ARCH AND GARCH PARAMETER ESTIMATION IN PRESENCE OF ADDITIVE NOISE USING PARTICLE METHODS

Saman Mousazadeh and Israel Cohen

Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel
{smzadeh@tx, icohen@ee}.technion.ac.il

ABSTRACT

In this paper, we propose a new method based on particle filters for maximum likelihood (ML) estimation of the parameters of autoregressive conditional heteroscedasticity (ARCH) and generalized autoregressive conditional heteroscedasticity (GARCH) models. Our method is based on gradient descend method and active set method for maximizing the likelihood function over parameters under stationarity constraints. The gradient of the likelihood function of observation given the parameters of the model, which is needed for gradient based optimization algorithm, is estimated using particle methods. Simulation results show the advantage of the proposed method over competing techniques.

***Index Terms*—** GARCH, ARCH, parameter estimation, noisy observations, particle methods

1. INTRODUCTION

Parameter estimation is the backbone of all model based signal processing algorithms. The generalized autoregressive conditional heteroscedasticity (GARCH) model was first introduced by Bollerslev [1] as an extension of the autoregressive conditional heteroscedasticity (ARCH) model developed by Engle [2] to model econometric data. Since then, many researchers have extended and used these models in several speech and image processing applications. Cohen [3] modeled the speech signal in the short time Fourier transform (STFT) domain as a complex GARCH process and used this model for speech enhancement. AR-GARCH model was utilized for modeling speech signal in the time domain and for developing voice activity detection (VAD) algorithms [4, 5]. Abdolahi et al. [6] used the parameters of the GARCH model for speech recognition in Persian isolated digits. Noiboar and Cohen [7] used the GARCH model for anomaly detection in sonar images. In all of the above mentioned uses of GARCH models in speech signal processing, the authors assume that the parameters of the GARCH model are known or can be estimated from a database.

This research was supported by the Israel Science Foundation (grant no. 1130/11).

Parameter estimation methods for ARCH and GARCH models, such as quasi maximum likelihood (QML) [8], two stage least squares (TSLS) [9], and constrained two stage least squares (CTSLS) [10], assume that the data are clean (without any additive noise). So in most speech signal processing algorithms these methods are often inapplicable due to presence of additive noise. Recently, Mousazadeh and Cohen [11] proposed an ML method for simultaneous parameter estimation and state smoothing of complex (G)ARCH processes in the presence of additive noise. In that work, it is assumed that the probability density functions (pdfs) of the additive corrupting noise and the process noise (v_k in the definition of GARCH (1)) are Gaussian. This assumption restricts the application of the method in general cases. Poyiadjis et al. [12] introduced a method for parameter estimation of general state-space models using particle methods. In their method, they find the ML estimate of the parameters. The drawback of their method is that it does not consider the cases that some constraints such as stationarity are imposed on the model. ARCH and GARCH parameter estimation in presence of noise is one of these cases.

In this paper, we propose a gradient based technique for estimating the parameters of ARCH and GARCH models in presence of additive noise. More specifically, we use a gradient descent method together with active set method to obtain the maximum of the likelihood function. The gradient of the likelihood function of noisy observations given the parameters which is needed for gradient based optimization algorithm, is obtained by particle method. The reminder of this paper is as follows. In Section 2, we formulate the problem and introduce a novel technique for estimating the parameters of ARCH and GARCH models in the presence of noise considering the stationarity constraints. Simulation results and performance comparison are presented in Section 3. We conclude the paper in Section 4.

2. PROBLEM FORMULATION AND THE PROPOSED METHOD FOR PARAMETER ESTIMATION

GARCH model of order (q, p) is defined as [1]

$$z_k = \sigma_k v_k \tag{1}$$

$$\sigma_k = \left(c_0 + \sum_{i=1}^{q} b_i z_{k-i}^2 + \sum_{i=1}^{p} a_i \sigma_{k-i}^2 \right)^{1/2} \tag{2}$$

where $\boldsymbol{\theta} = [c_0, b_1, \cdots, b_q, a_1, \cdots, a_p]$ is the vector of parameters of the GARCH model, $(\cdot)^T$ is the transpose operator and v_k's are independent identically distributed random variables with zero mean and unit variance. ARCH model is a special case of GARCH model with $a_i = 0, \forall i$. Necessary and sufficient conditions for stationarity with finite second order moment are [1]

$$\begin{aligned}
&c_0 > 0 \\
&b_i \geq 0 \qquad ; \qquad i = 1, 2, \cdots, q \\
&a_i \geq 0 \qquad ; \qquad i = 1, 2, \cdots, p \\
&\sum_{i=1}^{q} b_i + \sum_{i=1}^{p} a_i < 1.
\end{aligned} \tag{3}$$

