A ROBUST IDENTIFICATION ALGORITHM FOR TRAFFIC MODELS IN TELECOMMUNICATIONS

Omar Ait-Hellal[1], **Eitan Altman**[1]

INRIA, 2004 Route des Lucioles,
06902 Sophia Antipolis Cedex, France.
Tel : (+33) 04 92 38 77 86
Email: {oaithel,altman}@sophia.inria.fr

Tamer Başar[2]
Coordinated Science Laboratory,
University of Illinois, 1308 West Main Street,
Urbana, IL 61801-2307, USA.
Tel: (217) 333-3607; Fax: (217) 244-1653;
Email: tbasar@decision.csl.uiuc.edu.

ABSTRACT

The goal of this paper is to present a robust identification algorithm for the parameters of AR stochastic processes, and employ this in analyzing telecommunication traffic data, mainly video traffic (VBR) in ATM networks.

1. Introduction

There has been growing interest in MA, AR and AR-MA models of traffic in telecommunications networks in recent years. Such models allow not only for the identification of the characteristics and understanding of the behavior of traffic, but also for queueing performance analysis (e.g. [1]) as well as optimization and control issues.

ARMA models have been used for characterizing LAN traffic [5], video codec sources in ATM networks [7, 14], and ATM traffic [9]. In the latter, the ARMA traffic model has been shown to perform better than a two-state MMPP (Markov Modulated Poisson Process). Comparisons with other traffic models can be found in [13].

An important feature of ARMA models is that they can easily be used in control design. We cite in this context [15], which uses an auto-regressive approach for designing a predictive congestion control in the presence of other real time traffic. Two other references, [2] and [3], have analyzed the stability of rate based flow control of ABR (Available Bit Rate) traffic in ATM in the presence of exogenous VBR traffic. In [4], a robust control approach has been used for the design of rate-based flow controls for ABR traffic, in the presence of exogenous traffic modeled by ARMA processes.

The goal of this paper is to present a robust identification algorithm for the parameters of AR stochastic processes, and employ this in analyzing telecommunication traffic data.

1. The work by the first two authors was supported by the France-Telecom C.N.E.T. through research grant CTI 97 1B 206

2. Research supported by Grants ANI 9813710 and INT-9804950 from the National Science Foundation.

2. The model

We consider in this paper the estimation of the parameters $\ell_1, \ldots, \ell_d$ of the AR process

$$\xi_n = \sum_{i=1}^{d} \ell_i \xi_{n-i} + k\phi_n \,, \tag{1}$$

where k is taken to be a constant, and equal to 1 without any loss of generality. For the special case when the driving noise ϕ_n is i.i.d. Gaussian, the least squares approach can be used to obtain an optimal estimator for both its variance (in case it is unknown) as well as the coefficients ℓ; see e.g. [10, 12, 16].

We shall adopt here, instead, a *robust* identification approach. This technique often leads to better performance when the noise is correlated and/or non-stationary, and it is more robust to modeling imprecisions. These advantages were already observed in the design of optimal controllers with noisy information and dynamics driven by some "noise", see [4] for examples of dynamic flow control in high speed telecommunications networks. Our derivation here will follow closely the methodology introduced in [6] where the robust identification problem was addressed in a continuous-time setting. We use the following vector notation of the dynamics (1):

$$\xi_{n+1} = \eta_n' \ell + \phi_n \tag{2}$$

where $\ell \stackrel{\text{def}}{=} (\ell_d, \ldots, \ell_1)'$ and $\eta_n \stackrel{\text{def}}{=} (\xi_{n-d+1}, \xi_{n-d+2}, \cdots \xi_n)'$. ϕ_n is an unknown noise sequence (satisfying some persistency of excitation conditions, to be made precise later).

We wish to obtain a sequence of (progressively improving) estimates for ℓ, to be denoted $\hat{\ell}_n$ at step n, so that $\hat{\ell}_n$ would depend on all the past and present values of $\xi_{(\cdot)}$, i.e. $\hat{\ell}_n = \hat{\ell}(\xi_n, \xi_{n-1}, \ldots, \xi_0)$. The criterion to be used (which is to be minimized) is of the H$^\infty$ type, which is the gain from the energy of the unknowns to a weighted quadratic identification error: $J(\{\hat{\ell}_n\}_{n=0}^{\infty}) :=$

$$\sup_{\{\phi_n\}_{n=0}^{\infty}, \ell} \frac{\sum_{n=0}^{\infty} (\ell - \hat{\ell}_n)' Q_n (\ell - \hat{\ell}_n)}{\sum_{n=0}^{\infty} |\phi_n|^2 \; + \; (\ell - \overline{\ell}_0)' \overline{Q}_0 (\ell - \overline{\ell}_0)} \tag{3}$$

