Abstract

Real-world multistage stochastic optimization problems are often characterized by the fact that the decision maker may take actions only at specific points in time, even if relevant data can be observed much more frequently. In such a case there are not only multiple decision stages present but also several observation periods between consecutive decisions where profits/costs occur contingent on the stochastic evolution of the uncertainty factors. We refer to such multistage decision problems with encapsulated multiperiod random costs, as multiscale stochastic optimization problems. In this article we present a tailor-made data structure for the numerical solution of such problems. We first propose a new method for the generation of scenario lattices and then incorporate the multiscale feature by leveraging the theory of stochastic bridge processes. All necessary ingredients to our proposed framework are elaborated explicitly for various popular modeling choices, including both diffusion and jump models.

1 Introduction

Optimization models over a larger time period can be classified into two types:

- Multiperiod models: The decisions are made at the very beginning whereas the consequences of the decisions depend on the development of a process over time. A typical example is a buy-and-hold portfolio strategy.
- Multistage models: Decisions can be made at regular moments in time. Typical examples are active portfolio strategies.

The purpose of this paper is to introduce multistage models, which incorporate the properties of both, multiperiod models (for the development between decision stages) and multistage models. Stochastic multiperiod models are simple from their structure. In contrast, multistage stochastic models are objects of intensive research. The book of Pflug and Pichler [28] gives a general overview on this topic. As mentioned, the model classes studied in the present article contain multiple decision stages but also multiperiod processes on a finer scale between the decision stages. Examples for such problems include:

- Department of Statistics and Operations Research, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria
- International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
• Supply network extension problems, where major decisions (such as whether to defer, to stage, to mothball, or to abandon a certain infrastructure investment opportunity; cf. [25]) can only be made at strategic time points (say, once every few years), but resulting profits/costs are subject to daily fluctuations of market prices.

• Inventory control problems with limited storage capacity and backlogged/lost demand due to out-of-stock events, where procurement of goods is restricted by logistical constraints/time efforts.

• Structured portfolio investment problems, where rebalancing is possible only at given time points (say, once every few weeks due to product terms and conditions), but contained barrier features make profits/losses depend on the full trajectory of asset prices.

• Power plant management problems, where operating plans need to be fixed for a certain period ahead (say, once every few days due to physical constraints avoiding instant reaction to market conditions), but actual profits/losses depend on each tick of the energy market.

To the best of our knowledge, the existing literature does not offer a computational framework designed specifically towards the solution of such multistage stochastic optimization problems, where different time scales related to one and the same underlying stochastic process are present.

A solution approach for a class of problems which are of a similar flavour, yet crucially different in nature, is called multi-horizon stochastic programming. This idea was originally proposed by Kaut et al. [18]. In a series of papers by Tomasgard and various coauthors, the multi-horizon approach has then been used for applications in the modeling of energy markets [36, 37, 42], as well as investigated from a theoretical perspective [24, 40]. Infrastructure planning problems, being the original motivation, typically involve (rarely happening) strategic decisions as well as operational tasks (daily business). To overcome the memory issue resulting from frequently branching scenario trees, the authors of [18] suggest to start with a tree for the strategic scale only. In a second step, they attach another tree to each node of the strategic-scale tree. The key assumption for the multi-horizon stochastic programming approach to work is that the strategic scale and all operational scales are independent from each other. In contrast, the approach suggested in the present paper ensures that different scenarios in between consecutive decisions are eventually bundled in one node. This is achieved by leveraging the theory of stochastic bridges. We consider this as the strength of our model, that it respects the stochasticity in between decisions but keeps the problem computationally tractable.

The approach suggested in this article consists of two parts, each dealing with one of the two time scales. Neglecting for a moment the presence of the finer observation scale and looking only at the coarser decision scale, the requirements to the discrete structure are the same as for any standard multistage stochastic optimization problem. Due to the exponential growth of scenario trees, their usage is computationally not feasible for problems with many decision stages.¹ Therefore, one typically reverts to scenario lattices in such cases.

¹For instance, on a binary tree (which typically will be a rather poor uncertainty model), hourly decisions for a time horizon of one day will correspond to about 17 million nodes, daily
However, while the literature on the construction of scenario trees is relatively rich (see, e.g., [14, 15, 20, 28, 29]), the lattice construction literature is rather sparse. Bally and Pagès [3] originally suggested an optimal quantization algorithm for simulatable Markov chains. We consider such an approach based on the minimization of a distance measure (between the targeted distribution and its discretization) as the standard and also state-of-the-art approach. A review on approximation methods for probability distributions, including optimal quantization algorithms, can be found in the book of Pflug and Pichler [28, Ch. 4.1]. Löündorf and Wozabal [22] combine optimal quantization with moment matching and thereby obtain a scenario lattice with conditional expectations matching those of the underlying continuous state process.

Typically, the generation of scenarios out of a sample of observed data works in the following two-step procedure:

- **Data**
- **Modelling & Estimation**
- **Simulation/Quantization**

In this article, we suggest a lattice generation method for the case when a time-homogeneous Markovian diffusion model is selected in the first step. The approach is purely based on the infinitesimal drift and diffusion coefficient functions of the model and directly provides a scenario lattice, without requiring a simulation/quantization procedure. While the suggested technique is in fact not novel but rather well-known in the Markov processes literature (see, e.g., the classical book of Karlin and Taylor [17, Chp. 15]), it seems that these discretization concepts have not been studied yet in the stochastic optimization literature (cf. the review article of Löündorf [21]). Our main contribution in this respect is a stability result for the optimal value of a generic multistage stochastic optimization problem. It is shown that the approximation error w.r.t. the optimal value in the continuous (state space) diffusion model can be controlled when the suggested lattice generation method is applied.

A coherent approach for the finer scale requires an interpolation which respects the laws of the stochastic process discretized by the given scenario lattice. This brings us to the theory of stochastic bridges, i.e., processes pinned to given values at the beginning and at the end of some interval of time. While the Brownian bridge is typically addressed in any textbook on stochastic processes and commonly applied to improve the efficiency of Monte-Carlo simulations, the stochastic processes literature on more general stochastic bridges seems rather sparse. A seminal work in this respect is the work of Fitzsimmons et al. [12]. In the words of the authors of [12], it is part of Markovian folklore that the right way to define bridges in some generality is by a suitable Doob $h$-transform of the space-time process. We orientate ourselves by their guidelines and the derivations in [23] to establish the dynamics of a general multi-dimensional diffusion process conditioned on the prescribed value of its solution at some future time. These dynamics will involve the gradient of the transition density, which is in general not available in closed form. However, in special cases which include some rather popular modeling choices, the bridge process dynamics simplify to a tractable form. Then, efficient simulation is possible by means of standard

decisions for one month will give about 1 billion of nodes, and weekly decisions for one year will result in a magnitude of $10^{15}$ nodes.
discretization schemes for stochastic differential equations. We elaborate several such special cases explicitly. Moreover, we present a simulation scheme for the example of geometric Brownian motion (GBM) which operates directly on generated paths from the unconditioned process and thus enables an even more efficient generation of bridge process trajectories. If the cost function is particularly amenable (e.g., linear), a simulation might not even be required, as expected costs can be computed analytically in some models.

There are some articles on simulation methods (i.e., mainly acceptance-rejection methods) for general diffusion bridges in the statistical analysis literature since the early 2000’s, for example the work of Papaspiliopoulos and Roberts [26], or Bladt and Sørensen [7, 8] on the simulation of diffusion bridges, as well as Gonçalves and Roberts [13] or Pollock [32] for the extension to Monte Carlo methods for jump-diffusion bridges. We also include jump processes in our analysis, as we propose a simulation algorithm for compound Poisson bridges in the case of Normally, Exponentially, or Gamma distributed jump sizes. In particular, we derive the conditional distribution of each jump-size given both the final value of the process as well as the number of jumps in the interval.

The general contribution of this article is to propose a scenario generation method which allows for a computational approach to numerically solve multi-scale stochastic optimization problems. To be more precise, the contribution of the paper is twofold. First, it comprises the presentation and analysis of a new method for the construction of scenario lattices when the underlying model is of the diffusion type; this is purely related to the coarser decision time scale. Second, as regards the interpolation on the finer scale, we elaborate the details for a number of popular modeling choices. We treat several diffusion models and touch upon jump processes by suggesting a method to simulate from compound Poisson bridges with three different jump-size distributions. The problem class and the suggested approach are illustrated in Fig. 1.

The outline of the paper is as follows. Section 2 deals with the construction of scenario lattices as a model for the information flow over the decision stages. In Sect. 3 we present the details on the interpolation of the information flow through the intermediate observation points. Section 4 illustrates the suggested approach with a simple illustrative multiscale inventory control problem. We conclude in Sect. 5.

