LEAST SQUARES DETECTION OF MULTIPLE CHANGES IN FRACTIONAL ARIMA PROCESSES

Martial Coulon
INP-ENSEEIHT / TESA
Toulouse, FRANCE
martial.coulon@tesa.prd.fr

Ananthram Swami
Army Research Laboratory
Adelphi, MD, USA
a.swami@ieee.org

ABSTRACT

We address the problem of estimating changes in fractional integrated ARMA (FARIMA) processes. These changes may be in the Long Range Dependence (LRD) parameter or the ARMA parameters. The signal is divided into "elementary" segments: the objective is then to estimate the segments in which the changes occur. This estimation is achieved by minimizing a penalized least-squares criterion based on the parameter estimates computed in each segment. The optimization problem is then solved using a dynamic programming algorithm. Simulation results on synthetic data are reported.

1. INTRODUCTION

Recent studies have shown the self-similar nature of ethernet or internet packet traffic data [8], [12]. These studies suggest new approaches for analysis and understanding of traffic data, and in particular for traffic monitoring. One of the key problems is the accurate estimation of the Hurst parameter, which describes the degree of Long-Range Dependence (LRD) in the data. This problem has been addressed in many articles (see for instance [1] and references therein). In particular, Abry and Veitch [11] have developed a wavelet-based estimator which is unbiased, consistent, and with "quasi-minimal" variance. Important advantages of this estimator are: 1) it is robust to polynomial trends, as well as to level shifts in mean and/or variance of the processes [10]; 2) its low computational cost ($O(n)$) allows real-time applications; 3) it is useful for the detection of a *single* change in the LRD parameter.

However, it seems unlikely that the network traffic can be accurately modeled by one or two parameters (i.e., the Hurst parameter, and the magnitude of LRD effects). FARIMA models have been suggested to be good models for bursty data such as variable bit rate video (VBR) traffic [3]; the FARIMA model with heavy-tailed innovations was used in [5], and some buffer allocation schemes are studied in [7]. The FARIMA model captures the LRD nature of the data; other authors have used Gaussian innovations [7].

Here, we model traffic data as a FARIMA process; we use extensions of the Abry-Veitch estimator in order to detect *multiple* changes in the process parameters. Adaptive estimation of parameters and tracking are useful for dynamic bandwidth allocation, which is important since practical communication systems are constrained by finite bandwidth.

2. PROBLEM FORMULATION

The observed signal $Y_N := \{y(n),\ n = 1, \cdots, N\}$ is modeled as

$$y(n) = y_i(n)\ ,\ n \in \{t_{i-1}+1; t_i\}\ ,\ i \in \{1, \ldots, M\} \quad (1)$$

with $t_0 = 0$ and $t_M = N$. In (1), $y_i(n)$ is a FARIMA(p, d_i, q) process with LRD parameter d_i, and ARMA parameter vectors $\underline{a}_i \triangleq [a_i(0), \cdots, a_i(p)]$, and $\underline{b}_i \triangleq [b_i(0), \cdots, b_i(q)]$. Hence, process $y_i(n)$ satisfies the difference equation:

$$\Phi_i(z^{-1})(1 - z^{-1})^{d_i} y_i(n) = \Psi_i(z^{-1})\varepsilon_i(n) \qquad (2)$$

where $\Phi_i(z^{-1}) = \sum_{k=0}^{p} a_i(k) z^{-k}$, $\Psi_i(z^{-1}) = \sum_{k=0}^{q} b_i(k)$ z^{-k}, $(1 - z^{-1})^{d_i} = \sum_{k=0}^{+\infty} \frac{\Gamma(d_i+1)}{k!\Gamma(d_i-k+1)} z^{-k}$, and $\varepsilon_i(n)$ is an i.i.d. possibly non-Gaussian sequence with finite mean and variance. Moreover, it is assumed that $0 < d_i < 1/2$ to ensure LRD [2]. The process $y_i(n)$ can then be considered as an i.i.d. input successively filtered by the FARIMA$(0, d_i, 0)$ filter and the ARMA(p, q) filter with parameter vectors $\underline{a}_i$ and $\underline{b}_i$ [2].

The problem addressed in this paper is the estimation of the changepoints $\{t_i \,| i = 1, \cdots, M-1\}$, given the data Y_N.

It is assumed that the parameters d_i, $\underline{a}_i$ and $\underline{b}_i$, as well as the distribution of the processes $y_i(n)$, $i \in \{1, \ldots, M\}$ are unknown. In particular, these processes may have different distributions, e.g., $y_1(n)$ may be Gaussian, $y_2(n)$ exponential, etc. This allows us to deal with a large scale of data traffic processes, which are often modeled as non-Gaussian processes [5]. In particular, the estimation procedure is applicable to the case where the change affects the distribution of the process. It is assumed that the model orders p and q are known and identical for all $y_i(n)$. The case of unknown and changing model orders is discussed in Section 6.

