j=0}^{i-1} n^{-1} \exp \left\{ \sum_{j=0}^{n-1} n^{-1} \right\}
\end{aligned}$$

$$\leq c_{q,\mu,\sigma} \left\{ |X_0|^{2q} + 1 \right\} (1 + e).$$

To conclude the proof we can notice that using the Jensen Inequality and the sub-additivity of the root we have that

$$\begin{aligned} \mathbb{E} \left[\max_{i \leq n} |X_i^n|^q \right] &\leq \sqrt{\mathbb{E} \left[\max_{i \leq n} |X_i^n|^{2q} \right]} \\ &\leq \sqrt{c_{q,\mu,\sigma} \left\{ |X_0|^{2q} + 1 \right\} (1 + e)} \leq c_{q,\mu,\sigma} \left\{ |X_0|^q + 1 \right\} (1 + e). \end{aligned}$$

□

We state here the discrete Grönwall's Lemma that we have used in some previous proofs of this work.

Lemma 4. [Discrete Grönwall's Lemma [26]] Let $y_n, f_n,$ and g_n be non-negative sequences such that for $n > 0, y_n \leq f_n + \sum_{k=0}^n g_k y_k,$ then

$$y_n \leq f_n + \sum_{k=0}^n f_k g_k \exp \left(\sum_{j=k}^n g_j \right).$$

B Appendix - Details on Experiments

In both the experiments we decided on the following strategy

- (a) try to solve (2) using Algorithm 2 in [6];
- (b) if after a specified maximal number of iterations, step (a) has not returned a solution, solve the minimization problem (iii) and then we apply Algorithm 3 in [6], given the sequence of weights solving (iii) and so (2). Remember that Algorithm 3 in [6] is a combination of the Algorithms in [53] and [6].

The results in [6] can also help to reduce further the cardinality of the state space. [6, Theorem 3, item 3] tells us that given a solution of (2) we could possibly modify it changing the points (and therefore the weights) if the new points satisfy a specified condition. The idea is thus to search if there are points $|\bar{l}_j| \leq c, \bar{l}_j \in \mathcal{Y}\mathbb{Z}^2$ such that \bar{l}_j satisfies the condition explicited in [6, Theorem 3, item 3] and \bar{l}_j have been already used in the past - indeed let us recall that we build the tree recursively starting from the initial point. Briefly, we indicate $\{l_j^*, \{w_j^*\}$ a solution of (2), when the Caratheodory's reduction has already taken place, and $C(\{\cdot\})$ indicates the cone built from the points $\{\cdot\}$. [6, Theorem 3, item 3] claims that any point into the cone $C(\{-l_j^*\} \setminus -l_1^*)$ can be exchanged with l_1^* to have a new solution to (2). Namely, considering that we want to find a solution in $\mathcal{Y}\mathbb{Z}^2,$ we can look for any point \bar{l} in

$$\{|l| < c, l \in \mathcal{Y}\mathbb{Z}^2\} \cap \{l \in \mathcal{Y}\mathbb{Z}^2 : x + l \text{ has been used to build previous r.v.}\} \cap C(\{-l_j^*\} \setminus -l_1^*).$$

The points $\bar{l} \cup \{l_j^*\} \setminus l_1^*$ constitute another solution to (2), changing accordingly the weights. This can be generalized for any $l_i^*,$ not only l_1^* and can be iteratively done for all of them. This procedure can help us to reduce further the total number of states necessary to approximate the considered SDE, being able to re-use points already used, and therefore increasing the “level of recombination” of the tree. As a drawback, this procedure leads to a higher computational cost.