2 Scenario lattice generation for decision stages

Numerical solutions of stochastic optimization problems require discrete structures. For multistage problems, one typically uses scenario trees to model the evolution of uncertainty over time. Scenario trees allow for general path-dependent solutions, as for each node there exists a unique path to the root of the tree. However, scenario trees grow exponentially in the number of stages, a fact that easily overwhelms any computer’s memory. Therefore, in the case of a “sufficiently Markovian” decision structure one typically reverts to a special case of scenario trees, namely scenario lattices. For the latter structure, a node does not necessarily have a unique predecessor but different paths may coincide in a certain node of some stage. In this way, one can obtain a rich set of paths with relatively few nodes.

\[\text{See, e.g., the book of Kloeden and Platen [19] for a detailed treatment.}\]
Figure 1: Multiscale stochastic optimization problems: each decision stage involves multiple observation periods, where actual costs resulting from the proceeding decision are realized. The objective depends on the whole trajectory, not only on the value of the process at the decision points.

Left: A lattice model for the decision stages - node values and transition probabilities are estimated. Probabilities are indicated by the different sizes of the nodes. Right: Simulation of the interpolating bridge process between consecutive decision stages to determine realized costs.

The construction of scenario lattices typically works in a two-step procedure. First, one discretizes the marginal distributions for all stages. In a second step, one decides about allowed state transitions and determines conditional transition probabilities in between consecutive stages. The state-of-the-art method for such a lattice generation procedure is based on the (stagewise) minimization of the Wasserstein distance between the (modeled) underlying distribution – which is typically continuous – and its discretization on the lattice. A detailed description of this approach can be found in [22, Sect. 3.2].

We will now propose an alternative lattice generation approach, which is not based on optimal quantization theory but rather relies on Markov chain approximation results.

2.1 Markov chain approximation for diffusion processes

*Birth-and-death Markov chains* are discrete stochastic processes defined on the integer grid, where each transition depends only on the current state and allows for three possibilities: to remain in the current state, to move one up, or to move one down. Many Markov chains can be approximated by a diffusion process. It works by a transformation of the time scale and a renormalization of the state variable. The idea is, e.g., explained in the book of Karlin and Taylor [17, Ch. 15]. Pflug and Swietanowski [30] have looked at the problem from the converse perspective. They elaborate that any diffusion process which
possesses a stationary distribution can be approximated by a birth-and-death Markov chain in the following way.

Consider a one-dimensional recurrent Markov process \( X_t \), as defined by

\[
\begin{align*}
\frac{dX_t}{dt} &= \mu(X_t) \, dt + \sigma(X_t) \, dW_t \\
X_0 &= x_0
\end{align*}
\]

where \( W \) denotes a standard Brownian motion, and the initial value \( x_0 \) is a given constant. Throughout the paper, the coefficient functions \( \mu(\cdot) \) and \( \sigma(\cdot) \) are (as it is usual) assumed to be square-integrable functions satisfying the following growth conditions:

- \(|\mu(x) - \mu(y)| \leq L \cdot |x - y|,\)
- \(|\sigma(x) - \sigma(y)| \leq L \cdot |x - y|,\)
- \(\mu^2(x) \leq L^2 \cdot (1 + x^2),\)
- \(\sigma^2(x) \leq L^2 \cdot (1 + x^2),\)

for some \( L > 0 \). Notice that the Lipschitz-continuity implies that one may specify a constant \( L_\mu \) such that \( \mu(x) \leq L_\mu + L \cdot |x| \).

**Algorithm 2.1** (Markov chain approximation method for diffusion processes).

For a diffusion process \( X \) as given by (1), define its \( N \)-th Markov chain approximation as the process constructed along the following scheme.

1. Choose a strictly monotonic, three times differentiable function \( H(x) \) with \( H''(0) \leq M < \infty \), for some constant \( M \), as well as functions \( g(x) \) and \( \tau(x) \) with \( |\tau(x)| \leq 1 \) for all \( x \), in such a way that the drift and diffusion coefficient functions in (1) are matched:

\[
\begin{align*}
\mu(H(x)) &= H'(x)g(H(x)) + \frac{1}{2}H''(x)\tau^2(H(x)) \\
\sigma(H(x)) &= H'(x)\tau(H(x)).
\end{align*}
\]

2. Determine the initial state \( i_0 \) such that \( H(i_0) = x_0 \).

3. Define the transition probabilities

\[
\begin{align*}
\pi_u^i, N &= \left[ \frac{1}{2} \left( \tau^2 \left( H \left( \frac{i}{2N} \right) \right) + \frac{1}{2N} g \left( H \left( \frac{i}{2N} \right) \right) \right) \right]_0^1, \\
\pi_d^i, N &= \left[ \frac{1}{2} \left( \tau^2 \left( H \left( \frac{i}{2N} \right) \right) - \frac{1}{2N} g \left( H \left( \frac{i}{2N} \right) \right) \right) \right]_0^1, \\
\pi_r^i, N &= 1 - \pi_u^i, N - \pi_d^i, N,
\end{align*}
\]

where \( \lfloor x \rfloor_0 := \min\{\max\{x, 0\}, 1\} \), for jumping up, down, and remaining in its state, respectively.

4. Define the piecewise constant (continuous time) process \( \tilde{X}_t^N \), where \( \tilde{X}_t^N := \tilde{X}_{\lfloor 2Nt \rfloor}^N \) lives in the states \( H \left( \frac{i}{2N} \right) \); the floor function being denoted by \( \lfloor \cdot \rfloor \).

We now want to show the stability of the optimal solution of a generic multistage stochastic optimization problem, when the underlying diffusion model is approximated by the method of Alg. 2.1 above. Before we are ready to do so, we first provide some required results.
Proposition 2.1. Define the process $X^z: [0,T] \times \Omega \to \mathbb{R}$ as the process $X$ defined in (1) but starting in $z \in \{x,y\}$. Assume that for all $t \in [0,T]$ the condition

$$
\int_0^t |\mu(X^z_s)| + |\sigma(X^z_s)|^2 + \frac{|X^z_s - X^y_s| |\mu(X^z_s) - \mu(X^y_s)| + |\sigma(X^z_s) - \sigma(X^y_s)|^2}{|X^z_s - X^y_s|^p} ds < \infty
$$

(2)

is satisfied a.e. Then, the following stability of the diffusion process with respect to its starting value holds:

$$
\|X^z_T - X^y_T\|_{L^q} \leq |x - y| \cdot \left\| e^{\int_0^T \frac{(x^z_s - x^y_s)(\mu(x^z_s) - \mu(x^y_s)) + \frac{1}{p}(x^z_s - x^y_s)^2}{|x^z_s - x^y_s|^p} ds} \right\|_{L^p} \quad (3)
$$

for any $p$ and $q$ such that $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. See Cox, Hutzenthaler, Jentzen [11, Cor. 2.19] □

Definition 2.1 (GenMSP). Define a generic multistage stochastic optimization problem (GenMSP) to be of the following form:

$$
\inf_x \mathbb{E}^Q \left[ \sum_{t=0}^T C_t (\xi_t, x_t) \right] \quad \text{s.t. } x_t \in X_t \forall t = 0, \ldots, T
$$

(4)

for some convex sets $X_t$ and a scenario process $\xi \in L^1(\mathbb{R}, \mathbb{Q})$, generating the filtration $\sigma(\xi)$. Moreover, assume the cost function $C_t(\cdot, \cdot)$ convex in the decisions (for any fixed scenario), and Lipschitz continuous with constant $L$ w.r.t. the scenario process (for any fixed decision policy). Denote the optimal value of (4), as a function of the underlying probability model, by $v^*(\mathbb{Q})$.

Proposition 2.2. Consider a GenMSP as defined in Def. 2.1 above. Let the distance between two paths $\xi^{(1)}_{0:T}$ and $\xi^{(2)}_{0:T}$ up to time $t \leq T$ be defined by $\|\xi^{(1)}_{0:t} - \xi^{(2)}_{0:t}\| := \sum_{s=0}^t \|\xi^{(1)}_s - \xi^{(2)}_s\|_1$. Let $Q \in \{\hat{\mathbb{P}}, \hat{\mathbb{P}}\}$ for two (d-dimensional) Markovian multistage distributions $\hat{P}$ and $\hat{P}$, both defined on some $\Omega \subseteq \mathbb{R}^{d \times T}$. Assume that, for all $t = 0, \ldots, T-1$, there exist constants $\kappa_t$ and $\varepsilon_t$ such that the Wasserstein distances $\mathcal{W}$ of the corresponding single-stage conditional transition probability measures $P_{t+1}(\cdot | \xi_t)$ and $\hat{P}_{t+1}(\cdot | \xi_t)$ satisfy the conditions

$$
\mathcal{W}
\left( P_{t+1} \left( \cdot | \xi^{(1)}_t \right), \hat{P}_{t+1} \left( \cdot | \xi^{(1)}_t \right) \right) \leq \kappa_t \cdot \|\xi^{(1)}_{0:t} - \xi^{(2)}_{0:t}\|, \quad (5)
$$

$$
\mathcal{W}
\left( P_{t+1} \left( \cdot | \xi^{(2)}_t \right), \hat{P}_{t+1} \left( \cdot | \xi^{(2)}_t \right) \right) \leq \varepsilon_t, \quad (6)
$$

uniformly for all paths $\xi^{(1)}_{0:T}, \xi^{(2)}_{0:T}$. Then, the following upper bound for the difference between the optimal values $v^*(\hat{\mathbb{P}})$ and $v^*(\hat{\mathbb{P}})$ holds:

$$
\left| v^* (\hat{\mathbb{P}}) - v^* (\hat{\mathbb{P}}) \right| \leq L \cdot \sum_{t=0}^T \varepsilon_t \prod_{s=t+1}^T (1 + \kappa_s). \quad (7)
$$

The definition of the Wasserstein distance can be found in the Appendix.
Proof. Follows immediately from [27, Thm. 6.1] and [28, Lem. 4.27].

