Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: http://www.tandfonline.com/loi/rquf20

Pricing via recursive quantization in stochastic volatility models

Giorgia Callegaro, Lucio Fiorin & Martino Grasselli

To cite this article: Giorgia Callegaro, Lucio Fiorin & Martino Grasselli (2017) Pricing via recursive quantization in stochastic volatility models, Quantitative Finance, 17:6, 855-872, DOI: 10.1080/14697688.2016.1255348

To link to this article: https://doi.org/10.1080/14697688.2016.1255348

Quantitative Finance, 2017
Vol. 17, No. 6, 855–872, https://doi.org/10.1080/14697688.2016.1255348

Pricing via recursive quantization in stochastic volatility models

GIORGIA CALLEGARO†, LUCIO FIORIN† and MARTINO GRASSELLI*†‡

†Department of Mathematics, University of Padova, via Trieste 63, 35121 Padova, Italy
‡De Vinci Research Center (DVRC), Finance Group, 92916 Paris La Défense, France

(*Received 13 February 2016; accepted 21 October 2016; published online 20 December 2016*)

We provide the first recursive quantization-based approach for pricing options in the presence of stochastic volatility. This method can be applied to any model for which an Euler scheme is available for the underlying price process and it allows one to price vanillas, as well as exotics, thanks to the knowledge of the transition probabilities for the discretized stock process. We apply the methodology to some celebrated stochastic volatility models, including the Stein and Stein [*Rev. Financ. Stud.* 1991, (4), 727–752] model and the SABR model introduced in Hagan *et al.* [*Wilmott Mag.*, 2002, 84–108]. A numerical exercise shows that the pricing of vanillas turns out to be accurate; in addition, when applied to some exotics like equity-volatility options, the quantization-based method overperforms by far the Monte Carlo simulation.

Keywords: Quantization; Monte Carlo method; Stochastic volatility model; Vanilla options; Equity volatility option

JEL Classification: C63, G13

1. Introduction

The problem of pricing derivative contracts in a stochastic volatility framework has been deeply investigated in the literature. Starting from the pioneering works of Hull and White (1987), Chesney and Scott (1989), Stein and Stein (1991) and Heston (1993), researchers have introduced different stochastic models for the spot volatility process, both with continuous trajectories and including jumps, in order to catch the stylized facts of the implied volatility surface, namely the smile and skew effects. The positivity of the volatility process is another crucial feature: in order to ensure it, Hull and White (1987) assume that the volatility itself follows a Geometric Brownian motion. Chesney and Scott (1989) instead propose to model the volatility as the exponential of a (stationary) Ornstein–Uhlenbeck process. Stein and Stein (1991) assume that the volatility is itself a stationary Ornstein–Uhlenbeck process. This assumption allows one to recover the information about the risk neutral density of the underlying through a Fourier-based method and to price vanillas in a very fast and efficient way. On the other hand, the Gaussian distribution of the volatility process opens delicate issues on the positivity. Heston (1993) assumes that the volatility is a CIR process, which remains positive under some parameter restrictions (the so called Feller condition). Both the models of Stein and Stein (1991) and Heston (1993) belong to the class of affine models, for which Duffie *et al.* (2000) developed a systematic approach in order to perform the Fourier methodology, including the presence of jumps in both the underlying and the volatility process. Examples of models including jumps that have been successfully calibrated in both a single factor and a multi factor setting for the volatility process are e.g. Bates (1996), Jacobs and Li (2008) and Christoffersen *et al.* (2009).

Besides the undoubtful quality of the Fast Fourier Transform (henceforth FFT) methodology for pricing vanillas in the affine class of models, it is noteworthy that in the case of exotic derivatives, the pricing problem still represents a challenge and the Monte Carlo simulation seems to be the only possibility (despite its potentially high computational cost), apart from approximations in the spirit of Sesana *et al.* (2014) and Date and Islyaev (2015). In addition, the pricing of equity volatility products, like Corridor Variance Swaps, may reveal some delicate numerical issues while applying the FFT approach, as it requires a rigorous check of the integrability of the corresponding characteristic function (see e.g. Da Fonseca *et al.* 2015). What is more, the FFT approach clearly cannot be applied when dealing with models whose characteristic function is not known explicitly. Nevertheless, in some cases, approximations are available, as in the case of the celebrated SABR model of Hagan *et al.* (2002), representing a standard in the banking industry, for which we know the shape of the implied volatility

*Corresponding author. Email: grassell@math.unipd.it

e.g. for small time to maturity or small vol of vol. The same holds true for the α-hypergeometric volatility model of Da Fonseca and Martini (2016), for which a Mellin transform and some asymptotics are available. Remarkably, in the last model, the volatility remains positive as it is assumed to be the exponential of a mean reverting process.

In this paper, we introduce an alternative pricing method based on quantization. The method is model free, in that it only requires the knowledge of an Euler scheme for the stochastic equation modeling the evolution of the underlying process. It does not require any affine property of the model and we will show in the numerical section that our results over-perform by far the Monte Carlo simulation.

Quantizing a random variable (or vector) taking values in a discrete (resp., continuous) set consists in approximating it by a random variable that takes values in a discrete subset (resp., set). Hence, quantization can be seen as an alternative to discretization methods for stochastic processes and random variables. The birth of quantization dates back to the 1950s, when in the Bell laboratories *ad hoc* signal discretization procedures were developed for signal transmission. Quantization has been later widely used in many fields, such as: operations research, information theory, cluster analysis, pattern and speech recognition, numerical integration, data mining and numerical probability. More recently, in the 1990s, the results obtained in numerical probability have originated new applications to mathematical finance. Quantization of random vectors (the so-called vector quantization, to be distinguished from functional quantization, which is used to approximate stochastic processes by discretizing the trajectories' space) can be seen as a discretization of the probability space, providing, in some sense, the best possible approximation to the original distribution. It is therefore crucial to optimally choose the discretization grid and to measure the error we make by approximating the continuous signal via a discrete one. A grid minimizing the L^2 distance between the original signal and its discrete approximation is called an optimal quadratic quantizer. Many numerical procedures have been developed to obtain optimal quadratic quantizers of the Gaussian (and even non-Gaussian) distribution in high dimension, mostly based on stochastic optimization algorithms. For a comprehensive introduction to optimal vector quantization and its applications, we refer to Pagès (2014) and references therein. For recent successful applications of vector quantization on machine learning for bankruptcy prediction and pattern recognition see, for instance, Charalambous *et al.* (2000) and Thangavel and Ashok Kumar (2006). While theoretically sound and deeply investigated, optimal quantization typically suffers from the numerical burden that the algorithms involve (see e.g. Pagès and Printems 2005). This is mainly due to the highly time-consuming procedure required to obtain the optimal grid, especially in the multi-dimensional case, where stochastic algorithms are necessary. Nevertheless, a new type of quantization that seems to be very promising, called *recursive marginal quantization*, or fast quantization, has been introduced in Pagès and Sagna (2015). It provides sub-optimal (stationary) quantizers in a very fast way (if compared to the time required by standard quantization algorithms to produce optimal grids), to the point that e.g. it has been recently possible to successfully calibrate a local volatility model using this technology, as illustrated in Callegaro *et al.* (2015). Another recent application of recursive marginal quantization to pricing of exotic options in local volatility settings, via a backward Monte Carlo procedure can be found in Bormetti *et al.* (2015).

Starting from the results in Callegaro *et al.* (2015), we push forward the methodology to the case where the volatility is stochastic. Of course, the passage from local to stochastic volatility models requires additional care, from both the theoretical and the numerical point of view. For example, a direct application of the argument presented in Pagès and Sagna (2015) is not possible in the multi dimensional case.

We will focus on the case where the volatility process is driven by a one-dimensional stochastic factor. Nevertheless, most of our results can be easily extended to a multi dimensional setting, with the only drawback of a relevant increase of notational burden. We emphasize that the approach presented here only requires that an Euler scheme is available for the stochastic model, which is the case for most popular stochastic volatility models.

In the numerical illustration, we apply the new methodology to some stochastic volatility models:

- the Stein and Stein (1991) model, which was successively extended by Schöbel and Zhu (1999). This model is affine, so that a Fourier-based approach is available. Prices obtained via Monte Carlo and via Fourier techniques will be our benchmark values when testing the precision of the proposed quantization method.
- the celebrated SABR model, introduced in Hagan *et al.* (2002) and subsequently extended and analysed by dozens of researchers both in financial institutions and in academia. Indeed, this model represents the standard reference in the banking industry for most financial models, ranging from equity, interest rates, FX and commodities. Although closed form formulas are not available for vanillas, excellent approximations have been provided in this setting for small maturities and/or small vol of vol. Benchmark prices will be obtained here via standard Monte Carlo.
- the α-hypergeometric stochastic volatility model recently introduced by Da Fonseca and Martini (2016). In this case, no explicit formulas for vanillas are available, so that also here benchmark prices will be obtained via standard Monte Carlo.

We will apply the recursive marginal quantization method to all these models and we will compare prices of vanillas with the corresponding benchmark prices. Moreover, we will see that the quantization-based approach can also be applied to the pricing of non-vanilla options. In order to give an idea of the flexibility of our approach, we consider a particular class of exotics, namely equity volatility options that are receiving a growing attention in the financial community. Of course, our methodology can be easily applied as well to other stochastic volatility models, like e.g. the celebrated Heston (1993) one, for which a different technique, still based on Pagès and Sagna (2015), has been developed in Fiorin *et al.* (2015) in a general multi-dimensional setting. However, the approach of Fiorin *et al.* (2015) is not suitable for pricing equity volatility products, as the volatility has to be quantized in a very parsimonious number of points in order to be numerically tractable.

Pricing of equity volatility options represents a challenging topic, especially in the case of non affine models, for which Monte Carlo represents the only alternative. We are going to show that our quantization-based algorithm is much more performing than the standard Monte Carlo approach.

The paper is organized as follows: in section 2 we recall some basic notions on quantization, with a particular focus on recursive marginal quantization. In section 3, we develop the quantization-based approach in a general stochastic volatility setting. In section 4, we apply the methodology to some stochastic volatility models: we show the performance of the recursive marginal quantization in the pricing of vanillas as well as in the case of equity volatility options, with particular focus on corridor variance swaps. Section 5 concludes, while we gather in the Appendix some technical proofs and additional material on quantization which may be useful for the reader in view of the implementation.

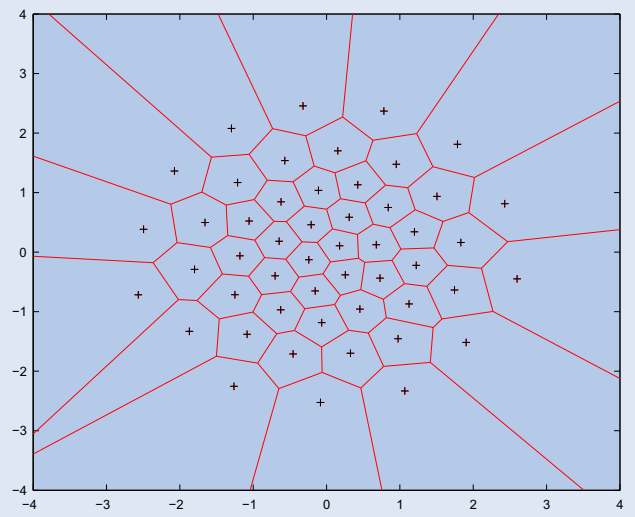

Figure 1. Example of optimal grid for a bivariate (standard) Gaussian distribution with $N = 50$.

2. Essentials on quantization

We start by providing some technical details on quantization of a random variable. For a more extensive introduction to quantization, we refer to Graf and Luschgy (2000) and to Pagès (2014).†

Let X be a one-dimensional real valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite rth moment (with $r > 0$) and probability distribution $\mathbb{P}_X$. For a given $N \in \mathbb{N}$, N-quantizing the random variable X, taking infinitely many values, consists in approximating it by a one-dimensional discrete random variable taking values in a set of cardinality N, that we denote e.g. by $\Gamma = \{\gamma_1, \ldots, \gamma_N\}$, with $\gamma_i \in \mathbb{R}$, $1 \leq i \leq N$.

In practice, quantizing X on a given grid $\Gamma = \{\gamma_1, \ldots, \gamma_N\}$ corresponds to projecting X on the grid Γ following the closest neighbour rule. An *N-quantizer* is a Borel function $q_N : \mathbb{R} \to \Gamma \subset \mathbb{R}$ projecting X on Γ and the projection of X on Γ, $q_N(X)$, is called *the quantization of X*. Typically, the notation $\widehat{X}^\Gamma$, or, for simplicity, $\widehat{X}$, is used for the quantization of X, as an alternative to $q_N(X)$. The induced mean L^r-error is called *L^r-mean quantization error* and it is given by

$$\|X - \widehat{X}\|_r = \| \min_{1 \leq i \leq N} |X - \gamma_i| \|_r$$

where $||X||_r := [\mathbb{E}(|X|^r)]^{1/r}$. As a function of Γ, the L^r-mean quantization error is continuous and reaches a minimum over all the grids with size at most N. A grid $\Gamma^\star$ minimizing the L^r-mean quantization error over all the grids with size at most N is called an *L^r-optimal quantizer* and in the case $r = 2$ the corresponding grid is an optimal *quadratic* quantizer.‡

An optimal quantizer is associated to an optimal grid $\Gamma^\star$ and to an optimal Borel partition of the space $\mathbb{R}$, $(C_i(\Gamma^\star))_{1 \leq i \leq N}$, and vice versa, so that the quantizer is defined as follows

$$\widehat{X}^{\Gamma^\star} = \widehat{X} = \sum_{i=1}^{N} \gamma_i \mathbb{1}_{C_i(\Gamma^\star)}(X)$$

where the partition $\{C_i(\Gamma^\star)\}_{i=1,\ldots,N}$, with $C_i(\Gamma^\star) \subset \{\xi \in \mathbb{R} : |\xi - \gamma_i| = \min_{1 \leq j \leq N} |\xi - \gamma_j|\}$, is called the *Voronoi partition*, or *tessellation* induced by $\Gamma^\star$. Moreover, the L^r-mean quantization error vanishes as the grid size N goes to infinity and the convergence rate is given by the celebrated Zador theorem (see Graf and Luschgy 2000), that when X is in dimension one reads:

$$\min_{\Gamma,\ |\Gamma| = N} \|X - \widehat{X}^\Gamma\|_r = Q_r(\mathbb{P}_X) N^{-1} + o\big(N^{-1}\big)$$

where $Q_r(\mathbb{P}_X)$ is a non negative constant. An example of optimal 50-dimensional quantization grid for the bivariate Gaussian distribution is given in figure 1.

From a numerical point of view, finding an optimal quantizer may be a very challenging task. This motivates the introduction of sub-optimal criteria. We then introduce the notion of *stationary quantizer*:

Definition 2.1 An N-quantizer $\Gamma = \{\gamma_1, \ldots, \gamma_N\}$ inducing the quantization $\widehat{X}^\Gamma$ of X is said to be stationary if

$$\mathbb{E}[X | \widehat{X}^\Gamma] = \widehat{X}^\Gamma.$$

Remark 2.2 A L^r-optimal quantizer is stationary, but the viceversa does not hold true in general.

The result behind the birth of recursive marginal quantization is the following: if we introduce the distortion function associated with $\Gamma = \{\gamma_1, \ldots, \gamma_N\}$

$$D(\Gamma) := \mathbb{E}\left[|X - \hat{X}^\Gamma|^2\right] = \sum_{i=1}^{N} \int_{C_i(\Gamma)} |x - \gamma_i|^2 \mathrm{d}\mathbb{P}_X(x), \quad (1)$$

then the distortion function is continuously differentiable (as a function of Γ) and it can be proved that stationary quantizers are critical points of the distortion function (see e.g. proposition

†We also mention the optimal quantization website: http://www.quantize.maths-fi.com, where one can download the optimized quadratic quantization grids of the d-dimensional Gaussian distributions $\mathcal{N}(0; I_d)$, for $N = 1$ up to 10^4 and for $d = 1, \ldots, 10$.

