

FAST FILTERING WITH NEW SPARSE TRANSITION MARKOV CHAINS

Ivan Gorynin, Emmanuel Monfrini and Wojciech Pieczynski

SAMOVAR, Telecom Sudparis, CNRS, Université Paris-Saclay, 9, rue Charles Fourier, Evry, France

ABSTRACT

We put forward a novel Markov chain approximation method with regard to the filtering problem. The novelty consists in making use of the sparse grid theory to deal with the curse of dimensionality. Our method imitates the marginal distribution of the latent continuous process with a discrete probability distribution on a sparse grid. The grid points may be seen as the states of a Markov chain. We construct such a Markov chain to imitate the whole process. The transition probabilities are then chosen to preserve the joint moments of the underlying continuous process. We provide a simulation study of a multivariate stochastic volatility filtering problem where we compare the proposed methodology with a similar technique and with the particle filtering.

Index Terms— Markov chain approximation method, Hidden Markov models, Stochastic volatility, Maximum entropy principle, Non-linear propagation.

1. INTRODUCTION

The hidden Markov models (HMM) [1] are an universal tool for modeling time series. They include a \mathbb{R}^d -valued discrete time Markov process $\{\mathbf{X}_n\}_{n \geq 1}$ such that \mathbf{X}_{n+1} conditional on \mathbf{X}_n arises from the probability density function $f(\mathbf{x}_{n+1}|\mathbf{x}_n)$. We consider estimating $\{\mathbf{X}_n\}_{n \geq 1}$ but only have access to a discrete time $\mathbb{R}^{d'}$ -valued process $\{\mathbf{Y}_n\}_{n \geq 1}$. The latter is such that \mathbf{Y}_n conditional on \mathbf{X}_n arises from a probability density function $g(\mathbf{y}_n|\mathbf{x}_n)$.

The autoregressive stochastic volatility model (ARSV) [2, 3] is an example of HMM. This model, among others, allows describing the volatility of the spot price of a financial asset. Let $\{\mathbf{Y}_n\}_{n \geq 1}$ be a time series of continuously compounded returns, the ARSV model reads for $n \geq 1$:

$$\mathbf{X}_{n+1} = \mu + \phi(\mathbf{X}_n - \mu) + \sigma \mathbf{U}_{n+1}; \quad (1a)$$

$$\mathbf{Y}_n = \exp(\mathbf{X}_n / 2) \mathbf{V}_n, \quad (1b)$$

where $\{\mathbf{U}_n\}_{n \geq 1}$, $\{\mathbf{V}_n\}_{n \geq 1}$ are standard Gaussian white noises and $\mathbf{X}_1 \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)$. Here $\{\mathbf{X}_n\}_{n \geq 1}$ is the underlying latent log-variance process which is supposed to be autoregressive (mean-reverting). The parameters (μ, ϕ, σ) are fixed and relates to $\{\mathbf{X}_n\}_{n \geq 1}$. We have μ which is the

long-term mean, ϕ measures the slowness of the mean reversion and refers to the volatility clustering phenomenon *cf.* [4]. Finally, σ is the standard deviation of innovations and designs how volatile the log-variance is.

The exact estimation of \mathbf{X}_n given $\{\mathbf{Y}_t\}_{t=1}^n$, *i.e.* the computation of $E[\mathbf{X}_n | \mathbf{Y}_1, \dots, \mathbf{Y}_n]$, is generally intractable in an HMM [1]. Existing filtering methodologies, such as the Extended Kalman Filter (EKF) [5] and the Unscented Kalman Filter (UKF) [6] allow obtaining approximate biased solutions. The particle filter [7] is asymptotically optimal, but can be computationally expensive, especially if the state space is multivariate.

In this paper, we develop a novel design for filtering in the high dimensional state space. We first consider using our new Markov chain approximation (MCA) method to replace $\{\mathbf{X}_n\}_{n \geq 1}$ with a Markov chain. Then we run an exact optimal filtering algorithm in the hidden Markov chain framework [5]. The MCA methods are quite general and widely usable numerical approximations to stochastic processes [8, 9, 10, 11, 12].

