SEQUENTIAL MONTE CARLO SIMULATION OF DYNAMICAL MODELS WITH SLOWLY VARYING PARAMETERS: AN EXTENSION

William Fong and Simon Godsill

Signal Processing Group, University of Cambridge,
Cambridge, CB2 1PZ, U.K.
wnwf2@eng.cam.ac.uk sjg@eng.cam.ac.uk

ABSTRACT

In this paper, we improve on the slow time-varying partial correlation (STV-PARCOR) model recently suggested by us [2] to include any deterministic interpolator. We then suggest a modification to the on-line filtering algorithm to accomodate the changes. It is believed that the modification will improve on the simulation results as it takes into account the underlying trend of the parameter evolution. The suggested algorithm is tested with real speech data and preliminary results are shown and compared with those generated using existing approaches.

1. INTRODUCTION

Many real world data analysis problems involve sequential estimation of filtering distribution $p(x_t|y_{1:t})$, where x_t is the unobserved state of the system at time t and $y_{1:t} \triangleq \{y_1, \ldots, y_t\}$ are observations made over some time interval $t \in \{1, \ldots, T\}$. In most cases, the data structures can be very complex, typically involving elements of non-Gaussianity, high-dimensionality and non-linearity, which may not be solvable analytically. Sequential Monte Carlo methods, also known as Particle Filters (PF), have been proposed to overcome these problems. Refer to [1] for an up-to-date survey of the field. Within the particle filter framework, the filtering distribution is approximated with an empirical distribution formed from point masses, or particles,

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta(x_t - x_t^{(i)}), \quad \sum_{i=1}^{N} w_t^{(i)} = 1,\ w_t^{(i)} \geq 0$$

where $\delta(.)$ is the Dirac delta function and $w_t^{(i)}$ is a weight attached to particle $x_t^{(i)}$.

In recent years, various approaches have been developed to apply the sequential Monte Carlo filtering strategies for the purpose of audio signal enhancement (see, for example, [3] and citations therein). These approaches assume a Gaussian random walk model for the system parameter evolution at every time step, which may not give a sufficiently slow or smooth variation with time. This makes the standard PF inefficient as it is known that the filter becomes highly degenerate for random walks with very low variance. Recently, Fong and Godsill [2] propose a slow time-varying partial correlation (STV-PARCOR) model to solve this problem. In their work, the system coefficients are considered to evolve stochastically on a block-to-block basis and all coefficients in-between are found by linear interpolator. Based on the STV-PARCOR model and under the particle filtering framework, an algorithm for on-line joint estimation of system parameters signal is developed.

This work serves as an extension to the work of [2]. In this paper, we generalised the STV-PARCOR particle filter, so that any deterministic interpolator can be used. We then describe the modified algorithm for the generation of "delayed" state realisations. Finally, preliminary simulation results are shown.

2. STATE-SPACE REPRESENTATION AND AUDIO MODEL

In this section, we describe the model adopted in this paper — the STV-PARCOR model. A lengthy time series is divided into R non-overlapping blocks. If β is the block size, we define $\vec{x}_{\tau+1} \triangleq \{x_{\tau\beta+1}, \ldots, x_{\tau\beta+\beta}\}$ as a group of unobserved states of the system and $\vec{y}_{\tau+1} \triangleq \{y_{\tau\beta+1}, \ldots, y_{\tau\beta+\beta}\}$ as observations made over some blocks $\tau \in \{0, \ldots, R-1\}$.

Assuming a Markovian structure for the model, the problem can then be formulated in a state-space form as follows,

$$\begin{array}{lll} \vec{x}_{\tau+1} & \sim & f(\vec{x}_{\tau+1}|\vec{x}_{\tau}) \quad \text{State evolution density} \\ \vec{y}_{\tau+1} & \sim & g(\vec{y}_{\tau+1}|\vec{x}_{\tau+1}) \quad \text{Observation density} \end{array} \tag{1}$$

where $f(.|.)$ and $g(.|.)$ are pre-specified state evolution and observation densities. It should be noted that the state-space model adopted here is different from the standard one [5], which relates the unobserved states $\{x_t\}$ and observations $\{y_t\}$ made over a time interval $t \in \{1, \ldots, T\}$. (1), however, defines the state evolution between different blocks.

