ADAPTIVE CONTROL IN THE PRESENCE OF SENSOR MEASURE OUTLIERS

J. M. Lemos*, **R. N. Silva**†, **J. S. Marques**‡

* INESC-ID/IST, R. Alves Redol 9, 1000-029 Lisboa, Portugal
e-mail: jlml@inesc.pt

† FCT/UNL
e-mail: rns@mail.fct.unl.pt

‡ ISR/IST
email: jsm@isr.ist.utl.pt

Keywords: Adaptive control, robust estimation, outliers, median filter, Bayes inference.

Abstract

Adaptive control of plants whose sensor measurements are corrupted by outliers is considered. Outliers are large deviations of the signal being measured, only occurring in a few percent of the observations. For adaptive controllers relying on an implicit Gaussian assumption, both the identification and the underlying control law are yielded by the minimization of quadratic losses. Therefore, although rare, outliers heavily degrade performance due to their large amplitude. This problem is tackled in this paper. Methods for outlier removal which are suitable for adaptive control applications are reviewed and illustrated through an application to position control in the ball and beam plant. [1]

1 Introduction.

A crucial factor for the success of feedback control is the availability of a reliable measure, within the bandwidth of the reference to track, of the variable to be controlled. For feedforward control a similar fact also holds, concerning the measure of the disturbance to reject. Since sensors introduce noise, a filtering of some kind is needed in order to prevent unnecessary errors (in particular those falling outside the servo bandwidth) to propagate to control decisions, causing these to be erroneous. For this sake, a filter processing sensor measures is inserted, acting as an estimator of the true value of the variable.

In most cases a Gaussian noise assumption is either explicitly or implicitly made and the estimate is a mean of the observations. This corresponds to minimize an average quadratic loss, given by

$$J(\hat{x}) = E[(x - \hat{x})^2] \tag{1}$$

where x is the variable of interest and $\hat{x}$ is its estimate.

Recall the statistical concept of *asymptotic relative efficiency* of an estimator [18]. This is defined as the ratio between the lowest achievable variance for the estimated parameters (the Cramer-Rao bound) and the actual variance provided by the estimator when a large data sample is considered. The efficiency depends on the noise model. Under Gaussian noise the mean estimator has an asymptotic efficiency of 1, therefore achieving the optimum value, while the median estimator efficiency is only 0.637. If the noise is Gaussian, the mean should thus be preferred.

The situation is however different if the noise is non-Gaussian, in particular if it presents outliers. Outliers [12] are large deviations of the signal being

[1]The work of the first and third authors has been performed within the project *FATOL – POSI/SRI/39643/2001*. The work was also performed within the POSI program of the 3rd EC Framework. The authors thank prof. Mário Figueiredo of IST for calling their attention to reference [5].

measured, only occurring in a few percent of the observed samples, but enough probable not to be explained by the tails of the Gaussian distribution. Outliers are due to unknown causes and they can hardly, if at all, be modelled from first principles.

Although rare in time, due to their large amplitude, outliers heavily contaminate the output of noise removing filters designed on the basis of a Gaussian assumption. As a consequence, the performance of an adaptive controller may be seriously degraded because of two factors:

- First, the outliers are seen as fake disturbances; while trying to counteract them, the controller actually induces undesirable changes in the plant output.
- Furthermore, the plant model identifier block on which adaptation relies is mislead (if, *e. g.*, based on recursive least squares, causing controller gains to be severely detuned. This may even lead to an unstable or highly oscillating closed-loop.

The point is that, due to its very high amplitude, an outlier will have a major influence on the quadratic criteria (1) because the square will amplify it [12].

Since outliers are due to unknown causes and seldom occur, the strategy followed in this paper is to remove them from sensor data before processing this signal by the controller.

Outlier detection and removal is a major issue in statistics [6, 17]. A robust method, with respect to deviations from normality, is defined as one which is nearly as efficient as the classical procedure (based on sample mean and variance) for a normal distribution, but is considerably more efficient overall for non-normal distributions [17]. While there is a rich bibliography on robust methods for applications such as time series [16, 1, 3, 8], computer vision [11] or system identification [**?**], robustness with respect to outliers has apparently not received much attention in the control literature. Nevertheless, the subject is of importance for plants for which reliable sensors are hard to build. Internal combustion motors and biomedic systems provide examples.