‡When X is in dimension one, the uniqueness of the optimal N-quantizer is guaranteed if X admits a log-concave density function, see Graf and Luschgy (2000).

2.2 in Pagès (2014)), i.e. a stationary quantizer $\widetilde{\Gamma}$ satisfies

$$\nabla D(\widetilde{\Gamma}) = 0.$$

Hence, from a numerical point of view, stationary quantizers are interesting, insofar they can be found through zero search recursive procedures, that can be efficiently performed. Several algorithms have been proposed to obtain stationary quantizers and the probabilities associated to every grid point. These numerical procedures can be essentially divided into two categories: gradient-based methods and fixed-point methods. The former class includes Newton-Raphson algorithm, which requires the computation of the gradient and of the Hessian of the distortion function, whereas the second category includes the Lloyd I algorithm, a fixed-point algorithm based on the definition of stationary quantizer.

In the next section, we give a brief introduction to recursive marginal quantization, showing how it is possible to translate the quantization of a stochastic process into the quantization of a (suitable) family of random variables.

2.1. Recursive marginal quantization

On a probability space representing all the randomness of our economic context, we consider a continuous-time Markov diffusion V, whose evolution is specified by the following SDE:

$$\mathrm{d}V_t = b(t, V_t)\mathrm{d}t + a(t, V_t)\mathrm{d}W_t, \quad V_0 = v_0 > 0 \in \mathbb{R} \tag{2}$$

where W is a standard Brownian motion and we suppose that b and a are such that a strong solution to the above SDE exists (more detailed hypotheses required for the existence of a strong solution to the SDE in the two dimension case will be given in section 3 in a stochastic volatility setting).

Having fixed a time horizon $T > 0$ and a time discretization grid $\{0 = t_0, t_1, \ldots, t_M = T\}$, with constant step size $\Delta := \frac{T}{M}$, such that $t_k = k\Delta$ for $k = 0, \ldots, M$, the Euler scheme for the process V reads: $\widetilde{V}_{t_0} = \widetilde{V}_0 = v_0$ and

$$\widetilde{V}_{t_{k+1}} = \widetilde{V}_{t_k} + b(t_k, \widetilde{V}_{t_k})\Delta + a(t_k, \widetilde{V}_{t_k})\sqrt{\Delta}Z_k \tag{3}$$

where Z_k is a centered Normal random variable with variance equal to one. The following result will be crucial for the development of recursive marginal quantization.

LEMMA 2.3 *For every $0 \le k \le M-1$, conditionally on the event $\{\widetilde{V}_{t_k} = x\}$, the random variable $\widetilde{V}_{t_{k+1}}$ is Gaussian:*

$$\mathcal{L}\left(\widetilde{V}_{t_{k+1}} | \{\widetilde{V}_{t_k} = x\}\right) = \mathcal{N}\left(m_k(x), \sigma_k^2(x)\right), \tag{4}$$

where

$$m_k(x) = x + b(t_k, x)\Delta$$
$$\sigma_k^2(x) = a^2(t_k, x)\Delta.$$

Proof It follows immediately from equation (3), given that $\sqrt{\Delta}Z_k \sim \mathcal{N}(0, \Delta)$. □

The idea behind recursive marginal quantization, which has been recently introduced in Pagès and Sagna (2015), is quantizing the stochastic process V in equation (2) in an approximated way, by working on the (marginal) random variables $(\widetilde{V}_{t_k})_{0 \le k \le M}$ in equation (3). Hence, we will apply vector quantization theory to every (one dimensional) random variable $\widetilde{V}_{t_k}$ in order to find one stationary quantizer for every $1 \le k \le M$. Notice that the distribution of $\widetilde{V}_{t_{k+1}}$ is in general unknown. Nevertheless, thanks to lemma 2.3, we know its distribution conditional on $\widetilde{V}_{t_k}$. Hence, this will lead to a recursive marginal (vector) quantization procedure.

Notation: for notational simplicity, we indicate by $\widehat{V}_{t_k}$ (instead of $\widehat{\widetilde{V}}_{t_k}$) the quantization of the random variable $\widetilde{V}_{t_k}$, $0 \le k \le M$.

So, we are now looking for stationary quantizers for every random variable $\widetilde{V}_{t_k}$, for $k = 0, 1, \ldots, M$.

Remark 2.4 Obviously $\widehat{V}_{t_0} = v_0$, i.e. the grid at time t_0 consists of a single point.

Recall that stationary quantizers are zeros of the gradient of the distortion function. We denote by D_{k+1} (recall equation (1)) the distortion function relative to $\widetilde{V}_{t_{k+1}}$, namely:

$$D_{k+1}(\Gamma^{N_{k+1}^V}) = \sum_{i=1}^{N_{k+1}^V} \int_{C_i(\Gamma^{N_{k+1}^V})} (v_{k+1} - \gamma_i^{N_{k+1}^V})^2 \, \mathbb{P}(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}) \tag{5}$$

where $\Gamma^{N_{k+1}^V} = \{\gamma_1^{N_{k+1}^V}, \gamma_2^{N_{k+1}^V}, \ldots, \gamma_{N_{k+1}^V}^{N_{k+1}^V}\}$ is the N_{k+1}^V-quantizer at time t_{k+1} relative to $\widetilde{V}_{t_{k+1}}$.

By definition, the distortion $D_{k+1}(\Gamma^{N_{k+1}^V})$ depends on the distribution of $\widetilde{V}_{t_{k+1}}$, which is, in general, unknown. Nevertheless, thanks to lemma 2.3, we can rewrite the distortion function (5) in terms of the conditional law of $\widetilde{V}_{t_{k+1}}$ given $\widetilde{V}_{t_k}$ and of the law of $\widetilde{V}_{t_k}$, thus obtaining a recursive formula to compute the stationary quantizer. In fact, the distribution function of $\widetilde{V}_{t_k}$ can be written as follows, for $k \ge 1$,

$$\begin{aligned}\mathbb{P}(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}) &= \mathrm{d}v_{k+1} \int_{\mathbb{R}} \phi_{m_k(v_k),\sigma_k(v_k)}(v_{k+1}) \mathbb{P}(\widetilde{V}_{t_k} \in \mathrm{d}v_k) \\ &= \mathrm{d}v_{k+1} \mathbb{E}\left[\phi_{m_k(\widetilde{V}_{t_k}),\sigma_k(\widetilde{V}_{t_k})}(v_{k+1})\right]\end{aligned} \tag{6}$$

where $\phi_{m,\sigma}$ denotes the density function associated with a Normal distribution $\mathcal{N}(m, \sigma^2)$. Inserting equation (6) in equation (5) we find:

$$\begin{aligned}D_{k+1}(\Gamma^{N_{k+1}^V}) = \sum_{i=1}^{N_{k+1}^V} \int_{C_i(\Gamma^{N_{k+1}^V})} &\left(v_{k+1} - \gamma_i^{N_{k+1}^V}\right)^2 \\ &\times \mathbb{E}\left[\phi_{m_k(\widetilde{V}_{t_k}),\sigma_k(\widetilde{V}_{t_k})}(v_{k+1})\right] \mathrm{d}v_{k+1}\end{aligned} \tag{7}$$

and now, replacing $\widetilde{V}$ with $\widehat{V}$ in the above equation, we have the following discrete version of the distortion function, that we denote by $\widehat{D}$:

$$\begin{aligned}\widehat{D}_{k+1}(\Gamma^{N_{k+1}^V}) &= \sum_{i=1}^{N_{k+1}^V} \int_{C_i(\Gamma^{N_{k+1}^V})} \left(v_{k+1} - \gamma_i^{N_{k+1}^V}\right)^2 \\ &\quad \times \mathbb{E}\left[\phi_{m_k(\widehat{V}_{t_k}),\sigma_k(\widehat{V}_{t_k})}(v_{k+1})\right] \mathrm{d}v_{k+1} \\ &= \sum_{i=1}^{N_{k+1}^V} \int_{C_i(\Gamma^{N_{k+1}^V})} \left(v_{k+1} - \gamma_i^{N_{k+1}^V}\right)^2\end{aligned}$$

$$\times \sum_{j=1}^{N_k^V} \phi_{m_k(\gamma_j^{N_k^V}),\sigma_k(\gamma_j^{N_k^V})}(v_{k+1}) \times \mathbb{P}(\widehat{V}_{t_k} \in C_j(\Gamma^{N_k^V}))\, \mathrm{d}v_{k+1}$$

where $\Gamma^{N_k^V} = \{\gamma_1^{N_k^V}, \gamma_2^{N_k^V}, \ldots, \gamma_{N_k^V}^{N_k^V}\}$ is the N_k^V-quantizer at time t_k relative to $\widetilde{V}_{t_k}$. The distortion function is now characterized recursively and it is possible to compute its gradient and Hessian matrix, hence to obtain stationary quantizers for $k = 1$ and so on, via e.g. the Newton-Raphson procedure.

We refer the interested reader to Pagès and Sagna (2015) for a complete background on recursive marginal quantization, in the one-dimensional setting, that includes the error analysis.

3. Quantization of a stochastic volatility model

We now extend the one dimensional approach developed in the previous section to the case where there is a one-dimensional stochastic volatility factor. We shall see that this extension is not trivial.

3.1. Model dynamics and Euler discretization

Let us consider a general stochastic volatility model described by the following dynamics:

$$\begin{cases} \mathrm{d}S_t = rS_t\mathrm{d}t + \alpha(S_t, V_t)\left(\rho \mathrm{d}B_t + \sqrt{1-\rho^2}\mathrm{d}B_t^{\perp}\right) \\ \mathrm{d}V_t = b(t, V_t)\mathrm{d}t + a(t, V_t)\mathrm{d}B_t. \end{cases} \tag{8}$$

with $S_0 = s_0 > 0$, $V_0 = v_0 > 0$. Here S is the price process, V is the variance process (more precisely, the term $\alpha(S_t, V_t)/S_t$ is usually referred as to the volatility process) and B and $B^{\perp}$ are independent one-dimensional Brownian motions under the risk neutral probability measure $\mathbb{Q}$ and $\rho \in (-1, 1)$. The assumption of a constant short interest rate r can be easily relaxed.

The functions $\alpha : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $b : [0, T] \times \mathbb{R} \to \mathbb{R}$ and $a : [0, T] \times \mathbb{R} \to \mathbb{R}$ are measurable and they satisfy the boundedness and Lipschitz conditions which ensure the existence of a strong solution of the above SDE for every $s_0, v_0 \in \mathbb{R}$ (see e.g.(Rogers and Williams 2000, Ch. V, theorem 11.2):

ASSUMPTION 3.1 *We suppose that the following conditions hold:*

(i) *there exists C_T such that, for every $t \in [0, T]$*

$$|b(t, 0)| + |a(t, 0)| \le C_T;$$

(ii) *there exists a real positive constant $[b]_{Lip}$ such that for every $t \in [0, T]$ and every $x, y \in \mathbb{R}$, $|b(t, x) - b(t, y)| \le [b]_{Lip}|x - y|$;*

(iii) *there exists a real positive constant $[\Sigma]_{Lip}$ such that for every $t \in [0, T]$ and every $u, w \in \mathbb{R}^2$, with $u = (u_1, u_2)$, $v = (w_1, w_2)$, $||\Sigma(t, u) - \Sigma(t, w)|| \le [\Sigma]_{Lip}|u - w|$, where*

$$\Sigma(t, u) := \begin{pmatrix} \rho\alpha(u_1, u_2) & \sqrt{1-\rho^2}\alpha(u_1, u_2) \\ a(t, u_2) & 0 \end{pmatrix}. \tag{9}$$

Having fixed a time horizon $T > 0$ and a time discretization grid $\{0 = t_0, t_1, \ldots, t_M = T\}$, with constant time step $\Delta := T/M$, we consider the Euler schemes for S and V: $\widetilde{S}_0 = s_0$, $\widetilde{V}_0 = v_0$ and for $k \ge 1$

$$\widetilde{S}_{t_{k+1}} = \widetilde{S}_{t_k} + r\widetilde{S}_{t_k}\Delta + \alpha(\widetilde{S}_{t_k}, \widetilde{V}_{t_k}) \times \left(\rho\sqrt{\Delta}Z_k + \sqrt{1-\rho^2}\sqrt{\Delta}Z_k^{\perp}\right) \tag{10}$$

$$\widetilde{V}_{t_{k+1}} = \widetilde{V}_{t_k} + b(t_k, \widetilde{V}_{t_k})\Delta + a(t_k, \widetilde{V}_{t_k})\sqrt{\Delta}Z_k, \tag{11}$$

where $t_k = k\Delta$ and Z_k and $Z_k^{\perp}, k = 0, \ldots, M$ are i.i.d. standard Gaussian random variables. From the Euler scheme of the process V in equation (11) we obtain†

$$\sqrt{\Delta}Z_k = \frac{\widetilde{V}_{t_{k+1}} - \widetilde{V}_{t_k} - b(t_k, \widetilde{V}_{t_k})\Delta}{a(t_k, \widetilde{V}_{t_k})}. \tag{12}$$

Hence, the stochastic component in equation (10), given (12), becomes

$$\rho\sqrt{\Delta}Z_k + \sqrt{1-\rho^2}\sqrt{\Delta}Z_k^{\perp} = \rho\left(\frac{\widetilde{V}_{t_{k+1}} - \widetilde{V}_{t_k} - b(t_k, \widetilde{V}_{t_k})\Delta}{a(t_k, \widetilde{V}_{t_k})}\right) + \sqrt{1-\rho^2}\sqrt{\Delta}Z_k^{\perp}, \tag{13}$$

which gives an equivalent formulation of the Euler scheme for the price process S in equation (10). For the reader's ease, we rewrite the Euler scheme for the pair (S, V):

$$\widetilde{S}_{t_{k+1}} = \widetilde{S}_{t_k} + r\widetilde{S}_{t_k}\Delta + \rho\alpha(\widetilde{S}_{t_k}, \widetilde{V}_{t_k}) \times \left(\frac{\widetilde{V}_{t_{k+1}} - \widetilde{V}_{t_k} - b(t_k, \widetilde{V}_{t_k})\Delta}{a(t_k, \widetilde{V}_{t_k})}\right) + \sqrt{1-\rho^2}\alpha(\widetilde{S}_{t_k}, \widetilde{V}_{t_k})\sqrt{\Delta}Z_k^{\perp} \tag{14}$$

$$\widetilde{V}_{t_{k+1}} = \widetilde{V}_{t_k} + b(t_k, \widetilde{V}_{t_k})\Delta + a(t_k, \widetilde{V}_{t_k})\sqrt{\Delta}Z_k. \tag{15}$$

The following Lemma is an immediate consequence of equations (14) and (11) and it represents the natural extension of lemma 2.3 to our bi-dimensional setting.

LEMMA 3.2 *For every $0 \le k \le M - 1$, conditionally on the event $\{\widetilde{S}_{t_k} = s_k, \widetilde{V}_{t_{k+1}} = v_{k+1}, \widetilde{V}_{t_k} = v_k\}$, the random variable $\widetilde{S}_{t_{k+1}}$ is Gaussian:*

$$\mathcal{L}\left(\widetilde{S}_{t_{k+1}} \middle| \{\widetilde{S}_{t_k} = s_k, \widetilde{V}_{t_{k+1}} = v_{k+1}, \widetilde{V}_{t_k} = v_k\}\right) = \mathcal{N}\left(\bar{m}_k(s_k, v_{k+1}, v_k), \bar{\sigma}_k^2(s_k, v_k)\right), \tag{16}$$

with

$$\bar{m}_k(s_k, v_{k+1}, v_k) := s_k + rs_k\Delta + \rho\alpha(s_k, v_k) \times \left(\frac{v_{k+1} - v_k - b(t_k, v_k)\Delta}{a(t_k, v_k)}\right)$$

$$\bar{\sigma}_k^2(s_k, v_k) := (1-\rho^2)(\alpha(s_k, v_k))^2\Delta.$$

In the following subsections, we are going to explain how the bi-dimensional stochastic process (S, V) can be quantized via recursive marginal quantization. This will consist in quantizing the family of random variables $(\widetilde{S}_{t_k}, \widetilde{V}_{t_k})_{k\in\{0,M\}}$ in equations (14) and (11) in a recursive way, thanks to lemma 3.2. The idea is the following:

†We implicitly assume that the model is non degenerate, i.e. $a(t_k, \widetilde{V}_{t_k}) \neq 0$ with probability 1, for every k.

