PARTICLE FILTERING BASED RECOVERY OF NOISY GARCH PROCESSES

Tomer Michaeli and Israel Cohen

Department of Electrical Engineering
Technion–Israel Institute of Technology, Haifa, Israel
{tomermic@tx, icohen@ee}.technion.ac.il

ABSTRACT

In this paper, we address the problem of enhancement of a noisy GARCH process using a particle filter. We compare our approach experimentally to a previously developed recursive estimation scheme. Simulations indicate that a significant gain in performance is obtained, at the cost of higher sensitivity to errors in the GARCH parameters. The proposed method allows tackling arbitrary driving noise distributions as well as arbitrary fidelity criteria.

Index Terms— Particle filtering, GARCH, dynamic estimation.

1. INTRODUCTION

Recently, an approach for statistically modeling speech signals in the STFT domain based on generalized autoregressive conditional heteroscedasticity (GARCH) processes was proposed [1, 2]. The GARCH model, which was adopted from the field of financial time series analysis, is characterized by heavy-tailed distributions and volatility clustering, properties which are also characteristics of speech STFT coefficients. In the signal processing community, the GARCH model was employed for voice activity detection [3, 4], speech recognition [5] and speech enhancement [6], among other tasks.

In [6] a recursive estimation algorithm was proposed for recovering a complex GARCH process x_t from its noisy version $y_t = x_t + w_t$, where w_t is an iid complex Gaussian noise component independent of x_t. This algorithm comprises a *propagation step*, in which the estimate of x_t based on the noisy measurements up to time t is processed to yield an estimate of x_{t+1}, followed by an *update step* in which the measurement y_{t+1} is used to update the estimate of x_{t+1}. It was shown in [6] that this approach, when applied to the STFT coefficients of noisy speech signals, leads to better enhancement results than the decision-directed method [7] in terms of log-spectral distortion (LSD) and perceptual evaluation of speech quality scores (PESQ, ITU-T P.862).

These findings support the suitability of the GARCH model to speech signals. Nevertheless, the recursive estimator of [6] is suboptimal in the sense that it *does not* minimize the desired distortion measure at time t given the entire history of measurements up to time t. Specifically, in speech enhancement applications, one is usually interested in minimizing the mean squared error (MSE) of the spectral amplitude or of the log-spectral amplitude (LSA). The method of [6] relies on several approximations and thus does not generally coincide with the minimum MSE (MMSE) or the MMSE-LSA estimators. An important open question, then, is whether this approach can be improved by using alternative recursive estimation schemes.

In this work, we address the problem of recovering a complex GARCH process from its noisy version using a particle filter. This sequential Monte Carlo approach allows to approximate the optimal estimator (under any chosen fidelity criterion) as accurately as desired by increasing the number of particles. Furthermore, the particle filtering methodology allows to tackle arbitrary distributions. We consequently use this method to assess the proximity of the algorithm of [6] to the optimal solution in various situations. Simulations show that the performance of the proposed algorithm is better for the range of GARCH parameter values that are typical to speech, especially in low SNR scenarios.

The paper is organized as follows. Section 2 presents the GARCH model and its use in speech enhancement. In Section 3 we briefly review the recursive estimation algorithm of [6] and propose an alternative particle filter based approach to tackle the recovery of a noisy GARCH process. Finally, in Section 4 we present simulation results, confirming the advantage of the proposed method.

2. THE GARCH SPEECH MODEL

A common assumption underlying many speech enhancement methods is that distinct frequency bins of the STFT of a speech signal are independent random processes. Consequently, the processing of the STFT coefficients is carried out separately for each frequency bin, which allows to omit the bin subscript from our notation. Let x_t denote the STFT expansion coefficient of a speech signal at time t in some frequency bin. We model x_t as a complex GARCH process of order $(1, 1)$ defined by

$$x_t = \sigma_t v_t, \tag{1}$$

where $\{v_t\}$ are statistically independent complex random variables with zero mean and unit variance

$$E[v_t] = 0, \quad E\left[|v_t|^2\right] = 1, \tag{2}$$

and the conditional variance σ_t^2 itself is a random process, which evolves as

$$\sigma_t^2 = \sigma_{\min}^2 + \mu|x_{t-1}|^2 + \delta\left(\sigma_{t-1}^2 - \sigma_{\min}^2\right). \tag{3}$$