Here $\overline{\ell}_0$ is some initial estimate for ℓ, $\overline{Q}_0 > 0$ is a fixed weighting matrix, and $Q_n \geq 0$, $n = 0, 1, \ldots$ is a sequence of weighting matrices, allowed to depend causally on $\{\xi_n\}$, which will be specified further in due course. Our objective is to find a sequence $\hat{\ell}^* := \{\hat{\ell}_n^*\}_{n=0}^{\infty}$, with the properties that

$$J(\hat{\ell}^*) = \inf_{\hat{\ell}=\{\ell_n\}_{n=0}^{\infty}} J(\hat{\ell}) =: (\gamma_{\text{id}}^*)^2 \tag{4}$$

and $\lim_{n\to\infty} \hat{\ell}_n^* = \ell\,, \quad \forall$ admissible $\{\phi_n\}$.

We will in fact obtain a parameterized family of such identifiers, $\hat{\ell}^\gamma := \{\ell_n^\gamma\}_{n=0}^{\infty}$, such that $\hat{\ell}^\gamma$ achieves the finite upper bound for the zero-sum dynamic game with soft-constrained kernel

$$\tilde{L}_\gamma := \sum_{n=0}^{\infty} |\ell - \hat{\ell}_n|^2_{Q_n} - \gamma^2 \sum_{n=0}^{\infty} |\phi_n|^2 - \gamma^2 |\ell - \overline{\ell}_0|^2_{\overline{Q}_0} \tag{5}$$

where the maximizing variables are $\{\phi_n\}_{n=0}^{\infty}$ and ℓ, and the minimizing variables are $\{\hat{\ell}_n\}_{n=0}^{\infty}$.

3. Robust estimation of the parameters ℓ

Theorem 1 below provides the complete solution to this identification problem. First we introduce some notation, and some relevant expressions.

Let $\Sigma_n, n = 1, 2, \ldots$ be a sequence of $d \times d$-dimensional positive-definite matrices, generated by

$$\begin{aligned} \Sigma_{n+1} &= \Sigma_n + \eta_{n-1}\eta_{n-1}' - \frac{1}{\gamma^2} Q_n\,; \\ \Sigma_1 &= \overline{Q}_0 - \gamma^{-2} Q_0 \end{aligned} \tag{6}$$

where γ is assumed to be chosen large enough so that $\Sigma_n > 0$ for $n \geq 1$.[1] Further, let $r_n, n \geq 0$, be a sequence of d-dimensional vectors generated by

$$\begin{aligned} r_{n+1} &= r_n + (\Sigma_{n+1} + \eta_n \eta_n')^{-1} (\xi_{n+1} - \eta_n' r_n)\, \eta_n\,, \\ r_0 &= \overline{\ell}_0 \end{aligned} \tag{7}$$

Finally, we characterize the class of admissible noise sequences $\{\phi_n\}_{n=0}^{\infty}$, in terms of the two conditions below:

Condition 1 : *Boundedness*: $\sum_{n=0}^{\infty} (A_n - \gamma^{-2} B_n) < \infty$, where

$$\begin{aligned} A_n \;:=\; & \left(\Sigma_{n+1}\tilde{\ell}_n + \phi_n \eta_n\right)' (\Sigma_{n+1} + \eta_n \eta_n')^{-1} \\ & \times \left(\Sigma_{n+1}\tilde{l}_n + \phi_n \eta_n\right) - \tilde{l}_n' \Sigma_{n+1} \tilde{\ell}_n \end{aligned} \tag{8}$$

$$\begin{aligned} B_n \;:=\; & \left(\Sigma_{n+1}\tilde{\ell}_n + \phi_n \eta_n\right)' (\Sigma_{n+1} + \eta_n \eta_n')^{-1} Q_n \\ & \times (\Sigma_{n+1} + \eta_n \eta_n')^{-1} \left(\Sigma_{n+1}\tilde{\ell}_n + \phi_n \eta_n\right) \end{aligned} \tag{9}$$

and

$$\begin{aligned} \tilde{l}_{n+1} &= (\Sigma_{n+1} + \eta_n \eta_n')^{-1} \left[\Sigma_{n+1}\tilde{l}_n - \phi_n \eta_n\right] \\ \tilde{\ell}_0 &= \ell - \overline{\ell}_0 \end{aligned} \tag{10}$$

Condition 2 : *Persistency of excitation*: $\lim_{n\to\infty} \lambda_{\min}\left(\sum_{i=0}^{n} Q_i\right) = \infty$.

[1] Note that since $\overline{Q}_0 > 0$, this condition is guaranteed to hold at $\gamma^{-2} = 0$, and by continuity for some range of finite values for γ, provided that Q_n is picked appropriately. For example, picking $Q_n = \eta_{n-1}\eta_{n-1}'$, we can assure satisfaction of this condition for all $\gamma \geq 1$, if $\overline{Q}_0 - Q_0 > 0$.