**Lemma 2.1.** Consider a probability measure $P$ and a sequence of probability measures $(P_n)$ on the compact interval $[-K, K]$, for some $K > 0$. Then, weak convergence implies convergence in Wasserstein distance:

$$ P_n \xrightarrow{w} P \implies \mathcal{W}(P_n, P) \rightarrow 0. $$

**Proof.** By Billingsley and Topsøe [6], the weak convergence $P_n \xrightarrow{w} P$ implies the following equivalence: $\sup_{g \in G} |\int g dP_n - \int g dP| \rightarrow 0$ holds if and only if both

$$ \sup_{g \in G} \sup_{x,y} |g(x) - g(y)| < \infty \quad (8) $$

and

$$ \lim_{\delta \rightarrow 0} \sup_{g \in G} P\{x : \sup_{|x-y| \leq \delta} |g(x) - g(y)| > \varepsilon\} \rightarrow 0 \quad \forall \varepsilon > 0 \quad (9) $$

hold. Notice that $G = \{g \in \text{Lip}(1) : g(0) = 0\}$ on $[-K, K]$ fulfills (8) and (9). As $P_n$ and $P$ have bounded support, it holds $\mathcal{W}(P_n, P) = \sup_{g \in \text{Lip}(1)} \int g \, dP_n - \int g \, dP$. Thus, it follows $\mathcal{W}(P_n, P) \rightarrow 0$.

**Theorem 2.1.** Consider a continuous probability measure $P$ and a sequence of probability measures $(P_n)$. Suppose that the conditions

$$ \int x^2 \, dP \leq M \quad (10) $$

and

$$ \int x^2 \, dP_n \leq M \quad \forall n \quad (11) $$

hold, for some constant $M < \infty$. Then, weak convergence implies convergence in Wasserstein distance:

$$ P_n \xrightarrow{w} P \implies \mathcal{W}(P_n, P) \rightarrow 0. $$

**Proof.** Denote the cdf of $P_n$ by $F_n$ and that of $P$ by $F$. Notice that, by a version of Čebyšev’s inequality, for $K > 0$ it holds that

$$ \int_{-\infty}^{-K} F_n(x) \, dx = \int_{-\infty}^{-K} P_n(-\infty, x] \, dx \leq \int_{-\infty}^{-K} \frac{M}{x^2} \, dx = \frac{M}{K}, \quad (12) $$

and similarly

$$ \int_{-\infty}^{-K} F(x) \, dx \leq \frac{M}{K}, \quad \int_{K}^{\infty} (1 - F_n(x)) \, dx \leq \frac{M}{K}, \quad \int_{K}^{\infty} (1 - F(x)) \, dx \leq \frac{M}{K}. \quad (13) $$

Now choose $K$ large enough such that $\frac{M}{K} \leq \varepsilon$ holds. Then,

$$ \int_{-\infty}^{-K} |F_n(x) - F(x)| \, dx \leq \int_{-\infty}^{-K} F_n(x) \, dx + \int_{-\infty}^{-K} F(x) \, dx \leq 2\varepsilon, $$

8
and
\[ \int_{K}^{\infty} |F_n(x) - F(x)| \, dx \leq \int_{K}^{\infty} (1 - F_n(x)) \, dx + \int_{K}^{\infty} (1 - F(x)) \, dx \leq 2\varepsilon. \]

Define the probability measure \( P^K_n \) as \( P_n \) conditioned on the interval \([-K,K]\), where we know \( P_n([-K,K]) \geq 1 - 2\varepsilon \) by (12) and (13). Define \( P^K \) analogously.

By Lemma 2.1, it holds
\[ \int |F^K_n(x) - F^K(x)| \, dx \to 0, \]
as \( n \to \infty \). Let \( c_n := F_n(K) - F_n(-K) \) and \( c := F(K) - F(-K) \). Since \( F \) is continuous, \( c_n \to c \). Using
\[ F^K_n(x) = \frac{F_n(x) - F_n(-K)}{c_n} \quad \text{and} \quad F^K(x) = \frac{F(x) - F(-K)}{c}, \]
we get
\[ \int_{-K}^{K} |F^K_n(x) - F^K(x)| \, dx \leq \int_{-K}^{K} |c_n F^K_n(x) - c F^K(x)| \, dx + 2K|F_n(-K) - F(-K)| \]
\[ \leq c \int_{-K}^{K} |F^K_n(x) - F^K(x)| \, dx + 2K|c_n - c| + 2K|F_n(-K) - F(-K)|. \]

Now, for any \( \varepsilon > 0 \), we can make \( n \) such large that
\[ 2K|c_n - c| \leq \varepsilon, \quad 2K|F_n(-K) - F(-K)| \leq \varepsilon, \quad c \int_{-K}^{K} |F^K_n(x) - F^K(x)| \, dx \leq \varepsilon. \]

Then, in total, we get \( \int_{-K}^{K} |F^K_n(x) - F^K(x)| \, dx \leq 3\varepsilon \) and finally
\[ \mathbb{W}(P_n, P) = \int |F_n(x) - F(x)| \, dx \leq 7\varepsilon. \]

\[ \square \]

Lemma 2.2. Let \( \tilde{X}_t^N \) be constructed according to Alg. 2.1 and starting in \( x_0 \) at time \( t_0 \). Then, for \( t_1 \geq t_0 \), the following bound for the second moment holds:
\[ \mathbb{E} \left[ \left( \tilde{X}_{t_1}^N \right)^2 \right] \leq x_0^2 \cdot e^{K_1(t_1-t_0)} + o \left( K_2 \cdot (t_1 - t_0) \cdot e^{K_1(t_1-t_0)} \right), \]
where \( K_1, K_2 \in \mathbb{R} \) depend only on the Lipschitz(-like) constants controlling the growth of the coefficient functions in (1).
Proof. The conditional expected increment of the squared process is given by

\[
\mathbb{E}\left[ (\tilde{X}_{n+1}^N)^2 - (\tilde{X}_n^N)^2 \right] = H^2\left( \frac{i}{2N} \right) - H^2\left( \frac{i-1}{2N} \right) \cdot p_i^{(N)} + \left[ H^2\left( \frac{i}{2N} \right) - H^2\left( \frac{i-1}{2N} \right) \right] \cdot q_i^{(N)}
\]

Applying this iteration scheme, we finally obtain (using the shorthand notation \(\tilde{X}_0^N\)) the targeted estimate

\[
\mathbb{E}\left[ (\tilde{X}_{n+1}^N)^2 - (\tilde{X}_n^N)^2 \right] \leq \frac{1}{22N} K_1 H^2\left( \frac{i}{2N} \right) + o\left( \frac{K_2}{22N} \right),
\]

where \(K_1 := 2L_\mu + 2L + L^2\) and \(K_2 := L^2 + 2L_\mu\). Then, by the tower property of the expected value, we get

\[
\mathbb{E}\left[ (\tilde{X}_{n+1}^N)^2 \right] \leq \mathbb{E}\left[ (\tilde{X}_n^N)^2 \right] \cdot \left( 1 + \frac{K_1}{22N} \right) + o\left( \frac{K_2}{22N} \right).
\]

Applying this iteration scheme, we finally obtain (using the shorthand notation \(\tilde{N} := 2^N(t_1 - t_0)\)) the targeted estimate

\[
\mathbb{E}\left[ (\tilde{X}_{n+1}^N)^2 \right] \leq \mathbb{E}\left[ (\tilde{X}_n^N)^2 \right] \cdot \left( 1 + \frac{K_1}{22N} \right)^{\tilde{N}} + o\left( \frac{K_2}{22N} \right) \sum_{i=0}^{\tilde{N}-1} \left( 1 + \frac{K_1}{22N} \right)^i
\]

Proposition 2.3. For any even integer \(p \geq 2\), the moments of the solution of (1) satisfy the following estimate:

\[
\mathbb{E}\left[ X_p^t \right] \leq (1 + x_0^p) e^{(p+1)L^2t}
\]

Proof. See the book of Platen and Heath [31, Lem. 7.8.1].
Theorem 2.2. Consider a GenMSP according to Def. 2.1. Let the uncertainty process \( \xi \) be modelled by a diffusion according to (1). Assume that the coefficient functions satisfy the regularity condition (2). Observe \( \xi \) in all decision stages \( t = 0, \ldots, T \) of GenMSP and denote the resulting discrete-time continuous state-space model by \( \mathbb{P} \). Let \( \xi \) be discretized according to the Markov chain approximation method given in Alg. 2.1 and denote the discrete model resulting from the \( N \)-th approximation by \( \mathbb{P}^N \). Then, the optimal value \( v^*(\mathbb{P}) \) of the approximate problem tends to the optimal value \( v^*(\mathbb{P}) \) of the original problem, as \( N \to \infty \). For fixed \( N \), a stability estimate of the form (7) holds.