A GARCH$(1,1)$ process has a finite unconditional variance $E[|x_t|^2]$ if its parameters satisfy

$$\sigma_{\min}^2 > 0, \quad \mu \geq 0, \quad \delta \geq 0, \quad \mu + \delta < 1. \tag{4}$$

The parameters μ and δ control the typical duration of clusters of small and large magnitudes, whereas the parameter $\sigma_{\min}^2$ affects the unconditional variance $E[|x_t|^2]$ of the process:

$$E\left[|x_t|^2\right] = \sigma_{\min}^2 \frac{1-\delta}{1-\mu-\delta}. \tag{5}$$

The STFT expansion coefficients of speech signals are characterized by long periods of small magnitudes separated by short bursts of large magnitudes, which correspond to speech presence. Such a behavior can be obtained in the GARCH model by choosing δ relatively small and $\mu + \delta$ close to 1. The characteristic dynamic range of the process x_t can be controlled by tuning $p_v(v)$, the distribution of the process v_t. Common models include the Gaussian, Gamma and Laplace distributions [6].

3. RECURSIVE ESTIMATION FROM NOISY MEASUREMENTS

Assume that a GARCH$(1,1)$ process is observed through additive noise:

$$y_t = x_t + w_t, \tag{6}$$

where $w_t \sim \mathcal{CN}(0, \sigma^2)$ are statistically independent complex circular Gaussian random variables. Our goal is to produce at time t an estimate $\hat{x}_t$ based on the set of measurements $\{y_\tau\}_{\tau=1}^t$ such that the expected distortion $E[d(x_t, \hat{x}_t)]$ is minimized. In speech enhancement applications, one is typically interested in minimizing the MSE of the spectral amplitude

$$d(x, \hat{x}) = (|x| - |\hat{x}|)^2 \tag{7}$$

or of the LSA

$$d(x, \hat{x}) = (\log|x| - \log|\hat{x}|)^2. \tag{8}$$

3.1. Approximate Recursive Recovery

An important property of the GARCH model is that the random variables $\{x_t\}$ are conditionally independent given $\{\sigma_t^2\}$. Therefore, had σ_t^2 been known at time t, the optimal estimate of x_t given $\{y_\tau\}_{\tau=1}^t$ would be only a function of the current measurement y_t. Furthermore, a closed form expression for this estimate is available under the MSE criterion in many interesting situations, including the cases where $x_t|\sigma_t^2$ has a Gaussian, Gamma, or Laplacian distribution [6]. An analytic formula for the Gaussian speech model is also available under the MSE-LSA criterion. Based on this observation, it was proposed in [6] to recursively estimate σ_t^2 given the measurements $\{y_\tau\}_{\tau=1}^t$ in an MMSE sense, and substitute the estimate $\widehat{\sigma_t^2}$ in the formula for the estimator of x_t given y_t.

The algorithm proposed in [6] is suboptimal for two reasons. First, substituting the MMSE estimate of σ_t^2 in the formula for the estimator of x_t, generally does not lead to the minimal value of $E[d(x_t, \hat{x}_t)]$. Second, the recursive scheme of [6] for estimating the conditional variance is only an approximation of the MMSE estimate of σ_t^2 given the entire past $\{y_\tau\}_{\tau=1}^t$. To asses the accuracy of these approximations, we now propose using a particle filter, which can approximate the optimal estimate as accurately as desired, by employing a large number of particles.

3.2. Recovery via the CONDENSATION Algorithm

The CONDENSATION algorithm [8] is one of a class of particle filtering methods, in which the conditional density of the state x_t given the measurements $\{y_\tau\}_{\tau=1}^t$ is approximated by a weighted combination of delta functions:

$$p\left(x_t \left| \{y_\tau\}_{\tau=1}^t \right.\right) \approx \sum_{i=1}^N \pi_t^i \delta(x_t - x_t^i). \tag{9}$$

The "particles" $\{x_t^i\}_{i=1}^N$ and weights $\{\pi_t^i\}_{i=1}^N$ are propagated in time according to the evolution of the state and measurements, which are assumed to be of the form [9]

$$x_t = f(x_{t-1}, v_t) \tag{10}$$

$$y_t = h(x_t, w_t) \tag{11}$$

for arbitrary functions $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$. An important assumption underlying particle filtering methods is that the noise processes v_t and w_t are iid and mutually independent.