Theorem 1 *Consider the identification problem formulated in (4), and let Conditions 1 and 2 be satisfied. Then γ_{id}^* defined by (4) is finite, and for all $\gamma > \gamma_{\text{id}}^*$ we have :*
1) The matrices Σ_n, $n = 1, 2, \ldots$, generated by (6) are positive definite.
2) There exists an identification policy that achieves the (zero) upper value in the game defined by kernel (5), which is given by $\hat{\ell}_n^\gamma = r_n\,, \quad n = 0, 1, \ldots$ where the sequence $\{r_n\}$ is generated by (7).
3) For all noise sequences satisfying Conditions 1 and 2, the sequence $\{\hat{\ell}_n^\gamma\}$ converges to the true value of ℓ, i.e. $\lim_{n\to\infty} \hat{\ell}_n^\gamma = \ell$.
4) Conditions 1 and 2 are satisfied with probability 1 if $\{\phi_n\}$ is a sample of a martingale difference process, or simply of a (discrete-time) white noise process.

If $\gamma < \gamma_{\text{id}}^*$, then (6) does not generate a nonnegative-definite solution. ⋄

Remark 1 *As hinted in the earlier footnote, a special choice of Q_n that leads to an explicit value for γ_{id}^* is, for some $\lambda > 1$,*

$$\overline{Q}_0 = \lambda Q_0\,; \quad Q_n = \eta_{n-1}\eta_{n-1}'\,,\; n \geq 0 \tag{11}$$

in which case $\gamma_{\text{id}}^* = 1$. ⋄

Remark 2 If we let $\gamma \to \infty$, then $\{\hat{\ell}_n^\gamma\}$ yields the *least-squares* (LS) estimator for ℓ, in which case (6) becomes

$$\Sigma_{n+1}^{\infty} = \Sigma_n^{\infty} + \eta_{n-1}\eta_{n-1}'\,, \Sigma_1^{\infty} = \overline{Q}_0 \tag{12}$$

As another extreme case, let $\gamma = \gamma_{\text{id}}^* = 1$, when Q_0 and $Q_n, n \geq 1$ are chosen as in (11). Then, we have the *least-mean square* (LMS) estimator [8, 10] for ℓ, in which case (6) admits the simple solution $\Sigma_n^* = (\lambda - 1)\overline{Q}_0$, $n \geq 0$, and hence (7) becomes

$$\begin{aligned} r_{n+1} &= r_n + \left((\lambda - 1)\overline{Q}_0 + \eta_n \eta_n'\right)^{-1} (\xi_{n+1} - \eta_n' r_n)\, \eta_n, \\ r_0 &= \overline{\ell}_0 \end{aligned} \tag{13}$$

For related properties in the case of continuous time, see [6]. ⋄

Proof of Theorem 1. Parts of the proof parallels its continuous-time counterpart in [6], and hence the arguments in some cases will be kept brief here. Especially, proofs of the positive-definiteness of Σ_n's for all $\gamma > \gamma_{id}^*$, and the fact that they develop negative eigenvalues for $\gamma < \gamma_{id}^*$ are identical with their counterparts in the continuous time, as given in [6]. We therefore proceed here with the proof of part (2).

Consider first a finite-horizon truncation of (5): For $n \geq 1$:

$$W_n^\gamma := \sum_{i=0}^{n} |\ell - \hat{\ell}_i|^2_{Q_i} - \gamma^2 \sum_{i=0}^{n-1} |\phi_i|^2 - \gamma^2 |\ell - \overline{\ell}_0|^2_{\overline{Q}_0}$$

For $n = 0$:

$$W_0^\gamma := |\ell - \hat{\ell}_0|^2_{Q_0} - \gamma^2 |\ell - \overline{\ell}_0|^2_{\overline{Q}_0}$$

After substituting for ϕ_i from (2), and completing the squares in ℓ, we obtain the following equivalent representation for W_n^γ :

$$W_n^\gamma \equiv -\gamma^2|\ell - p_n|^2_{\Sigma_{n+1}} + m_{n+1}\,, \quad n \geq 0 \qquad (14)$$

where Σ_n is generated by (6), and p_i, m_{i+1}, $i \geq 0$, are generated by

$$\begin{aligned} p_i &= \Sigma_{i+1}^{-1}\left[\Sigma_i p_{i-1} + \xi_i\eta_{i-1} - \frac{1}{\gamma^2}Q_i\hat{\ell}_i\right]; \quad (15)\\ p_0 &= \Sigma_1^{-1}[\overline{Q}_0\overline{\ell}_0 - \frac{1}{\gamma^2}Q_0\hat{\ell}_0]\,, \end{aligned}$$

and

$$\begin{aligned} m_{i+1} &= m_i - \gamma^2 s_i + |\hat{\ell}_i - \overline{\ell}_i|^2_{(\Sigma_i+\eta_{i-1}\eta'_{i-1})\Sigma_{i+1}^{-1}Q_i}\,;\\ m_1 &= |\hat{\ell}_0 - \overline{\ell}_0|^2_{\overline{Q}_0\Sigma_1^{-1}Q_0} \qquad (16) \end{aligned}$$