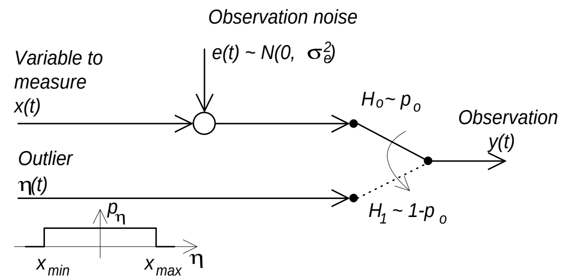

Figure 1: Model of measurement with outliers.

[15] describe how the use of a multiple model based Kalman-Bucy filter may be employed for monitoring (and ultimately making decisions) of patient condition after renal transplants using measurements corrupted by outliers. Other examples are ill defined process variables obtained from indirect measurements. In this paper, the ball and beam laboratory pilot plant is taken as an example in which outliers play a major role.

The contributions of this paper are the following: First, by means of an example (position control of the ball and beam plant) the difficulties facing adaptive algorithms relying on quadratic criteria are demonstrated. Methods which are appropriate for on-line outlier removal in adaptive control applications are then considered. One of these methods, which amounts to the inclusion of a median filter applied to an over-sampled signal, is then tested with an adaptive controller on the ball and beam plant. It is shown that, without the outlier removal filter, the controller gains converge to values leading to an highly oscillating closed-loop.

2 On-line sensor outlier removal.

Two different approaches for the design of outlier removal algorithms are considered. The first relies on Bayes inference while the second is based on the modification of the quadratic loss function of (1).

2.1 Bayes inference approach.

One possibility for modelling outliers is to assume that the observations (sensor measures) are made according to the model represented in fig. 1.

Let $y(t)$ denote the observation made at discrete time t and $x(t)$ be the corresponding true value of the variable to measure. It is assumed that $x_{min} \leq x(t) \leq x_{max}$.

Under hypothesis H_o, occurring with probability p_o, close to 1, the observation $y(t)$ is equal to the value of the variable to measure, $x(t)$, added by zero mean white Gaussian noise of (constant) variance σ_e^2, denoted $e(t)$:

$$y(t) = x(t) + e(t) \tag{2}$$

Under hypothesis H_1, which occurs with probability $1 - p_o$, close to zero, an outlier occurs. In this case, the observation is no longer related to the variable being measured but, instead, is given by a random variable $\eta(t)$ with a probability density function (p.d.f.) which is uniform in the range of measurement, from x_{min} to x_{max}.

According to a Bayesian approach, in order to detect that a given observation is actually an outlier, the probability of both hypothesis given the observations is computed. For H_o this is

$$P(H_o|y(t),\, Y^{t-1}) = C \cdot p(y(t)|H_o,\, Y^{t-1}) p_o \tag{3}$$

where Y^{t-1} is the set of observations up to $t-1$ and C is a normalizing constant. Given the model of the observations when H_o holds, (3) reads

$$P(H_o|y(t),\, Y^{t-1}) = C \frac{1}{\sigma_e \sqrt{2\pi}} e^{-\frac{(y(t)-x(t))^2}{2\sigma_e^2}} p_o \tag{4}$$

For computing (4), the value of $x(t)$ is needed. Since this is unknown, it is replaced by a convenient estimate.

For H_1,

$$P(H_1|y(t),\, Y^{t-1}) = C \frac{1}{x_{max} - x_{min}} (1 - p_o) \tag{5}$$

with C the same constant as in (4).

Both probabilities $P(H_o|y(t),\, Y^{t-1})$ and $P(H_1|y(t),\, Y^{t-1})$ are then compared. If

$$\frac{P(H_1|y(t),\, Y^{t-1})}{P(H_o|y(t),\, Y^{t-1})} > 1 \tag{6}$$

it is decided that an outlier has occurred.

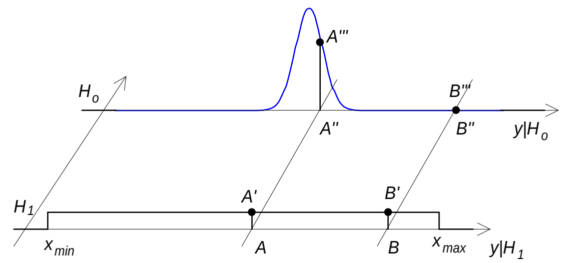

Figure 2: Probability density function of the observations in the presence of outliers.