Substituting (3) into (1), it can be seen that the evolution of a GARCH$(1,1)$ process x_t does not follow the form (10), rendering direct use of the CONDENSATION algorithm inapplicable. An alternative, then, is to regard σ_t^2 as the state variable to be estimated. However, substituting (1) into (6) leads to a measurement equation not in the form of (11).

To overcome these difficulties, we define an augmented state vector

$$\tilde{\boldsymbol{x}}_t = \left(x_t, \sigma_t^2\right)^T. \tag{12}$$

This allows writing both the state evolution and the measure-

ment equation in the form of (10) and (11), where

$$f\left(\tilde{\boldsymbol{x}}_{t-1}, v_t\right) = \begin{pmatrix} v_t\sqrt{\sigma^2_{\min} + \mu|x_{t-1}|^2 + \delta(\sigma^2_{t-1} - \sigma^2_{\min})} \\ \sigma^2_{\min} + \mu|x_{t-1}|^2 + \delta(\sigma^2_{t-1} - \sigma^2_{\min}) \end{pmatrix}, \tag{13}$$

$$h\left(\tilde{\boldsymbol{x}}, w_t\right) = \begin{pmatrix} 1 & 0 \end{pmatrix} \tilde{\boldsymbol{x}}_t + w_t. \tag{14}$$

We therefore propose maintaining a set of 2D particles $\{\tilde{\boldsymbol{x}}^i_t\}_{i=1}^N = \{(x^i_t, (\sigma^2_t)^i)^T\}_{i=1}^N$.

The CONDENSATION algorithm comprises three stages applied successively each time a new measurement becomes available. At time t, a new set of particles is generated by resampling from $\{\tilde{\boldsymbol{x}}^i_{t-1}\}_{i=1}^N$ based on the weights $\{\pi^i_{t-1}\}_{i=1}^N$. Then, each of the new particles is propagated using (13), with a random variable v^i_t drawn from $p_v(v)$, to obtain the set $\{\tilde{\boldsymbol{x}}^i_t\}_{i=1}^N$. Finally, the weights are updated using $\pi^i_t = p_{y_t|\tilde{\boldsymbol{x}}_t}(y_t|\tilde{\boldsymbol{x}}_t = \tilde{\boldsymbol{x}}^i_t)$, which in our case reduces to $\pi^i_t = p_w(y_t - x^i_t)$, and normalized such that $\sum_{i=1}^N \pi^i_t = 1$.

Once the particles and weights have been updated, the MMSE estimate, $E[x_t|\{y_\tau\}_{\tau=1}^t]$, can be approximated by

$$\hat{x}^{\text{MSE}}_t = \frac{1}{N}\sum_{i=1}^N \pi^i_t x^i_t. \tag{15}$$

Similarly, the MMSE-LSA estimate, which is given by $\exp\{E[\log(|x_t|)|\{y_\tau\}_{\tau=1}^t]\}e^{j\angle y_t}$ [6], can be computed as

$$\hat{x}^{\text{LSA}}_t = \exp\left\{\frac{1}{N}\sum_{i=1}^N \pi^i_t \log(|x^i_t|)\right\} e^{j\angle y_t}. \tag{16}$$

4. SIMULATIONS

We now compare the performance of the recursive estimator of [6] with that of the CONDENSATION algorithm outlined above, in simulations.

We begin by examining the performance of both algorithms in the task of MSE minimization. Figures 1 and 2 depict the difference between the output SNR of the recursive method of [6] and the particle filter proposed above for input SNRs of 10dB and 1dB respectively and for a range of values of the parameters δ and μ. Specifically, in this experiment we generated Gaussian-noise-driven GARCH$(1,1)$ processes with various values of the parameter δ in the range $[0.05, 0.95]$ and contaminated them with white Gaussian noise with variance $\sigma^2 = 1$. For each value of δ, the parameter μ was tuned such that $\delta + \mu = 0.999$ to obtain a behavior which is typical of speech STFT coefficients, and $\sigma^2_{\min}$ was calculated using (5) to yield the desired input SNR. The results were averaged over a set of 10 realizations per set of parameters.