where

$$\begin{aligned} \overline{\ell}_i &:= p_{i-1} \qquad (17)\\ &\quad + \left(\Sigma_i + \eta_{i-1}\eta'_{i-1}\right)^{-1}\left[\xi_i - \eta'_{i-1}p_{i-1}\right]\eta_{i-1}\,, \end{aligned}$$

$$\begin{aligned} s_{i+1} &= p'_i\Sigma_{i+1}p_i + \xi_{i+1}^2 - (\Sigma_{i+1}p_i + \xi_{i+1}\eta_i)' \quad (18)\\ &\times (\Sigma_{i+1} + \eta_i\eta'_i)^{-1}(\Sigma_{i+1}p_i + \xi_{i+1}\eta_i) \geq 0\,. \end{aligned}$$

Note that since $\Sigma_i > 0$ by hypothesis, the first term in (14) is negative and the third term in (16) is positive, and hence by choosing

$$\hat{\ell}_i = \overline{\ell}_i\,, \quad 0 \leq i \leq n, \qquad (19)$$

a nonpositive value for the game can be guaranteed. But, since the upper value cannot be negative (by construction of the game), it follows that (19) is indeed the (unique) optimal choice. Now, substituting for $\hat{\ell}_i = \overline{\ell}_i$, $i \geq 0$, in (15), and comparing the resulting expression with (17), we observe that $p_i = \overline{\ell}_i$, and using this fact in (15) we obtain (under (19)): $p_i = r_i$, $i \geq 1$; $p_0 = \overline{\ell}_0$, where r_i is generated by (7). This completes the proof of part (2), as the construction above was in forward time and independent of the length of the horizon in the finite truncation. The proof of the fact that $\Sigma_n > 0 \;\forall n$ if, and only if, $\gamma > \gamma^*_{\text{id}}$, is identical to that in the continuous-time case (see, [6], Theorem 2), and is therefore omitted.

For part (3), we consider a candidate Lyapunov function: $V(n;\tilde{\ell}) = \tilde{\ell}'\Sigma_{n+1}\tilde{\ell}$, in terms of the parameter estimation error $\tilde{\ell}_n := \ell - \hat{\ell}_n$, which has the dynamics described by (10), driven by the noise sequence $\{\phi_n\}_{n\geq 0}$. Some simple manipulations lead to

$$V(n+1;\tilde{\ell}_{n+1}) - V(n;\tilde{\ell}_n) = A_n - \gamma^{-2}B_n\,,$$

where A_n and B_n were defined by (8) and (9), respectively. Summing both sides from $n=0$ to $n=N-1$ gives

$$\begin{aligned} V(N;\tilde{\ell}_N) &\equiv \tilde{\ell}'_N\Sigma_{N+1}\tilde{\ell}_N \qquad (20)\\ &= \tilde{\ell}'_0\Sigma_1\tilde{\ell}_0 + \sum_{n=0}^{N-1}\left(A_n - \gamma^{-2}B_n\right)\,. \end{aligned}$$

Under Condition 1 the right-hand-side of (20) is bounded, whereas under Condition 2, $\lim_{N\to\infty}\lambda_{\min}(\Sigma_N) = \infty$. Hence, $\lim_{N\to\infty}\tilde{\ell}_N = 0$, for all admissible disturbances.

Finally, part (4) follows from a standard result on the extended least squares algorithm (corresponding to the case $\gamma^{-1} = 0$), by appropriately incorporating the second (negative) term (-$\gamma^{-2}B_n$) into the standard proof for that result [11]. This completes the proof of the theorem. $\diamond$

4. Estimation of the variance

In addition to the estimation of the parameters ℓ, one may also wish to obtain an estimate of the variance σ^2 of ϕ_n. Let $\hat{\sigma}_n^2$ be the estimator of σ^2 at step n of an iterative algorithm.

Let us define the autocorrelation function estimator at step n as

$$\hat{R}^n_{\xi\xi}[k] = \frac{1}{n+1}\sum_{i=k}^{n}\xi_i\xi_{i-k}, \qquad k = 0,1,...,d \quad (21)$$

We have (from the autocorrelation method for AR spectral estimation [10, 12]), and adopting the convention that $\xi_m = 0$ for $m < 0$,