- the (one dimensional) process V can be (independently) quantized as explained in section 2.1. This will be done in section 3.2;
- for what concerns S, it will be necessary to develop an *ad hoc* procedure, since for every $k \geq 0$, the stationary quantizer for $\widetilde{S}_{t_{k+1}}$ will depend on those of $\widetilde{S}_{t_k}$, $\widetilde{V}_{t_k}$ and $\widetilde{V}_{t_{k+1}}$ as the above lemma 3.2 suggests. The quantization of S will be the object of section 3.3.

Notation: Since we will deal with quantization grids for both S and V, from now on, for $k = 0, \ldots, M-1$ we will denote by $\widehat{\mathbf{v}}^{N_{k+1}^V} = (\widehat{v}_1^{N_{k+1}^V}, \ldots, \widehat{v}_{N_{k+1}^V}^{N_{k+1}^V})$ the quantizer for $\widetilde{V}_{t_{k+1}}$ and by $\widehat{\mathbf{s}}^{N_{k+1}^S} = (\widehat{s}_1^{N_{k+1}^S}, \ldots, \widehat{s}_{N_{k+1}^S}^{N_{k+1}^S})$ the quantizer for $\widetilde{S}_{t_{k+1}}$.

3.2. *Quantization of the process V*

Let us consider equation (11): the quantization of the random variables $(\widetilde{V}_{t_k})_{k\in\{0,\ldots,M\}}$ follows the same lines as in Pagès and Sagna (2015), as we explained in section 2.1. We briefly recall here, for reader's convenience, the main steps in the quantization procedure, focusing on the procedure at time t_{k+1}, hence supposing that the first k steps of the recursive quantization algorithm have already been performed.

- **Obtaining a Stationary Grid for $\widetilde{V}_{t_{k+1}}$**
  Suppose we are at time t_{k+1} and we aim at quantizing $\widetilde{V}_{t_{k+1}}$. In the recursive marginal quantization setting, this means that we have already quantized all the random variables $\widetilde{V}_{t_j}$ for $j = 0, \ldots, k$.
  Our final target is finding a stationary quantizer for $\widetilde{V}_{t_{k+1}}$, i.e. a critical point of the distortion function D_{k+1} in equation (7), namely:

$$D_{k+1}(\widehat{\mathbf{v}}^{N_{k+1}^V}) = \sum_{i=1}^{N_{k+1}^V} \int_{C_i(\widehat{\mathbf{v}}^{N_{k+1}^V})} \left(v_{k+1} - \widehat{v}_i^{N_{k+1}^V}\right)^2 \times \mathbb{E}\left[\phi_{m_k(\widetilde{V}_{t_k}),\sigma_k(\widetilde{V}_{t_k})}(v_{k+1})\right] \mathrm{d}v_{k+1}.$$

  As explained in Section 2.1, if we replace $\widetilde{V}$ by $\widehat{V}$, then the discrete distortion reads:

$$\begin{aligned}\widehat{D}_{k+1}(\widehat{\mathbf{v}}^{N_{k+1}^V}) &= \sum_{i=1}^{N_{k+1}^V} \int_{C_i(\widehat{\mathbf{v}}^{N_{k+1}^V})} \left(v_{k+1} - \widehat{v}_i^{N_{k+1}^V}\right)^2 \\ &\quad \times \sum_{j=1}^{N_k^V} \phi_{m_k(\widehat{v}_j^{N_k^V}),\sigma_k(\widehat{v}_j^{N_k^V})}(v_{k+1}) \\ &\quad \times \mathbb{P}(\widehat{V}_{t_k} \in C_j(\widehat{\mathbf{v}}^{N_k^V}))\, \mathrm{d}v_{k+1} \end{aligned} \tag{17}$$

  where $\widehat{\mathbf{v}}^{N_k^V} = \{\widehat{v}_1^{N_k^V}, \widehat{v}_2^{N_k^V}, \ldots, \widehat{v}_{N_k^V}^{N_k^V}\}$ is the N_k^V-quantizer at time t_k relative to $\widetilde{V}_{t_k}$, which is already known, together with the weights $\mathbb{P}(\widehat{V}_{t_k} \in C_j(\widehat{\mathbf{v}}^{N_k^V}))$, $j = 1, \ldots, N_k^V$, associated to every point of the grid.
  It is, then, possible to compute the gradient and the Hessian matrix of $\widehat{D}_{k+1}$, hence to obtain a stationary quantizer at time t_{k+1} via, e.g. the Newton-Raphson procedure.
  For reader's convenience, we provide in appendix 1 the gradient and the Hessian matrix for $\widehat{D}_{k+1}$.
- **Approximating the distribution of $\widetilde{V}_{t_{k+1}}$**
  Remember that

$$\mathbb{P}\left(\widehat{V}_{t_{k+1}} = \widehat{v}_i^{N_{k+1}^V}\right) = \mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in C_i(\widehat{\mathbf{v}}^{N_{k+1}^V})\right).$$

  Once the quantizer $\widehat{\mathbf{v}}^{N_{k+1}^V}$ has been obtained, the probabilities (weights) associated to every point of the grid can be approximated as follows:

$$\begin{aligned}&\mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in C_i(\widehat{\mathbf{v}}^{N_{k+1}^V})\right) \\ &\quad \approx \sum_{j=1}^{N_k^V} \left[\Phi_{0,1}\left(\widehat{v}_{k+1,i^+}(\widehat{v}_j^{N_k})\right) - \Phi_{0,1}\left(\widehat{v}_{k+1,i^-}(\widehat{v}_j^{N_k})\right)\right] \mathbb{P}(\widetilde{V}_{t_k} \in C_j(\widehat{\mathbf{v}}^{N_k^V})),\end{aligned} \tag{18}$$

  where $\Phi_{0,1}$ denotes the cumulative distribution function of a standard Gaussian random variable and where

$$\begin{aligned}\widehat{v}_{k+1,i^+}(\widehat{v}_j^{N_k^V}) &:= \frac{\widehat{v}_i^{N_{k+1}^V} + \widehat{v}_{i+1}^{N_{k+1}^V} - 2m_k(\widehat{v}_j^{N_k^V})}{2\sigma_k(\widehat{v}_j^{N_k^V})}, \\ \widehat{v}_{k+1,i^-}(\widehat{v}_j^{N_k^V}) &:= \frac{\widehat{v}_{i-1}^{N_{k+1}^V} + \widehat{v}_i^{N_{k+1}^V} - 2m_k(\widehat{v}_j^{N_k^V})}{2\sigma_k(\widehat{v}_j^{N_k^V})}.\end{aligned} \tag{19}$$

- **Transition probabilities from t_k to t_{k+1} for the process $\widetilde{V}$**
  In the quantization phase, we also obtain, for $i \in \{1, \ldots, N_{k+1}^V\}$ and $j \in \{1, \ldots, N_k^V\}$, the approximated transition probabilities:

$$\begin{aligned}&\mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in C_i(\widehat{\mathbf{v}}^{N_{k+1}^V}) \,\middle|\, \widetilde{V}_{t_k} \in C_j(\widehat{\mathbf{v}}^{N_k^V})\right) \\ &\quad = \Phi_{0,1}\left(\widehat{v}_{k+1,i^+}(\widehat{v}_j^{N_k^V})\right) - \Phi_{0,1}\left(\widehat{v}_{k+1,i^-}(\widehat{v}_j^{N_k^V})\right).\end{aligned}$$

3.3. *Quantization of the price process S*

We now focus on the quantization of S at time t_{k+1}, i.e. on the recursive marginal quantization of $\widetilde{S}_{t_{k+1}}$, supposing the first k steps of the recursive quantization algorithm have been performed. As done in previous section, we will highlight the main steps of the procedure in the following subsections, which constitute the main contribution of the paper.

3.3.1. Obtaining a stationary grid for $\widetilde{S}_{t_{k+1}}$. Recall that we denote by $\widehat{\mathbf{s}}^{N_{k+1}} = (\widehat{s}_1^{N_{k+1}}, \ldots, \widehat{s}_{N_{k+1}}^{N_{k+1}}))$ an N_{k+1} quantizer of $\widetilde{S}_{t_{k+1}}$ at time t_{k+1}. The distortion function D_{k+1} takes here the following form (recall equation (5)):

$$D_{k+1}(\widehat{\mathbf{s}}^{N_{k+1}}) = \sum_{i=1}^{N_{k+1}} \int_{C_i(\widehat{\mathbf{s}}^{N_{k+1}})} \left(s_{k+1} - \widehat{s}_i^{N_{k+1}}\right)^2 \mathbb{P}(\widetilde{S}_{t_{k+1}} \in \mathrm{d}s_{k+1}).$$

(20)

Thanks to lemma 3.2, we know that in our stochastic volatility setting the distribution of $\widetilde{S}_{t_{k+1}}$ depends not only on $\widetilde{S}_{t_k}$, but also on $\widetilde{V}_{t_k}$ and $\widetilde{V}_{t_{k+1}}$. So, in order to obtain a stationary quantizer for $\widetilde{S}_{t_{k+1}}$ by looking for the critical points of D_{k+1}, first of all it is necessary to make explicit $\mathbb{P}(\widetilde{S}_{t_{k+1}} \in \mathrm{d}s_{k+1})$: this is done in the following Lemma whose proof is given in appendix 2.

LEMMA 3.3 *For every $k \in \{0, \dots, M-1\}$ we have:*

$$\begin{aligned}\mathbb{P}(\widetilde{S}_{t_{k+1}} \in \mathrm{d}s_{k+1}) &= \int_{\mathbb{R}^3} \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k) \\ &\quad \times \phi_{\bar{m}_k(s_k, v_{k+1}, v_k), \bar{\sigma}_k(s_k, v_k)}(s_{k+1})\mathrm{d}s_{k+1}, \end{aligned} \tag{21}$$

where $\bar{m}_k(s_k, v_{k+1}, v_k)$ and $\bar{\sigma}_k(s_k, v_k)$ were defined in lemma 3.2.

With this result, it is now possible to compute the gradient and the Hessian matrix of the distortion function in equation (20) in a very efficient way. The interested reader can find in the appendix 3 the gradient and Hessian matrix of the distortion function at time t_{k+1}. Moreover, in the appendix 4, we compute the gradient and the Hessian matrix for the discrete version of the distortion function, to be used in the Newton-Raphson procedure to find a stationary quantizer for $\widetilde{S}_{t_{k+1}}$.

3.3.2. Approximating the distribution of $\widetilde{S}_{t_{k+1}}$. We now have all the ingredients to prove the main result of the paper, namely the formula giving the joint probability $\mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}))$. This is a non trivial step in the recursive procedure, due to the presence of $\widetilde{V}$ in the dynamics of $\widetilde{S}$. As a corollary, we will obtain the weights associated to every point in the quantizer $\widehat{\mathbf{s}}^{N_{k+1}^S}$.

Notation: We will denote by N_k^S the size of a quantizer for $\widetilde{S}_{t_k}$, while N_k^V will be the size of a quantizer for $\widetilde{V}_{t_k}$.

PROPOSITION 3.4 *Let $\widehat{\mathbf{s}}^{N_{k-1}^S}$, $\widehat{\mathbf{v}}^{N_k^V}$ and $\widehat{\mathbf{v}}^{N_{k-1}^V}$ be stationary quantizers.*

The approximated joint probability of $(\widetilde{S}_{t_k}, \widetilde{V}_{t_{k+1}}, \widetilde{V}_{t_k})$ is given by the following recursive formula, for $k = 1, \dots, M$:

$$\begin{aligned}&\mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V})) \\ &\approx \left(\Phi_{0,1}\big(\widehat{v}_{k+1,b^+}(v_c^{N_k^V})\big) - \Phi_{0,1}\big(\widehat{v}_{k+1,b^-}(v_c^{N_k^V})\big)\right) \\ &\quad \times \left(\sum_{d=1}^{N_{k-1}^S}\sum_{e=1}^{N_{k-1}^V}\left[\Phi_{0,1}\big(\widehat{s}_{k,a^+}(s_d^{N_{k-1}^S}, v_c^{N_k^V}, v_e^{N_{k-1}^V})\big)\right.\right. \\ &\quad \left.\left. - \Phi_{0,1}\big(\widehat{s}_{k,a^-}(s_d^{N_{k-1}^S}, v_c^{N_k^V}, v_e^{N_{k-1}^V})\big)\right] \right. \\ &\quad \left. \cdot \mathbb{P}(\widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \right. \\ &\quad \left. \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V}))\right), \end{aligned} \tag{22}$$

where $\widehat{v}_{k+1,b^-}(v_c^{N_k^V})$ are defined in (19) and

$$\begin{aligned}&\widehat{s}_{k+1,j^+}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V}) \\ &\quad = \frac{s_j^{N_{k+1}^S} + s_{j+1}^{N_{k+1}^S} - 2\bar{m}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})}{2\bar{\sigma}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_c^{N_k^V})} \\ &\widehat{s}_{k+1,j^-}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V}) \\ &\quad = \frac{s_{j-1}^{N_{k+1}^S} + s_j^{N_{k+1}^S} - 2\bar{m}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})}{2\bar{\sigma}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_c^{N_k^V})}. \end{aligned} \tag{23}$$

Moreover, for $k = 0$, i.e. at the first step of the quantization phase, we have that

$$\begin{aligned}&\mathbb{P}(\widetilde{S}_{t_0} \in C_a(\widehat{\mathbf{s}}^{N_0^S}), \widetilde{V}_{t_1} \in C_b(\widehat{\mathbf{v}}^{N_1^V}), \widetilde{V}_{t_0} \in C_c(\widehat{\mathbf{v}}^{N_0^V})) \\ &\quad = \left[\Phi_{0,1}\left(\widehat{v}_{1,b^+}(\widehat{v}_c^{N_0^V})\right) - \Phi_{0,1}\left(\widehat{v}_{1,b^-}(\widehat{v}_c^{N_0^V})\right)\right] \\ &\quad \cdot \mathbb{P}(\widetilde{S}_{t_0} \in C_a(\widehat{\mathbf{s}}^{N_0^S}))\mathbb{P}(\widetilde{V}_{t_0} \in C_c(\widehat{\mathbf{v}}^{N_0^V})). \end{aligned} \tag{24}$$

Proof See appendix 5. □

The following corollary, which is a direct consequence of the previous proposition, explains how to compute the weights of the stationary quantization grid (recall that $\mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S})) = \mathbb{P}(\widehat{S}_{t_k} = \widehat{s}_a^{N_k^S})$).

COROLLARY 3.5 *The weights associated to the points in the stationary quantizer $\widehat{\mathbf{s}}^{N_k^S}$ of $\widetilde{S}_{t_k}$ can be computed using the following formula:*

$$\begin{aligned}&\mathbb{P}\left(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S})\right) \\ &\quad = \sum_{b=1}^{N_{k+1}^V}\sum_{c=1}^{N_k^V}\mathbb{P}\left(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \right. \\ &\qquad \left. \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V})\right). \end{aligned} \tag{25}$$

3.3.3. Transition probabilities for $\tilde{S}$. From proposition 3.4, we can easily compute the transition probabilities also for the price process $\tilde{S}$.