The main drawback of the existing *grid-based* approaches is that the number of grid points grows exponentially with the dimension d . This work presents a way to use the sparse grids which are better suited for the high-dimensional problems. The number of points in a sparse grid grows polynomially with the dimension. There exists a recent method of filtering in the HMM which also uses the sparse grid theory [13]. However, our method is different since we do not make any Gaussian assumption in the state space nor in the observation space. This paper is also complementary to our prior works [14, 15, 16] where we represent some original system by a triplet model for which there exist fast filtering and smoothing algorithms.

In the context of this work, we assume that the distribution of $(\mathbf{X}_n, \mathbf{X}_{n+1})$ is independent from n . For the sake of simplicity, we denote $f(\mathbf{x}_{n+1}|\mathbf{x}_n)$ by $p(\mathbf{x}_{n+1}|\mathbf{x}_n)$ and $g(\mathbf{y}_n|\mathbf{x}_n)$ by $p(\mathbf{y}_n|\mathbf{x}_n)$. Similarly, we denote by $p(\mathbf{x}_n, \mathbf{x}_{n+1})$ the distribution of $(\mathbf{X}_n, \mathbf{X}_{n+1})$ and so on.

The next section is a brief overview of the quadrature rules and of the sparse grids. The third section details the contribution of the paper, which is an algorithm for representing a high-dimensional continuous process with a sparse transition Markov chain (STMC). The fourth section contains experiments and the last one contains conclusions and perspectives.

Algorithm 1 Approximating $\{\mathbf{X}_n\}_{n \geq 1}$ by a sparse transition Markov chain (STMC).

1. For each j and k such that $1 \leq j \leq d$ and $1 \leq k \leq C$, compute the k -points Gaussian quadrature $\mathcal{G}_{(j)}(k)$ over the j^{th} dimension of \mathbf{X} . Let us remember that one obtains $\mathcal{G}_{(j)}(k)$ by running a procedure that requires the unconditional moments $\{M(\gamma; j)\}_{\gamma=1}^{\gamma=2^C-1}$ cf. Section 2;
2. Define the sparse grid $\mathcal{Z} \subset \mathbb{R}^{2d}$ for \mathbf{Z} :

$$\mathcal{Z} = \bigcup_{q=2d}^{C+2d-1} \bigcup_{s=d}^{q-d} \mathcal{X}(s) \otimes \mathcal{X}(q-s),$$

where

$$\mathcal{X}(s) = \left\{ \mathcal{G}_{(1)}(k_1) \otimes \dots \otimes \mathcal{G}_{(d)}(k_d) \left| \sum_{j=1}^d k_j = s \right. \right\}$$

for each $s \geq d$;

3. Attribute the weights $\{\pi(\mathbf{z})\}_{\mathbf{z} \in \mathcal{Z}}$ as follows:

$$\pi(\mathbf{z}) \propto \exp \left(\sum_{\gamma \in \Upsilon} \lambda_0(\gamma) T(\mathbf{z}; \gamma) \right), \sum_{\mathbf{z} \in \mathcal{Z}} \pi(\mathbf{z}) = 1, \quad (5)$$

where $\{\lambda_0(\gamma)\}_{\gamma \in \Upsilon}$ are Lagrange multipliers. They relate to the optimization problem

$$\begin{aligned} \{\lambda_0(\gamma) \mid \gamma \in \Upsilon\} = & \underset{\{\lambda(\gamma) \mid \gamma \in \Upsilon\}}{\operatorname{argmin}} \left[- \sum_{\gamma \in \Upsilon} \lambda(\gamma) M(\gamma) + \right. \\ & \left. + \log \left(\sum_{\mathbf{z} \in \mathcal{Z}} \exp \left(\sum_{\gamma \in \Upsilon} \lambda(\gamma) T(\mathbf{z}; \gamma) \right) \right) \right]; \end{aligned}$$