Fig. 2 provides a graphic explanation, showing the p.d.f. of the last observation given each of both hypothesis. If the observation $y(t)$ falls at, say, point A, the length of the segment $[A,\, A']$ is smaller than the length of $[A'',\, A''']$ and H_o is selected. The opposite happens if the observation falls at point B, in which H_1 (existence of an outlier) is decided.

If it is decided that an outlier has occurred, the observation $y(t)$ is discarded and replaced by a forecast $\hat{x}(t)$ of the true value $x(t)$, made from previous observations. For computing this forecast, several possibilities may be envisaged. If a model of the mechanism producing $x(t)$ is available, Kalman-Bucy filtering [9] (or equivalent based on i/o models) might be used. Here, since the interest is on adaptive control, no model is assumed available. The choice is then to compute $\hat{x}(t)$ by linear extrapolation of the two previous estimates, yielding

$$\hat{x}(t) = 2\hat{x}(t-1) - \hat{x}(t-2) \tag{7}$$

Another possibility in this line would have been to ressort to an $\alpha - \beta$ filter [7].

2.2 Non-quadratic loss criteria.

An alternative approach is to assume that (2) always holds but that the variance σ_e^2 of the noise e is not constant but has fluctuations which cause the outliers [17, 5]. In particular the approach of [5], leading to a modification of the loss (1) is considered.

Following [5], write the likelihood function

$$p(y|x) = \sqrt{\frac{\beta}{\pi}} e^{-\beta(y-x)^2} \tag{8}$$

with

$$\beta \triangleq \frac{1}{\sigma_e^2} \tag{9}$$

Small values of β (corresponding to noise with a large variation) will model the outliers, while large values model data which may be processed by a filter designed using a quadratic criterion. Model σ_e, and therefore β, as a random variable with p.d.f. $p(\beta)$ and such that $p(\beta) = 0$ for $\beta = 0$. In this setting, (8) actually represents $p(y|x, \beta)$, the actual likelihood function being

$$p^*(y|x) = \int_0^\infty p(y|x, \beta)p(\beta)d\beta \tag{10}$$

or

$$p^*(y|x) = \frac{1}{\sqrt{\pi}} \int_o^\infty e^{-\beta(y-x)^2} \sqrt{\beta} p(\beta)d\beta \tag{11}$$

Given an observation y, the MAP estimate [18] of x is obtained by maximizing $p^*(y|x)$ with respect to x. Taking the negative of its logarithm, this reads (for just one observation, compare with (1):

$$\hat{x} = \min_x V(y - x) \tag{12}$$

with the *effective potential* [5] given by

$$V(z) = -ln \int_0^\infty e^{-\beta z^2} \sqrt{\beta} p(\beta)d\beta \tag{13}$$

If the noise variance has a precise value, $p(\beta)$ is a Dyrac delta function at that value and the quadratic criterion (1) is recovered. If, instead, $p(\beta)$ is such that it models the occurrence of outliers, modifications of the loss function (1) are yielded, leading to robust estimation algorithms. The two following examples, provided in [5], are relevant for the work to follow.

2.2.1 Two different variances.

Assume first that the inverse of the variance, β, may take only two different values, β_1 and β_2. Let $\beta_1 > \beta_2$ and assume that β_2 is very small (*i. e.* the corresponding value of the variance $\sigma_{e,2} = 1/\beta_2$ is very large) such as to model the existence of outliers. Furthermore, let β_2 occur with a small probability ε. The p.d.f. of the inverse of variance is therefore

$$p(\beta) = (1 - \varepsilon)\delta(\beta - \beta_1) + \varepsilon\delta(\beta - \beta_2) \tag{14}$$

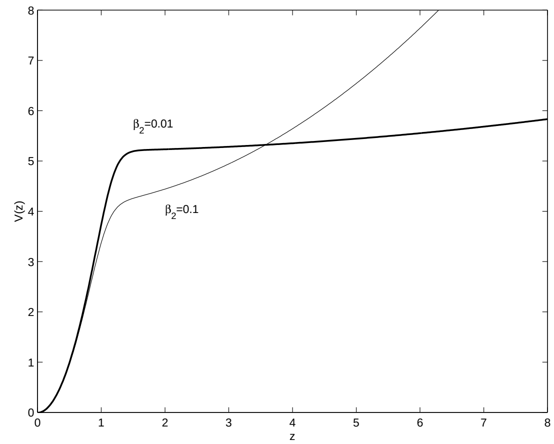

Figure 3: The loss corresponding to a p.d.f. of the noise variance with two values (according to [5]).