COROLLARY 3.6 *The transition probabilities from time t_{k-1} to time t_k can be computed using the following formula:*

$$\begin{aligned}&\mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}) | \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S})) \\ &\approx \frac{1}{\mathbb{P}(\widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}))} \\ &\quad \times \sum_{c=1}^{N_k^V}\sum_{e=1}^{N_{k-1}^V}\left[\Phi_{0,1}\big(\widehat{s}_{k,a^+}(s_d^{N_{k-1}^S}, v_c^{N_k^V}, v_e^{N_{k-1}^V})\big)\right. \\ &\quad \left. - \Phi_{0,1}\big(\widehat{s}_{k,a^-}(s_d^{N_{k-1}^S}, v_c^{N_k^V}, v_e^{N_{k-1}^V})\big)\right] \\ &\quad \times \mathbb{P}(\widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})). \end{aligned}$$

In conclusion, the quantization of $\widetilde{S}_{t_{k+1}}$ requires the knowledge of the quantizers of $\widetilde{S}_{t_k}$, of $\widetilde{V}_{t_{k+1}}$ and of $\widetilde{V}_{t_k}$. The quantizers of $\widetilde{V}$ can be computed independently using the formulas in section 3.2 and they can be kept off-line. Then, stationary quantizers for $\widetilde{S}$ can be obtained recursively in a very fast way.

3.4. Error analysis

Using quantization to approximate a random variable X allows to estimate an expected value of a function of X (i.e. $\mathbb{E}[f(X)]$) by the finite sum $\mathbb{E}[f(\widehat{X})]$, where $\widehat{X}$ is a quantization of X. In

the case when we focus on prices of vanilla options, that is when $X = S_T$, with S in equation (14), and f is the payoff function, the following result (see e.g. Pagès and Printems 2005, Pagès and Sagna 2015) is crucial: if f is Lipschitz continuous, then

$$\left|\mathbb{E}[f(X)] - \mathbb{E}[f(\widehat{X})]\right| \leq [f]_{\text{Lip}} ||X - \widehat{X}||_2,$$

where $[f]_{\text{Lip}}$ is the Lipschitz constant of f and $||X - \widehat{X}||_2$ is the quadratic quantization error. The above inequality suggests that quantization is a good approximation method, as long as the quantization error $||X - \widehat{X}||_2$ can be bounded. This motivates the error analysis developed in this subsection.

We denote by Y the pair of price and variance processes introduced in section 3, namely $Y := (S, V)$.

PROPOSITION 3.7 *Consider the model in equation (8) and recall the definition of Σ in equation (9). Moreover, suppose that assumption 3.1 holds true. Given a time horizon $T > 0$ and a time discretization grid $\{0 = t_0, \ldots, t_M = T\}$, for every $k \in \{1, \ldots, M\}$ and for any $\eta \in (0, 1]$ we have*

$$||Y_{t_k} - \widehat{Y}_{t_k}||_2 \leq \sqrt{\frac{C_T}{M}} + K_{2,2,\eta} \sum_{\ell=1}^{k} a_\ell \times (b, \Sigma, t_k, \Delta, y_0, L, 2+\eta) \frac{1}{\sqrt{N_k^V}},$$

where C_T is a positive constant depending only on T, $K_{2,2,\eta}$ is a universal constant, $L := \max\left\{[b]_{Lip}, \max_{s\in[0,T]} b(s,0), [\Sigma]_{Lip}, ||\Sigma(\cdot,0)||_{sup}\right\}$ and a_ℓ are real constants depending on $[b]_{Lip}$ and $[\Sigma]_{Lip}$, that do not explode when M goes to infinity.

Proof We have

$$||Y_{t_k} - \widehat{Y}_{t_k}||_2 \leq ||Y_{t_k} - \widetilde{Y}_{t_k}||_2 + ||\widetilde{Y}_{t_k} - \widehat{Y}_{t_k}||_2$$

where we recall that $\widetilde{Y}_{t_k}$ denotes the Euler scheme at time t_k relative to the pair (S, V) in equations (14) and (11). Thus, the recursive marginal quantization error is two-sided, being the sum of the error due to the Euler scheme and the one due to the vector quantization of $\widetilde{Y}_{t_k}$ via $\widehat{Y}_{t_k}$. For what concerns the former, under our hypotheses we have

$$\mathbb{E}\left(\sup_{k\in\{0,\ldots,M\}} |Y_{t_k} - \widetilde{Y}_{t_k}|^2\right) \leq \frac{C_T}{M},$$

where C_T only depends on T. Hence, for every $k \in \{1, \ldots, M\}$:

$$||Y_{t_k} - \widetilde{Y}_{t_k}||_2 \leq \sqrt{\frac{C_T}{M}}.$$

On the other hand, the error coming from the quantization of $\widetilde{Y}_{t_k}$, can be controlled thanks to theorem 3.1 in Pagès and Sagna (2015). Indeed, applying their result in our two-dimensional setting we find:

$$||\widetilde{Y}_{t_k} - \widehat{Y}_{t_k}||_2 \leq K_{2,2,\eta} \sum_{\ell=1}^{k} a_\ell (b, \Sigma, t_k, \Delta, y_0, L, 2+\eta) \frac{1}{\sqrt{N_k^V}}$$

and the proof is complete. □

There are still some questions that deserve further investigation. For example, beside the Stein and Stein model, the stochastic volatility models considered in our numerical illustration, present some potential issues as the coefficients are not globally Lipschitz, so that the global error results presented in this Section have to be tailored to those settings. This issue is beyond the scope of this paper and we leave it for future research.

4. Numerical illustration

In this numerical section we apply the methodology to some stochastic volatility models. We first consider the Stein and Stein (1991) affine model, for which the Fourier based method is available for pricing. Of course our methodology can be easily applied as well to other affine stochastic volatility models like e.g. the Heston (1993) one, for which a slightly different approach has been developed in Fiorin *et al.* (2015). Then we consider some non affine models that do not admit closed form formulas for the price of vanillas, like the celebrated SABR model introduced by Hagan *et al.* (2002) and the α-hypergeometric volatility model recently introduced by Da Fonseca and Martini (2016). In these cases, only approximated formulas based on asymptotic results are available, besides of course the Monte Carlo approach.

For the pricing of vanillas, we recall that having computed the grids $\mathbf{s}_T = (s_T^i)_{i=1,\ldots,N_s}$ and the weights $\mathbb{P}(S_T \in C_i(\mathbf{s}_T))$, the price at time 0 of a Call option with maturity T and strike price K can be straightforwardly approximated as follows:

$$C_{0,T} = e^{-rT}\mathbb{E}[(S_T - K)^+] \approx e^{-rT} \sum_{i=1}^{N_s} \max(s_T^i - K, 0)\mathbb{P}(S_T \in C_i(\mathbf{s}_T)).$$

All the computations have been performed using *Matlab2015* on a CPU 2.4 GHz and 8 Gb memory computer. The *Matlab* code is available upon request.

4.1. Vanillas in the Stein and Stein (1991) model

Stein and Stein (1991) assume that the instantaneous variance follows an Ornstein–Uhlenbeck process. The model is described by the following dynamics:

$$\mathrm{d}S_t = rS_t\mathrm{d}t + V_tS_t(\rho\mathrm{d}B_t + \sqrt{1-\rho^2}\mathrm{d}B_t^\perp)$$
$$\mathrm{d}V_t = \kappa(\theta - V_t)\mathrm{d}t + \xi\mathrm{d}B_t.$$

Let $t_k = k\Delta$, $k = 0, \ldots, M$ and denote S_{t_k} by S_k and V_{t_k} by V_k. The Euler scheme reads

$$\widetilde{S}_{k+1} = \widetilde{S}_k + r\widetilde{S}_k\Delta + \rho\widetilde{V}_k\widetilde{S}_k\sqrt{\Delta}Z_k + \sqrt{1-\rho^2}\widetilde{V}_k\widetilde{S}_k\sqrt{\Delta}Z_k^\perp,$$
$$\widetilde{V}_{k+1} = \widetilde{V}_k + \kappa(\theta - \widetilde{V}_k)\Delta + \xi\sqrt{\Delta}Z_k,$$

where $Z_k = B_{k+1} - B_k$ and $Z_k^\perp = B_{k+1}^\perp - B_k^\perp$ are i.i.d. Gaussian random variables.

Exploiting the Cholesky decomposition, we get

$$\widetilde{S}_{k+1} = \widetilde{S}_k + \frac{\rho}{\xi}\left[\widetilde{V}_{k+1} - \widetilde{V}_k - \kappa(\theta - \widetilde{V}_k)\Delta\right]\widetilde{V}_k\widetilde{S}_k + \sqrt{1-\rho^2}\sqrt{\Delta}\widetilde{V}_k\widetilde{S}_kZ_k^\perp$$
$$\widetilde{V}_{k+1} = \widetilde{V}_k + \kappa(\theta - \widetilde{V}_k)\Delta + \xi\sqrt{\Delta}Z_k.$$

Table 1. Comparison between pricing Call and Put options via closed form formulas and quantization in the Stein and Stein (1991) model. When the strike is lower or equal than 100 we are pricing a Call option, when it is greater than 100 we are pricing a Put option. The inverse of the Hessian matrix, which is tridiagonal and symmetric, is calculated using the LU-decomposition. Computational time for the quantization of the volatility process: 0.079 s. Computational time for the quantization of the price process: 10.9689 s.

Strike	Benchmark price	Quantization price	Relative error (%)
$K = 60$	40.1947	40.2294	0.0864
$K = 65$	35.4500	35.4804	0.0858
$K = 70$	30.8727	30.9078	0.1138
$K = 75$	26.5234	26.5582	0.1311
$K = 80$	22.4604	22.4744	0.0623
$K = 85$	18.7361	18.7488	0.0677
$K = 90$	15.3903	15.3825	0.0506
$K = 95$	12.4451	12.4277	0.1403
$K = 100$	9.9070	9.8681	0.3925
$K = 105$	12.8284	12.6995	1.0053
$K = 110$	16.0608	15.9153	0.9058
$K = 115$	19.6277	19.4646	0.8308
$K = 120$	23.4893	23.3295	0.6801
$K = 125$	27.6002	27.4349	0.5990
$K = 130$	31.9187	31.7473	0.5370
$K = 135$	36.4042	36.2324	0.4719
$K = 140$	41.0213	40.8566	0.4016

We consider the following set of parameters for the Stein and Stein (1991) model:

$$r = 0.04, \quad \xi = 0.05, \quad \kappa = 3, \quad \theta = 0.3,$$
$$\rho = -0.5, \quad T = 1, \quad V_0 = 0.25, \quad S_0 = 100,$$

while for the time discretization and for the quantization we take

$$\Delta = \frac{1}{20}, \quad N_k^V = 30, \quad N_k^S = 30, \quad k = 1, \dots, M,$$

where N_k^V and N_k^S are the sizes, respectively, of the grids for the volatility and for the price.

We compare the price given by the quantization with a benchmark price given by the closed formulas as in the paper of Schöbel and Zhu (1999). The computation of the grids is made using a Newton-Raphson procedure both for the volatility and the price processes.

From table 1 it turns out that pricing of vanillas via quantization is accurate. The quantization of the volatility process is almost immediate, while for the underlying it takes about 10 s, which is much less than what we can obtain with a classic quantization method based on stochastic algorithms. In table 2 we provide relative errors in percentage for longer maturities up to 5 years. We emphasize that, although pricing via quantization cannot be comparable with the (faster) Fourier approach available for the affine models, quantization can be applied also to non vanilla products as we are going to show in the next section. What is more, quantization can be applied also to non affine models, as we will show in the next subsection. In figure 2 we report the implied volatility smile induced by the Stein and Stein (1991) model of our example. Figure 3 displays the quantization grids used in the procedure.

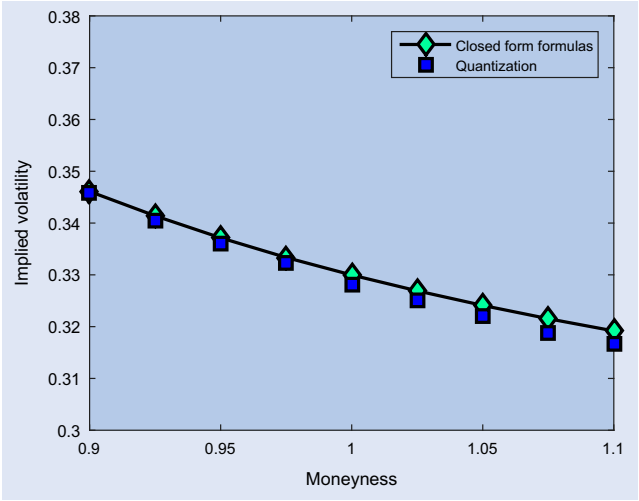

Figure 2. Example of fit of the implied volatility smile in the Stein and Stein (1991) model. We consider 9 strikes with maturity 1 year.

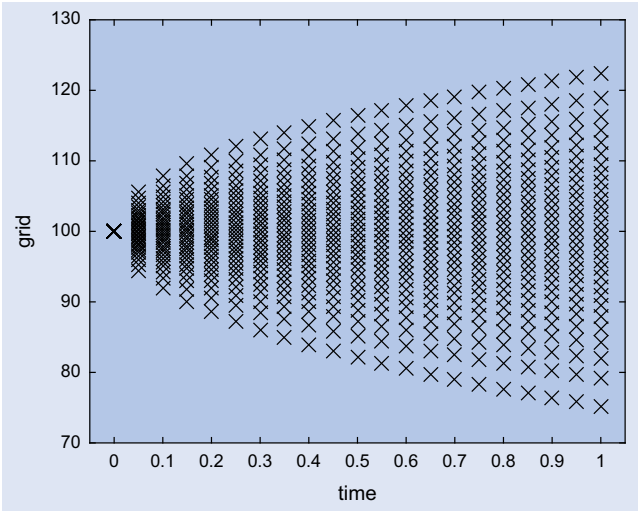

Figure 3. Quantization grids for the Stein and Stein (1991) model. Time step is taken as $\Delta = \frac{1}{20}$, while $N_k^V = N_k^S = 30$, $k = 1, \dots, M$ is the size of the grids for the volatility and the underlying.

4.2. Vanillas in the SABR model of Hagan et al. (2002)

The Hagan *et al.* (2002) (non affine) model is characterized by the following dynamics (here S denotes the forward price under the forward measure):

$$\mathrm{d}S_t = V_t S_t^\beta (\rho \mathrm{d}B_t + \sqrt{1-\rho^2}\mathrm{d}B_t^\perp)$$
$$\mathrm{d}V_t = \alpha V_t \mathrm{d}B_t.$$

The Euler scheme is given by

$$\widetilde{S}_{k+1} = \widetilde{S}_k + \rho \widetilde{S}_k^\beta \widetilde{V}_k \sqrt{\Delta} Z_k + \sqrt{1-\rho^2}\widetilde{S}_k^\beta \widetilde{V}_k \sqrt{\Delta} Z_k^\perp$$
$$\widetilde{V}_{k+1} = \widetilde{V}_k + \alpha \widetilde{V}_k \sqrt{\Delta} Z_k$$

and exploiting the Cholesky decomposition we get

$$\widetilde{S}_{k+1} = \widetilde{S}_k + \frac{\rho}{\alpha}(\widetilde{V}_{k+1} - \widetilde{V}_k)\widetilde{S}_k^\beta + \sqrt{1-\rho^2}\sqrt{\Delta}\widetilde{S}_k^\beta \widetilde{V}_k Z_k^\perp$$
$$\widetilde{V}_{k+1} = \widetilde{V}_k + \alpha \widetilde{V}_k \sqrt{\Delta} Z_k.$$

Table 2. Relative error in percentage between pricing Call and Put options via closed form formulas and quantization in the Stein and Stein (1991) model for $T = 1/2, 2/3, 1, 2, 3, 5$ years. When the strike is smaller than 100 we consider Calls, while when it is greater than 100 we take Puts.