3. HYPOTHESES

Because of the LRD, the effects of a change in the Hurst parameter cannot be observed just after the change. In other words, it is difficult (and even useless) to detect changes with an accuracy equal to one sample. Therefore, this study is restricted to the location of changes within small segments (hence it is implicitly assumed that two changes cannot be separately detected if they occur in the same segment). More precisely, consider a segmentation of $\{n = 1, ..., N\}$ into K "elementary" segments

$$I_k \triangleq [kN_s + 1; (k+1)N_s] \, (k \in \{1, \ldots, K\})$$

of N_s samples (i.e., $N = K \times N_s$). Then, the actual signal $y(n)$ is accurately approximated by the signal $\widetilde{y}(n)$, which is equal to $(y(n))_{n=1,...,N}$, except in the elementary segments where the changes occur: in these segments, denoted by (I_{l_i}), $1 \leq i \leq M-1$, $\widetilde{y}(n)$ has the FARIMA structure with parameters $(d_i, \underline{a}_i, \underline{b}_i)$ over the whole segment, whereas the FARIMA parameters of $y(n)$ change from $(d_i, \underline{a}_i, \underline{b}_i)$ to $\left(d_{i+1}, \underline{a}_{i+1}, \underline{b}_{i+1}\right)$ inside the segment I_{l_i}. Consequently, the signal $y(n)$ is considered as if it were the signal $\widetilde{y}(n)$ (note however that the algorithm operates on the actual data $y(n)$).

Notation:

- $\theta_k \triangleq \left(d_k, \underline{a}_k^T, \underline{b}_k^T\right)^T$: the true parameter vector of the FARIMA process $\widetilde{y}(n)$ over the segment I_k, $k = 1, \ldots, K$;
- l_i, $i = 0, \cdots, M$: the number of the elementary segment where the ith change occurs, i.e., the last segment of the signal $\widetilde{y}_i(n)$; here, $l_0 \triangleq 0$.
- θ_i^*, $i = 1, \ldots, M$: the true parameter vector of the process $\widetilde{y}_i(n)$, over the segments $(I_k)_{l_{i-1}+1 \leq k \leq l_i}$.

Then, one should have:

$$\theta_k = \theta_i^* \, , k = l_{i-1}+1, \ldots, l_i, \; i = 1, \ldots, M.$$

The method derived in this paper consists of estimating the FARIMA parameters over these K segments, yielding estimated vectors $\widehat{\theta}_k \triangleq \left(\widehat{d}_k, \widehat{\underline{a}}_k, \widehat{\underline{b}}_k\right)$ for $k = 1, \ldots, K$, and of statistically detecting changes in the estimated parameters, $\left(\widehat{\theta}_k\right)$. The problem reduces to detecting changes in the mean of $\left(\widehat{\theta}_k\right)$ provided that the estimates $\widehat{\theta}_k$ of θ_k are (asymptotically) unbiased. In this case, a change detected at the lth lag of $\left(\widehat{\theta}_k\right)$ means that a change occurred *during* segment I_k in the process $y(n)$. However, it is not possible to refine the change location within this segment. Thus, the resolution of the change detector in $y(n)$ is equal to the length of the "elementary" segments, N_s. Consequently, the choice of N_s must be a trade-off between high/low resolution and the sufficiently large number of samples required to accurately estimate the FARIMA parameters.

4. PARAMETER ESTIMATION

We propose a three step procedure to estimate the FARIMA parameters.

1. The estimate of the LRD parameter, $\widehat{d}_k$, is obtained via the Abry-Veitch algorithm [11];
2. The data $(y(n))\, , n \in I_k$ are filtered by the FARIMA $(0, -\widehat{d}_k, 0)$ filter, yielding the signal $(z_k(n))$, $n = 1, \ldots, N_s$. If the estimate of $\widehat{d}_k$ is sufficiently accurate, then $z_k(n)$ is an ARMA process.
3. The ARMA parameters are estimated from $(z_k(n))$, $n = 1, \ldots, N_s$ using standard techniques. For example, one could use the Yule-Walker equations for the AR estimate, and the long AR method for the MA estimate [9].