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n+1}\sum_{i=0}^{n}|\phi_i|^2\\ &= \frac{1}{n+1}\sum_{i=0}^{n}\left|\xi_i - \sum_{j=1}^{d}\hat{\mathrm{l}}_j^n\xi_{i-j}\right|^2 \qquad (22)\\ &= \frac{1}{n+1}\sum_{i=0}^{n}\left(\xi_i - \sum_{j=1}^{d}\hat{\mathrm{l}}_j^n\xi_{i-j}\right)\xi_i\\ &- \frac{1}{n+1}\sum_{i=0}^{n}\left(\xi_i\sum_{j=1}^{d}\hat{\mathrm{l}}_j^n\xi_{i-j}\right)\sum_{j=1}^{d}\hat{\mathrm{l}}_j\xi_{i-j} \quad (23) \end{aligned}$$

Above, $\hat{\mathrm{l}}^n$ is the least squares estimate of ℓ at time n (see [10, 12]) (which is r_n generated by (7) where Σ is generated by (6) with $\gamma = \infty$), i.e.

$$\hat{\mathrm{l}}^{n+1} = \hat{\mathrm{l}}^n + \left(\sum_{i=0}^{\infty}\eta_i\eta'_i + \overline{Q}_0\right)^{-1}(\xi_{n+1} - \eta'_n\hat{\mathrm{l}}_n)\hat{\mathrm{l}}_n,$$

$\hat{\mathrm{l}}_0 = \overline{\mathrm{l}}_0$. The last term in (23) is equal to zero, since $\hat{\mathrm{l}}^n$ is the optimal estimator (i.e. obtained by differentiating (22) with respect to $\hat{\mathrm{l}}$, and setting it to zero); hence, from (21) we have

$$\hat{\sigma}_n^2 = \hat{R}^n_{\xi\xi}[0] - \sum_{i=1}^{d}\hat{\ell}^n l_i\hat{R}^n_{\xi\xi}[i],$$

where $\hat{R}^n_{\xi\xi}[k]$ is generated by

$$\hat{R}^{n+1}_{\xi\xi}[k] = \frac{n+1}{n+2}\hat{R}^n_{\xi\xi}[k] + \frac{1}{n+2}\xi_{n+1}\xi_{n+1-k},$$

$k = 0, ..., d$.

5. Numerical results and simulation

In the first set of examples, we compare the method presented here with the autocorrelation one [10, 12]. We will see that, for γ sufficiently large, in particular when the spectrum of $\overline{Q}_0$ is very small (e.g. $\overline{Q}_0 = \epsilon I$, where I is the identity matrix), this method converges to the autocorrelation. The latter method is known to give the best performance when the number of samples is very large.

The choice of $\overline{Q}_0$ and γ determine the quality of the estimation. As will be seen in the examples we analyzed, the case $\gamma = 1$ gave the worst performances of the identification. This is due to the fact that $\gamma = 1$ is the smallest possible value of γ, and we may thus have numerical instability at this value.

We chose $\overline{Q}_0$ as the identity matrix, and in view of Remark 1, we have set the weighting matrix $Q_n = \eta_{n-1}\eta'_{n-1}$. Recall that the choice of $\overline{\ell}_0$ has practically no effect on the results; thus we have set it equal to the zero-vector. We chose three values for γ; the intermediate case $\gamma = 3$ and the two extreme cases where $\gamma = 1$ and γ is sufficiently large ($\gamma = 100$).

We first study the performance of this method under large and small numbers of samples, and compare them with the autocorrelation method. In the first set of results given in Tables 1 and 2, a small number of samples (64 samples in each example) were generated from a zero-mean Gaussian distribution with variance σ_w, while in the second set (Tables 3-4), a large number of samples (1000 samples in each example) were generated with the same variances. These samples were taken to be the input (white noise) of an ARMA process whose order is considered, initially, to be known ($d = 3$) and the parameters $\ell_i,\ i = 1, 2, 3$ were as given in the Tables. The case where the order is unknown will be discussed later. We then applied the algorithm described in the paper to the output of the ARMA process in order to estimate the different parameters.

Tables 1-2 show the real parameters and the estimated ones for the examples with 64 samples. The method described here as well as the autocorrelation (AC) method do not estimate the real parameters, although the output process, when considering the estimated ARMA process, is practically the same as the original one. We can also see that the estimated white noise variance is very precise, except for the case where $\gamma = 1$.

Tables 3-4, however, confirm that this method as well as the autocorrelation (AC) one gives better performances when analyzing a large number of samples. The particular case where $\gamma = 1$ may result in unstable parameters as reflected in the figures given in Tables 3-4.

We next focus on the case where the order is unknown. There exist different criteria in the literature for order estimation. The one which appears to be in more general usage, is the *Akaike Information Criterion (AIC)*, which is defined as [10]

$$AIC(k) = N \ \ln \hat{\sigma}_k + 2k$$

where the optimal order (k) is the integer that minimizes the above quantity, N is the number of samples, and $\hat{\sigma}_k$ is the estimated noise variance for the considered order k. Note that this method is more precise for a large number of samples (large N). It is not practical, however, to use this criterion in our case, since the method presented here aims, in particular, to estimate the parameters in a dynamic way (real time), thus making it very expensive in terms of time and computations.