Strike	6 months	8 months	1 year	2 years	3 years	5 years
$K = 80$	0.0163	0.4872	0.0623	0.2878	0.7504	0.1215
$K = 85$	0.0919	0.6252	0.0677	0.2429	0.7759	0.0288
$K = 90$	0.2047	0.7637	0.0506	0.1525	0.7373	0.0196
$K = 95$	0.3347	1.0095	0.1403	0.0033	0.6099	0.2327
$K = 100$	0.5438	1.3913	0.3925	0.2134	0.6213	0.1305
$K = 105$	0.6965	0.6958	1.0053	1.0965	0.5548	0.4547
$K = 110$	0.5546	0.4912	0.9058	1.0417	0.5707	0.3362
$K = 115$	0.4218	0.3275	0.8308	0.9646	0.6164	0.2715
$K = 120$	0.3410	0.2650	0.6801	0.8993	0.6939	0.0645

Table 3. Comparison between pricing a Call via closed form (approximation) formulas and quantization in the SABR model of Hagan *et al.* (2002). When the strike is lower or equal than 100, we are pricing a Call option, when it is greater than 100 we are pricing a Put option. The inverse of the Hessian matrix, which is tridiagonal and symmetric, is calculated using the LU-decomposition. Computational time for the quantization of the volatility process: 0.0796 s. Computational time for the price process: 12.1071 s.

Strike	Benchmark price	Quantization price	Relative error (%)
$K = 60$	40.0475	40.0195	0.0699
$K = 65$	35.1042	35.0694	0.0992
$K = 70$	30.2179	30.1716	0.1531
$K = 75$	25.4346	25.3742	0.2375
$K = 80$	20.8257	20.7607	0.3120
$K = 85$	16.4909	16.4208	0.4252
$K = 90$	12.5524	12.4800	0.5766
$K = 95$	9.1346	9.0628	0.7859
$K = 100$	6.3321	6.2564	1.1946
$K = 105$	9.1780	9.1159	0.6768
$K = 110$	12.6315	12.5798	0.4098
$K = 115$	16.5922	16.5446	0.2869
$K = 120$	20.9336	20.8837	0.2386
$K = 125$	25.5358	25.5008	0.1371
$K = 130$	30.3040	30.2700	0.1121
$K = 135$	35.1719	35.1424	0.0838
$K = 140$	40.0975	40.0770	0.0511

The parameters chosen for the pricing are

$$\alpha = 0.224 \quad \beta = 0.75 \quad \rho = -0.824$$
$$T = 1 \quad V_0 = 0.0719 \quad S_0 = 100,$$

and for the time discretization and the quantization we use as before:

$$\Delta = \frac{1}{20}, \quad N_k^V = 30, \quad N_k^S = 30, \quad k = 1, \dots, M.$$

The benchmark price is given by a Monte Carlo procedure with 10^6 paths. The quantization grids are obtained with a Newton-Raphson procedure both for the volatility and the price processes.

Table 3 displays the prices given by quantization and by the approximation given by Hagan *et al.* (2002). Also in this case we see that pricing is accurate even if not immediate when compared with the approximation formulas. Note, however, that pricing should be compared with the true benchmark given by the Monte Carlo simulation that takes around 30 s with 5×10^5 paths. In table 4, we provide relative errors in percentage for longer maturities up to 5 years.

4.3. *Vanillas in the α-hypergeometric model of Da Fonseca and Martini (2016)*

The α-hypergeometric model of Da Fonseca and Martini (2016) is described by the following dynamics:

$$\mathrm{d}S_t = S_t e^{V_t}(\rho \mathrm{d}B_t + \sqrt{1-\rho^2}\mathrm{d}B_t^{\perp}),$$
$$\mathrm{d}V_t = (a - be^{\alpha V_t})\mathrm{d}t + \sigma \mathrm{d}B_t.$$

Note that the volatility remains strictly positive in this model. The Euler scheme of the model is given by

$$\widetilde{S}_{k+1} = \widetilde{S}_k + \rho \widetilde{S}_k e^{\widetilde{V}_k}\sqrt{\Delta}Z_k + \sqrt{1-\rho^2}\widetilde{S}_k e^{\widetilde{V}_k}\sqrt{\Delta}Z_k^{\perp}$$
$$\widetilde{V}_{k+1} = \widetilde{V}_k + (a - be^{\alpha \widetilde{V}_k})\Delta + \sigma\sqrt{\Delta}Z_k$$

from which we get

$$\widetilde{S}_{k+1} = \widetilde{S}_k + \frac{\rho}{\sigma}\big(\widetilde{V}_{k+1} - \widetilde{V}_k - (a - be^{\alpha \widetilde{V}_k})\Delta\big)\widetilde{S}_k e^{\widetilde{V}_k} + \sqrt{1-\rho^2}\sqrt{\Delta}\widetilde{S}_k e^{\widetilde{V}_k} Z_k^{\perp}$$
$$\widetilde{V}_{k+1} = \widetilde{V}_k + (a - be^{\alpha \widetilde{V}_k})\Delta + \sigma\sqrt{\Delta}Z_k.$$

The parameters selected for pricing are

$$a = 1, \quad b = 1, \quad \alpha = 2, \quad \sigma = 1, \quad \rho = -0.2,$$
$$T = 1, \quad V_0 = \log(0.25), \quad S_0 = 100,$$

and for the time discretization and the quantization we use as before:

$$\Delta = \frac{1}{20}, \quad N_k^V = 30, \quad N_k^S = 30, \quad k = 1, \dots, M.$$

The benchmark price is given by a Monte Carlo procedure with 10^6 paths, due to the fact that, up to our knowledge, no closed form formulas are available for pricing vanillas in this setting. The quantization grids are obtained with a Newton-Raphson procedure both for the volatility and the price processes.

Table 5 shows the comparison between the quantization prices and the Monte Carlo ones for the α-hypergeometric model of Da Fonseca and Martini (2016). Here, the computational time required for the quantization of the underlying

Table 4. Relative error in percentage between pricing a Call via closed form (approximation) formulas and quantization in the SABR model of Hagan *et al*. (2002) for $T = 1/2, 2/3, 1, 2, 3, 5$ years. When the strike is smaller than 100 we consider Calls, while when it is greater than 100 we take Puts.

Strike	6 months	8 months	1 year	2 years	3 years	5 years
$K = 80$	0.1442	0.2010	0.3120	0.6003	0.9684	1.4529
$K = 85$	0.2365	0.3124	0.4252	0.7643	1.0255	1.5444
$K = 90$	0.4489	0.4702	0.5766	0.8719	1.0788	1.6065
$K = 95$	0.7356	0.6999	0.7859	1.0474	1.2036	1.6463
$K = 100$	1.0909	1.1682	1.1946	1.1475	1.5285	1.6887
$K = 105$	0.5121	0.6236	0.6768	0.8248	0.8341	1.0884
$K = 110$	0.2909	0.3296	0.4098	0.4961	0.5877	0.8479
$K = 115$	0.2128	0.2330	0.2869	0.4475	0.5004	0.7162
$K = 120$	0.1134	0.1472	0.2386	0.3428	0.4650	0.6570

Table 5. Comparison between pricing a Call via closed form formulas and quantization in the α-hypergeometric model of Da Fonseca and Martini (2016). When the strike is lower or equal than 100 we are pricing a Call option, when it is greater than 100 we are pricing a Put option. The inverse of the Hessian matrix, which is tridiagonal and symmetric, is calculated using the LU-decomposition. Computational time for the quantization of the volatility process: 0.1109 s. Computational time for the price process: 6.8141 s.

Strike	Benchmark price	Quantization price	MC price	Rel. error (%)
$K = 60$	44.1617	44.4653	[42.7957, 45.5276]	0.6876
$K = 65$	40.2496	40.5186	[38.9443, 41.5548]	0.6684
$K = 70$	36.5165	36.7643	[35.2690, 37.7640]	0.6788
$K = 75$	32.9795	33.2305	[31.7867, 34.1723]	0.7611
$K = 80$	29.6571	29.8535	[28.5158, 30.7984]	0.6623
$K = 85$	26.5673	26.7371	[25.4743, 27.6603]	0.6391
$K = 90$	23.7232	23.8925	[22.6753, 24.7711]	0.7134
$K = 95$	21.1375	21.2545	[20.1315, 22.1435]	0.5537
$K = 100$	18.8159	18.8704	[17.8488, 19.7831]	0.2896
$K = 105$	21.8264	21.7331	[21.1575, 22.4954]	0.4276
$K = 110$	25.0123	24.8636	[24.2792, 25.7453]	0.5944
$K = 115$	28.4277	28.2293	[27.6277, 29.2277]	0.6980
$K = 120$	32.0491	31.8511	[31.1796, 32.9187]	0.6180
$K = 125$	35.8509	35.5602	[34.9094, 36.7924]	0.8109
$K = 130$	39.8070	39.5573	[38.7913, 40.8227]	0.6272
$K = 135$	43.8971	43.5695	[42.8051, 44.9891]	0.7462
$K = 140$	48.1015	47.8111	[46.9314, 49.2716]	0.6037

Table 6. Relative error in percentage between pricing a Call via closed form formulas and quantization in the α-hypergeometric model of Da Fonseca and Martini (2016) for $T = 1/2, 2/3, 1, 2, 3, 5$ years. When the strike is smaller than 100 we consider Calls, while when it is greater than 100 we take Puts.

Strike	6 months	8 months	1 year	2 years	3 years	5 years
$K = 80$	0.1727	0.3272	0.6128	0.6376	1.3259	0.8820
$K = 85$	0.1991	0.2761	0.5919	3.3031	0.9801	1.1970
$K = 90$	0.2367	0.2067	0.6676	0.6973	0.6082	0.9134
$K = 95$	0.4040	0.1172	0.5123	0.5973	1.2570	0.7741
$K = 100$	0.8696	0.0282	0.2568	0.6090	0.6490	0.7868
$K = 105$	0.9646	1.1712	0.2504	0.7204	1.0005	0.7519
$K = 110$	0.7996	1.1538	0.4478	1.7931	0.9324	0.7088
$K = 115$	0.7227	1.1119	0.5725	1.5081	0.8718	0.8042
$K = 120$	0.6479	0.9285	0.5040	1.0315	0.4906	0.8892

process is lower than in the previous models, but it still remains quite high (about 7 s). On the other hand, a Monte Carlo simulation with 10^5 paths takes about 10 s and gives prices which are much less accurate (see column 'MC price' in table 5). In table 6, we provide relative errors in percentage for longer maturities up to 5 years.

4.4. Pricing of an exotic: Volatility Corridor Swap

Here we consider the pricing of a Volatility Corridor Swap, that is a non vanilla contract belonging to the class of equity-volatility derivatives. A Corridor Variance Swap is a generalization of a standard variance swap in that the volatility is

Table 7. Pricing of a Volatility Corridor Swap in the SABR model for different values of the corridor. We fixed $\alpha = 0.25$, $\beta = 0.75$, $\rho = -0.75$.

Corridor	Benchmark price	Quantization price	MC price	Relative error (%)
$L = 80, H = 120$	0.2487	0.2484	[0.2397, 0.2576]	0.0973
$L = 85, H = 115$	0.2447	0.2451	[0.2344, 0.2549]	0.1673
$L = 90, H = 110$	0.2278	0.2297	[0.2131, 0.2430]	0.8551

Table 8. Pricing of a Volatility Corridor Swap in the SABR model for different values of the parameter β. We fixed $\alpha = 0.25$, $\rho = -0.75$, $L = 80$, $H = 120$.

Parameter	Benchmark price	Quantization price	MC price	Relative error (%)
$\beta = 0.5$	0.2498	0.2492	[0.2413, 0.2588]	0.2296
$\beta = 0.7$	0.2498	0.2491	[0.2408, 0.2583]	0.2738
$\beta = 0.9$	0.2282	0.2301	[0.2139, 0.2437]	0.8554

Table 9. Pricing of a Volatility Corridor Swap in the SABR model for different values of the parameter α. We fixed $\beta = 0.75$, $\rho = -0.75$, $L = 80$, $H = 120$.

Parameter	Benchmark price	Quantization price	MC price	Relative error (%)
$\alpha = 0.1$	0.2494	0.2496	[0.2453, 0.2535]	0.0751
$\alpha = 0.3$	0.2482	0.2476	[0.2377, 0.2590]	0.2138
$\alpha = 0.5$	0.2462	0.2440	[0.2290, 0.2635]	0.8969

Table 10. Pricing of a Volatility Corridor Swap in the SABR model for different values of the parameter ρ. We fixed $\alpha = 0.25$, $\beta = 0.75$, $L = 80$, $H = 120$.

Parameter	Benchmark price	Quantization price	MC price	Relative error (%)
$\rho = -0.4$	0.2490	0.2486	[0.2401, 0.2581]	0.1522
$\rho = -0.6$	0.2490	0.2486	[0.2399, 0.2579]	0.1552
$\rho = -0.8$	0.2487	0.2484	[0.2396, 0.2574]	0.0861

Table 11. Pricing of a Volatility Corridor Swap in the α-Hypergeometric model for different values of the parameter α. We fixed $a = 1$, $b = 1$, $\sigma = 1$, $\rho = -0.2$, $L = 80$, $H = 120$.

Parameter	Benchmark price	Quantization price	MC price	Relative error (%)
$\alpha = 1$	0.2344	0.2342	[0.2131, 0.2557]	0.0773
$\alpha = 1.5$	0.2396	0.2396	[0.2176, 0.2617]	0.0012
$\alpha = 2$	0.2427	0.2442	[0.2203, 0.2651]	0.6362

Table 12. Pricing of a Volatility Corridor Swap in the α-Hypergeometric model for different values of the parameter ρ. We fixed $a = 1$, $b = 1$, $\sigma = 1$, $\alpha = 1$, $L = 80$, $H = 120$.

Parameter	Benchmark price	Quantization price	MC price	Relative error (%)
$\rho = -0.4$	0.2343	0.2348	[0.2135, 0.2550]	0.2415
$\rho = -0.6$	0.2349	0.2372	[0.2150, 0.2548]	0.9588
$\rho = -0.8$	0.2361	0.2384	[0.2173, 0.2549]	0.9623

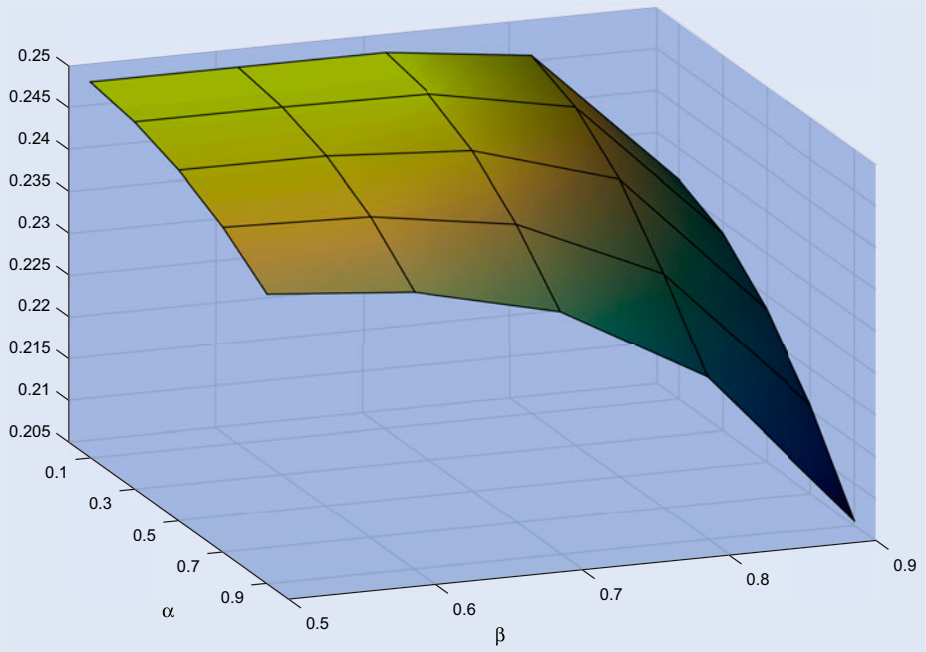

Figure 4. Price of a Volatility Corridor Swap in the SABR model as a function of parameters α, β.

accumulated only when the underlying stock is within a prespecified interval $[L, H]$, see Carr and Lewis (2004). For the affine models, a Fourier approach has been developed for these products (see e.g. Da Fonseca *et al.* (2015) and references therein). In this subsection, we focus on the SABR and the α- hypergeometric models for which such Fourier approach cannot be applied and we show that the quantization method we propose is very efficient when compared with the Monte Carlo benchmark.