Proof. We want to show that \( \mathbb{P} \) and \( \mathbb{P}^N \) satisfy the conditions (5) and (6), with \( \varepsilon_t \downarrow 0 \) as \( N \) increases. Then, the statement follows readily from Prop. 2.2.

The diffusion model satisfies condition (5) by Prop. 2.1. Moreover, since for \( N \) large enough it holds

\[
E \left[ \hat{X}_{n+1}^{(N)} - \hat{X}_n^{(N)} \left| H \left( \frac{i}{2N} \right) \right. \right] \\
= \left[ H \left( \frac{i+1}{2N} \right) - H \left( \frac{i}{2N} \right) \right] \cdot p_i^{(N)} + \left[ H \left( \frac{i}{2N} \right) - H \left( \frac{i-1}{2N} \right) \right] \cdot q_i^{(N)} \\
= \frac{1}{2N} H' \left( \frac{i}{2N} \right) \cdot H \left( \frac{i}{2N} \right) + \frac{1}{2N^2} H'' \left( \frac{i}{2N} \right) \cdot \frac{1}{2N} + o \left( \frac{1}{2N^2} \right),
\]

as well as

\[
E \left[ \left( \hat{X}_{n+1}^{(N)} - \hat{X}_n^{(N)} \right)^2 \left| H \left( \frac{i}{2N} \right) \right. \right] \\
= \left[ H \left( \frac{i+1}{2N} \right) - H \left( \frac{i}{2N} \right) \right]^2 p_i^{(N)} + \left[ H \left( \frac{i}{2N} \right) - H \left( \frac{i-1}{2N} \right) \right]^2 q_i^{(N)} \\
= \frac{1}{2N} \sigma^2 \left( H \left( \frac{i}{2N} \right) \right) + o \left( \frac{1}{2N^2} \right),
\]

and

\[
E \left[ \left( \hat{X}_{n+1}^{(N)} - \hat{X}_n^{(N)} \right)^4 \left| H \left( \frac{i}{2N} \right) \right. \right] = o \left( \frac{1}{2N^2} \right),
\]

it follows the convergence of the finite dimensional distributions

\[
\left( \hat{X}_{2Nt_1}^{N}, \hat{X}_{2Nt_2}^{N}, \ldots, \hat{X}_{2NT}^{N} \right) \xrightarrow{D} \left( X_{t_1}, X_{t_2}, \ldots, X_T \right),
\]

see [17, pg.169]. Since we have constructed the lattice in such a way, that each atom of the distribution of \( X_t^{N} \) is also an atom of the distribution of \( X_t^{M} \), for all \( M \geq N \), it follows also the weak convergence of all conditional probabilities. By Thm. 2.1, the latter implies convergence in Wasserstein distance, as the conditions (10) and (11) hold by Prop. 2.3 and Lem. 2.2, respectively.

Thus, condition (6) is shown to be satisfied.

Remark. The rescaling of time was necessary in the construction of Alg. 2.1 in order for Thm. 2.2 to hold. However, notice that the method in essence specifies a ternary transition rule. While blindly using the directly resulting ternary lattice

11
would not rely on any supporting theory, it might still be interesting to test
its performance, especially for problems with multiple observation periods but
relatively few decisions.

3 Interpolating bridge processes

3.1 Diffusion processes

Let us start with a generic multi-dimensional model with drift and multiple
factors. Afterwards, we derive the bridge process dynamics explicitly for several
special cases that are frequently used in the literature.

Theorem 3.1. Let \( X \) be a \( d \)-dimensional \( n \)-factor diffusion model, i.e.,
\[
\begin{align*}
    dX_t &= \mu(X_t) \, dt + \Sigma(X_t) \, dW(t),
\end{align*}
\]
where \( \mu(\cdot) : \mathbb{R}^n \to \mathbb{R}, \sigma(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n \times m} \), and \( W \) is an \( m \)-dimensional Brownian
motion. Then, for \( t \in [t_1, t_2] \), the dynamics of \( X \) conditioned on both, the
starting value \( x_1 \) at time \( t_1 \) and the final value \( x_2 \) at time \( t_2 \), are given by
\[
    d\hat{X}_t = \left( \mu(\hat{X}_t) + \left( \Sigma \Sigma^\top (\hat{X}_t) \right) \nabla_x \log f_{t_2}(x_2|\hat{X}_t, t) \right) dt + \Sigma(\hat{X}_t) \, dW(t),
\]
where \( f_{t_2} \) denotes the transition density of \( X \) at time \( t_2 \).

Proof. The proof can be found in the Appendix.

We subsequently focus on the one-dimensional case. Let \( X_{t_2} = x_2 \) be fixed
for all examples below.

General state-dependent parameters For a general 1D diffusion process
\( X \) described by the SDE
\[
    dX_t = \mu(X_t) \, dt + \sigma(X_t) \, dW_t,
\]
the dynamics of the associated bridge process are given by
\[
    d\hat{X}_t = \left( \mu(\hat{X}_t) + \sigma^2(\hat{X}_t) \frac{\partial}{\partial x} \log f_{t_2}(x_2|\hat{X}_t, t) \right) dt + \sigma(\hat{X}_t) \, dW_t,
\]
where \( f_t \) denotes the transition density of \( X \) at time \( t \).

Vašíček model / Ornstein–Uhlenbeck process The model presented in
Vašíček [39] is considered as the first stochastic model for the term structure of
interest rates. It is a one-factor model for the short rate, featuring mean rever-
sion. In Vašíček’s model, the instantaneous rate \( r \) is described by a Gaussian
Ornstein-Uhlenbeck process, i.e., as the solution of the SDE
\[
    dr_t = \kappa(\theta - r_t) \, dt + \sigma \, dW_t,
\]
process dynamics for the CIR model are given by Geometric Brownian motion (GBM) and is typically referred to as CIR model. The square root diffusion process is a cumbersome object but yet an analytic expression. Thus, the transition density of $r$ is that of a Normal distribution, i.e.

$$f_t(x|r_x, s) = \frac{1}{\sqrt{2\pi v(t-s)}} \exp \left( -\frac{(x - \theta + (\theta - r_s)e^{-\kappa(t-s)})^2}{2v(t-s)} \right),$$

where $v(\Delta t) := \frac{\sigma^2(1-e^{-2a\Delta t})}{2a}$. Hence, the derivative of the logarithmized transition density is a closed form expression and the bridge process $\hat{r}$ associated with $r$, pinned to the value $r_{t_2} = x_2$, is described by the dynamics

$$d\hat{r}_t = \left( \kappa(\theta - \hat{r}_t) + \frac{2\kappa (x_2 - \theta + e^{-\kappa(t-t_2)}(\theta - \hat{r}_t))}{1 - e^{-2\kappa(t-t_2)}} \right) dt + \sigma dW_t. \quad (16)$$

Cox–Ingersoll–Ross (CIR) model / Square–root diffusion The second classical interest rate term structure model was introduced in Cox et al. [10]. It is typically referred to as CIR model. The square root diffusion process

$$dr_t = \kappa(\theta - r_t) dt + \sigma\sqrt{r_t} dW_t$$

is used as an improvement of the Vašček model. The transition density $f(\cdot, \cdot)$ of the square–root diffusion process is a cumbersome object but yet an analytic expression. Hence, the bridge process associated with the CIR model is described by tractable dynamics. In particular, for $0 < s < t < T$, we get

$$\eta(t, x; x_s, s) := \frac{\partial}{\partial x} \log f_t(x | x_s, s)$$

$$= \left( \frac{1}{2\kappa (e^{\kappa(t-s)} - 1)^{2} x\sigma^2 \sqrt{xx_x} e^{\kappa(x+s)} I_q[\nu(x)]} \right)$$

$$\times \left\{ 2\kappa^2 x x_x \left( e^{\kappa(2t-s)} - e^{\kappa t} \right) \left(I_{q-1}[\nu(x)] + I_{q+1}[\nu(x)] \right) \right.$$  

$$- \left( e^{\kappa(t-s)} - 1 \right) \sqrt{xx_x} e^{\kappa(x+s)} I_q[\nu(x)]$$

$$\times (2\theta\kappa - \sigma^2 + 2e^{\kappa(t-s)}(4\kappa x - 2\theta\kappa + \sigma^2)) \right\},$$

where

$$\nu(x) := \frac{4\kappa \sqrt{xx_x} e^{\kappa(x+s)}}{\sigma^2(e^{\kappa s} - e^{\kappa x})}, \quad q := \frac{2\kappa \theta}{\sigma^2} - 1,$$

and $I_q(\cdot)$ denotes the modified Bessel function of the first kind. Then, the bridge process dynamics for the CIR model are given by

$$d\hat{r}_t = \left( \kappa(\theta - \hat{r}_t) + \hat{r}_t \sigma^2 \eta(t_2, x_2; \hat{r}_t, t) \right) dt + \sigma\sqrt{\hat{r}_t} dW_t.$$