For the choice of audio model, we suggest a time varying partial correlation (TV-PARCOR) model. The advantage of adopting such model is that approximate stability can easily be enforced, provided that the PARCOR coefficients $\{\rho_t\}$ vary sufficient slowly with time [3]. The audio signal process $\{u_t\}$ is then modelled as

$$u_t = \sum_{i=1}^{p_t} a_{t,i} u_{t-i} + e_t$$

where a_t is the TVAR coefficient at time t, which is found by transforming the PARCOR coefficient ρ_t via the Levinson-Durbin recursion. ϕ_{e_t} is the log-excitation variance and p_t is the time-varying model order. Refer to [3] for a detailed description of the audio model adopted here.

The full specification of the state-space model is as follows: at any time t, the state vector x_t is partitioned as $[z_t, \theta_t]'$ with $z_t \triangleq [u_{t-p_t+1}, \ldots, u_t]'$ and $\theta_t \triangleq [\rho_t, \phi_{e_t}, p_t]'$ being the signal state and the parameter state respectively.

In the setup of the particle filter, a proposal distribution [1, 3] similar to that of [2] has been adopted, which takes the form:

$$f(\vec{\theta}_\tau|\vec{\theta}_{\tau-1}) = f(\rho_{\tau\beta}|\rho_{(\tau-1)\beta}) f(p_{\tau\beta}|p_{(\tau-1)\beta}) \times \times f(\phi_{e_{\tau\beta}}|\phi_{e_{(\tau-1)\beta}}) \quad (2)$$

As in [2], the block variation of the log-excitation variance and model order take the form,

$$f(\phi_{e_{\tau\beta}}|\phi_{e_{(\tau-1)\beta}}) = \mathcal{N}(\phi_{e_{(\tau-1)\beta}}, \sigma^2_{\phi_e}) \quad (3)$$

$$f(p_{\tau\beta}|p_{(\tau-1)\beta}) = \sum_{k=-1}^{+1} P_k \delta\left(p_{\tau\beta} - (p_{(\tau-1)\beta} + k)\right) \quad (4)$$

where $\{\phi_{e_t}, p_t\} = \{\phi_{e_{\tau\beta}}, p_{\tau\beta}\}$ for $t \in \{(\tau-1)\beta+1, \ldots, \tau\beta\}$, *i.e.* both ϕ_{e_t} and p_t are assumed to be fixed within a block.

For the PARCOR coefficients, a constrained random walk model [3] is assumed for the block variation,

$$f(\rho_{\tau\beta}|\rho_{(\tau-1)\beta}) = \begin{cases} \mathcal{N}(\rho_{(\tau-1)\beta}, \sigma^2_\rho I) & \text{if } \max_i\{|\rho_{\tau\beta,i}|\} < 1 \\ 0 & \text{otherwise} \end{cases}$$

A constrained random walk model of this form will ensure approximate stability provided that the PARCOR coefficients vary sufficiently slowly. Having sampled the last PARCOR coefficient of the block τ, $\rho_{\tau\beta}$, all the intermediate PARCOR coefficients are found by some deterministic methods using previously sampled PARCOR coefficients $\{\rho_{(\tau-m)\beta}; m = 0, \ldots, \eta+1\}$,

$$\rho_t = h_t(\rho_{(\tau-\eta-1)\beta}, \ldots, \rho_{\tau\beta}) \quad (5)$$

where $t \in \{(\tau-\eta-1)\beta+1, \ldots, \tau\beta\}$. For the interpolator functions, h_t, the Legendre polynomials, Fourier basis function and B-splines are popular choices, all of which will ensure a slow and smooth evolution of the reflection coefficients [4]. In [2], we have implemented a linear interpolator, which should be considered as a special case of (5). We believe that (5) will give a better approximation than the simplified linear interpolator case as it takes into account the better underlying smooth trend of the coefficient evolution.