Now, suppose that the observation signal, y_k, is a GARCH process corrupted with an additive noise, i.e.

$$y_k = z_k + n_k \tag{4}$$

where n_k is the additive noise sequence. For the rest of the paper we assume that the probability density functions of v_k and n_k are completely known except for some unknown parameters such as the noise level. If we define the hidden state vector as

$$\boldsymbol{x}_k = \begin{bmatrix} z_k & \cdots & z_{k-q+1} & \sigma_k & \cdots & \sigma_{k-p+1} \end{bmatrix}^T \tag{5}$$

then we will have

$$\begin{aligned}
\boldsymbol{x}_{k+1} &= F\left(\boldsymbol{x}_k, v_k, \boldsymbol{\theta}\right) && (6)\\
y_k &= G\left(x_k, n_k, \boldsymbol{\theta}\right) && (7)
\end{aligned}$$

where

$$F\left(\boldsymbol{x}_k, v_k, \boldsymbol{\theta}\right) = \begin{bmatrix} \sqrt{c_0 + \sum_{i=1}^{q} b_i z_{k-i}^2 + \sum_{i=1}^{p} a_i \sigma_{k-i}^2} v_k \\ z_{k-1} \\ \vdots \\ z_{k-p+1} \\ \sqrt{c_0 + \sum_{i=1}^{q} b_i z_{k-i}^2 + \sum_{i=1}^{p} a_i \sigma_{k-i}^2} \\ \sigma_{k-1} \\ \vdots \\ \sigma_{k-p+1} \end{bmatrix} \tag{8}$$

$$G\left(\boldsymbol{x}_k, n_k, \boldsymbol{\theta}\right) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{x}_k + n_k. \tag{9}$$

Reducing this model to ARCH process corrupted with additive noise is straightforward. It is worth mentioning that if the probability density functions of v_k and n_k have some unknown parameters, we can modify the vector of parameters such that it includes these unknown parameters. Available methods for estimating the parameters in this general state space (i.e. equations) model are extended Kalman Filters (EKF), unscented Kalman filter (UKF) and particle filters (PF)[13]. Poyiadjis et al. [12], introduced a method for parameter estimation in the general state-space model given by (8) and (9), using particle methods. This algorithm is based on finding the likelihood function of the observations in terms of the vector of parameters θ i.e. $p(y_0, y_1, \cdots, y_{K-1}; \theta)$, using particle methods and finding the maximum of this function by a gradient decent method.

In our case, i.e. parameter estimation for ARCH and GARCH models in presence of additive noise, stationarity and finite second order moment constraints must be considered. Since the method presented in [12] gives an estimate of the gradient vector of the probability density function of the observations, we can use this estimate along with the active set and gradient projection methods [14] to solve our constrained maximization problem. The constraints on parameters which are given by (3) can be restated as follows

$$\begin{bmatrix} -\boldsymbol{I}_{p+q+1} \\ \boldsymbol{1}_{1\times(p+q+1)} \end{bmatrix} \begin{bmatrix} c_0 \\ b_1 \\ \vdots \\ b_q \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \boldsymbol{A\theta} \leq \begin{bmatrix} \epsilon \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{10}$$

where $\boldsymbol{I}_d$ is $d \times d$ identity matrix, $\boldsymbol{1}_{n\times m}$ is $n \times m$ matrix with all of its elements equal to one, ϵ is a very small number greater than zero. By $\boldsymbol{a} \leq \boldsymbol{b}$, we mean that each element in $\boldsymbol{a}$ is smaller than or equal to the corresponding element in $\boldsymbol{b}$. Our algorithm is proposed in Table 1. It is worth mentioning that the proposed algorithm can be used in any parameter estimation problem with linear constraints. It is worth mentioning that the proposed method can be modified to obtain a recursive ML algorithm, which can be applied to the case where the parameters are time varying, see [11] for further information.