$Method$	ℓ_1	ℓ_2	ℓ_3	$\hat{\ell}_1$	$\hat{\ell}_2$	$\hat{\ell}_3$
AC	0.2	-0.25	0.15	2.1636	-0.233033	0.108599
$\gamma = 1$	0.2	-0.25	0.15	0.534902	-0.135629	-0.246114
$\gamma = 3$	0.2	-0.25	0.15	0.042431	-0.230542	0.108594
$\gamma = 100$	0.2	-0.25	0.15	0.048174	-0.232192	0.1087082
AC	-0.2	0.1	0.12	-0.340112	0.019188	0.123230
$\gamma = 1$	-0.2	0.1	0.12	0.1909286	-0.010988	0.1448226
$\gamma = 3$	-0.2	0.1	0.12	-0.346932	0.020468	0.124073
$\gamma = 100$	-0.2	0.1	0.12	-0.340363	0.019262	0.123822

Table 1: the estimated ARMA parameters with 64 samples

$Method$	σ_w	$\hat{\sigma}_w$
AC	2.1636	2.1564
$\gamma = 1$	2.1636	2.2090
$\gamma = 3$	2.1636	2.1572
$\gamma = 100$	2.1636	2.1566
AC	432.736	431.755
$\gamma = 1$	432.736	473.911
$\gamma = 3$	432.736	431.143
$\gamma = 100$	432.736	431.720

Table 2: The estimated variance with 64 samples

In the following we consider the output of an ARMA process of order $d = 4$ whose input is a white Gaussian noise (the same example above (1000 samples)) with variance $\sigma_w = 389.66$. The parameters of the original ARMA process are $\ell_1 = 0.1$, $\ell_2 = -0.3$, $\ell_3 = 0.2$ and $\ell_4 = +0.15$. γ is set to 3. Our aim is to estimate the parameters (ℓ and σ_w) by considering different orders ($d = 6, 5, 4, 3, 2$). From Table 5 we see that in the case of overmodeling (i.e. when the considered order is larger than the real one, $d = 5$ *resp.* $d = 6$), the extra parameters ($\hat{\ell}_5$ *resp.* $\hat{\ell}_5, \hat{\ell}_6$) are very close to zero. We are thus interested in the undermodeling case. Note also, that from Tables 3-4

$Method$	ℓ_1	ℓ_2	ℓ_3	$\hat{\ell}_1$	$\hat{\ell}_2$	$\hat{\ell}_3$
AC	0.2	-0.25	0.15	0.20021	-0.22773	0.11313
$\gamma = 1$	0.2	-0.25	0.15	-0.02839	0.10727	1.33372
$\gamma = 3$	0.2	-0.25	0.15	0.20253	-0.22769	0.11354
$\gamma = 100$	0.2	-0.25	0.15	0.20033	-0.22789	0.11415
AC	-0.2	0.1	0.12	-0.20065	0.11233	0.11162
$\gamma = 1$	-0.2	0.1	0.12	-0.51881	0.0312	1.24733
$\gamma = 3$	-0.2	0.1	0.12	-0.19884	0.11420	0.10984
$\gamma = 100$	-0.2	0.1	0.12	-0.20077	0.112438	0.11155

Table 3: The estimated ARMA parameters with 1000 samples

Method	σ_w	$\hat{\sigma}_w$
AC	1.9483	1.9470
$\gamma = 1$	1.9483	1.9901
$\gamma = 3$	1.9483	1.9466
$\gamma = 100$	1.9483	1.9469
AC	389.66	389.65
$\gamma = 1$	389.66	362.43
$\gamma = 3$	389.66	389.70
$\gamma = 100$	389.66	389.649

Table 4: The estimated variance with 1000 samples

($\gamma = 3$) and Table 5 the estimation is more precise when the real order is not large, in particular for the white noise variance (that is the estimation when d was set equal to 3 was better than when it was set equal to 4).

	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
$\hat{\ell}_1$	0.07653	0.13170	0.107648	0.103262	0.102389
$\hat{\ell}_2$	-0.30247	-0.31645	-0.274472	-0.279605	-0.28347
$\hat{\ell}_3$	-	0.18250	0.164473	0.172835	0.167835
$\hat{\ell}_4$	-	-	0.133342	0.129899	0.138194
$\hat{\ell}_5$	-	-	-	0.030982	0.027879
$\hat{\ell}_6$	-	-	-	-	0.029354
$\hat{\sigma}_w$	399.043	392.385	388.982	388.792	388.598

Table 5: The estimation with unknown order

To conclude, when the model to be identified is a true AR model whose dimension is known, driven by i.i.d. noise, then we obtained good convergence by our robust identification algorithm, although a faster rate of convergence was obtained with the LS scheme (even for the case of uniformly distributed i.i.d. noise, instead of the Gaussian one). We obtained, however, better performance under the robust identification scheme (for particular values of γ) when the dimension of the AR process was undershot.