4. Consider the state space $\mathcal{X} \subset \mathbb{R}^d$ of the time-invariant Markov chain approximation:

$$\mathcal{X} = \bigcup_{s=d}^{C+d-1} \mathcal{X}(s).$$

For each consecutive states $\mathbf{r}_n, \mathbf{r}_{n+1}$ in \mathcal{X} , consider their joint probability:

$$p(\mathbf{r}_n, \mathbf{r}_{n+1}) = \begin{cases} \pi([\mathbf{r}_n, \mathbf{r}_{n+1}]) & \text{if } [\mathbf{r}_n, \mathbf{r}_{n+1}] \in \mathcal{Z} \\ 0 & \text{otherwise} \end{cases};$$

5. Finally, the marginal and the transition probabilities of $\{\mathbf{R}_n\}_{n \geq 1}$ are:

$$p(\mathbf{r}_n) = \sum_{\mathbf{r}_{n+1} \in \mathcal{X}} p(\mathbf{r}_n, \mathbf{r}_{n+1});$$

$$p(\mathbf{r}_{n+1} \mid \mathbf{r}_n) = \frac{p(\mathbf{r}_n, \mathbf{r}_{n+1})}{p(\mathbf{r}_n)}.$$

Remark 3 The equation (5) is due to Tanaka and Toda [20]. They show that the solution of the optimization problem in Step 3 maximizes the entropy of the weights subject to the condition of matching the mixed moments of \mathbf{Z} , i.e.

$$\begin{aligned} \{\pi(\mathbf{z}) \mid \mathbf{z} \in \mathcal{Z}\} = & \underset{\{\omega(\mathbf{z}) \mid \mathbf{z} \in \mathcal{Z}\}}{\operatorname{argmin}} \sum_{\mathbf{z} \in \mathcal{Z}} \omega(\mathbf{z}) \log(\omega(\mathbf{z})) \\ \text{subject to } & \begin{cases} \sum_{\mathbf{z} \in \mathcal{Z}} T(\mathbf{z}; \gamma) \omega(\mathbf{z}) = M(\gamma), \forall \gamma \in \Upsilon \\ \sum_{\mathbf{z} \in \mathcal{Z}} \omega(\mathbf{z}) = 1, \omega(\mathbf{z}) \geq 0, \forall \mathbf{z} \in \mathcal{Z} \end{cases} \end{aligned}$$

4. APPLICATION TO FILTERING VOLATILITY

In this section we consider an application of our Markov chain approximation method to the HMM filtering problem. We run *Algorithm 1* to incorporate the original latent dynamics into $\{\mathbf{R}_n\}_{n \geq 1}$, then we set $p(\mathbf{y}_n \mid \mathbf{r}_n) = g(\mathbf{y}_n \mid \mathbf{x}_n = \mathbf{r}_n)$. Thus, we define a hidden Markov chain. Next, the optimal filtering algorithm computes the posterior margins of all the hidden states, i.e. it computes $p(\mathbf{r}_n \mid \mathbf{y}_1, \dots, \mathbf{y}_n)$ for all $n > 0$. Finally, we estimate \mathbf{X}_n conditional on $\{\mathbf{Y}_t\}_{t=1}^n$ by

$$\hat{\mathbf{X}}_n = \sum_{\mathbf{r}_n \in \mathcal{X}} \mathbf{r}_n p(\mathbf{r}_n \mid \mathbf{y}_1, \dots, \mathbf{y}_n), \quad (6)$$

i.e. $\hat{\mathbf{X}}_n$ is a weighted centroid of the points in \mathcal{X} .