It is shown in [11] that the estimate $\widehat{d}_k$ of d_k is unbiased and consistent. Thus, for N_s large enough, the process obtained in step 2 should be close to the true ARMA part of the FARIMA process. Consequently, the techniques used in step 3 are well suited and should yield asymptotically unbiased and consistent estimators. Hence, the estimates $\widehat{\theta}_k$ of θ_k should be (asymptotically) unbiased and consistent, and the problem indeed amounts to detecting changes in the mean of $\left(\widehat{\theta}_k\right)_k$, as mentioned above. A detailed analysis of the consistency of the parameter estimates is beyond the scope of this paper, and will be addressed in another article.

Since the true parameters d_i, $\underline{a}_i$ and $\underline{b}_i$ are unknown for all segments, the true vector θ_i^* is also unknown, and the problem amounts to estimating changepoints of unknown amplitudes in the multi-dimensional signal $\left(\widehat{\theta}_k\right)_{k=1,...,K}$.

- In the case of a **known** number of changes, the estimates of the change points $\widehat{l} \triangleq \left(\widehat{l}_1, \ldots, \widehat{l}_{M-1}\right)$ in the signal $\left(\widehat{\theta}_k\right)$, $k = 1, \ldots, K$ can be obtained by minimizing the following least-squares criterion:

$$\widehat{l} = \underset{l=(l_1,...,l_{M-1})^T}{\arg\min} \sum_{i=1}^{M} \sum_{k=l_{i-1}+1}^{l_i} \left\| \widehat{\theta}_k - \theta_i \right\|^2 \qquad (3)$$

where $\overline{\theta}_i \triangleq \frac{1}{l_i - l_{i-1}} \sum\limits_{k=l_{i-1}+1}^{l_i} \widehat{\theta}_k$. The estimated change points in the signal $\widetilde{y}(n)$ are then given by $\widehat{t}_i = \widehat{l}_i N_s$.

- In the case of an **unknown** number of changes, the previous criterion has to be penalized, leading to:

$$\widehat{l} = \underset{l=(l_1,...,l_{M-1})^T}{\arg\min} \sum_{i=1}^{M} \sum_{k=l_{i-1}+1}^{l_i} \left\| \widehat{\theta}_k - \overline{\theta}_i \right\|^2 + \gamma(K) M \qquad (4)$$

where $\gamma(K)$ is a positive penalizing term which fixes the resolution of the estimate, i.e., the number of detected changes.

As shown above, the estimates are obtained by minimizing a function on a finite but generally very large set (except for M and/or K small). In such case, one generally resorts to MCMC algorithms, such as simulated annealing. However, in this particular case, one can generalize the approach used in [4] based on the dynamic programming algorithm. The main advantages of this algorithm compared with MCMC methods are: 1) it yields the exact solution to the minimization problem, and 2) its computational cost is much lower than that for simulated annealing.

5. SIMULATION RESULTS

Many simulations have been performed to validate the method presented above. We first present the results obtained on a synthetic signal $y(n)$ formed by the concatenation of six FARIMA$(1, d, 2)$ processes ($p = 1$ and $q = 2$). These processes are generated by filtering a zero-mean i.i.d. input successively through the FARIMA$(0, d, 0)$ filter and the ARMA(p, q) filter. All sub-processes$(y_i(n))$, $i = 1, \ldots, 6$ are next normalized, so that they all have the same mean and variance. Consequently, the change detection cannot be achieved by simply looking for changes in the mean or the variance of the process.

The significant parameters of the algorithm are: i) the length of the "elementary" segments N_s; ii) the position of the abrupt changes, and more precisely the distance between two successive changes (in order to visualize the resolution of the detector); iii) the parameter values of the different FARIMA processes: in particular, it is interesting to analyse what happens when two adjacent sub-processes have close parameters; iv) the distributions of the different processes, in particular: does it matter whether the process is Gaussian or non-Gaussian? v) the penalizing term in the case of an unknown number of changes.

Because of space limitations, only points ii), iii), and iv) are considered in this paper. In the following simulations, the number of samples is fixed to $N = 2^{15} = 32768$), and the elementary segment length to $N_s = 1024$. The true changepoints are $t = [$6200;12300;16400;24500;27800;32768$]$, so that these changes fall within the segments I_7, I_{13}, I_{17}, I_{24}, I_{28}, I_{32}. Note that these lengths are realistic for traffic data analysis; further there is no quasi-periodicity in the changepoint sequence. Finally, the input sequences $\varepsilon_i(n)$ of the FARIMA processes $y_i(n)$ are successively normally and exponentially distributed.

Two cases are considered: in the first one, the adjacent sub-signals $y_i(n)$ have very different parameters, and in the second, these parameters are closer.