In the algorithm given in Table 1, N is the number of particles, $y_0^{K-1} = \{y_0, y_1, \cdots, y_{K-1}\}$ and $\widetilde{\zeta}_{\boldsymbol{\theta}}(y_0^k, \chi_k^{(i)})$ and

$\widetilde{\nabla\zeta}_{\boldsymbol{\theta}}(y_0^k, \chi_k^{(i)})$ are computed by the following equations.

$$\widetilde{\zeta}_{\boldsymbol{\theta}}(y_0^k, \chi_k^{(i)}) = \sum_{i=1}^{N} \widetilde{a}_{k-1}^{(i)} g_{\boldsymbol{\theta}}(y_k|\chi_k^{(i)}) f_{\boldsymbol{\theta}}(\chi_k^{(i)}|\chi_{k-1}^{(i)}) \tag{11}$$

$$\widetilde{\nabla\zeta}_{\boldsymbol{\theta}}(y_0^k, \chi_k^{(i)}) = \sum_{i=1}^{N} \widetilde{a}_{k-1}^{(i)} g_{\boldsymbol{\theta}}(y_k|\chi_k^{(i)}) f_{\boldsymbol{\theta}}(\chi_k^{(i)}|\chi_{k-1}^{(i)}) \times \left[\nabla \left(\log g_{\boldsymbol{\theta}}(y_k|\chi_k^{(i)})\right) + \nabla \log \left(f_{\boldsymbol{\theta}}(\chi_k^{(i)}|\chi_{k-1}^{(i)})\right) + \widetilde{\beta}_{k-1}^{(i)}\right] \tag{12}$$

where $g_{\boldsymbol{\theta}}(y_k|\chi_k^{(i)})$ and $f_{\boldsymbol{\theta}}(\chi_k^{(i)}|\chi_{k-1}^{(i)})$ are conditional transition and observation distributions, respectively, i.e.,

$$y_k|\boldsymbol{x}_k \sim g_{\boldsymbol{\theta}}(y_k|\boldsymbol{x}_k) \tag{13}$$

$$\boldsymbol{x}_k|\boldsymbol{x}_{k-1} \sim f_{\boldsymbol{\theta}}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \tag{14}$$

In this algorithm $q_{\boldsymbol{\theta}_k}(.|y_0^k, \chi_{k-1}^{(i)})$ is the assumed distribution and is case dependent. The best choice of this distribution is $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, y_0^k; \boldsymbol{\theta})$ which is the conditional distribution of the current state, $\boldsymbol{x}_k$, conditioned on the previous state and y_0^k. In most problems, computation of this distribution and generating samples from it is almost impossible. So, the usual choice for the distribution is the transition distribution, i.e.,

$$q_{\boldsymbol{\theta}_k}(\cdot|y_0^k, \chi_{k-1}^{(i)}) = f_{\boldsymbol{\theta}_k}(\cdot|\chi_{k-1}^{(i)}). \tag{15}$$

This algorithm can be easily applied to ARCH and GARCH parameter estimation problems in presence of additive noise. The main point is obtaining the conditional observation and transition distributions as follows

$$g_{\boldsymbol{\theta}}(y_k|\boldsymbol{x}_k) = p_n(y_k - \boldsymbol{x}_k(1)) \tag{16}$$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \frac{1}{\boldsymbol{x}_k(q+1)} p_v\left(\frac{x_k(1)}{\boldsymbol{x}_k(q+1)}\right) \times \prod_{i=2, i\neq q+1}^{p+q} \delta(\boldsymbol{x}_k(i) - \boldsymbol{x}_{k-1}(i-1)) \times \delta\left(\boldsymbol{x}_k(q+1) - \sqrt{c_0 + \sum_{i=1}^{q} b_i \boldsymbol{x}_{k-1}^2(i) + \sum_{i=1}^{p} a_i \boldsymbol{x}_{k-1}^2(i+q)}\right) \tag{17}$$

where $\boldsymbol{x}_k(j)$ is the j-th element of $\boldsymbol{x}_k$, $\delta(\cdot)$ is the Dirac delta function and p_v and p_n are the distributions of v_k and n_k, respectively.