Let us consider an example of a stochastic volatility model in the multi-asset framework [21]. Let $\mathbf{Y}_n \in \mathbb{R}^2$ denote the log-returns of two correlated assets. We assume that

$$\mathbf{Y}_n \sim \mathcal{N}(0, \Sigma_n), \quad (7)$$

where $\Sigma_n \in \mathbb{R}^{2 \times 2}$ is the dynamic covariance of \mathbf{Y}_n . In turn, Σ_n follows a Wishart autoregressive process [21] and we set for our study

$$\Sigma_n = \mathbf{X}_n \mathbf{X}_n^\top + \mathbf{Q}; \quad (8a)$$

$$\mathbf{X}_{n+1} = \mathbf{A} \mathbf{X}_n + \mathbf{D} \mathbf{U}_n. \quad (8b)$$

In the equation above we have $\mathbf{X}_n \in \mathbb{R}^2$, \mathbf{Q} , \mathbf{A} and \mathbf{D} are fixed matrices in $\mathbb{R}^{2 \times 2}$, \mathbf{Q} is positive definite, $\{\mathbf{U}_n\}_{n \geq 1}$ is a standard Gaussian white noise process in \mathbb{R}^2 and $\mathbf{X}_0 = 0$.

We estimate Σ_n conditional on $\{\mathbf{Y}_t\}_{t=1}^n$ by

$$\hat{\Sigma}_n = \sum_{\mathbf{r}_n \in \mathcal{X}} (\mathbf{r}_n \mathbf{r}_n^\top + \mathbf{Q}) p(\mathbf{r}_n \mid \mathbf{y}_1, \dots, \mathbf{y}_n).$$

We evaluate the accuracy of an estimation procedure by using the mean S-divergence criterion (MSD):

$$MSD = \frac{1}{N} \sum_{n=1}^N \delta_S^2(\Sigma_n, \hat{\Sigma}_n),$$

where δ_S^2 is the S-divergence [22]. The S-divergence is a distance-like function which applies to the positive definite

matrices. That is, for some positive definite matrices \mathbf{S}_1 and \mathbf{S}_2 ,

$$\delta_S^2(\mathbf{S}_1, \mathbf{S}_2) = \log \det \left(\frac{\mathbf{S}_1 + \mathbf{S}_2}{2} \right) - \frac{1}{2} \log \det (\mathbf{S}_1 \mathbf{S}_2).$$

The accuracy of the filter (6) depends on the grid size parameter C and the number of unconditional moments of \mathbf{X} that we preserve. We maintain the joint moments of $(\mathbf{X}_n, \mathbf{X}_{n+1})$ up to the second order and the unconditional moments of \mathbf{X} up to the $(2C - 1)^{\text{th}}$ order. In Table 1 we provide the MSD values which we find for five different values of C . The parameters of the multivariate stochastic volatility model are in Table 2. Figure 1 illustrates an example of a realization of the multivariate stochastic volatility process and of the corresponding posterior estimation by our algorithm.

As a comparison with the existing filtering methodologies, we consider using the *Full Grid* approach instead of the *Sparse Grid* in Step 2 of *Algorithm 1*. Thus, for a fixed C , the complexity of our method is $\text{card}(\mathcal{Z})$ when using the *Sparse Grid* and C^4 if using the *Full Grid*. In our view, the complexity of filtering measures the number of elementary operations per filter iteration. We also consider the particle filtering with various numbers m of particles. The equivalent complexity of the particle filter is m . Figure 2 compares these methods in terms of both accuracy and complexity. For instance, we reach the level 0.035 of MSD by using the particle filter with 250 particles, or a *Full Grid* MCA with approximately 80 points, or a STMC with approximately 40 points. Therefore, our method allows reaching the desired accuracy with the lowest complexity, so with the highest speed among these approaches.