Geometric Brownian motion (GBM) For a GBM $X$, described by

$$dX_t = X_t(\mu dt + \sigma dW_t), \quad (17)$$

the transition density is available as an analytic expression. Thus, the dynamics of the associated bridge process take the explicit form

$$d\hat{X}_t = \hat{X}_t \left( \frac{\log(x_2) - \log(X_t) - (\mu - \frac{1}{2}\sigma^2)(t_2 - t)}{t_2 - t} \right) dt + \sigma dW_t.$$
Brownian motion with drift In the simplest case of a Brownian motion with constant drift and volatility, i.e.,
\[ dX_t = \mu dt + \sigma dW_t, \]
the associated bridge process is the well-known Brownian bridge. Its dynamics are given by
\[ d\hat{X}_t = \frac{x_2 - \hat{X}_t}{t_2 - t} dt + \sigma dW_t. \]

3.1.1 Pathwise construction of the bridge process for GBM

Proposition 3.1. Consider a GBM \( X_t \), as defined in \((17)\). Assume that it starts in \( X_{t_1} = x_1 \) and it shall be pinned to the value \( x_2 \) at time \( t_2 \). Then, for \( t_1 \leq t \leq t_2 \), the bridge process \( \hat{X}_t \) is given by
\[ \hat{X}_t = x_1 \cdot \exp \left\{ \sigma \left( \int_{t_1}^t dW_s - \frac{t - t_1}{t_2 - t_1} \int_{t_1}^{t_2} dW_s \right) + \frac{t - t_1}{t_2 - t_1} \log \left( \frac{x_2}{x_1} \right) \right\}. \]

Proof. Obviously, \( \hat{X}_{t_1} = x_1 \) as well as \( \hat{X}_{t_2} = x_2 \) do hold. Moreover, let us check that the structure of the original process is indeed preserved by this bridge process. Denote the exponent in \( (18) \) by \( Y_t \) and consider \( x_2 = e^{Y_{t_2}} \) as a lognormally distributed random variable, in particular \( \text{Var}(Y_{t_2}) = \sigma^2 (t_2 - t_1) \). Define \( \Sigma_{t_i}(t) := \sigma^2 (t - t_i) \). Then, for \( t_1 \leq s_1 \leq s_2 \leq t_2 \), it holds that
\[ \text{Cov}(Y_{s_1}, Y_{s_2}) = \Sigma_{t_i}(s_1) - \frac{\Sigma_{t_i}(s_1) \Sigma_{t_1}(s_2)}{\Sigma_{t_1}(t_2)} + \frac{\Sigma_{t_i}(s_2) \Sigma_{t_1}(s_1)}{\Sigma_{t_1}(t_2)} \Sigma_{t_1}(s_1) \]
\[ + \frac{\Sigma_{t_i}(s_1) \cdot \Sigma_{t_1}(s_2)}{(\Sigma_{t_1}(t_2))^2} \Sigma_{t_1}(t_2) + \frac{\Sigma_{t_1}(s_2) \cdot \Sigma_{t_1}(s_1)}{(\Sigma_{t_1}(t_2))^2} \Sigma_{t_1}(t_2) \]
\[ = \sigma^2 (s_2 - s_1), \]
which confirms that \( Y \) is again a Brownian motion and thus \( \hat{X} \) is a GBM.

Notice that Prop. 3.1 offers a straightforward simulation algorithm for the GBM bridge process, as only the Brownian increments, i.e. Gaussian random variables, need to be generated. Figure 2 shows a collection of sample paths simulated by the above method.

3.2 Jump processes

Stochastic processes that do not fluctuate in a continuous manner but rather by sudden jumps, are popular models for a variety of applications. The majority of typical jump models belongs to the class of Lévy processes. Lévy processes are stochastic processes characterized by independent and stationary increments as well as stochastically continuous sample paths.\(^4\) In addition to their prominence in the physical sciences,\(^5\) there is a particularly vast literature on Lévy processes as a model for the random evolution of variables present in the financial markets.\(^6\) As we are dealing with bridge processes here, let us mention the fact that the Markov property of Lévy bridges is inherited from the Markov property of Lévy processes [16, Proposition 2.3.1].

\(^4\)See, e.g., the books of Applebaum [1], Bertoin [5], or Sato [33] for Lévy processes theory.

\(^5\)See, e.g., the review article [41] on the subject and other articles contained in [4].

\(^6\)See, e.g., the books of Cont and Tankov [9], Schoutens [34], or Schoutens and Cariboni [35].
3.2.1 Compound Poisson bridges

The most fundamental and prominent jump process is the Poisson process, counting the number of occurrences of some random event. For the modeling of a situation where not only the number of those (quantifiable) events but also their size matters, the compound Poisson process is a natural extension. It is extensively used, e.g., for actuarial applications as insurance companies are naturally not only interested in the number of claims happening to their customers but even more importantly in the claim sizes.\footnote{The book of Albrecher and Asmussen \cite{2} includes a comprehensive treatment of the compound Poisson model in risk theory, including not only an exhaustive list of its properties but also a discussion of its wide range of applications. In particular, the problem studied in \cite[Chapter V, pg. 146]{2} is of a related flavor to the problem of this section: They characterize a sample path in the compound Poisson risk model given that it leads to ruin.}

We present a method to simulate sample paths from a compound Poisson bridge process, i.e., a compound Poisson process with given initial and final value (and time). For jump-size distribution families that are closed under convolution or where convolution results in another tractable parametric family, some ingredients to our simulation scheme can be derived analytically and thus efficient simulation is possible. We carry out this exercise for the most popular representatives of jump-size distributions, i.e., the Normal distribution, the Exponential distribution, and the Gamma distribution. For distributions that do not allow for a tractable representation of the required convolution objects, one will have to revert to statistical procedures such as acceptance-rejection methods.

Consider a compound Poisson process $X$ with intensity $\gamma$ and jump-size distribution given by the density $f$. To avoid notational conflicts, we reserve the lower index in $X_t$ to describe the process $X$ at time $t$. In contrast, we use an upper index to enumerate individual jumps (as random variables). The realization of an ‘$i$-th’ jump $X^i$ is denoted by $x_i$. Consider now the process $X_t$ in the interval $[t_1, t_2]$, where we are given the values $X_{t_1}$ and $X_{t_2}$. Define $c := X_{t_2} - X_{t_1}$. We suggest the simulation of the bridge process to be performed in the following three steps.

\[\text{Figure 2: Sample paths of GBM bridges (}\mu = 0.01, \sigma = 0.2).\]
\[\text{Left: 100 steps. Right: 20 steps.}\]
I: Simulation of the number of jumps  As a first step, simulate from the conditional Poisson process \( N \) given the value of the sum \( \sum_{i=1}^{N} X^i = c \). This yields a realization of the number of jumps occurring over the considered time interval \( [t_1, t_2] \). The probability function of this object is given by

\[
\mathbb{P} \left[ N = n \left| \sum_{i=1}^{N} X^i = c \right. \right] = \frac{f_{\sum_{i=1}^{N} X^i \mid N=n}(c) \cdot \mathbb{P}[N = n]}{\sum_{i=1}^{N} f_{X^i}(c)} = \frac{f^{*n}(c) \cdot (\lambda(t_2-t_1))^n e^{-\lambda(t_2-t_1)}}{\sum_{m=0}^{\infty} f^{*m}(c) \cdot (\lambda(t_2-t_1))^m \cdot m!} .
\]

(19)

For simulation purposes we cut the support of this conditional distribution of \( N \) to an interval \( [0, \bar{N}] \subseteq \mathbb{N}_0 \), in such a way that \( \mathbb{P}[N > \bar{N}] \sum_{i=1}^{N} X^i = c] < \varepsilon \), for some small value of \( \varepsilon \).

Consider the Normal distribution as a jump size distribution, i.e. \( X^i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2) \), and let \( c > 0 \). The convolution of \( j \) iid \( \mathcal{N}(\mu, \sigma^2) \) distributions is an \( \mathcal{N}(j\mu, j\sigma^2) \) distribution. Thus,

\[
\mathbb{P} \left[ N > \bar{N} \left| \sum_{i=1}^{N} X^i = c \right. \right] = \sum_{n=\bar{N}+1}^{\infty} \frac{f^{*n}(c) \cdot (\frac{\lambda(t_2-t_1))^n}{n!} e^{-\lambda(t_2-t_1)}}{\sum_{m=0}^{\infty} f^{*m}(c) \cdot \mathbb{P}[N = m]}
\]

\[
\leq \sum_{n=\bar{N}+1}^{\infty} f^{*n}(c) = \sum_{n=\bar{N}}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(c-n\mu)^2}{2n\sigma^2} \right)
\]

\[
\leq \sum_{n=\bar{N}+1}^{\infty} c_1 \cdot \exp(-c_2n) = c_1 \cdot \frac{e^{-c_2(\bar{N}+1)}}{1 - e^{-c_2}},
\]

where \( c_1 := \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{\mu^2}{2\sigma^2}) \), \( c_2 := \frac{\sigma^2}{2\sigma^2} \). To ensure this upper bound to be smaller than \( \varepsilon \), we therefore require \( \varepsilon \ll c_1 \) and

\[
\bar{N} > -\frac{\log (\varepsilon(1-e^{-c_2})/c_1)}{c_2} - 1.
\]