3. SEQUENTIAL AUDIO SIGNAL AND PARAMETER ESTIMATION

Based on the STV-PARCOR model, [2] describes a way for joint estimation of the signal and parameter state under the sequential Monte Carlo filter framework. The suggested algorithm proceeds as follows:

Random samples are to be drawn from the joint filtering distribution $p(\vec{z}_\tau, \vec{\theta}_\tau|\vec{y}_{1:\tau})$ which can be factorised as follows,

$$\begin{aligned} p(\vec{z}_\tau, \vec{\theta}_\tau|\vec{y}_{1:\tau}) &= p(\vec{z}_\tau|\vec{\theta}_\tau, \vec{y}_{1:\tau}) p(\vec{\theta}_\tau|\vec{y}_{1:\tau}) \quad (6) \\ &= \int p(\vec{z}_\tau|\vec{\theta}_{1:\tau}, \vec{y}_{1:\tau}) p(\vec{\theta}_{1:\tau}|\vec{y}_{1:\tau}) d\vec{\theta}_{1:\tau-1} \end{aligned}$$

Assume there exists a particulate approximation for the marginal parameter filtering distribution,

$$p(\vec{\theta}_{1:\tau}|\vec{y}_{1:\tau}) \approx \sum_{i=1}^{N} w_\tau^{(i)} \delta(\vec{\theta}_{1:\tau} - \vec{\theta}^{(i)}_{1:\tau}) \quad (7)$$

as it is assumed that the proposal distribution for the parameter state being the prior (2), the importance weight will simply take the form,

$$w_\tau^{(i)} \propto \prod_{t=(\tau-1)\beta+1}^{\tau\beta} p(y_t|\theta^{(i)}_{1:t}, y_{1:t-1}) \quad (8)$$

with $\{\theta_t^{(i)}; (\tau-\eta-1)\beta+1, \ldots, \tau\beta\}$ are parameter states recently updated using the deterministic interpolator (5). The joint filtering distribution (6) can then be approximated by

$$\begin{aligned} p(\vec{z}_\tau, \vec{\theta}_\tau|\vec{y}_{1:\tau}) &\approx \int p(\vec{z}_\tau|\vec{\theta}_{1:\tau}, \vec{y}_{1:\tau}) \times \\ &\quad \times \sum_{i=1}^{N} w_\tau^{(i)} \delta(\vec{\theta}^{(i)}_{1:\tau} - \vec{\theta}_{1:\tau}) d\vec{\theta}_{1:\tau-1} \\ &\approx \sum_{i=1}^{N} p(\vec{z}_\tau|\vec{\theta}^{(i)}_{1:\tau}, \vec{y}_{1:\tau}) w_\tau^{(i)} \delta(\vec{\theta}^{(i)}_\tau - \vec{\theta}_\tau) \end{aligned}$$

Hence, given $\theta^{(i)}_{1:\tau}$, signal realisations can be drawn from

$$\vec{z}^{(i)}_\tau \sim p(\vec{z}_\tau|\vec{\theta}^{(i)}_{1:\tau}, \vec{y}_{1:\tau})$$