Consider the formula giving the price at time 0 of a Volatility Corridor Swap (wlog we do not consider the presence of the strike price due to the linearity of the contract):

$$S(L,H,T) := \mathbb{E}\left[\frac{1}{T}\int_0^T V_u \mathbb{1}_{\{L<S_u<H\}}\mathrm{d}u\right],$$

that can be rewritten as

$$\frac{1}{T}\sum_{i=0}^{M-1}\mathbb{E}\int_{t_i}^{t_{i+1}} V_u \mathbb{1}_{\{L<S_u<H\}}\mathrm{d}u,$$

so that the expected value can be approximated using the quantization method, i.e.:

$$\begin{aligned}
&\frac{1}{T}\sum_{i=0}^{M-1}\mathbb{E}\int_{t_i}^{t_{i+1}} V_u \mathbb{1}_{\{L<S_u<H\}}\mathrm{d}u\\
&\quad\simeq \frac{1}{T}\sum_{i=0}^{M-1}\mathbb{E}\int_{t_i}^{t_{i+1}} \widehat{V}_u \mathbb{1}_{\{L<\widehat{S}_u<H\}}\mathrm{d}u\\
&\quad= \frac{1}{T}\sum_{i=0}^{M-1}\int_{t_i}^{t_{i+1}}\mathbb{E}\left(\widehat{V}_u \mathbb{1}_{\{L<\widehat{S}_u<H\}}\right)\mathrm{d}u.
\end{aligned}$$

Hence, we can approximate $S(L,H,T)$ by means of $\widehat{S}(L,H,T)$:

$$\begin{aligned}
&\widehat{S}(L,H,T)\\
&\quad= \frac{1}{T}\sum_{i=0}^{M-1}\int_{t_i}^{t_{i+1}}\sum_{j,\ell}\widehat{v}_j \mathbb{1}_{\{L<\widehat{s}_\ell<H\}}\\
&\qquad\times \mathbb{P}(\widehat{S}_u=\widehat{s}_\ell, \widehat{V}_u=\widehat{v}_j)\,\mathrm{d}u\\
&\quad= \frac{\Delta}{T}\sum_{i=0}^{M-1}\sum_{j=1}^{N_i^V}\sum_{\ell=1}^{N_i^S}\widehat{v}_j^{N_i^V}\mathbb{1}_{\{L<\widehat{s}_\ell^{N_i^S}<H\}}\\
&\qquad\times \mathbb{P}(\widehat{S}_{t_i}=\widehat{s}_\ell^{N_i^S}, \widehat{V}_{t_i}=\widehat{v}_j^{N_i^V})\\
&\quad= \frac{1}{M}\sum_{i=0}^{M-1}\sum_{j=1}^{N_i^V}\sum_{\ell=1}^{N_i^S}\widehat{v}_j^{N_i^V}\mathbb{1}_{\{L<\widehat{s}_\ell^{N_i^S}<H\}}\\
&\qquad\times \mathbb{P}(\widehat{S}_{t_i}=\widehat{s}_\ell^{N_i^S}, \widehat{V}_{t_i}=\widehat{v}_j^{N_i^V}).
\end{aligned}$$

We use the same grids used in the pricing of vanillas, so that the pricing is immediate (figure 4).

In tables 7–10, we display the price of the Volatility Corridor Swap for different values of the parameters within the SABR model. We notice that pricing is fast and accurate. What is more, once the SABR model has been quantized, pricing of vanillas or Volatility Corridor Swaps is almost immediate, while a Monte Carlo simulation still needs a computational cost (about 25 s with 1.5×10^5 paths for a Volatility Corridor Swap in order to reach a similar precision of the recursive marginal quantization approach). Here, the benchmark price is computed via a Monte Carlo simulation with 10^6 paths.

In tables 11 and 12, we repeat the same experiments for the α-hypergeometric model. The benchmark price is still computed via a Monte Carlo simulation with 10^6 paths. Also in this case, the pricing is fast and accurate. Moreover, once the model has been quantized, pricing of vanillas or Volatility Corridor Swaps is almost immediate, while a Monte Carlo simulation requires a higher computational cost (about 13 s with $\times 10^5$ paths for a Volatility Corridor Swap).

5. Conclusion

In this paper, we have presented the first recursive quantization based approach for pricing in the presence of stochastic volatility. Our framework is flexible enough to include most popular volatility models and it applies to affine as well as to non affine models. Pricing of vanilla options is accurate and, thanks to the knowledge of the transition density for the underlying, it is possible to price efficiently also non vanilla options. We tested our procedure on three different stochastic volatility settings with an increasing difficulty in terms of analytical tractability. Moreover, as an application to non vanilla options, we considered the pricing of a Corridor Volatility Swap under the SABR and the α-hypergeometric (non affine) models, where we showed that pricing is fast and accurate when compared with the Monte Carlo one. Recursive marginal quantization can be applied also to the pricing of other path dependent derivatives, for which some attempts have already been considered in the literature from an optimal quantization perspective (see e.g. Pagès and Wilbertz 2012, Bally *et al*. 2005 for American options, and Sagna (2012) for barrier options). Although existing results on marginal quantization are still preliminary, we believe they are promising enough to motivate further research in the field. In particular, we are confident to be able to improve the numerical performance achieved with classic techniques like in Longstaff and Schwartz (2001) for American options and Lipton and McGhee (2002) for Barrier options. Another room for improvement relies in the possibility to allow for more sophisticated discretization schemes, like e.g. Milstein, which are currently under investigation.

Let us also mention that hybrid models, such as stochastic local volatility (SLV) models, are becoming popular in practice, since they allow one to improve pricing accuracy for exotic options, including volatility options, while matching the whole implied volatility surface. For a comprehensive study on the application of hybrid SLV models to option pricing, we refer e.g. to the PhD thesis of Tian (2013). In general a hybrid SLV model is specified in the form

$$\begin{cases} \mathrm{d}S_t = \mu_1(S_t, t)\mathrm{d}t + L(S_t, t)\sigma_1(S_t, V_t, t)\mathrm{d}W_t^1 \\ \mathrm{d}V_t = \mu_2(V_t, t)\mathrm{d}t + \sigma_2(S_t, V_t, t)\mathrm{d}W_t^2, \end{cases}$$

with $\mathrm{d}W_t^1\mathrm{d}W_t^2 = \rho\mathrm{d}t$ and where L is known as a 'leverage function', which has to be carefully determined by market information. We immediately see that our lemma 3.2 and all the subsequent technical results in section 3, needed to perform the quantization algorithm, still hold true, obviously, as soon as the leverage function has been specified. Hence, our numerical procedure can be safely applied also to hybrid SLV models.

Finally, from a calibration perspective, we expect to gain in robustness, as in the case of local volatility models investigated in Callegaro *et al*. (2015), where they showed that marginal quantization performs very well and does not suffer the typical problems related to the choice of the starting point for the calibration.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Bally, V., Pagès, G. and Printems, J., A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance*, 2005, **15**(1), 119–168.

Bates, D.S., Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Rev. Financ. Stud.*, 1996, **9**, 69–107.

Bormetti, G., Callegaro, G., Livieri, G. and Pallavicini, A., A backward Monte Carlo approach to exotic option pricing. Working Paper, 2015. Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2686115.

Callegaro, G., Fiorin, L. and Grasselli, M., Quantized calibration in local volatility. *Risk Mag.*, 2015, **28**(4), 62–67.

Carr, P. and Lewis, K., Corridor variance swaps. *Risk Mag.*, 2004, February, 67–72.

Charalambous, C., Charitou, A. and Kaourou, F., Comparative analysis of artificial neural network models: Application in bankruptcy prediction. *Ann. Oper. Res.*, 2000, **99**, 403–425.

Chesney, M. and Scott, L., Pricing European currency options: A comparison of the modified Black–Scholes model and a random variance model. *J. Financ. Quant. Anal.*, 1989, **24**, 267–284.

Christoffersen, P., Heston, S. and Jacobs, K., The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Manage. Sci.*, 2009, **55**(12), 1914–1932.

Da Fonseca, J., Gnoatto, A. and Grasselli, M., Analytic pricing of volatility-equity options within affine models: An efficient conditioning technique. *Oper. Res. Lett.*, 2015, **43**(6), 601–607.

Da Fonseca, J. and Martini, C., The α-hypergeometric stochastic volatility model. *Stoch. Process. Appl.*, 2016, **5**, 1472–1502.

Date, P. and Islyaev, S., A fast calibrating volatility model for option pricing. *Eur. J. Oper. Res.*, 2015, **243**(2), 599–606.

Duffie, D., Pan, J. and Singleton, K., Transform analysis and asset pricing for affine jump-diffusion. *Econometrica*, 2000, **68**, 1343–1376.

Fiorin, L., Sagna, L. and Pagès, G., Componentwise and Markovian product quantization of an $\mathbb{R}^d$-valued Euler diffusion process with applications. Working Paper, 2015. Available online at: http://arxiv.org/abs/1511.01758.

Graf, S. and Luschgy, H., *Foundations of Quantization for Probability Distributions*, 2000 (Springer: New York).

Hagan, P.S., Kumar, D., Lesniewski, A.S. and Woodward, D.E., Managing smile risk. *Wilmott Mag.*, 2002, September, 84–108.

Heston, S.L., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 1993, **6**, 327–343.

Hull, J. and White, A., The pricing of options on assets with stochastic volatilities. *J. Finance*, 1987, **42**, 281–300.

Jacobs, K. and Li, X., Modeling the dynamics of credit spreads with stochastic volatility. *Manage. Sci.*, 2008, **54**(6), 1176–1188.

Lipton, A. and McGhee, W., Universal barriers. *Risk Mag.*, 2002, **15**, 81–85.

Longstaff, F.A. and Schwartz, E.S., Valuing American options by simulation: A simple least-squares approach. *Rev. Financ. Stud.*, 2001, **14**(1), 113–147.

Pagès, G., Introduction to optimal vector quantization and its applications for numerics, 2014. Preprint available online at: http://hal.archives-ouvertes.fr/INSMI/hal-01034196.

Pagès, G. and Printems, J., Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods Appl.*, 2005, **11**(4), 407–446.

Pagès, G. and Sagna, A., Recursive marginal quantization of the Euler scheme of a diffusion process. *Appl. Math. Finance*, 2015, **22**(5), 463–498.

Pagès, G. and Wilbertz, B., *Optimal Delaunay and Voronoi Quantization Schemes for Pricing American Style Options*, pp. 171–217, 2012 (Springer: Berlin Heidelberg).

Rogers, L. and Williams, D., *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*, 2000 (Cambridge University Press: Cambridge).

Sagna, A., Pricing of barrier options by marginal functional quantization. *Monte Carlo Methods Appl.*, 2012, **17**, 371–398.
Schöbel, R. and Zhu, J., Stochastic volatility with an Ornstein Uhlenbeck process: An extension. *Eur. Finance Rev.*, 1999, **3**, 23–46.
Sesana, D., Marazzina, D. and Fusai, G., Pricing exotic derivatives exploiting structure. *Eur. J. Oper. Res.*, 2014, **236**(1), 369–381.
Stein, E.M. and Stein, J.C., Stock price distributions with stochastic volatility: An analytic approach. *Rev. Financ. Stud.*, 1991, **4**, 727–752.
Thangavel, K. and Ashok Kumar, D., Optimization of code book in vector quantization. *Ann. Oper. Res.*, 2006, **143**, 317–325.
Tian, Y., The hybrid stochastic-local volatility model with applications in pricing FX options. PhD Thesis. School of Mathematical Sciences Monash University, Australia, 2013. Available online at SSRN 2399935.

Appendix 1. Gradient and Hessian for the quantization of $\widehat{V}_{t_{k+1}}$

The gradient and Hessian of the distortion function $\widehat{D}_{k+1}$ in equation (17) are:

$$\begin{aligned}
&\frac{\partial \widehat{D}_{k+1}}{\partial \widehat{v}_j^{N_{k+1}^V}}(\widehat{\mathbf{v}}^{N_{k+1}^V}) \\
&\quad = 2\sum_{i=1}^{N_k^V}\Bigg\{\big(\widehat{v}_j^{N_{k+1}^V} - m_k(\widehat{v}_i^{N_k^V})\big)\Big(\Phi_{0,1}\big(\widehat{v}_{k+1,j^+}(\widehat{v}_i^{N_k^V})\big) \\
&\qquad - \Phi_{0,1}\big(\widehat{v}_{k+1,j^-}(\widehat{v}_i^{N_k^V})\big)\Big) + \sigma_k(\widehat{v}_i^{N_k^V})\Big(\phi_{0,1}\big(\widehat{v}_{k+1,j^+}(\widehat{v}_i^{N_k^V})\big) \\
&\qquad - \phi_{0,1}\big(\widehat{v}_{k+1,j^-}(\widehat{v}_i^{N_k^V})\big)\Big)\Bigg\}\mathbb{P}(\widetilde{V}_{t_k} \in C_i(\widehat{\mathbf{v}}^{N_k^V})). \qquad \text{(A1)}
\end{aligned}$$

where

$$\begin{aligned}
\widehat{v}_{k+1,j^+}(\widehat{v}_i^{N_k^V}) &:= \frac{\widehat{v}_j^{N_{k+1}^V} + \widehat{v}_{j+1}^{N_{k+1}^V} - 2m_k(\widehat{v}_i^{N_k^V})}{2\sigma_k(\widehat{v}_i^{N_k^V})}, \\
\widehat{v}_{k+1,j^-}(\widehat{v}_i^{N_k^V}) &:= \frac{\widehat{v}_{j-1}^{N_{k+1}^V} + \widehat{v}_j^{N_{k+1}^V} - 2m_k(\widehat{v}_i^{N_k^V})}{2\sigma_k(\widehat{v}_i^{N_k^V})}, \qquad \text{(A2)}
\end{aligned}$$

and with $\Phi_{0,1}$ denoting the cumulative distribution function of a standard Gaussian random variable.