The convolution of \( j \) independent Exponential distributions with parameter \( \lambda \) gives an Erlang distribution with parameters \( \lambda \) and \( j \). The convolution of \( j \) Gamma(\( \alpha, \theta \)) distributions gives a Gamma(\( j\alpha, \theta \)) distribution. The Erlang distribution is a special case of the Gamma distribution, i.e. for integer-valued shape parameters. Hence, a criterion for \( \bar{N} \) associated with Gamma distributed jumps also gives a criterion associated with Exponentially distributed jumps. From requiring, for the Gamma distribution case,

\[
\mathbb{P} \left[ N > \bar{N} \left| \sum_{i=1}^{N} X^i = c \right. \right] \leq \sum_{n=\bar{N}+1}^{\infty} f^{*n}(c) = 1 - \frac{e^{-\theta c}}{c} \sum_{n=1}^{N} (\theta c)^n \frac{(\theta c)^{\alpha n}}{\Gamma(n\alpha)} \leq \varepsilon ,
\]

we get the condition

\[
\sum_{n=1}^{N} (\theta c)^{\alpha n} \frac{(\theta c)^{\alpha n}}{\Gamma(n\alpha)} \geq c(1 - \varepsilon)e^{\theta c},
\]

from which one obtains \( \bar{N} \) by running an elementary trial and error program.
Having determined the value $\bar{N}$, one can then easily compute an approximation of $\mathbb{P}[N = n| \sum_{i=1}^{N} X^i = c]$, for all $n = 1, \ldots, \bar{N}$, by cutting the sum in the denominator of (19) after the index $\bar{N}$. The cumulative distribution function is then easily obtained by summation of the single probabilities and inverse transform sampling gives a straightforward simulation scheme by applying its inverse to random draws from the uniform distribution on $[0, 1]$.

II: Simulation of the jumping times Suppose that some value $n$ for the number of jumps in the interval $[t_1, t_2]$ has been simulated by the method outlined above. Then, the precise jumping times are uniformly distributed over this interval. More precisely, the joint distribution of the jumping times $(\tau_1, \ldots, \tau_n)$ equals the law of the order statistics of $n$ independent uniform random variables on $[t_1, t_2]$ (cf., e.g., [9, Prop. 2.9]). Thus, the jumping times can easily be generated by another $n$ calls of a standard (pseudo) random number generator.

III: Simulation of the jump sizes As a last step, let us study the conditional distribution of the summands $X^i, i = 1, \ldots, n$, given the value of the sum $\sum_{i=1}^{n} X^i = c$, where $n$, the number of jumps, is fixed. The corresponding densities are given by

\begin{align*}
    f_{X^1|X^1+\ldots+X^n}(x | c) &= \frac{f(x) f^{*(n-1)}(c-x)}{f^{*n}(c)} \\
    f_{X^2|X^2+\ldots+X^n}(x | c - x_1) &= \frac{f(x) f^{*(n-2)}(c-x_1-x)}{f^{*(n-1)}(c-x_1)} \\
    &\vdots \\
    f_{X^{n-1}|X^{n-1}+\ldots+X^n}\left(x \bigg| \sum_{j=1}^{n-2} x_j \right) &= \frac{f(x) f^{n-2}(c-\sum_{j=1}^{n-2} x_j-x)}{f^{*n}(c)} \\
    f_{X^n|X^n}\left(x \bigg| \sum_{j=1}^{n-1} x_j \right) &= \frac{f(x) f^{n-1}(c-\sum_{j=1}^{n-1} x_j-x)}{f(c-\sum_{j=1}^{n-1} x_j)} = \delta_{c-\sum_{j=1}^{n-1} x_j}(x),
\end{align*}

where we use a version of Bayes rule and the decomposition

$$f_{S_1, S}(x, z) = f_{S_1}(x) f_{S_2}(z-x)$$

for the joint density of the summand $S_1$ and the sum $S = S_1 + S_2$ of two random variables $S_1$ and $S_2$. The $n$-fold convolution of the function $f$ with itself is denoted by $f^{*n}$.

**Remark.** Notice that for the last jump $X^n$ the conditional jump-size distribution is a Dirac distribution with all the mass centered in the remaining gap between the target value and the value of the process after the penultimate jump $x^{n-1}$. Hence, the proposed simulation procedure ends up in the targeted value with probability one.

For Proposition 3.2–3.4 below, denote the jump size distribution by $F$. We study the conditional distribution $F_k$ of the $k$-th jump $X^k$ given the value of the sum of the remaining $(n-k+1)$ jumps, i.e. $\sum_{i=k}^{n} X^i = c - \sum_{j=1}^{k-1} x_j =: C_k$. 

17
Proposition 3.2 (Gaussian jumps). Let \( F \sim \mathcal{N}(\mu, \sigma^2) \). Then, \( \hat{F}_k^N \) is a normal distribution with mean \( \frac{C_k}{n-k+1} \) and variance \( \left( \frac{n-k}{n-k+1} \right) \sigma^2 \), for any \( 1 \leq k \leq n \).

Proof. Using the convolution properties of the Normal distribution, \( \hat{F}_k^N \) is characterized by the density function

\[
\begin{align*}
    f_{X_k|X_1, \ldots, X_n}(x|C_k) &= \frac{f(x)f^{(n-k)}(C_k - x)}{f^{(n-k+1)}(C_k)} \\
    &= \frac{1}{\sqrt{2\pi \left( \frac{n-k}{n-k+1} \right) \sigma^2}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} - \frac{(C_k - x - (n-k)\mu)^2}{2(n-k)\sigma^2} + \frac{1}{2(n-k+1)\sigma^2} \right) \\
    &= \frac{1}{\sqrt{2\pi \left( \frac{n-k}{n-k+1} \right) \sigma^2}} \exp \left( -\frac{x - C_k}{\frac{n-k}{n-k+1}} \right)^2 \\
    &\text{for any } 1 \leq k \leq n.
\end{align*}
\]

\( \square \)

Proposition 3.3 (Exponentially distributed jumps). Let \( F \sim \text{Exp}(\lambda) \). Then, \( \hat{F}_k^\text{Exp} \) is a Lomax distribution, for any \( 1 \leq k \leq n \).\(^8\) In particular,

\[
\hat{F}_k^\text{Exp} \sim \text{Lomax}(k - n, -C_k).
\]

Proof. Using the convolution properties of the Exponential distribution, \( \hat{F}_k^\text{Exp} \) is characterized by the density function

\[
\begin{align*}
    f_{X_k|X_1, \ldots, X_n}(x|C_k) &= \frac{f(x)f^{(n-k)}(C_k - x)}{f^{(n-k+1)}(C_k)} \\
    &= \frac{\lambda e^{-\lambda x} \Delta^{n-k}(C_k - x)^{n-k-1} e^{-\lambda(C_k - x)}}{\Delta^{n-k+1}(C_k - x)^{n-k} e^{-\lambda C_k}} \\
    &= \frac{(n-k)(C_k - x)^{n-k-1}}{C_k^{n-k+1}} = n - k \left( \frac{x}{C_k} \right)^{n-k-1} \\
    &= f_{\text{Lom}}(x; -(n-k), -C_k).
\end{align*}
\]

\( \square \)

Remark. The Lomax distribution is a special case of the generalized Pareto distribution (GP), namely \( \text{Lomax}(\alpha, \beta) \sim \text{GP}(0, 1/\alpha, \lambda/\alpha) \). As such, it is typically contained in commercial software packages. Built in functions can then be used for straightforward simulation.

Remark. Observe in passing, as a quick cross-check of the above results, that in both the Normal and the Exponential distribution case, the derived conditional distributions \( \hat{F}_n \) of the last jump \( X_n \) have expectation \( C_n \) and variance 0.

\(^8\)The Lomax distribution \( \text{Lomax}(\alpha, \beta) \) corresponds to a shifted Pareto distribution and is thus occasionally referred to as Pareto type II distribution. Its density function is given by

\[
f_{\text{Lom}}(y; \alpha, \beta) = \frac{\alpha}{\beta} \left( 1 + \frac{y}{\beta} \right)^{-(\alpha+1)}
\]

where \( \alpha \) is a shape parameter and \( 1/\beta \) is a scale parameter.
Proposition 3.4 (Gamma distributed jumps). Let $F \sim \text{Gamma}(\alpha, \theta)$. Then, $\hat{F}_k$ is a generalized Beta distribution of first kind, for any $1 \leq k < n$.\footnote{The density function of the generalized Beta distribution of first kind is given by $f_{\text{GB1}}(y; a, b, p, q) = \frac{|p|y^{p-1} (1 - \left(\frac{y}{b}\right))^q - 1}{b^{p}B(p, q)}$, where $B(\cdot)$ denotes Euler’s Beta function.} In particular,

$$\hat{F}_k \sim \text{GB1}(1, C_k, \alpha, (n-k)\alpha).$$

Proof. Using the convolution properties of the Gamma distribution, $\hat{F}_k$ is characterized by the density function

$$f_{X^i|X^k,\ldots,X^n}(x|C_k) = \frac{f(x)f^{(n-k)}(C_k - x)}{f^{(n-k+1)}(C_k)} \frac{\Gamma(\alpha - \theta x)}{\Gamma((n-k)\alpha)} \frac{\Gamma(n-k-1)}{\Gamma(n-k+1)\alpha - 1} \frac{1}{C_k^{(n-k)\alpha - 1}} \cdot$$

Remark. As the Gamma function $\Gamma(\cdot)$ is only defined for strictly positive arguments, the case $k = n$ is not covered in Proposition 3.4 above. However, we have generally addressed the latter case before.