As seen in the figures, the CONDENSATION algorithm attains a significantly higher SNR than the recursive approach of [6] for small values of δ. As explained in Section 2, this range of values of δ is of particular interest when modeling speech. A comparison between Figures 1 and 2 reveals that the gain in performance is greater when the input SNR is low (Fig. 2). This indicates that the particle filter approach should perform better when applied to real speech signals, especially in the high frequency bins, where the SNR is usually low.

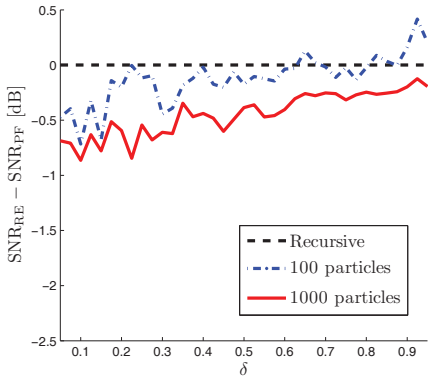

Fig. 1. *Particle filtering versus [6].* SNR $= 10$dB.

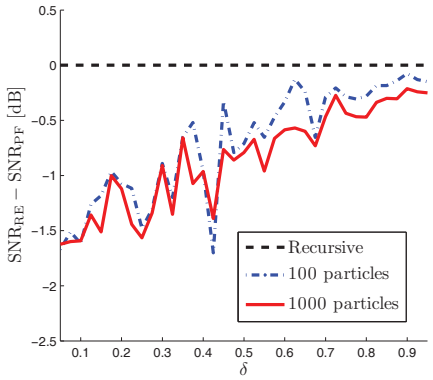

Fig. 2. *Particle filtering versus [6].* SNR $= 1$dB.

Figures 3 and 4 compare the performance of both algorithms in the task of MSE-LSA minimization. They depict the ratio between the MSE-LSA of the particle filter proposed above and that of the recursive method of [6] for input SNRs of 10dB and 1dB respectively. As can be seen, the CONDENSATION algorithm attains a significantly lower MSE-LSA than the recursive approach of [6] for small values of δ. Similar to the MSE experiment, here too the gain is more significant for low input SNRs.

The computational load of the CONDENSATION method is greater than that of [6] roughly by a factor of the number of particles N. Nevertheless, as N increases, the gain in performance is more significant. Examining the MSE-LSA experiment, which is of particular interest in speech enhancement, it can be seen that even a moderate amount of 10 particles leads to a large reduction in the MSE-LSA.

Finally, we compare both algorithms in a model mismatch scenario. Specifically, in practical applications the GARCH parameters $(\mu, \delta, \sigma^2_{\min})$ are usually not known in advance but rather need to be estimated from the noisy process y_t itself. We now address the question: to what extent is the perfor-

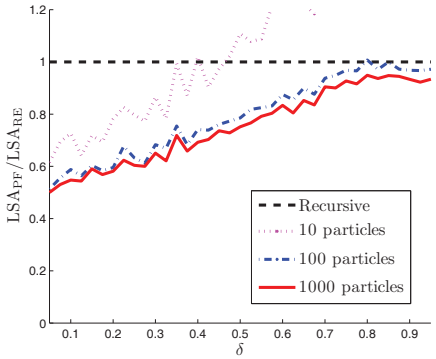

Fig. 3. *Particle filtering versus [6].* SNR $= 10$dB.