Focusing on the Hessian tri-diagonal matrix, the diagonal terms are:

$$\begin{aligned}
&\frac{\partial^2 \widehat{D}_{k+1}}{\partial^2 \widehat{v}_j^{N_{k+1}^V}}(\widehat{\mathbf{v}}^{N_{k+1}^V}) \\
&\quad = 2\sum_{i=1}^{N_k^V}\Bigg\{\Big(\Phi_{0,1}\big(\widehat{v}_{k+1,j^+}(\widehat{v}_i^{N_k^V})\big) - \Phi_{0,1}\big(\widehat{v}_{k+1,j^-}(\widehat{v}_i^{N_k^V})\big)\Big) \\
&\qquad - \frac{1}{4\sigma_k(\widehat{v}_i^{N_k^V})}\phi_{0,1}\big(\widehat{v}_{k+1,j^+}(\widehat{v}_i^{N_k^V})\big)\big(\widehat{v}_{j+1}^{N_{k+1}^V} - \widehat{v}_j^{N_{k+1}^V}\big) \\
&\qquad - \frac{1}{4\sigma_k(\widehat{v}_i^{N_k^V})}\phi_{0,1}\big(\widehat{v}_{k+1,j^-}(\widehat{v}_i^{N_k^V})\big)\big(\widehat{v}_j^{N_{k+1}^V} - \widehat{v}_{j-1}^{N_{k+1}^V}\big)\Bigg\} \\
&\qquad \times \mathbb{P}(\widetilde{V}_{t_k} \in C_i(\widehat{\mathbf{v}}^{N_k^V})). \qquad \text{(A3)}
\end{aligned}$$

The sub-diagonal elements are:

$$\begin{aligned}
&\frac{\partial^2 \widehat{D}_{k+1}}{\partial \widehat{v}_j^{N_{k+1}^V}\partial \widehat{v}_{j-1}^{N_{k+1}^V}}(\widehat{\mathbf{v}}^{N_{k+1}^V}) \\
&\quad = -\frac{1}{2}\sum_{i=1}^{N_k^V}\Bigg\{\phi_{0,1}\big(\widehat{v}_{k+1,j^-}(\widehat{v}_i^{N_k^V})\big) \\
&\qquad \times \left(\frac{\widehat{v}_j^{N_{k+1}^V} - \widehat{v}_{j-1}^{N_{k+1}^V}}{\sigma_k(\widehat{v}_i^{N_k^V})}\right)\Bigg\}\mathbb{P}(\widetilde{V}_{t_k} \in C_i(\widehat{\mathbf{v}}^{N_k^V})), \qquad \text{(A4)}
\end{aligned}$$

while the super-diagonal terms are given by

$$\begin{aligned}
&\frac{\partial^2 \widehat{D}_{k+1}}{\partial \widehat{v}_j^{N_{k+1}^V}\partial \widehat{v}_{j+1}^{N_{k+1}^V}}(\widehat{\mathbf{v}}^{N_{k+1}^V}) \\
&\quad = -\frac{1}{2}\sum_{i=1}^{N_k^V}\Bigg\{\phi_{0,1}\big(\widehat{v}_{k+1,j^+}(\widehat{v}_i^{N_k^V})\big) \\
&\qquad \times \left(\frac{\widehat{v}_{j+1}^{N_{k+1}^V} - \widehat{v}_j^{N_{k+1}^V}}{\sigma_k(\widehat{v}_i^{N_k^V})}\right)\Bigg\}\mathbb{P}(\widetilde{V}_{t_k} \in C_i(\widehat{\mathbf{v}}^{N_k^V})). \qquad \text{(A5)}
\end{aligned}$$

Appendix 2. Proof of lemma 3.3

Proof First of all, by using Bayes formula, Fubini Theorem and lemma 3.2 we have:

$$\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_{k+1}} \le x) \\
&\quad = \int_{-\infty}^{x}\mathbb{P}(\widetilde{S}_{t_{k+1}} \in \mathrm{d}s_{k+1}) \\
&\quad = \int_{-\infty}^{x}\int_{\mathbb{R}^3}\mathbb{P}(\widetilde{S}_{t_{k+1}} \in \mathrm{d}s_{k+1}|(\widetilde{S}_{t_k} = s_k,\, \widetilde{V}_{t_{k+1}} \\
&\quad = v_{k+1},\, \widetilde{V}_{t_k} = v_k))\mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k,\, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1},\, \widetilde{V}_{t_k} \in \mathrm{d}v_k) \\
&\quad = \int_{\mathbb{R}^3}\mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k,\, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1},\, \widetilde{V}_{t_k} \in \mathrm{d}v_k) \\
&\qquad \times \int_{-\infty}^{x}\phi_{\bar{m}_k(s_k,v_{k+1},v_k),\bar{\sigma}_k(s_k,v_k)}(s_{k+1})\mathrm{d}s_{k+1} \\
&\quad = \int_{\mathbb{R}^3}\mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k,\, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1},\, \widetilde{V}_{t_k} \in \mathrm{d}v_k) \\
&\qquad \times \Phi_{\bar{m}_k(s_k,v_{k+1},v_k),\bar{\sigma}_k(s_k,v_k)}(x), \qquad \text{(B1)}
\end{aligned}$$

where as usual the Gaussian density ϕ and cumulative Gaussian distribution function Φ are defined resp. as follows:

$$\begin{aligned}
&\phi_{\bar{m}_k(s_k,v_{k+1},v_k),\bar{\sigma}_k(s_k,v_k)}(x) \\
&\quad = \frac{1}{\bar{\sigma}_k(s_k,\, v_k)\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x - \bar{m}_k(s_k,\, v_{k+1},\, v_k)}{\bar{\sigma}_k(s_k,\, v_k)}\right)^2\right], \\
&\Phi_{\bar{m}_k(s_k,v_{k+1},v_k),\bar{\sigma}_k(s_k,v_k)}(x) \\
&\quad = \int_{-\infty}^{x}\phi_{\bar{m}_k(s_k,v_{k+1},v_k),\bar{\sigma}_k(s_k,v_k)}(s_{k+1})\mathrm{d}s_{k+1}.
\end{aligned}$$

We now focus on $\mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k,\, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1},\, \widetilde{V}_{t_k} \in \mathrm{d}v_k)$ that can be rewritten as follows:

$$\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k,\, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1},\, \widetilde{V}_{t_k} \in \mathrm{d}v_k) \\
&= \int_{\mathbb{R}^2}\underbrace{\mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k,\, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}|\widetilde{S}_{t_{k-1}} = s_{k-1},\, \widetilde{V}_{t_k} = v_k,\, \widetilde{V}_{t_{k-1}} = v_{k-1})}_{(a)} \\
&\cdot \underbrace{\mathbb{P}(\widetilde{S}_{t_{k-1}} \in \mathrm{d}s_{k-1},\, \widetilde{V}_{t_k} \in \mathrm{d}v_k,\, \widetilde{V}_{t_{k-1}} \in \mathrm{d}v_{k-1})}_{(b)}, \qquad \text{(B2)}
\end{aligned}$$

The term (b) in (B2) can be easily computed via the recursive algorithm, so let us focus on the term (a).

As it can be seen from equations (14) and (11), conditionally on $\{\widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k, \widetilde{V}_{t_{k-1}} = v_{k-1}\}$, S_{t_k} and $V_{t_{k+1}}$ are independent (non centered) Gaussian, since B and $B^{\perp}$ are. Then we

have that

$$\begin{aligned}(a) &= \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1} | \widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} \\ &\quad = v_k, \widetilde{V}_{t_{k-1}} = v_{k-1}) \\ &= \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k | \widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k, \widetilde{V}_{t_{k-1}} = v_{k-1}) \\ &\quad \times \mathbb{P}(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1} | \widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k, \widetilde{V}_{t_{k-1}} = v_{k-1}) \\ &= \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k | \widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k, \widetilde{V}_{t_{k-1}} = v_{k-1}) \\ &\quad \times \underbrace{\mathbb{P}(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1} | \widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k)}_{(c)},\end{aligned}$$

where the last equality follows from the Markov property of $\widetilde{V}$. We now focus on (c). Notice that $\widetilde{V}_{t_{k+1}}$ given $\widetilde{V}_{t_k}$ is an affine transformation of the standard Gaussian random variable Z_k (recall equation ((11))), hence it is independent of all the random variables Z_ℓ and $Z_\ell^\perp$, for $\ell = 0, \ldots, k-1$. In particular, this means that, given $\widetilde{V}_{t_k}$, $\widetilde{V}_{t_{k+1}}$ is independent of $\widetilde{S}_{t_{k-1}}$. Using this independence and the Bayes rule we have:

$$\begin{aligned}&\mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1} | \widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k\right) \\ &\quad = \frac{\mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k\right)}{\mathbb{P}\left(\widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k\right)} \\ &\quad = \frac{\mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{S}_{t_{k-1}} = s_{k-1} | \widetilde{V}_{t_k} = v_k\right) \mathbb{P}\left(\widetilde{V}_{t_k} = v_k\right)}{\mathbb{P}\left(\widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k\right)} \\ &\quad = \frac{\mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1} | \widetilde{V}_{t_k} = v_k\right) \mathbb{P}\left(\widetilde{S}_{t_{k-1}} = s_{k-1} | \widetilde{V}_{t_k} = v_k\right) \mathbb{P}\left(\widetilde{V}_{t_k} = v_k\right)}{\mathbb{P}\left(\widetilde{S}_{t_{k-1}} = s_{k-1}, \widetilde{V}_{t_k} = v_k\right)} \\ &\quad = \mathbb{P}\left(\widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1} | \widetilde{V}_{t_k} = v_k\right)\end{aligned}$$

So, from lemmas 2.3 and 3.2 it turns out that

$$\begin{aligned}(a) = \phi_{\bar{m}_k(s_{k-1}, v_k, v_{k-1}), \bar{\sigma}_k(s_{k-1}, v_{k-1})}(s_k)\mathrm{d}s_k \\ \cdot \phi_{m_k(v_k), \sigma_k(v_k)}(v_{k+1})\mathrm{d}v_{k+1}.\end{aligned} \tag{B3}$$

Therefore we have an iterative method to compute the distribution of the triplet:

$$\begin{aligned}&\mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k) \\ &\quad = \int_{\mathbb{R}^2} \left(\phi_{\bar{m}_k(s_{k-1}, v_k, v_{k-1}), \bar{\sigma}_k(s_{k-1}, v_{k-1})}(s_k)\mathrm{d}s_k\right) \\ &\quad \times \left(\phi_{m_k(v_k), \sigma_k(v_k)}(v_{k+1})\mathrm{d}v_{k+1}\right) \\ &\quad \times \mathbb{P}(\widetilde{S}_{t_{k-1}} \in \mathrm{d}s_{k-1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k, \widetilde{V}_{t_{k-1}} \in \mathrm{d}v_{k-1}).\end{aligned} \tag{B4}$$

Now, from (B1) we have

$$\begin{aligned}\mathbb{P}(\widetilde{S}_{t_{k+1}} \le x) = \int_{\mathbb{R}^3} \Phi_{\bar{m}_k(s_k, v_{k+1}, v_k), \bar{\sigma}_k(s_k, v_k)}(x)\mathbb{P}(\widetilde{S}_{t_k} \\ \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k),\end{aligned}$$

so that, using (B4) and differentiating, we get

$$\begin{aligned}\mathbb{P}(\widetilde{S}_{t_{k+1}} \in \mathrm{d}s_{k+1}) = \int_{\mathbb{R}^3} \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k) \\ \times \phi_{\bar{m}_k(s_k, v_{k+1}, v_k), \bar{\sigma}_k(s_k, v_k)}(s_{k+1})\mathrm{d}s_{k+1}\end{aligned} \tag{B5}$$

and the proof is completed. □

Appendix 3. Gradient and Hessian for the quantization of $\widetilde{S}_{t_{k+1}}$

The gradient and the Hessian of the distortion function $\widehat{D}_{k+1}$ in equation (20) are

$$\begin{aligned}&\frac{\partial D_{k+1}}{\partial x_j}(\mathbf{x}) \\ &\quad = \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{N_{k+1}} \int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}} (x_j - y)^2 \mathbb{P}(\widetilde{S}_{t_{k+1}} \in \mathrm{d}s_{k+1})\right) \\ &\quad = \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{N_{k+1}} \int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}} (x_j - s_{k+1})^2\right. \\ &\qquad \times \int_{\mathbb{R}^3} \phi_{\bar{m}_k(s_k, v_{k+1}, v_k), \bar{\sigma}_k(s_k, v_k)}(s_{k+1})\mathrm{d}s_{k+1} \\ &\qquad \left.\times \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k)\right) \\ &\quad = \int_{\mathbb{R}^3} \Bigg[\underbrace{\frac{\partial}{\partial x_j}\left(\sum_{i=1}^{N_{k+1}} \int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}} (x_j - s_{k+1})^2 \phi_{\bar{m}_k(s_k, v_{k+1}, v_k), \bar{\sigma}_k(s_k, v_k)}(s_{k+1})\mathrm{d}s_{k+1}\right)}_{(\star\star)} \\ &\qquad \times \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k)\Bigg].\end{aligned}$$

It can be easily deduced that the term $(\star\star)$ is given by

$$\begin{aligned}(\star\star) = 2\Big\{\big(x_j - \bar{m}_k(s_k, v_{k+1}, v_k)\big)\Big[\Phi_{0,1}\big(x_{k+1,j^+}(s_k, v_{k+1}, v_k)\big) \\ -\Phi_{0,1}\big(x_{k+1,j^-}(s_k, v_{k+1}, v_k)\big)\Big] + \bar{\sigma}_k(s_k, v_{k+1}, v_k) \\ \times \Big[\phi_{0,1}\big(x_{k+1,j^+}(s_k, v_{k+1}, v_k)\big) \\ -\phi_{0,1}\big(x_{k+1,j^-}(s_k, v_{k+1}, v_k)\big)\Big]\Big\},\end{aligned}$$

where

$$\begin{aligned}\bar{x}_{k+1,j^+}(s_k, v_{k+1}, v_k) &= \frac{x_j + x_{j+1} - 2\bar{m}_k(s_k, v_{k+1}, v_k)}{2\bar{\sigma}_k(s_k, v_k)}, \\ \bar{x}_{k+1,j^-}(s_k, v_{k+1}, v_k) &= \frac{x_{j-1} + x_j - 2\bar{m}_k(s_k, v_{k+1}, v_k)}{2\bar{\sigma}_k(s_k, v_k)},\end{aligned} \tag{C1}$$

so we get that

$$\begin{aligned}\frac{\partial D_{k+1}}{\partial x_j}(\mathbf{x}) = 2\int_{\mathbb{R}^3} \Big\{&\big(x_j - \bar{m}_k(s_k, v_{k+1}, v_k)\big) \\ &\times \Big[\Phi_{0,1}\big(\bar{x}_{k+1,j^+}(s_k, v_{k+1}, v_k)\big) \\ &- \Phi_{0,1}\big(\bar{x}_{k+1,j^-}(s_k, v_{k+1}, v_k)\big)\Big] \\ &+ \bar{\sigma}_k(s_k, v_k)\Big[\phi_{0,1}\big(\bar{x}_{k+1,j^+}(s_k, v_{k+1}, v_k)\big) \\ &- \phi_{0,1}\big(\bar{x}_{k+1,j^-}(s_k, v_{k+1}, v_k)\big)\Big]\Big\} \\ &\cdot \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k).\end{aligned} \tag{C2}$$

Similarly we can write the components of the (tri-diagonal and symmetric) Hessian matrix:

$$\begin{aligned}&\frac{\partial^2 D_{k+1}}{\partial^2 x_j}(\mathbf{x}) \\ &\quad = 2\int_{\mathbb{R}^3} \Bigg\{\Big(\Phi_{0,1}\big(\bar{x}_{k+1,j^+}(s_k, v_{k+1}, v_k)\big) \\ &\qquad - \Phi_{0,1}\big(\bar{x}_{k+1,j^-}(s_k, v_{k+1}, v_k)\big)\Big) \\ &\qquad - \frac{1}{4\bar{\sigma}_k(s_k, v_k)}\phi_{0,1}\big(\bar{x}_{k+1,j^+}(s_k, v_{k+1}, v_k)\big)\big(x_{j+1} - x_j\big)\end{aligned}$$

$$- \frac{1}{4\bar{\sigma}_k(s_k, v_k)}\phi_{0,1}\big(\bar{x}_{k+1,j^-}(s_k, v_{k+1}, v_k)\big)\big(x_j - x_{j-1}\big)\Big\}$$
$$\cdot \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k);$$

$$\frac{\partial^2 D_{k+1}}{\partial x_j \partial x_{j-1}}(\mathbf{x}) = -\frac{1}{2}\int_{\mathbb{R}^3}\left\{\phi_{0,1}\big(\bar{x}_{k+1,j^-}(s_k, v_{k+1}, v_k)\big)\left(\frac{x_j - x_{j-1}}{\bar{\sigma}_k(s_k, v_k)}\right)\right\} \times \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k);$$

$$\frac{\partial^2 D_{k+1}}{\partial x_j \partial x_{j+1}}(\mathbf{x}) = -\frac{1}{2}\int_{\mathbb{R}^3}\left\{\phi_{0,1}\big(\bar{x}_{k+1,j^+}(s_k, v_{k+1}, v_k)\big)\left(\frac{x_{j+1} - x_j}{\bar{\sigma}_k(s_k, v_k)}\right)\right\} \times \mathbb{P}(\widetilde{S}_{t_k} \in \mathrm{d}s_k, \widetilde{V}_{t_{k+1}} \in \mathrm{d}v_{k+1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k).$$