Table 1. Proposed algorithm for estimating the parameters of a linearly constrained model using particle methods

1. Choose an initial value ($\boldsymbol{\theta}_0$) for the parameter vector.
2. Find the set of active constraints and form the matrix $\boldsymbol{A}_c$ by removing the rows corresponding to inactive constraints from matrix $\boldsymbol{A}$.
3. For $k = 0$ to $k = K-1$ do the following
   (a) For $i = 1, 2, \cdots, N$ sample
   $\chi_k^{(i)} \sim q\left(\chi_k^{(i)}|\boldsymbol{x}_{k-1}, y_0^k; \boldsymbol{\theta}\right) = \sum_{i=1}^N \widetilde{a}_{k-1}^{(i)} q_{\boldsymbol{\theta}_k}\left(\cdot|y_0^k, \chi_{k-1}^{(i)}\right)$
   (b) Find the values of $a_k^{(i)}$ and $\rho_k^{(i)}$ using
   $a_k^{(i)} = \frac{\hat{\zeta}_{\boldsymbol{\theta}}\left(y_0^k, \chi_k^{(i)}\right)}{q(\chi_k^{(i)}|\boldsymbol{x}_{k-1}, y_0^k; \boldsymbol{\theta})}$
   $\rho_k^{(i)} = \frac{\widetilde{\nabla\zeta}(y_0^k, \chi_k^{(i)})}{q(\chi_k^{(i)}|\boldsymbol{x}_{k-1}, y_0^k; \boldsymbol{\theta})}$
   (c) Find the values of $\widetilde{a}_k^{(i)}$ and $\widetilde{\beta}_k^{(i)}$ using
   $\widetilde{a}_k^{(i)} = \frac{a_k^{(i)}}{\sum_{i=1}^N a_k^{(i)}}$
   $\widetilde{a}_k^{(i)} \widetilde{\beta}_k^{(i)} = \frac{\rho_k^{(i)}}{\sum_{i=1}^N a_k^{(i)}} + \widetilde{a}_k^{(i)} \frac{\sum_{i=1}^N \rho_k^{(i)}}{\sum_{i=1}^N a_k^{(i)}}$
4. Calculate the value of gradient of the pdf using
   $\boldsymbol{g} = \nabla \log\left(\hat{p}(y_0^{K-1}; \boldsymbol{\theta})\right) = \sum_{k=0}^{K-1} \frac{\sum_{i=1}^N \rho_k^{(i)}}{\sum_{i=1}^N a_k^{(i)}}$.
5. Calculate the following quantities
   $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{A}_c^T(\boldsymbol{A}_c\boldsymbol{A}_c^T)^{-1}\boldsymbol{A}_c$
   $\boldsymbol{d} = \boldsymbol{P}\boldsymbol{g}^T$
6. If $\boldsymbol{d} = 0$ then find α_1 and α_2 such that
   $\alpha_1 = \max\{\alpha| + \alpha d, \alpha \in \mathbb{R}^+\}$.
   Select a value for α_2 such that $0 \leq \alpha_2 \leq \alpha_1$.
   Set $\boldsymbol{\theta} = \boldsymbol{\theta} + \alpha_2\boldsymbol{d}$ and go to 2.
7. If $\boldsymbol{d} \neq 0$, compute
   $\boldsymbol{\lambda} = -\boldsymbol{A}_c^T(\boldsymbol{A}_c\boldsymbol{A}_c^T)^{-1}\boldsymbol{A}_c\boldsymbol{g}$ and do the following steps
   (a) If for all j corresponding to active constraints the j-th element of $\boldsymbol{\lambda}$ is greater or equal to zero, then $\boldsymbol{\theta}$ is optimum, and the algorithm is finished.
   (b) Else, eliminate the row corresponding to the most negative element of $\boldsymbol{\lambda}$ from the matrix $\boldsymbol{A}_c$
   and go to 3.

3. SIMULATION RESULTS

In this section we investigate the performance of the proposed method PCon (the algorithm depicted in Table 1) compared with four ML methods. The first one, denoted by MLC, employs the clean data (unavailable in practical situations) for estimating the parameters.The second one, denoted by MLN, employs the noisy data for estimating the parameters assuming that the data are clean. The third one, denoted by MLG, is the one proposed in [11], assuming both the process noise and corrupting noise are Gaussian. The last method, denoted by PUnc (the algorithm presented in [12]) employs particle methods without considering the stationarity constraints.