Although different, this model has a close relationship with the approach followed in section 2.1. Indeed, instead of a uniform p.d.f. under H_1 one could also assume a Gaussian p.d.f. with a very large variance. Due to the high variability of the resulting signal, adding the true signal would not change much the result in practical terms.

It is shown in [5] that the effective potential (13) becomes in the case of (14), apart from an unimportant constant term:

$$V(z) = \beta_1 z^2 - \ln[1 + \frac{\varepsilon}{1-\varepsilon}\sqrt{\frac{\beta_2}{\beta_1}} e^{z^2(\beta_1 - \beta_2)}] \tag{15}$$

This function is plotted in fig. 3 for $\varepsilon = 0.1$, $\beta_1 = 4$ and two values of β_2, $\beta_2 = 0.1$ and $\beta_2 = 0.01$. The resulting effective potential has clearly two zones: For small errors it behaves like a quadratic potential. Far from the origin the behaviour is much flat so that the large amplitude of outliers has not much weight on the criterion.

2.2.2 The modulus loss.

Another possibility is to model the noise standard deviation by

$$p(\sigma_e) = 2\sigma_e e^{-\frac{\sigma_e^2}{4}} \tag{16}$$

A plot of this function is seen in fig. 4. In this example, the most likely value of the variance is close to 1.5. However, higher values of σ_e may also be

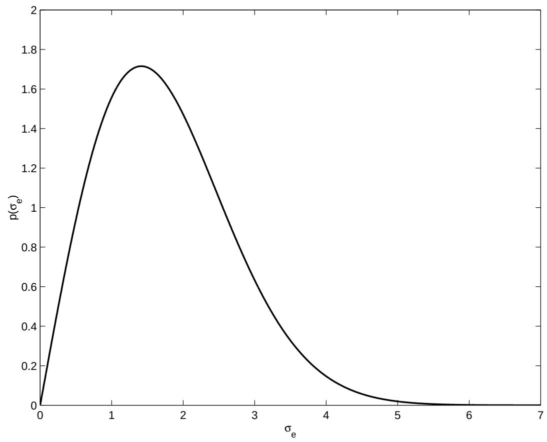

Figure 4: Probability density function of the noise variance leading to a modulus criterion (according to [5]).

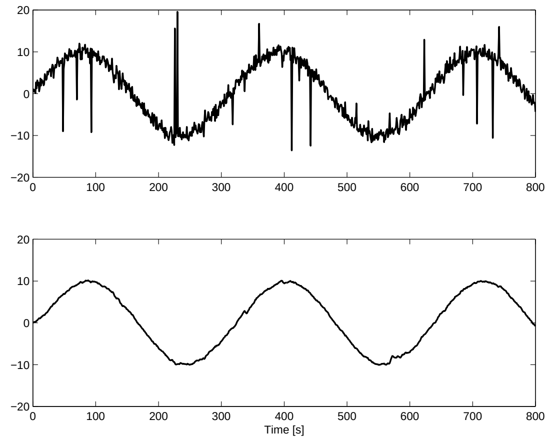

Figure 5: A sinusoid corrupted with gaussian noise and outliers (above) and its filtering with outliers removed by Bayesian inference.

produced, although with a much lower likelihood, thereby modeling outliers. For this situation, it is shown in [5] that the corresponding effective potential is the modulus of the error:

$$V(x - \hat{x}) = |x - \hat{x}| \tag{17}$$

This is a classical loss function [18], introduced heuristically in order to prevent large deviations from distorting the estimate. The result of [5] provides a link with a probabilistic model of the noise.

As is well known, the optimal estimate corresponding to (17) is the median of the signal to estimate [18].