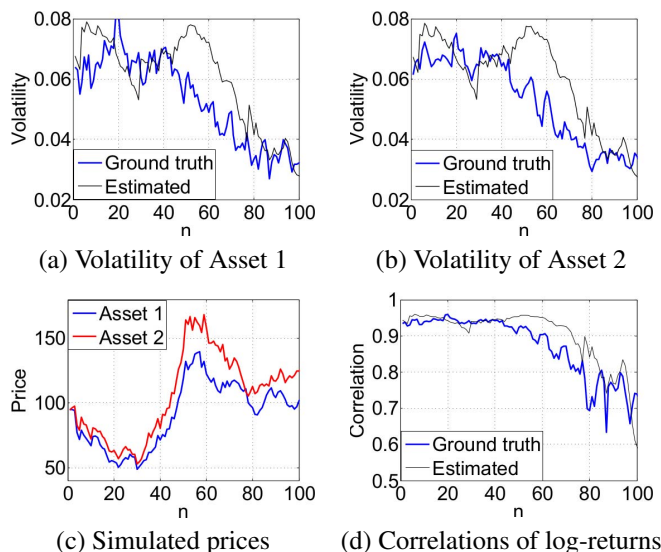


Fig. 1. A realization of the multivariate stochastic volatility process $\{(7), (8)\}$. In figures (a), (b) and (d), the black line plots the filtering estimates of the volatilities and correlations.

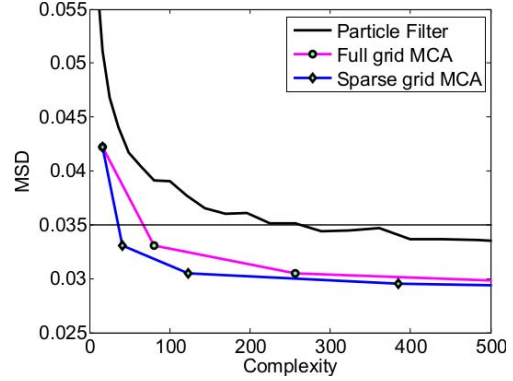


Fig. 2. MSD values versus the complexity of each method. The *Sparse Grid* approach has a better accuracy per complexity ratio because it provides a good coverage with a lower number of grid points.

Table 1. Accuracy of filtering by using the *Sparse Grid* MCA

C	2	3	4	5	6
$\text{card}(\mathcal{Z})$	16	41	123	385	953
MSD	0.042	0.033	0.030	0.029	0.029

5. CONCLUSION

We proposed a general method of filtering and applied it to a multivariate stochastic volatility model. The filtering solution is approximated by using a new Markov chain approximation method in order to cast the given HMM model to the HMC framework, in which exact optimal filter is feasible. The presented method is suitable for high-dimensional state spaces and may realize some speed-ups compared to the existing approaches. Simulation results confirm the effectiveness of our methodology.

As a perspective, we will extend this method to deal with more complex models. Such are, for example, the asymmetric stochastic volatility models [23]. Moreover, we will consider an application of this method to the estimation of the HMM parameters. We will also consider a comparison with the other methods on the real-world market data.

Table 2. The parameters of the volatility model $\{(7), (8)\}$

\mathbf{Q}		\mathbf{A}		\mathbf{DD}^T	
5	0	0.8	0.2	2	1
0	5	0.2	0.8	1	2
$\times 10^{-4}$				$\times 10^{-5}$	