For instance, if we assume a conditional Gaussian state-space model, then all the computations can be done under the framework of the Kalman filter and smoother [3]. *e.g.* for $t \in \{(\tau-1)\beta+1,\ldots,\tau\beta\}$, $p(y_t|\theta_{1:t}^{(i)}, y_{1:t-1})$ from (8) can be found by the prediction error decomposition [5] and the marginal signal filtering distribution can be rewritten as:

$$p(\vec{z}_\tau|\vec{\theta}_{1:\tau}^{(i)}, \vec{y}_{1:\tau}) = \prod_{t=(\tau-1)\beta+1}^{\tau\beta} p(z_t|z_{(\tau-1)\beta+1:t-1}, \theta_{1:t}^{(i)}, y_{1:\tau\beta})$$
$$= \prod_{t=(\tau-1)\beta+1}^{\tau\beta} \mathcal{N}(z_t; \zeta_{t|\tau\beta}^{(i)}, P_{t|\tau\beta}^{(i)}) \quad (9)$$

with $\zeta_{t|\tau\beta}^{(i)} \triangleq \mathrm{E}(z_t|y_{1:\tau\beta}, \theta_{1:t}^{(i)})$ and $P_{t|\tau\beta}^{(i)} \triangleq \mathrm{cov}(z_t - \zeta_{t|\tau\beta}^{(i)}, z_t - \zeta_{t|\tau\beta}^{(i)}|y_{1:\tau\beta}, \theta_{1:t}^{(i)})$ are sufficient statistics found by the Kalman smoother.

4. IMPLEMENTATION

We modify the STV-PARCOR particle filter suggested in [2] to facilitate the generalised STV-PARCOR model. Let $\tau_\eta = \tau - \eta - 1$, assuming that the parameter realisations $\{\theta_{\tau_\eta\beta}^{(i)}, \ldots, \theta_{(\tau-1)\beta}^{(i)}\}$ and the signal sufficient statistics ($\zeta_{\tau_\eta\beta|\tau_\eta\beta}^{(i)}$ and $P_{\tau_\eta\beta|\tau_\eta\beta}^{(i)}$) are available for $\{i = 1,\ldots,N\}$ from the previous iteration of the filter, state random samples can be drawn from the filtering distribution as follows:
For $i = 1,\ldots,N$,

- Generate random sample from $p_{\tau\beta}^{(i)} \sim f(p_{\tau\beta}|p_{(\tau-1)\beta}^{(i)})$ and $\phi_{e_{\tau\beta}}^{(i)} \sim f(\phi_{e_{\tau\beta}}|\phi_{e_{(\tau-1)\beta}}^{(i)})$. For each $p_{\tau\beta}^{(i)}$, sample $\rho_{\tau\beta}^{(i)}$ from $f(\rho_{\tau\beta}|\tilde{\rho}_{(\tau-1)\beta}^{(i)})$, where

$$\tilde{\rho}_{(\tau-1)\beta}^{(i)} = \begin{cases} \begin{bmatrix} \rho_{(\tau-1)\beta}^{(i)} \\ 0 \end{bmatrix} & \text{if } p_{\tau\beta}^{(i)} > p_{(\tau-1)\beta}^{(i)} \\ \begin{bmatrix} \rho_{(\tau-1)\beta,1\ldots p_{\tau\beta}^{(i)}}^{(i)} \end{bmatrix}' & \text{otherwise} \end{cases}$$

- Make each of $\{\rho_{\tau_\eta\beta}^{(i)}, \ldots, \rho_{\tau\beta}^{(i)}\}$ the same size by appending 0 to the vector if necessary. Using these fixed grids, $\{\rho_t^{(i)}; \tau_\eta\beta < t < \tau\beta, t \neq n\beta\}$ for n being integer are found by deterministic interpolator (5). In our simulations, we have implemented the cubic spline with $\eta = 3$, *i.e.* for $j = 1,\ldots,p_{\tau\beta}^{(i)}$,

$$\rho_{t,j}^{(i)} = \sum_{k=\tau_\eta}^{\tau} c_{k,j}^{(i)} \mathcal{B}^3(t - k\beta)$$

where $c_{k,j}$ is the spline coefficient to be determined and $\mathcal{B}^3(t)$ is the 3^{rd} order B-spline basis function.