2.3 Comparison.

As is apparent from the previous discussion, although developed from different points of view, the approaches for outlier removal considered have a number of similarities. The key issue is that outliers are modelled in both cases by ”long tails” of a pdf. Furthermore, in both cases the methods end up by relying on the replacement of a quadratic loss by a function which yields smaller values for high amplitudes of the independent variable.

Both figs. 5 and 6 show (above) a low frequency sinusoid corrupted with additive gaussian noise and outliers generated according to the model of fig. 1, and (below) the result of outliers suppressing with the Bayesian inference algorithm described in section 2.1 (fig. 5) and with a median filter (fig. 6).

Both approaches are much equivalent in practice, with the median filter being a bit simpler to implement. It should be remarked that, if the ”true” signal is lost for significant periods of time, the algorithm relying on Bayes inference may be used with advantage. In this case the procedure described in [9] may be employed, but a model better then just (7) is to be used.

When aiming at control applications, it should be kept in mind that both classes of methods introduce delay in the filtered signal. In order to prevent relative stability degradation of the control loop, the signal to filter should be sampled at a rate high enough to allow recovering by decimation. In this respect, it is also convenient to resort to an adaptive control algorithm which is robust with respect to uncertainty in delay.

3 The Ball and Beam Plant.

The ball and beam plant is a classic benchmark [2]. It consists (fig. 7) of a rail (the ”beam”) over which a small metallic sphere (the ”ball”) is to be equilibrated. The angle of the rail with respect to the horizontal is changed by an electric DC motor, whose

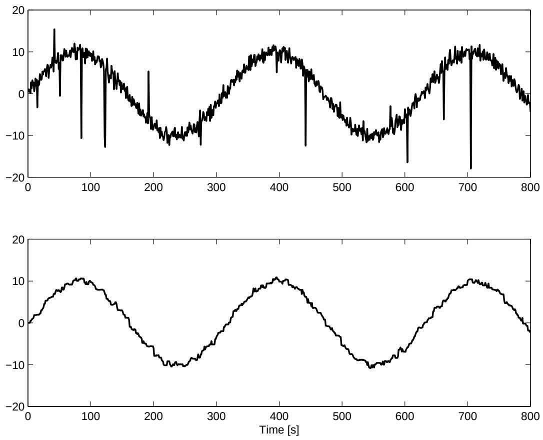

Figure 6: A sinusoid corrupted with gaussian noise and outliers (above) and its filtering with a median filter.

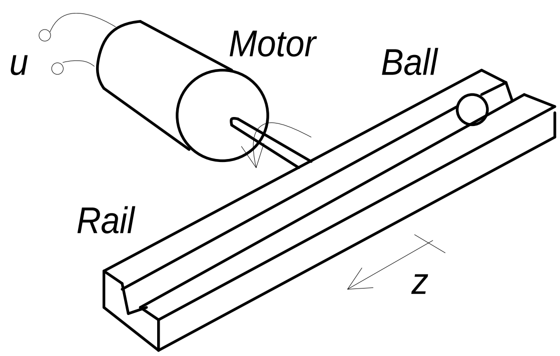

Figure 7: The ball and beam system.

rotor tension is the manipulated variable. The position y of the ball, as measured along the beam, is the plant output. This position is sensed by measuring the resistance of two resistive wires which are short-cut by the metallic ball. Outliers arise because of contact irregularities when the ball rotates along the wires.

In a first approximation the movement of the ball may be modelled by a double integrator in series with a sine function nonlinearity [2]. This basic model may be improved by taking into account the motor dynamics and changes in the moment of inertia due to the ball movement.

4 Adaptive Control.

Although the outlier removal methods described can be coupled with a wide variety of control algorithms, it is convenient to use a controller which is insensitive to plant delay. Furthermore, considering the difficulties associated with the ball and beam plant, the controller chosen should be able to tackle issues such as time varying and unmodelled plant dynamics and low stability margins.

Bearing these difficulties in mind, the MUSMAR control algorithm [10] was selected. This algorithm has been used with advantage in a number of industrial processes such as superheated steam temperature control in an industrial boiler [14], oil temperature control in a distributed collector solar field [4] and rate of cooling control in arc welding [13]. A key issue is the fact that MUSMAR relies on multiple identifiers, the redundancy thereby introduced being the key for its robustness properties [10].