As a specific example, we considered a 3rd order AR model with $l_1 = 0.1, l_2 = -0.2, l_3 = 0.15$. We estimated the parameters of a 2nd order AR model that corresponds to the output from this 3rd order one. We obtained under the robust identification scheme (with $\gamma = 1.1$), $\hat{\ell}_1 = 0.0616923268, \hat{\ell}_2 = -0.213499012$. For the LS scheme, we obtained $\hat{\ell}_1 = 0.0649439124, \hat{\ell}_2 = -0.195600381$. For a reasonable comparison, we simulated 3 AR models with the same input noise (i.i.d. standard Gaussian) over 10000 time units:
(1) $X_t =$ the ideal signal i.e. the one obtained with the true AR parameters,
(2) $Y_t =$ the output of a 2nd order AR model with the parameters estimated using robust estimation, and
(3) $Z_t =$ the output of a 2nd order AR model with the parameters estimated using the LS scheme.

We compared the L^2 norms of the differences between the output trajectories, and obtained a better fit by using robust identification:

$$\begin{aligned} 5.531687 &= T^{-1} \sum_{t=1}^{T} |Y_t - X_t|^2 \\ &< T^{-1} \sum_{t=1}^{T} |Z_t - X_t|^2 = 5.882055. \end{aligned}$$

6. An Application

The main objective of parameter identification is to approximate a given process (or traffic) by an ARMA process whose input is a white Gaussian noise. This is of great importance mainly when estimating the necessary bandwidth for a highly variable traffic, such as a VBR traffic in ATM networks or inversely the leftover CBR VBR bandwidth, which allows to better control the ABR traffic [4]. Our purpose in this section is to use this identification algorithm to estimate the ARMA process parameters approximating a given VBR traffic. We proceed as follows: The JPEG *Star Wars* traces are considered as the output of an ARMA process. We use the identification algorithm ($\gamma = 3$) to estimate the parameters and variance of the white noise, which we consider to be Gaussian. Different orders are considered ($d = 2,\ 4,\ 8$). The second step consists of generating samples from a zero-mean Gaussian distribution with the estimated variance, which will be considered as the input process of the estimated ARMA filter. We compare then the output process (traffic) to the *Star Wars* traces.

To compare the two traffics, we considered a simplified network which consists of a source (VBR) sending cells (*Star War* and the estimated traffic) to a destination via a single bottleneck node (switch). We are, thus, interested in the average ($\bar{Q}$) and the maximum queue size (Q_{max}), at the switch, due to each traffic. The frames are divided into cells and one frame is sent each 30 ms. Only the first 3500 frames are taken into account which correspond, approximately, to 100 sec simulation.

Table 6 illustrates these performances for different traffic. The traces we obtained have zero mean, hence we added 83828.65 (Bytes) to all the traces to arrive at positive means. This value corresponds to the smallest estimated value obtained for $d = 2$. The buffering capacity of the queue is infinite and the rate is about 22.35 $Mbits/s$. The average rate for the *Star War and the estimated traffic* during the 100 sec simulation is about 21.345 $Mbits/s$.

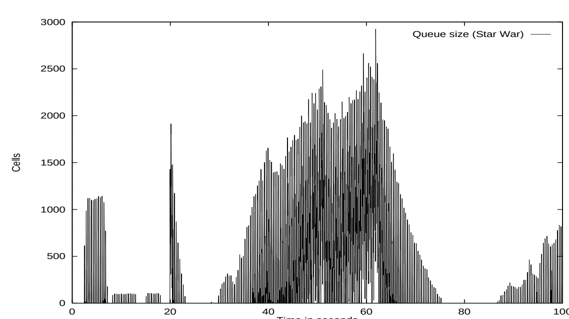

Figure 1: Star War

	Star War	ARMA $d=2$	ARMA $d=4$	ARMA $d=8$
$\hat{\ell}_1$	-	-0.189405	-0.074643	-0.043709
$\hat{\ell}_2$	-	-0.188241	-0.111603	-0.041235
$\hat{\ell}_3$	-	-	0.436488	0.287181
$\hat{\ell}_4$	-	-	-0.070675	-0.042038
$\hat{\ell}_5$	-	-	-	-0.044164
$\hat{\ell}_6$	-	-	-	0.313494
$\hat{\ell}_7$	-	-	-	-0.042470
$\hat{\ell}_8$	-	-	-	-0.042545
$\hat{\sigma}_w$	-	21960.266	19650.553	18551.793
Q	236.361	604.11	1078.24	1319.87
Q_{max}	2924	4099	8414	11188