Appendix 4. Towards the Newton algorithm to quantize $\widetilde{S}_{t_{k+1}}$

In this Appendix, we compute the gradient and the Hessian matrix for the discrete version of the distortion function, to be used in the Newton-Raphson procedure to find a stationary quantizer for $\widetilde{S}_{t_{k+1}}$. We denote by $\widehat{\mathbf{s}}^{N_{k+1}^S} = (\widehat{s}_1^{N_{k+1}^S}, \ldots, \widehat{s}_{N_{k+1}^S}^{N_{k+1}^S})$ and by $\widehat{\mathbf{s}}^{N_k^S} = (\widehat{s}_1^{N_{k+1}^S}, \ldots, \widehat{s}_{N_k^S}^{N_k^S})$ resp. the N_{k+1}^S and N_k^S-dimensional quantizers for $\widetilde{S}_{t_{k+1}}$ and $\widetilde{S}_{t_k}$ and by $\widehat{\mathbf{v}}^{N_{k+1}^V} = (\widehat{v}_1^{N_{k+1}^V}, \ldots, \widehat{v}_{N_{k+1}^V}^{N_{k+1}^V})$ and $\widehat{\mathbf{v}}^{N_k} = (\widehat{v}_1^{N_{k+1}^V}, \ldots, \widehat{v}_{N_k}^{N_k})$ the N_{k+1}^V and N_k^V-dimensional quantizers for $\widetilde{V}_{t_{k+1}}$ and $\widetilde{V}_{t_k}$, respectively. Taking the gradient of the discrete distortion function $\widehat{D}_{k+1}(\widehat{\mathbf{s}}^{N_{k+1}^S})$ gives

$$\frac{\partial \widehat{D}_{k+1}}{\partial \widehat{s}_j^{N_{k+1}^S}}(\widehat{\mathbf{s}}^{N_{k+1}^S}) = 2\sum_{a=1}^{N_k^S}\sum_{b=1}^{N_{k+1}^V}\sum_{c=1}^{N_k^V}\Big\{\big(\widehat{s}_j^{N_{k+1}^V} - \bar{m}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) \times \Big[\Phi_{0,1}\big(\widehat{s}_{k+1,j^+}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) - \Phi_{0,1}\big(\widehat{s}_{k+1,j^-}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big)\Big] + \bar{\sigma}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_c^{N_k^V}) \cdot \Big[\phi_{0,1}\big(\widehat{s}_{k+1,j^+}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) - \phi_{0,1}\big(\widehat{s}_{k+1,j^-}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big)\Big]\Big\} \cdot \mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V})),$$

where $\widehat{s}_{k+1,j^+}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})$ and $\widehat{s}_{k+1,j^-}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})$ have been defined in (23). The components of the tridiagonal discrete distortion Hessian are

$$\frac{\partial^2 \widehat{D}_{k+1}}{\partial^2 \widehat{s}_j^{N_{k+1}^S}}(\widehat{\mathbf{s}}^{N_{k+1}^S}) = 2\sum_{a=1}^{N_k^S}\sum_{b=1}^{N_{k+1}^V}\sum_{c=1}^{N_k^V}\Big\{\Big(\Phi_{0,1}\big(\widehat{s}_{k+1,j^+}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) - \Phi_{0,1}\big(\widehat{s}_{k+1,j^-}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big)\Big)$$
$$- \frac{1}{4\bar{\sigma}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_c^{N_k^V})}\phi_{0,1}\big(\widehat{s}_{k+1,j^+}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) \times \big(s_{j+1}^{N_{k+1}^S} - s_j^{N_{k+1}^S}\big)$$
$$- \frac{1}{4\bar{\sigma}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_c^{N_k^V})}\phi_{0,1}\big(\widehat{s}_{k+1,j^-}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) \times \big(s_j^{N_{k+1}^S} - s_{j-1}^{N_{k+1}^S}\big)\Big\} \cdot \mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}));$$

$$\frac{\partial^2 \widehat{D}_{k+1}}{\partial \widehat{s}_j^{N_{k+1}^S} \partial \widehat{s}_{j-1}^{N_{k+1}^S}}(\widehat{\mathbf{s}}^{N_{k+1}^S}) = -\frac{1}{2}\sum_{a=1}^{N_k^S}\sum_{b=1}^{N_{k+1}^V}\sum_{c=1}^{N_k^V}\left\{\phi_{0,1}\big(\widehat{s}_{k+1,j^-}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) \times \left(\frac{\widehat{s}_j^{N_{k+1}^V} - \widehat{s}_{j-1}^{N_{k+1}^V}}{\bar{\sigma}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_c^{N_k^V})}\right)\right\} \cdot \mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}));$$

$$\frac{\partial^2 \widehat{D}_{k+1}}{\partial \widehat{s}_j^{N_{k+1}^S} \partial \widehat{s}_{j+1}^{N_{k+1}^S}}(\widehat{\mathbf{s}}^{N_{k+1}^S}) = -\frac{1}{2}\sum_{a=1}^{N_k^S}\sum_{b=1}^{N_{k+1}^V}\sum_{c=1}^{N_k^V}\left\{\phi_{0,1}\big(\widehat{s}_{k+1,j^+}(\widehat{s}_a^{N_k^S}, \widehat{v}_b^{N_{k+1}^V}, \widehat{v}_c^{N_k^V})\big) \times \left(\frac{\widehat{s}_{j+1}^{N_{k+1}^V} - \widehat{s}_j^{N_{k+1}^V}}{\bar{\sigma}_k(\widehat{s}_a^{N_k^S}, \widehat{v}_c^{N_k^V})}\right)\right\} \cdot \mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V})).$$

Appendix 5. Proof of proposition 3.4

Proof We can decompose the initial joint distribution in this way:

$$\mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V})) = \sum_{d=1}^{N_{k-1}^S}\sum_{e=1}^{N_{k-1}^V}\mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})),$$

where we can rewrite the generic element of the summation as a function of the extremes of the Voronoi regions $C_a(\widehat{\mathbf{s}}^{N_k^S})$ and $C_b(\widehat{\mathbf{v}}^{N_{k+1}^V})$ as follows:

$$\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_k} \in C_a(\widehat{\mathbf{s}}^{N_k^S}), \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})) \\
&= \mathbb{P}(\widetilde{S}_{t_k} \le s_{a+1/2}^{N_k^S}, \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})) \\
&\quad - \mathbb{P}(\widetilde{S}_{t_k} \le s_{a-1/2}^{N_k^S}, \widetilde{V}_{t_{k+1}} \in C_b(\widehat{\mathbf{v}}^{N_{k+1}^V}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})) \\
&= \mathbb{P}(\widetilde{S}_{t_k} \le s_{a+1/2}^{N_k^S}, \widetilde{V}_{t_{k+1}} \le v_{b+1/2}^{N_{k+1}^V}, \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V}))
\end{aligned}$$

$$
\begin{aligned}
&- \mathbb{P}(\widetilde{S}_{t_k} \le s^{N_k^S}_{a+1/2}, \widetilde{V}_{t_{k+1}} \le v^{N_{k+1}^V}_{b-1/2}, \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}),\\
&\quad \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V}))\\
&- \mathbb{P}(\widetilde{S}_{t_k} \le s^{N_k^S}_{a-1/2}, \widetilde{V}_{t_{k+1}} \le v^{N_{k+1}^V}_{b+1/2}, \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}),\\
&\quad \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V}))\\
&+ \mathbb{P}(\widetilde{S}_{t_k} \le s^{N_k^S}_{a-1/2}, \widetilde{V}_{t_{k+1}} \le v^{N_{k+1}^V}_{b-1/2}, \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}),\\
&\quad \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})),
\end{aligned}
$$

with

$$
s^{N_k^S}_{a+1/2} = \frac{s^{N_k^S}_a + s^{N_k^S}_{a+1}}{2}; \quad s^{N_k^S}_{a-1/2} = \frac{s^{N_k^S}_{a-1} + s^{N_k^S}_a}{2};
$$
$$
v^{N_{k+1}^V}_{b+1/2} = \frac{v_b^{N_{k+1}} + v_{b+1}^{N_{k+1}}}{2}; \quad v^{N_{k+1}^V}_{b-1/2} = \frac{v_{b-1}^{N_{k+1}} + v_b^{N_{k+1}}}{2}.
$$

We now approximate the following term

$$
\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_k} \le s_k, \widetilde{V}_{t_{k+1}} \le v_{k+1}, \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}),\\
&\quad \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})),
\end{aligned}
$$

for general $s_k, v_{k+1} \in \mathbb{R}$.

Remember that, from (B4),

$$
\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_k} \le s_k, \widetilde{V}_{t_{k+1}} \le v_{k+1}, \widetilde{V}_{t_k} \le v_k)\\
&\quad = \int_{-\infty}^{s_k}\int_{-\infty}^{v_{k+1}}\int_{-\infty}^{v_k}\int_{\mathbb{R}^2} \phi_{\bar{m}_k(s_{k-1},v_k,v_{k-1}),\bar{\sigma}_k(s_{k-1},v_{k-1})}\\
&\qquad \times (s_k)\mathrm{d}s_k \phi_{m_k(v_k),\sigma_k(v_k)}(v_{k+1})\mathrm{d}v_{k+1}\\
&\qquad \cdot \mathbb{P}(\widetilde{S}_{t_{k-1}} \in \mathrm{d}s_{k-1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k, \widetilde{V}_{t_{k-1}} \in \mathrm{d}v_{k-1}),
\end{aligned}
$$

so that we have

$$
\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_k} \le s_k, \widetilde{V}_{t_{k+1}} \le v_{k+1}, \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}),\\
&\quad \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V}))\\
&\quad = \int_{-\infty}^{s_k}\int_{-\infty}^{v_{k+1}}\int_{C_c(\widehat{\mathbf{v}}^{N_k})}\int_{C_d(\widehat{\mathbf{s}}^{N_{k-1}^S})}\int_{C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})} \phi_{\bar{m}_k(s_{k-1},v_k,v_{k-1}),\bar{\sigma}_k(s_{k-1},v_{k-1})}\\
&\qquad \times (s_k)\mathrm{d}s_k \phi_{m_k(v_k),\sigma_k(v_k)}(v_{k+1})\mathrm{d}v_{k+1}\\
&\qquad \cdot \mathbb{P}(\widetilde{S}_{t_{k-1}} \in \mathrm{d}s_{k-1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k, \widetilde{V}_{t_{k-1}} \in \mathrm{d}v_{k-1})\\
&\quad = \int_{C_c(\widehat{\mathbf{v}}^{N_k})}\int_{C_d(\widehat{\mathbf{s}}^{N_{k-1}^S})}\int_{C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})}\left(\int_{-\infty}^{s_k} \phi_{\bar{m}_k(s_{k-1},v_k,v_{k-1}),\bar{\sigma}_k(s_{k-1},v_{k-1})}(s_k)\mathrm{d}s_k\right)\\
&\qquad \times \left(\int_{-\infty}^{v_{k+1}} \phi_{m_k(v_k),\sigma_k(v_k)}(v_{k+1})\mathrm{d}v_{k+1}\right)\\
&\qquad \cdot \mathbb{P}(\widetilde{S}_{t_{k-1}} \in \mathrm{d}s_{k-1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k, \widetilde{V}_{t_{k-1}} \in \mathrm{d}v_{k-1})\\
&\quad = \int_{C_c(\widehat{\mathbf{v}}^{N_k})}\int_{C_d(\widehat{\mathbf{s}}^{N_{k-1}^S})}\int_{C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})} \Phi_{\bar{m}_k(s_{k-1},v_k,v_{k-1}),\bar{\sigma}_k(s_{k-1},v_{k-1})}\\
&\qquad \times (s_k)\Phi_{m_k(v_k),\sigma_k(v_k)}(v_{k+1})\\
&\qquad \cdot \mathbb{P}(\widetilde{S}_{t_{k-1}} \in \mathrm{d}s_{k-1}, \widetilde{V}_{t_k} \in \mathrm{d}v_k, \widetilde{V}_{t_{k-1}} \in \mathrm{d}v_{k-1}).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_k} \le s_k, \widetilde{V}_{t_{k+1}} \le v_{k+1}, \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}),\\
&\quad \widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V}))\\
&\approx \Phi^{(s_k)}_{\bar{m}_k(s_d^{N_{k-1}^S}, v_c^{N_k^V}, v_e^{N_{k-1}^V}),\bar{\sigma}_k(s_d^{N_{k-1}^S}, v_e^{N_{k-1}^V})}\\
&\qquad \times \Phi^{(v_{k+1})}_{m_k(v_c^{N_k^V}),\sigma_k(v_c^{N_k^V})}\\
&\qquad \cdot \mathbb{P}(\widetilde{S}_{t_{k-1}} \in C_d(\widehat{\mathbf{s}}^{N_{k-1}^S}), \widetilde{V}_{t_k} \in C_c(\widehat{\mathbf{v}}^{N_k^V}), \widetilde{V}_{t_{k-1}} \in C_e(\widehat{\mathbf{v}}^{N_{k-1}^V})).
\end{aligned}
$$

Now we normalize the c.d.f.s of the Gaussian distributions, we sum up all the terms and we get the result.

Let us now spend a couple of words about the first step of the algorithm, for $k = 0$, concerning the computation of the probability $\mathbb{P}(\widetilde{S}_{t_0} \in C_a(\widehat{\mathbf{s}}^{N_0^S}), \widetilde{V}_{t_1} \in C_b(\widehat{\mathbf{v}}^{N_1}), \widetilde{V}_{t_0} \in C_c(\widehat{\mathbf{v}}^{N_0}))$.

Given that $\widetilde{S}_{t_0} = S_{t_0} = s_0$ and that $\widetilde{V}_{t_0} = V_{t_0} = v_0$, it follows that

$$
\begin{aligned}
&\mathbb{P}(\widetilde{S}_{t_0} \in C_a(\widehat{\mathbf{s}}^{N_0^S}), \widetilde{V}_{t_1} \in C_b(\widehat{\mathbf{v}}^{N_1}), \widetilde{V}_{t_0} \in C_c(\widehat{\mathbf{v}}^{N_0}))\\
&\quad = \mathbb{P}(\widetilde{S}_{t_0} \in C_a(\widehat{\mathbf{s}}^{N_0^S}))\mathbb{P}(\widetilde{V}_{t_1} \in C_b(\widehat{\mathbf{v}}^{N_1}), \widetilde{V}_{t_0} \in C_c(\widehat{\mathbf{v}}^{N_0})).
\end{aligned}
$$

Using the conditional Gaussian distribution of the variance process in lemma 2.3, we have

$$
\begin{aligned}
&\mathbb{P}(\widetilde{V}_{t_1} \in C_b(\widehat{\mathbf{v}}^{N_1}), \widetilde{V}_{t_0} \in C_c(\widehat{\mathbf{v}}^{N_0}))\\
&= \left[\Phi_{0,1}\left(\widehat{v}_{1,b^+}(\widehat{v}_c^{N_0})\right) - \Phi_{0,1}\left(\widehat{v}_{1,b^-}(\widehat{v}_c^{N_0})\right)\right]\mathbb{P}(\widetilde{V}_{t_0} \in C_c(\widehat{\mathbf{v}}^{N_0})),
\end{aligned}
$$

where $\widehat{v}_{1,b^\pm}(\widehat{v}_c^{N_0})$ are defined in (19). □