4.1 The MUSMAR algorithm.

MUSMAR relies on the minimization of the multi-step quadratic cost function:

$$J_T = E\left[\frac{1}{T}\sum_{i=0}^{T-1}(y(t+i+1)-r)^2 + \rho.u^2(t+i) \mid O^t\right] \tag{18}$$

where $E[\cdot|O^t]$ denotes the mean conditioned on the $\sigma-$algebra O^t induced by the observations made up to time t, T is an integer hereafter referred as the "prediction horizon" and $\rho \geq 0$ is a penalty in the manipulated variable effort. The variables u and y denote the plant manipulated variable and output and r denotes the reference to track.

For the sake of minimizing (18), the plant is described by the set of predictive models:

$$\begin{gathered}\hat{y}(t+i \mid t) = \theta_i u(t) + \Psi_i' s(t)\\ \hat{u}(t+i-1 \mid t) = \mu_{i-1} u(t) + \Phi_{i-1}' s(t)\\ i = 1,\, ...,\, T\end{gathered} \tag{19}$$

where $\hat{y}(t+i \mid t)$ and $\hat{u}(t+i-1 \mid t)$ are predictors in least squares sense given O^t, of, respectively $y(t+i)$ and $u(t+i-1)$, and $s(t)$ is the so called pseudo-state defined by:

$$s(t) = [y(t) \; ... \; y(t-n_a+1) \; u(t-1) \; ... \; u(t-n_b) \; r(t)]' \qquad (20)$$

where n_a and n_b are integers to be selected. The vector $s(t)$ is called "pseudo-state" [10] because, although it is not a state, it is a sufficient statistic for computing the control, even in the presence of correlated noise. The entries of $s(t)$ define the structure of the controller.

The coefficients θ_i, μ_{i-1} and vectors Ψ_i, Φ_{i-1} in equation (19) are parameters to be online estimated from measured plant input/output data by using Recursive Least Squares (RLS). For that sake, the following equations are used:

$$\hat{\Theta}(k) = \hat{\Theta}(k-1) + K(k)[y(k) - \hat{\Theta}(k-1)'\varphi(k-1)]$$

$$K(k) = \frac{P(k-1)\varphi(k-1)}{1 + \varphi'(k-1)P(k-1)\varphi(k-1)[1-\alpha(k)]}$$

$$P(k) = [I - K(k)\varphi'(k-1)(1-\alpha(k))]P(k-1)$$

Here, $\hat{\Theta}$ represents the vector of parameter estimates given, for each predictor j by

$$\hat{\Theta} = [\theta_j \; \psi_j']'$$

and $\varphi(k)$ represents the regressor, common to all predictors, given by

$$\varphi(k) = [u(k) \; s'(k)]'$$

The amount of information to forget in the direction of the current regressor, $\varphi(k)$, is given by

$$\alpha(k) = 1 - \lambda + \frac{1-\lambda}{\varphi'(k-1)P(k-1)\varphi(k-1)}$$

where λ is the directional forgetting factor, a number between 0 and 1.

For each i, $i = 1, \ldots, T$, the above set of parameters describes the dynamic behavior of the process output and manipulated variable from t to $t+i$, assuming a constant feedback of the pseudo-state over the prediction horizon. Minimization of (18) assuming (19) yields the control law:

$$u(t) = F's(t) + \eta(t) \qquad (21)$$

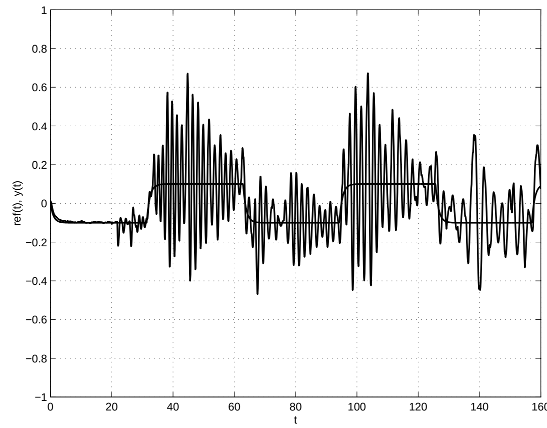

Figure 8: Adaptive control of the ball and beam plant in the presence of outliers (ball position).

with the vector of optimal gains given by:

$$F = -\frac{\sum_{i=1}^{T} \theta_i \Psi_i + \rho \sum_{i=1}^{T-1} \mu_i \Phi_i}{\sum_{i=1}^{T} \theta_i^2 + \rho \left(1 + \sum_{i=1}^{T-1} \mu_i^2\right)} \qquad (22)$$

and $\{\eta\}$ a low power (with respect to the power of $\{e\}$) dither noise injected in order to fulfil a persistency of excitation condition.