Table 6: *Star war* and its approximating ARMA process

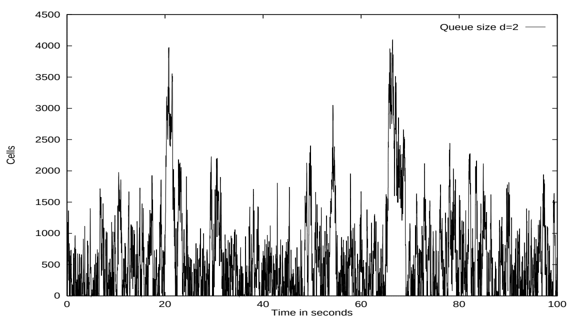

Figure 2: Approximation of Star War with ARMA $d = 2$

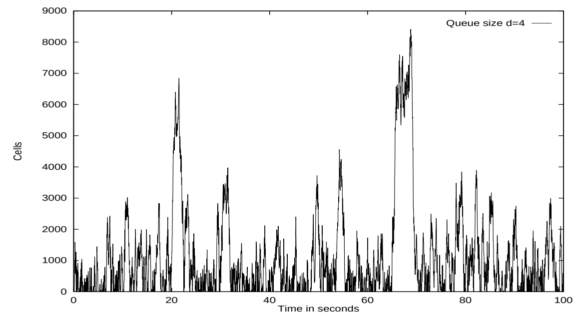

Figure 3: Queue size *vs.* time for 100 seconds simulation, for *Star War* traffic. ARMA model with $d = 4$

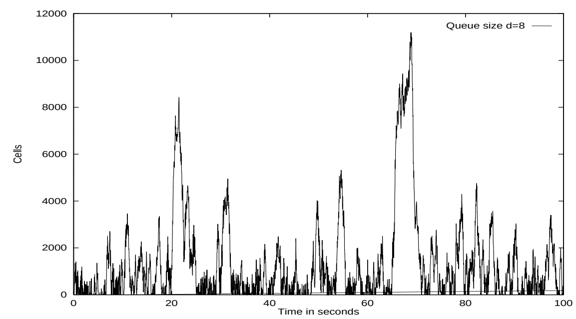

Figure 4: Queue size *vs.* time for 100 seconds simulation, for the approximation of the *Star War* traffic. ARMA model with $d = 8$

7. References

[1] R. G. Addie and M. Zukerman, "An approximation for performance evaluation of stationary single server queues," *IEEE Infocom '93*, San Francisco, 835–842, 1993.

[2] O. Ait-Hellal, E. Altman and T. Başar, "Rate based flow control with bandwidth information", *European Trans. on Telecom.*, pp. 55-66, 1997.

[3] E. Altman, F. Baccelli and J. C. Bolot, "Discrete-time analysis of adaptive rate control mechanisms", *High Speed Networks and their performance*, H.G. Perros and Y. Viniotis Eds., North Holland, 121-140, 1994.

[4] E. Altman and T. Başar, "Optimal Rate Control for High Speed Telecommunication Networks", *34th IEEE CDC*, Dec. 1995, New Orleans, Louisiana.

[5] S. Basu, A. Mukherjee, and S. Klivansky, "Time series models for internet traffic," in *IEEE Infocom '96*, San Fransisco, California, 611-619, Mar. 1996.

[6] G. Didinsky, Z. Pan and T. Başar, "Parameter identification for uncertain plants using H^∞ methods," *Automatica*, **31**, no. 9, 1227-1250, Sept. 1995.

[7] R. Grñenfelder, J. Cosmas, S. Manthorpe, and A. Odinma-Okafor, "Measurement and ARMA Model of Video Codecs in an ATM Environment," *13th ITC*, **14**, Copenhagen, 981-985, June 1991.

[8] B. Hassaibi, H. Sayed and T. Kailath, "LMS is H^∞ optimal", *Proc. IEEE 32nd Conf. on Decision and Control*, 74-79, San Antonio, Texas, USA, 1993.

[9] D. Heyman, A. Tabatabai, and T. V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks," *IEEE Trans. on Circuits and Systems for Video Technology*, **2**, 49-59, 1992.

[10] Steven M. Kay, "Modern Spectral Estimation: Theory and Application," *Prentice Hall,* 1988.

[11] T. L. Lai and C. Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *The Annals of Statistics*, **10**, no. 1, 154-166, 1982.

[12] S. Lawrence Marple Jr, "Digital Spectral Analysis with Applications," *Prentice Hall,* 1987.

[13] B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski, "TES-Based video source modeling for performance evaluation of integrated networks," *IEEE Trans. on Communications*, **42**, Oct. 1994.

[14] M. Nomura, T. Fujii, and N. Ohta, "Basic characteristics of variable rate video coding in ATM environments," *IEEE JSAC*, **7**, 752–760, 1989.

[15] G. Ramamurthy and B. Sengupta, "A predictive hop-by-hop congestion control policy for high-speed networks", *IEEE Infocom*, **3**, San Francisco, 1033-1041, 1993.

[16] S. S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness.* Prentice-Hall, Englewood Cliffs, NJ.