5.1. "DISTANT" PARAMETERS

The different parameters for the six sub-signals are:

	y_1	y_2	y_3	y_4	y_5	y_6
d	0.3	0.45	0.2	0.35	0.15	0.4
a_1	-0.5	0.4	0.8	-0.2	-0.5	0.9
b_1	0.8	0.5	-0.3	-0.9	-0.6	0.9
b_2	0.4	-0.7	0.2	-0.2	0.4	0.7

Fig. 1 shows the mean of the estimated abrupt changes computed over the 100 trials. The top figure presents the data Y_N and the actual change locations. The bottom figure presents the estimated change locations. These estimates are quite good: indeed, the pairs of estimated changes are separated by a time interval equal to N_s, i.e., the algorithm finds either l_i or $l_i + 1$ as the actual change point. This is quite satisfying, since we have only 32 samples in $\left(\widehat{\theta}_k\right)$ and six changes to estimate (i.e., we have only about 5 samples of $\left(\widehat{\theta}_k\right)$ per segment).

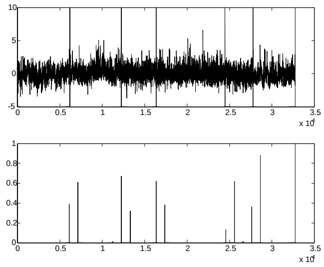

Fig. 1: mean of the estimated changes computed from 100 runs. $N_s = 1024$ - "distant" parameters.

5.2. "CLOSE" PARAMETERS

Now the parameters in some adjacent intervals are closer, so that the changes are smaller. The parameters are:

	y_1	y_2	y_3	y_4	y_5	y_6
d	0.3	0.25	0.2	0.45	0.4	0.2
a_1	-0.5	-0.4	-0.1	0.8	0.7	0.4
b_1	0.4	0.3	-0.2	-0.4	-0.6	0.5
b_2	0.8	0.7	0.8	0.2	0.4	0.6

The mean of the estimated abrupt changes (computed over the 100 trials) is shown in fig. 2.

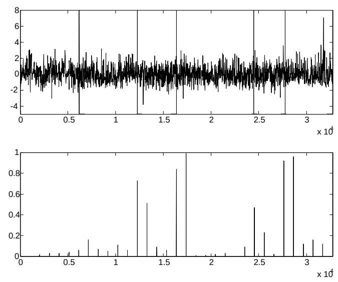

Fig. 2: mean of the estimated changes computed from 100 runs. $N_s = 1024$ - "close" parameters.

As expected, the algorithm has more difficulty in finding the exact changes (no change is detected at $l = 7$), in particular for those with small changes ($l = 7$ and $l = 24$). In this case, we should allow a greater range of possible numbers of changes, with a suitable penalizing term γ.

The lower performance of the estimator in this case can be understood from figs. 3-6, which present the parameter estimates corresponding to these simulations. Indeed, we can see fig. 3 that the LRD parameter estimation is not so accurate around $l = 7$ and $l = 24$, where the jumps are not really significant. This poor accuracy may explain why the changes are hardly detected for these points. However, it is surprising to note that, despite this poor LRD parameter estimation, the ARMA parameter estimation remains satisfying around $l = 7$ and $l = 24$. This observation tends to show that the ARMA estimation (performed after LRD estimation and filtering) is quite robust to bias in the LRD estimation.

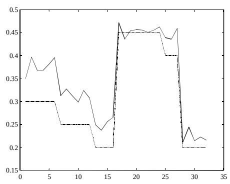

Fig. 3: LRD parameter.

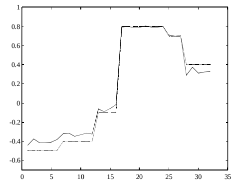

Fig. 4: AR parameter.

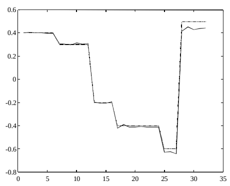

Fig. 5: first MA parameter.

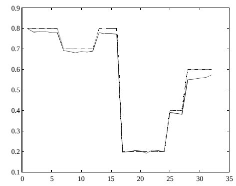

Fig. 6: second MA parameter.

Fig. 3 to 6: estimation of the parameters over each of the 32 elementary segments.
Solid line: mean over the 100 runs - Dotted line: true parameter. $N_s = 1024$ - "close" parameters.