The simulation scheme for compound Poisson processes, that has been elaborated in this section, is summarized in Algorithm 1 below. It includes the Normal, the Exponential and the Gamma distribution for the jump size. Figure 3 visualizes sample paths generated on the basis of this algorithm.

Remark. While efficient simulation of trajectories of compound Poisson bridges is indeed possible (given a tractable jump-size distribution), the distribution of the bridge process for some time $t \in (t_1, t_2)$ is generally an intractable object. Its cdf consists of the following terms:

$$\hat{F}_k(x) = \mathbb{P}[X_t \leq x | X_{t_2} - X_{t_1} = c]$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\sum_{i=0}^{m} X^i \leq x | X_{t_2} - X_{t_1} = c] \cdot \mathbb{P}[N_i = m | N_T = n] \cdot \mathbb{P}[N_T = n]$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{n} \left( \int_0^{x-m-2} \cdots \int_0^{x-y_1} f_1(y_1) \cdots f_{m-1}(y_{m-1}) dy_1 \cdots dy_{m-1} \right) \cdot \binom{n}{m} \left( \frac{t}{t_2 - t_1} \right)^m \left( 1 - \frac{t}{t_2 - t_1} \right)^{n-m} \frac{(\lambda(t_2 - t_1))^n}{n!} e^{-\lambda(t_2 - t_1)},$$

i.e., an infinite sum of the product of a Poisson distribution with parameter $\lambda(t_2 - t_1)$, a Binomial distribution with probability parameter $t/(t_2 - t_1)$, and
Figure 3: 5 sample paths of compound Poisson bridge processes ($\lambda = 3$).
Left: Normal jumps with parameters $\mu = 0.5, \sigma = 1$. Right: Exponential jumps with mean $\gamma = 0.5$.

Algorithm 1 A simulation scheme for compound Poisson bridge processes.

Given: $X_{t_1}, X_{t_2}$

Define: $c := X_{t_2} - X_{t_1}$, $C_k := c - \sum_{j=1}^{k-1} x_j$

I: Simulate a value $n$ for the number of jumps $N$

1: Determine $\bar{N}$ such that

\[
\frac{\lambda}{\sqrt{2\pi\sigma^2}} \bar{N} > -\log \left( \frac{\lambda (1 - e^{-c^2/2})}{c_2} \right) - 1,
\]
\[
c_1 = \frac{\mu}{\sqrt{2\pi\sigma^2}}, c_2 = \frac{\mu^2}{2\sigma^2}
\]

2: For $n = 1, \ldots, \bar{N}$ compute the conditional probabilities

\[
P[N = n | X_{t_2} - X_{t_1} = c] = \frac{f^{* n}(c) \cdot \lambda (t_2 - t_1)^n}{\sum_{m=1}^{N} f^{* m}(c) \cdot \lambda (t_2 - t_1)^m},
\]

using the convolution properties $N(\mu, \sigma^2) * N(\mu, \sigma^2) = N(2\mu, 2\sigma^2)$, $\text{Exp}(\gamma) * \text{Exp}(\gamma) = \text{Erl}(\gamma, 2)$, $\text{Gam}(\alpha, \theta) * \text{Gam}(\alpha, \theta) = \text{Gam}(2\alpha, \theta)$.

3: Obtain a value $n$ by simulating from the resulting distribution, i.e., by computing its cdf and applying the inverse to a sample from the Uniform distribution on $[0, 1]$

II: Simulate the jumping times

1: Simulate $n$ random variables from the Uniform distribution on $[t_1, t_2]$

III: Simulate the jump sizes

1: For $k = 1, \ldots, n$, simulate from the following distributions:

\[
\begin{array}{c|c|c}
N(\mu, \sigma^2) & \text{Exp}(\gamma) & \text{Gam}(\alpha, \theta) \\
\hline
\mathcal{N} \left( \frac{C_k}{n-k+1}, \left( \frac{n-k}{n-k+1} \right)^2 \right) & \text{Lomax}(k-n, -C_k) & \text{GB1}(1, C_k, \alpha, (n-k)\alpha) \\
\end{array}
\]
a complicated multidimensional integral over the conditional densities (using a shorthand notation) given in (20).

**Remark** (Further Lévy processes). For most Lévy processes, the density function at a given future time is not available in (semi-)closed form. However, in some special cases, bridge processes turn out to be of a surprisingly tractable nature. In the dissertation of Hoyle [16], one can find results for 1/2-stable processes, Inverse Gaussian processes and Cauchy processes, which imply that a simulation of associated bridges can be performed in a straightforward way: In the first two cases, by applying a deterministic function to a random draw from the standard Normal distribution; in the third case, the cumulative distribution function is given in terms of standard functions.

### 4 An illustrative example

We will now illustrate the suggested approach by a simple inventory control problem. Consider a business where some (perishable) goods can be sold for a unit price $a$. The stock can be replenished each Monday morning for the price $b$ per unit. During the week, the products are sold but the stock cannot be replenished. The demand varies. If the business runs out of stock, then costs $c$ occur depending on the remaining time until the next opportunity to fill the stock. For products left in stock at the end of the week, we assume that only 30% can still be used for the next week, but 70% need to be thrown away.

As a model for the demand, we use the Vasicek model (see (15) in Sect. 3.1). In particular, we do not consider any seasonal patterns for simplicity. Let the parameters of the Vasicek model be given by $\theta = 105, \kappa = 0.5, \sigma = 10$, and the starting value $x_0 = 100$. Three-stage problems are the smallest instances involving all issues associated with multistage decision making under uncertainty. Hence, it is our objective, for this purely illustrative example, to maximize expected profits over two upcoming weeks.

#### 4.1 Discretization of decision stages

Following the approach suggested in Sect. 2.1, we choose the functions $H(x) = \sigma x, g(x) = \frac{\theta}{\sigma}(\theta - x), \tau(x) = 1$ in case of the Vasicek model. Then, the resulting lattice corresponding to the $N$-th iteration of Alg. 2.1, discretizing the weekly decision stages $t = 0, 1, 2$, would be given by the discrete random variable $\tilde{X}^{(N)}_{t+4N}$ and the corresponding number of nodes on the lattice, as given in Tab. 1. Notice that this lattice construction serves for the discretization of the decision stages only. Hence, the probabilities of different paths between two stages, which end up in the same node, can be summed up and the intermediate nodes and paths do not have to be stored. If one wanted to store the full lattice construction, this would correspond to $(t \cdot 4^N + 1)^2$ nodes, $\frac{1}{2}(3^t \cdot 4^N + 1)$ conditional probabilities, and a total of $3^t \cdot 4^N$ paths up to time $t$, for the $N$-th iteration.

For ease of exposition, we keep the discrete model as small as possible for the current example and focus on illustrating the suggested methodology regarding the multiscale issue. Thus, we use a simple binary tree for the decision stages, which we obtain by a standard optimal quantization algorithm. The tree is visualized in Fig. 4. For real-world applications, trees/lattices with $10^5$
Table 1: Lattice via Alg. 2.1 - number of nodes at stage $t$, $N$-th iteration

<table>
<thead>
<tr>
<th>$t$</th>
<th>$N$</th>
<th>$X^{(1)}_1$</th>
<th>$X^{(2)}_2$</th>
<th>$X^{(3)}_3$</th>
<th>$X^{(4)}_4$</th>
<th>$X^{(5)}_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$\tilde{X}^{(1)}_1$</td>
<td>9</td>
<td>$\tilde{X}^{(2)}_2$</td>
<td>33</td>
<td>$\tilde{X}^{(3)}_3$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$\tilde{X}^{(1)}_8$</td>
<td>17</td>
<td>$\tilde{X}^{(2)}_3$</td>
<td>65</td>
<td>$\tilde{X}^{(3)}_2$</td>
</tr>
</tbody>
</table>

Figure 4: A simple binary tree as a discrete model for the decision stages. The Vašček model has been discretized by an optimal quantization algorithm.