- Given the sufficient statistics ($\zeta_{\tau_\eta\beta|\tau_\eta\beta}^{(i)}$ and $P_{\tau_\eta\beta|\tau_\eta\beta}^{(i)}$) and $\{\rho_t^{(i)}; \tau_\eta\beta < t \leq \tau\beta\}$, we run a forward sweep of the Kalman filter over blocks $\tau - \eta$ to τ. We then evaluate the importance weight $w_\tau^{(i)}$ according to (8).

Having generated the parameter set $\{\vec{\theta}_\tau^{(i)}; i = 1,\ldots,N\}$, we then resample (see [1] for details) it N times with replacement according to $w_\tau^{(i)}$. For the resampled parameter set $\{\vec{\theta}_\tau^{(i)}; i = 1,\ldots,N\}$, we run a backward sweep of the Kalman smoother and generate $\{\zeta_{t|\tau\beta}^{(i)}, P_{t|\tau\beta}^{(i)}; t = \tau_\eta\beta + 1,\ldots,\tau\beta\}$.

Theoretically, signal realisations $\{z_t^{(i)}; i = 1,\ldots,N, t = \tau_\eta\beta+1,\ldots,\tau\beta\}$ can be generated according to (9). However, as $\{\rho_t^{(i)}; (\tau_\eta+1)\beta < t < \tau\beta, t \neq n\beta\}$ are going to change in the next iteration, only $\{z_t; t \in \{\tau_\eta\beta+1,\ldots,(\tau_\eta+1)\beta\}$ will be drawn. Hence, the suggested algorithm will only give "delayed" state realisations.

In addition, continuity can be ensured by taking into account $z_{t-1}^{(i)}$ in the sampling of $z_t^{(i)}$, as there is only one free variable owing to overlapping betweem $z_{t-1}^{(i)}$ and $z_t^{(i)}$.

5. EXPERIMENTAL RESULTS

Experiments are conducted to investigate the effectiveness of the suggested algorithm (STV-PARCOR PF) for the purpose of audio noise reduction. In particular, we would like to verify our suggestion that the generalised STV-PARCOR model is a better model for slow time-varying processes (*e.g.* speech) then the TVAR model. Preliminary simulation results are shown and compared with those generated using the standard extended Kalman smoother (eKS) [5]. The clean speech clips used in this experiment are:

S1: Good service should be rewarded by big tips
S2: Draw every outer line first, then fill in the interior

The experiment setup is as follows, the clean speech signal is assumed to be submerged in white Gaussian noise (WGN) with known variance σ_v^2, *i.e.* $g(y_t|x_t) = \mathcal{N}(y_t; u_t, \sigma_v^2)$. The output SNR from different algorithms are recorded and compared. Owing to the stochastic nature of the Monte Carlo algorithm, simulation results for the STV-PARCOR PF are found by averaging the SNR improvement over five independent applications of the algorithm.

In our simulations, we have chosen a value of $N = 50$, the block size β is fixed to 100 for the STV-PARCOR PF. The hyperparameters (σ_ρ^2, $\sigma_{\phi_e}^2$ and $\{P_k; k = -1, 0, +1\}$) adopted here are assumed to be known and fixed. In consideration of the computational cost, it is assumed that the model order $p_t \in \{1,\ldots,p_{\max}\}$ with $p_{\max}$ being limited to 20. We note that $N = 50$ is extremely small for Monte Carlo simulation but the preliminary simulation results suggest that the algorithm works pretty well in such a case. As in other applications of the Particle Filter, simulation results improve as N increases.

For the eKS, a Gaussian random walk is assumed directly on the AR coefficients and the model order is fixed to 10. This model is employed as the TV-PARCOR model and time-varying model order is not straight forward to implement with the extended Kalman smoother. The hyperparameters adopted are adjusted so that the system parameters will cover the same range as the generalised STV-PARCOR model in β time steps. We note that this may not be the optimum setup for the eKS, however, this will give a fair comparison for both algorithms.