6. CONCLUSION AND DISCUSSION

The detection of multiple changes in FARIMA processes was studied. The objective was to localize changepoints in elementary segments. This was achieved by developing a three-step algorithm based on the Abry-Veitch estimate of the LRD parameter. Simulations on synthetic data showed good performance of the detector. However, many problems still remain and need to be investigated:

1. What happens when the model orders p and q are unknown and varying? It is obviously possible to estimate them when dealing with the ARMA parameter estimation. Indeed, we could use classical MDL or Akaike criteria. Now, the problem is that one would obtain parameter vectors $\left(\widehat{\theta}_k\right)_k$ with different dimensions. The criteria (3) and (4) would no longer be valid; these criteria need to be generalized. Moreover, due to possibly inaccurate estimation, a decrease in the performance of the detector should be expected.

2. In criteria (3) and (4), the standard L^2 norm on the set $\mathbb{R}^{p+q+1}$ is used . However, d must lie in $]0; 0.5[$ for LRD, whereas the range of the ARMA parameters (particularly MA) is not restricted (AR part is assumed to be minimum-phase). It should be interesting to use a weighted norm, which would give higher weights to the LRD parameter. How should these weights be chosen? Would it significantly change the performance of the detector?

3. We now provide a heuristic argument of consistency of the changepoint detector: 1) the LRD and ARMA parameter estimators are consistent; 2) the least-squares estimator of changes in the mean in the case of unknown amplitudes is consistent (see for instance [6]). The term "consistent" in our case should be made precise: indeed, only discrete time changepoints are considered, so that the notion of neighborhood cannot be defined as in the continuous case. Using the approach given in [6]: the discrete change points (t_i) , $i = 1, \ldots, M-1$ in the signal $(y(n))$, $n = 1, \ldots, N$ correspond to continuous change points $(\tau_i)_{i=1,\ldots,M-1}$ such that $t_i = [N\tau_i]$, $i = 1, \ldots, M-1$. It is assumed that there exists $\Delta > 0$ such that $\tau_i - \tau_{i-1} \geq \Delta$, $i = 1, \ldots, M-1$. In this case, one can deal with the consistency of $(\tau_i)_{i=1,\ldots,M-1}$, instead of that of $(t_i)_{i=1,\ldots,M-1}$, which is not rigorously defined.

4. Intuitively, it may be possible to generalize this approach to the estimation of non-abrupt changes, i.e., (almost) continuous variations of the different parameters. The first step of the detector, i.e., the estimation of the FARIMA parameters, would be unchanged. Only the least-squares estimation of $\left(\widehat{\theta}_k\right)$ would have to be modified to take this structure into account. It may be sufficient to introduce parameters related to the expected variations (e.g., 2 parameters per segment for linear variations, 3 parameters for quadratic variations,...). The minimization would then be achieved with respect to the changepoints as well as these parameters.

7. REFERENCES

[1] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch, "Wavelets for the Analysis, Estimation and Synthesis of Scaling Data", in Self-Similar Network Traffic and Performance Evaluation, K. Park and W. Willinger, Eds., Wiley Int., 2000.

[2] J. Beran, Statistics for Long-Memory Processes, Chapman&Hall, New York, 1959.

[3] J. Beran, R. Sherman, M. Taqqu and W. Willinger, "LRD in VBR traffic", IEEE Trans. Commun., 43, 1566–79, 1995.

[4] M. Coulon and J.-Y. Tourneret, "MAP and LS Estimation of Abrupt Changes in Multiplicative Noise using Dynamic Programming," Proc. of EUSIPCO'2000, Tampere, Finland, September 2000.

[5] J. Ilow, "Forecasting Network Traffic using FARIMA Models with Heavy Tailed Innovations", Proc. of ICASSP'2000, Istanbul, Turkey, May 2000.

[6] M. Lavielle et E. Moulines, "Least Squares estimation of an unknown number of shifts in a time series," to appear in J. Time Series Analysis, 1998.

[7] K. Nagarajan and G.T. Zhou, "A new resource allocation scheme for VBR video sources", Proc. Asilomar,Nov 2000.

[8] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," IEEE/ACM Transactions on Networking, Vol.2, N^o 1, pp. 1-15, Feb. 1994.

[9] B. Porat, Digital Processing of Random Signals : Theory and Methods, Prentice-Hall, 1994.

[10] M. Roughan and D. Veitch, "Measuring Long-Range Dependence under Changing Traffic Conditions", available at http://www.emulab.ee.oz.au/~darryl.

[11] D. Veitch and P. Abry, "A Statistical Test for the Time Constancy of Scaling Exponents", Preprint, available on http://www.emulab.ee.oz.au/~darryl..

[12] W. Willinger, M.S. Taqqu, R. Sherman, and D.V. Wilson, "Self-Similarity through High-Variability: Statisitical Analysis of Ethernet LAN Traffic at the Source Level", IEEE/ACM Transactions on Networking, Vol.5, N^o 1, pp. 71-96, Feb. 1997.