6. REFERENCES

- [1] Olivier Cappé, Eric Moulines, and Tobias Rydén, *Inference in hidden Markov models*, Springer-Verlag, 2005.
- [2] Eric Jacquier, Nicholas Polson, and Peter Rossi, “Bayesian analysis of stochastic volatility models,” *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 69–87, 2002.
- [3] Sangjoon Kim, Neil Shephard, and Siddhartha Chib, “Stochastic volatility: likelihood inference and comparison with arch models,” *The Review of Economic Studies*, vol. 65, no. 3, pp. 361–393, 1998.
- [4] Thomas Lux and Michele Marchesi, “Volatility clustering in financial markets: a microsimulation of interacting agents,” *International journal of theoretical and applied finance*, vol. 3, no. 04, pp. 675–702, 2000.
- [5] Brian Anderson and John Moore, *Optimal filtering*, Courier Corporation, 2012.
- [6] Simon Julier and Jeffrey Uhlmann, “New extension of the kalman filter to nonlinear systems,” in *AeroSense’97*. International Society for Optics and Photonics, 1997, pp. 182–193.
- [7] Arnaud Doucet and Adam Johansen, “A tutorial on particle filtering and smoothing: Fifteen years later,” *Handbook of Nonlinear Filtering*, vol. 12, no. 656-704, pp. 3, 2009.
- [8] Nikolay Gospodinov and Damba Lkhagvasuren, “A moment-matching method for approximating vector autoregressive processes by finite-state markov chains,” *Journal of Applied Econometrics*, vol. 29, no. 5, pp. 843–859, 2014.
- [9] Leland Farmer and Alexis Akira Toda, “Discretizing stochastic processes with exact conditional moments,” *Available at SSRN 2585859*, 2015.
- [10] Stephen Terry and Edward Knotek, “Markov-chain approximations of vector autoregressions: Application of general multivariate-normal integration techniques,” *Economics Letters*, vol. 110, no. 1, pp. 4–6, 2011.
- [11] George Tauchen, “Finite state markov-chain approximations to univariate and vector autoregressions,” *Economics letters*, vol. 20, no. 2, pp. 177–181, 1986.
- [12] Chia Chun Lo, Konstantinos Skindilias, and Andreas Karathanasopoulos, “Forecasting latent volatility through a markov chain approximation filter,” *Journal of Forecasting*, vol. 35, no. 1, pp. 54–69, 2016.
- [13] Bin Jia, Ming Xin, and Yang Cheng, “Sparse-grid quadrature nonlinear filtering,” *Automatica*, vol. 48, no. 2, pp. 327–341, 2012.
- [14] Ivan Gorynin, Stéphane Derrode, Emmanuel Monfrini, and Wojciech Pieczynski, “Exact fast smoothing in switching models with application to stochastic volatility,” in *Proc. of the EUSIPCO’15*, Nice, France, Aug. 2015, pp. 924–928.
- [15] Stéphane Derrode and Wojciech Pieczynski, “Fast filter in nonlinear systems with application to stochastic volatility model,” in *Proc. of the EUSIPCO’14*, Lisbon, Portugal, Sept. 2014, pp. 290–294.
- [16] Ivan Gorynin, Stéphane Derrode, Emmanuel Monfrini, and Wojciech Pieczynski, “Fast filtering in switching approximations of non-linear Markov systems with applications to stochastic volatility,” *IEEE Transactions on Automatic Control Submission*, accepted, May, 2016.
- [17] Gene Golub and John Welsch, “Calculation of gauss quadrature rules,” *Mathematics of computation*, vol. 23, no. 106, pp. 221–230, 1969.
- [18] Alberto Luceno, “Discrete approximations to continuous univariate distributions - an alternative to simulation,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 2, pp. 345–352, 1999.
- [19] Hans-Joachim Bungartz and Michael Griebel, “Sparse grids,” *Acta numerica*, vol. 13, no. 1, pp. 147–269, 2004.
- [20] Kenichiro Tanaka and Alexis Akira Toda, “Discrete approximations of continuous distributions by maximum entropy,” *Economics Letters*, vol. 118, no. 3, pp. 445–450, 2013.
- [21] Christian Gouriéroux, Joann Jasiak, and Razvan Sufana, “The Wishart autoregressive process of multivariate stochastic volatility,” *Journal of Econometrics*, vol. 150, no. 2, pp. 167–181, 2009.
- [22] Suvrit Sra, “A new metric on the manifold of kernel matrices with application to matrix geometric means,” in *Advances in Neural Information Processing Systems*, 2012, pp. 144–152.
- [23] Andrew Harvey and Neil Shephard, “Estimation of an asymmetric stochastic volatility model for asset returns,” *Journal of Business & Economic Statistics*, vol. 14, no. 4, pp. 429–434, 1996.