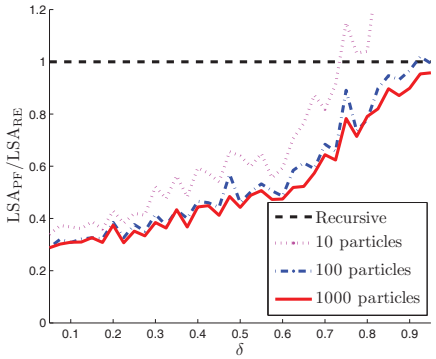

Fig. 4. *Particle filtering versus [6].* SNR $= 1$dB.

mance of the algorithms affected by using an incorrect set of parameters. From our experiments, both algorithms are barely sensitive to a mismatch in μ and δ. However, using an incorrect value for $\sigma^2_{\min}$, severely effects the results. Figure 5 shows the MSE-LSA attained by both algorithms when using different values of $\sigma^2_{\min}$ in the range $[0.01\sigma_0^2, 100\sigma_0^2]$, where σ_0^2 is the true value of $\sigma^2_{\min}$. In this experiment the input SNR was 10dB and both algorithms were provided with the true values of $(\delta, \mu) = (0.2, 0.799)$. As can be seen, the particle filter attains its minimal LSA at the true value of $\sigma^2_{\min}$. In contrast, the recursive estimator [6] benefits from an underestimate of $\sigma^2_{\min}$ and attains its minimum roughly at $0.1\sigma_0^2$. The LSA in this point is only 14% higher than the minimal LSA of the particle filter. Furthermore, we observe that the CONDENSATION approach is more sensitive to a mismatch in $\sigma^2_{\min}$. Specifically, its LSA is lower than that of [6] in the range $[0.3\sigma_0^2, 30\sigma_0^2]$ and higher otherwise.

5. CONCLUSIONS

We have proposed a particle-filtering approach for recovering a complex GARCH$(1,1)$ process contaminated by noise. The method can be used for GARCH signals with arbitrary driving noise distributions, as well as under arbitrary fidelity criteria. We showed through simulations that this algorithm is superior to the method developed in [6], most notably for values

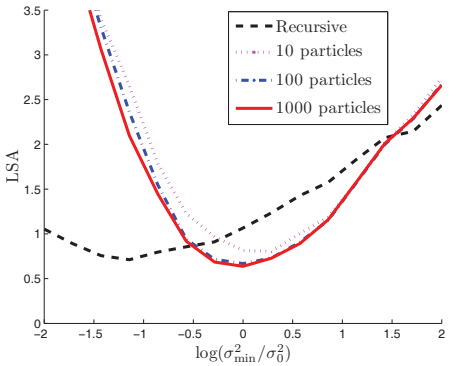

Fig. 5. *Performance under mismatch in* $\sigma^2_{\min}$. SNR $= 10$dB.

of the parameters that are typical to speech signals, under the MSE-LSA criterion, and in low input SNR scenarios. The disadvantage of the approach is that it is more sensitive to errors in the parameters. Future work will be concerned with fusing both methods to enhance speech signals. In time-frequency bins where the estimated parameters are expected to be inaccurate, the algorithm of [6] should be applied, whereas our method would be used in the rest of the time-frequency plane.

6. REFERENCES

[1] I. Cohen, "Modeling speech signals in the time–frequency domain using GARCH," *Signal Processing*, vol. 84, no. 12, pp. 2453–2459, 2004.

[2] I. Cohen, "From Volatility Modeling of Financial Time-Series to Stochastic Modeling and Enhancement of Speech Signals," in *Speech enhancement*, S. Makino J. Benesty and J. Chen, Eds., chapter 5, pp. 97–114. New York: Springer, 2005.

[3] H. K. Solvang, K. Ishizuka, and M. Fujimoto, "Voice activity detection based on adjustable linear prediction and GARCH models," *Speech Communication*, 2008.

[4] R. Tahmasbi and S. Rezaei, "Change point detection in GARCH models for voice activity detection," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 1038–1046, 2008.

[5] M. Abdolahi and H. Amindavar, "GARCH coefficients as feature for speech recognition in Persian isolated digit," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 17–21.

[6] I. Cohen, "Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models," *Signal processing*, vol. 86, no. 4, pp. 698–709, 2006.

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[8] M. Isard and A. Blake, "CONDENSATION–conditional density propagation for visual tracking," *International journal of computer vision*, vol. 29, no. 1, pp. 5–28, 1998.

[9] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, D. Sci, T. Organ, and S. A. Adelaide, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.