Table 2. NRMSE in parameter estimation for different methods

NMSE	MLC	MLN	MLG	PUnc	PCon
c_0	0.2436	4.2774	0.4688	0.3945	0.2948
a_1	0.2116	0.6594	0.4718	0.4175	0.2475
b_1	0.2991	0.7414	0.6758	0.5432	0.3864

For this simulation, we used a GARCH$(1, 1)$ process with Laplace process noise, corrupted by a zero mean Gaussian white noise with known variance. The signal to noise ratio (SNR) was 10dB, the number of available data was $K = 2000$, and the number of particles was $N = 100$. The value of the parameter α_2 was selected such that $\|\alpha_2 \boldsymbol{d}\| = \min(\|\alpha_1 \boldsymbol{d}\|, 0.01)$, where $\|\alpha\|$ is the length of the vector α. The parameter of the GARCH process is $\boldsymbol{\theta} = [2.00\ 0.50\ 0.20]$. For evaluating the performance of our method we used the normalized root means quare error (NRMSE) in parameter estimation which is obtained using 2000 Monte-Carlo realizations. The simulation results are depicted in Table 2. It is apparent that the proposed method which considers the stationarity condition has a better performance than that of competing methods except MLC which is not applicable to noisy data. The main drawback of our method is its high computational load. Some techniques are available to decrease this computational load drastically. For more information on these methods, see [15].

4. CONCLUSION AND REMARKS

We have proposed an algorithm based on particle methods for parameter estimation of ARCH and GARCH models in presence of additive noise. The main advantages of our method over the competing parameter estimation methods are dealing with contaminating noise and non-Gaussian distributions of the process noise and corrupting noise. The proposed method also guarantees a stationary model. Our method can be used in other state space parameter estimation problems having linear constraints. It is worth mentioning that the proposed method can be generalized such that it handles nonlinear constraints as well. The main drawback of our method is its high computational load. Some techniques are available to decrease this computational load.

5. REFERENCES

[1] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.

[2] R.F. Engle, "Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation," *Econometrica*, vol. 50, pp. 987–1008, 1982.

[3] I. Cohen, "Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models," *signal processing*, vol. 86, pp. 698–709, 2006.

[4] H. K. Solvanga, K. Ishizukac, and M. Fujimotoc, "Voice activity detection based on adjustable linear prediction and GARCH models," *Speech Communication*, vol. 50, pp. 476–486, 2008.

[5] S. Mousazadeh and I. Cohen, "AR-GARCH in presence of noise: Parameter estimation and its application to voice activity detection," *IEEE Trans. Audio, Speech and Language Processing*, vol. 19, no. 4, pp. 916–926, 2011.

[6] M. Abdolahi and H. Amindavar, "GARCH coefficients as feature for speech recognition in persian isolated digit," in *Proc. 30th IEEE Int. Conf. Acoust., Speech Signal Process.,(ICASSP)*, 2005, pp. 957–960.

[7] A. Noiboar and I. Cohen, "Anomaly detection based on wavelet domain GARCH random field modeling," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, pp. 1361–1373, 2007.

[8] A. A. Weiss, "Asymptotic theory of ARCH models estimation and testing," *Econometric Theory*, vol. 2, pp. 107–131, 1989.

[9] A. Bose and K. Mukherjee, "Estimating the ARCH parameters by solving linear equations," *Journal of Time series analysis*, vol. 24, pp. 127–136, 2003.

[10] S. Mousazadeh and M. Karimi, "ARCH parameter estimation via constrained two stage least squares method," in *Proc .IEEE International Symposium on Signal Processing and its Application, (ISSPA)*, 2007, pp. 1–4.

[11] S. Mousazadeh and I. Cohen, "Simultaneous parameter estimation and state smoothing of complex GARCH process in the presence of additive noise," *signal processing*, vol. 90-11, pp. 2947–2953, 2010.

[12] G. Poyiadjis, A. Doucet, and S. S. Singh, "Particle methods for optimal filter derivative: Application to parameter estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, 2005, pp. v/925– v/928.

[13] S. S. Haykin, *Kalman Filtering and Neural Networks*, John Wiley & Sons, Inc., New York, NY, USA, 2001.

[14] D. G. Luenberger, *Linear and nonlinear programming*, Addison- Wesley Publishing Company, Inc, 1989.

[15] A. G. Gray, "Nonparametric density estimation: toward computational tractability," in *SIAM International Conference on Data Mining*, 2003, pp. 203–211.