4.2 Solution of the problem

Denote the demand at time $s$ by (the continuous random variable) $X_s$, the (deterministic) stock level by $S_s$, and the remaining stock level at the beginning of week $t$ by $R_t$. In particular, $S_t = 0.3 \cdot R_{t-1} + \pi_t$ where we optimize over $\pi_t$. We set $R_{-1} = 0$. The profit during week $t \in \{1, 2\}$ is then given by

$$f_t(\pi_{t-1}, X_{t-1:t}) = \begin{cases} a \cdot (S_t - R_t) - b \cdot S_{t-1} & \text{if } R_t > 0 \\ (a - b) \cdot S_{t-1} - (1 - \tau_t) \cdot c & \text{if } R_t = 0, \end{cases}$$

where

$$\tau_{t+1} := \inf \left\{ s \in [0, 1] : \int_t^{t+s} X_u \, du > S_t \right\}$$

represents the first time of week $t$ when the business runs out of stock. Remaining stock at the end $T$ of the planning horizon enter the model with their value, i.e., we add $0.3 \cdot R_2 \cdot a$ to $f_T(\pi_{T-1})$. The problem can then be summarized as

$$\sup_{\pi} \sum_{t=1}^2 \mathbb{E}[f_t(\pi_{t-1})]$$

s.t. $\pi \in \sigma(X)$,

where $\sigma(X)$ denotes the filtration generated by the demand process $X$. As a concrete example, we set the problem parameters to $a = 10, b = 7, c = 1000$. We observe the demand on an hourly basis, 24/7.

The key observation for this illustrative example is the fact that profits depend in a highly nonlinear fashion on the whole demand trajectory. This is
due to the presence of the stopping times $\tau_t$ in the objective. Therefore, we apply the suggested methodology and generate a collection of paths between each pair of consecutive decision nodes. In such a way, expected profits during the week can be computed by a simple Monte Carlo simulation. The SDE describing the Vasicek bridge process is given in (16).

We have obtained the following numerical results. The optimal value is given by 517.65. The optimal strategy is given by the vector $[97.67, 102.02, 94.96]$, where the first value refers to the optimal decision in the root node, the second value refers to the optimal decision for the second week, given that the demand will be 106.46, and similarly for the third value and the node 93.98.

5 Conclusion

In this article, we have suggested a computational framework for multistage stochastic optimization problems with embedded multiperiod problems. We have named the suggested approach multiscale stochastic programming. It is based on a separation between the (standard) multistage decision problem, and the problem of determining path-dependent costs between two consecutive decisions. The paper contains a contribution to both aspects. One section was dedicated to the construction of scenario lattices as a discrete structure representing a time-homogeneous Markovian diffusion model. In particular, we examined a Markov-chain approximation approach and showed that the approximation error w.r.t. the optimal value of a generic multistage stochastic optimization problem can be controlled with the suggested methodology. In a second part, we suggested to leverage the theory of stochastic bridges to tackle the embedded multiperiod problem, which takes place on a much finer time-scale than the decision scale. An illustrative multiscale inventory control problem finally served as a proof of concept for the proposed methodology.

References


Appendix

The Wasserstein distance

**Definition 5.1 (Wasserstein distance).** Let \( P \) and \( \tilde{P} \) be two probability measures on \( \mathbb{R}^m \). Then, their Wasserstein distance is defined as

\[
W(P, \tilde{P}) := \inf_{\pi \in \mathcal{X}(P, \tilde{P})} \int_{\mathbb{R}^m \times \mathbb{R}^m} \|v - w\| \, \pi(dv, dw),
\]

where \( \mathcal{X}(P, \tilde{P}) \) denotes the set of all probability measures on \( \mathbb{R}^m \times \mathbb{R}^m \) with marginals \( P \) and \( \tilde{P} \).

If \( P \) and \( \tilde{P} \) are probability measures on the real line, then their Wasserstein distance can be represented in terms of their cdf’s \( F \) and \( \tilde{F} \) (Vallender [38]):

\[
W(P, \tilde{P}) = \int |F(x) - \tilde{F}(x)| \, dx.
\]

**Proof of Theorem 3.1**

**Proof.** For any \( t \in [t_1, t_2] \), denote the density function of the random variable \( X_t \) by \( f_t(x|X_{t_1} = x_1, X_{t_2} = x_2) \). Due to Bayes’ Theorem and the fact that solutions of SDEs are Markov processes, we may rewrite this function as

\[
\hat{f}_t(x|X_{t_1} = x_1, X_{t_2} = x_2) = \frac{f_{t_2}(x_2|x_{t_2} = x_{t_2}) \cdot f_t(x|X_{t_1} = x_1)}{f_{t_2}(x_2|X_{t_1} = x_1)}.
\]  

Then, for any Lipschitz-continuous function \( h : \mathbb{R}^d \to \mathbb{R} \), it holds that

\[
\mathbb{E} \left[ h(X_1) \Big| X_{t_1} = x_1, X_{t_2} = x_2 \right] = \int_{\mathbb{R}^d} h(x) \cdot \hat{f}_t(x|X_{t_1} = x_1, X_{t_2} = x_2) \, dx
\]

\[
= \frac{1}{f_{t_2}(x_2|X_{t_1} = x_1)} \mathbb{E} \left[ \int_{t_1}^{t} \left\{ \left( h(X_s) \frac{\partial}{\partial x} f_{t_2}(x_2|X_s, s) \right) \right. \right.
\]

\[
+ \sum_{i=1}^{d} \left( \frac{\partial}{\partial x^i} [h(X_s) \cdot f_{t_2}(x_2|X_s, s)] \right) \cdot \mu_i(X_s)
\]

\[
+ \left. \frac{1}{2} \sum_{i,j=1}^{d} \left( \frac{\partial^2}{\partial x^i \partial x^j} [h(X_s) \cdot f_{t_2}(x_2|X_s, s)] \right) \right\} \left[ \Sigma \Sigma^T (X_i) \right]_{i,j} \right] \, ds \Big| X_{t_1} = x_1
\]

\[
= \frac{1}{f_{t_2}(x_2|X_{t_1} = x_1)} \mathbb{E} \left[ \int_{t_1}^{t} \sum_{i=1}^{d} \mu_i(X_s) \left( f_{t_2}(x_2|X_s, s) \frac{\partial}{\partial x_i} h(X_s) \right) 
\]

\[
+ \frac{1}{2} \sum_{i,j=1}^{d} \left( f_{t_2}(x_2|X_s, s) \cdot \frac{\partial^2}{\partial x^i \partial x^j} h(X_s) + 2 \frac{\partial}{\partial x^i} h(X_s) \frac{\partial}{\partial x^j} f_{t_2}(x_2|X_s, s) \right)
\]

\[
\times \left[ \Sigma \Sigma^T (X_i) \right]_{i,j} \right] \, ds \Big| X_{t_1} = x_1
\],

by the multi-dimensional Itô lemma and the Kolmogorov backward equation, which ensures

\[
\frac{\partial}{\partial s} f_{t_2}(x_2|X_s, s) + \sum_{i=0}^{d} \mu_i(X_s) \frac{\partial}{\partial x_i} f_{t_2}(x_2|X_s, s) + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x^i \partial x^j} f_{t_2}(x_2|X_s, s) = 0.
\]
Differentiation with respect to the time parameter gives

$$\frac{\partial}{\partial t} \mathbb{E}[h(X_t) | X_{t_1} = x_1, X_{t_2} = x_2] = \int_{\mathbb{R}^d} h(x) \frac{\partial}{\partial t} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \, dx$$

$$= \int_{\mathbb{R}^d} \left\{ \sum_{i=1}^d \mu_i(x) \frac{\partial}{\partial x_i} h(x) + \frac{1}{2} \sum_{i,j=1}^d [\Sigma \Sigma^\top(x)]_{i,j} \frac{\partial^2}{\partial x^i \partial x^j} h(x) \right. $$

$$\left. + \sum_{i,j=1}^d [\Sigma \Sigma^\top(x)]_{i,j} \frac{\partial}{\partial x^i} h(x) \frac{\partial}{\partial x^j} \log f_{t_2}(x_2|x,t) \right\} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \, dx.$$

(22)

The function $h$ is Lipschitz by assumption and thus its gradient is bounded. It can be seen from (21) that $\hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \to 0$, $\frac{\partial}{\partial t} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \to 0$, as any $x^i \to \pm \infty$. Therefore, integrating (22) twice by parts gives

$$\int_{\mathbb{R}^d} h(x) \frac{\partial}{\partial t} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \, dx$$

$$= \int_{\mathbb{R}^d} \left\{ h(x) \left( - \sum_{i=1}^d \frac{\partial}{\partial x_i} \mu_i(x) \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \right) $$

$$- \sum_{i,j=1}^d \frac{\partial}{\partial x_i} [\Sigma \Sigma^\top(x)]_{i,j} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \frac{\partial}{\partial x^j} \log f_{t_2}(x_2|x,t) \right) $$

$$- \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x^i \partial x^j} h(x) \left[ \Sigma \Sigma^\top(x) \right]_{i,j} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \right\} \, dx,$$

from which we can deduce

$$\frac{\partial}{\partial t} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) = - \sum_{i=1}^d \frac{\partial}{\partial x^i} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \cdot \nu_i(x,t) $$

$$+ \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x^i \partial x^j} \hat{f}_t(x|X_{t_1}, t_1, X_{t_2}, t_2) \cdot [\Sigma \Sigma^\top(x)]_{i,j},$$

(23)

with

$$\nu : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$$

$$(x,t) \mapsto \mu(x) + [\Sigma \Sigma^\top(x)] \nabla_x \log f_{t_2}(x_2|x,t).$$

Equation (23) corresponds to the Fokker-Planck equation of the diffusion process

$$dX_t = \nu(X_t, t) \, dt + \Sigma(X_t) \, dW_t.$$