Figure 1 and Figure 2 show the 3D histogram plots of the first reflection coefficients ($\rho_{1,t}$) at different input SNRs (SNR_{in}) for the word “‘reward” in S1 using the STV-PARCOR PF. The plots are generated by grouping all PARCOR coefficients particles from five independent simulations. As shown in the plots, the suggested algorithm gives consistent results at different noise levels.

We then compare the performance of the suggested algorithm with the the eKS. The SNR improvements for different clips at different noise levels are summarised below:

Clip	SNR_{in}	STV-PARCOR PF	eKS
S1	0*dB*	3.86*dB*	1.92*dB*
S1	10*dB*	2.54*dB*	0.99*dB*
S1	20*dB*	1.08*dB*	0.87*dB*
S2	0*dB*	4.31*dB*	2.21*dB*
S2	10*dB*	2.80*dB*	1.57*dB*
S2	20*dB*	1.35*dB*	1.09*dB*

Audio outputs can be found at *http://www-sigproc.eng.cam.ac.uk/∼wnwf2/Eusipco2002.html*. Comparing the SNR improvements, the suggested STV-PARCOR PF consistently outperforms the eKS, which justifies using it in practice, even it induces a much heavier computational load when compare with the eKS.

6. CONCLUSION

We propose a generalisation to the STV-PARCOR model recently suggested by us to include any deterministic interpolator functions, h_t. We then describe an adaptation to the algorithm for joint estimation for signal and parameter. The algorithm is tested on real speech signals and compared with other standard approaches. Encouraging results are obtained. Further simulations will be conducted to investigate the effects of different interpolator functions and different lags, η. The results will be published in due course.

7. REFERENCES

[1] A. Doucet, N. de Freitas, and N. J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

[2] W. Fong and S. Godsill. Sequential Monte Carlo simulation of dynamical models with slowly varying parameters: Application to audio. In *Proceedings of the IEEE ICASSP*, 2002. To appear.

[3] W. Fong, S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing with application to audio signal enhancement. *IEEE Transactions on Signal Processing, Special Issue*, 50(2):438–449, February 2002.

[4] Y. Grenier. Time-dependent ARMA modeling of nonstationary signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(4):899–911, 1983.

[5] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.

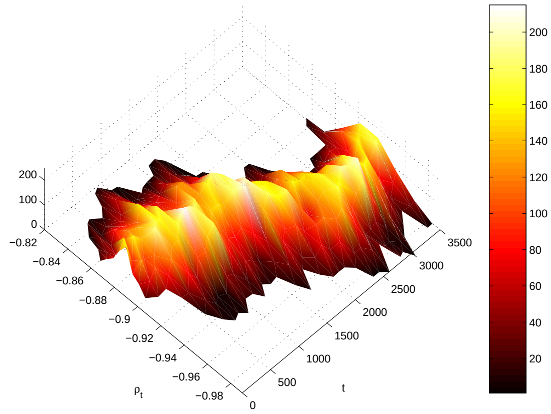

Figure 1: 3D histogram plot of $\rho_{t,1}$ for the word “reward” using the STV-PARCOR particle filter for SNR_{in}=0*dB*

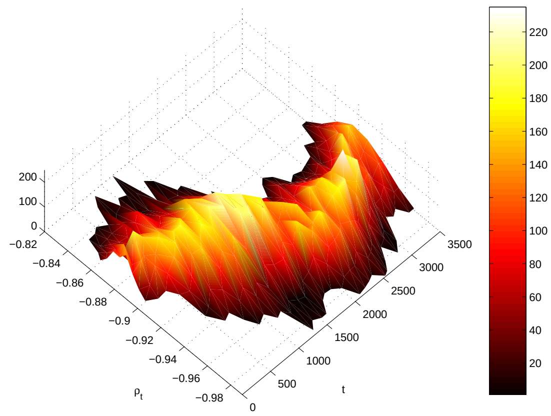

Figure 2: 3D histogram plot of $\rho_{t,1}$ for the word “reward” using the STV-PARCOR particle filter for SNR_{in}=20*dB*