4.2 Simulation results.

Figs. 8 and 9 show the result obtained when controlling the ball and beam plant with MUSMAR in the presence of sensor outliers. Fig. 8 shows the plant output while tracking a filtered square wave reference and fig. 9 plots the controller gains.

In order to prevent start-up adaptation transients, the controller gains have been initialized at values which lead to good performance. However, as seen in fig. 9, outliers force the RLS identifiers to be mislead and the gains become severely detuned. As a consequence a major oscillation arises in the output (8).

If, instead, a median filter is used to remove the outliers from the position signal fed to MUSMAR, the gains are kept at the correct values (fig. 11) and the output follows the reference 10).

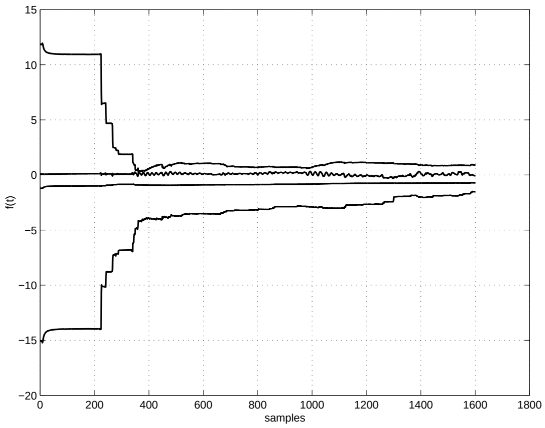

Figure 9: Adaptive control of the ball and beam plant in the presence of outliers (controller gains).

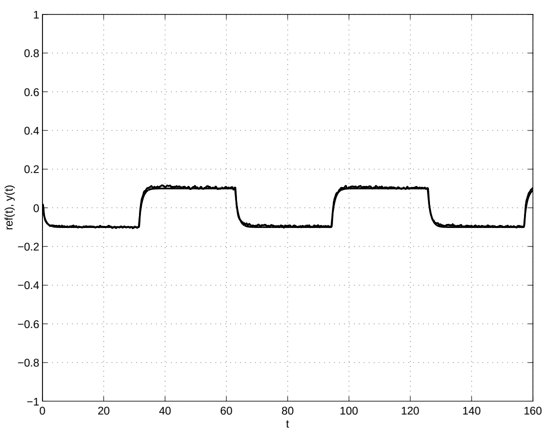

Figure 10: Adaptive control of the ball and beam plant with outliers removed by a median filter (ball position).

5 Conclusion.

The paper addressed the problem of controlling plants whose sensor measurements are corrupted by outliers. It has been shown that outliers may seriously disturb adaptive controllers but that this may be overcome by including an outlier removing filter. It should be remarked that the work reported in the paper suggests another approach for tackling outliers, *viz.* to redesign the adaptive controller with a ”robustifying” modification on the cost. This is achieved by replacing the quadratic function in the cost by an approximation of the function defined in fig. 3 made of piecewise quadratic functions. This is

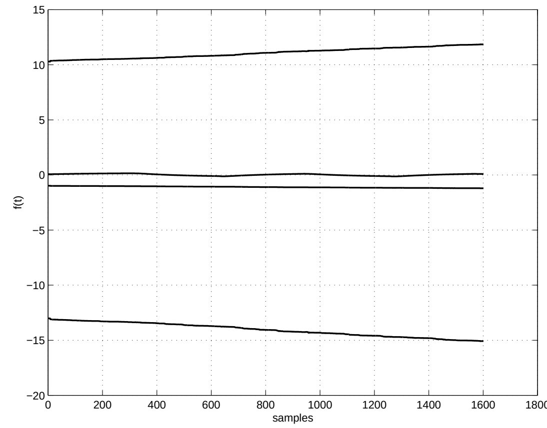

Figure 11: Adaptive control of the ball and beam plant with outliers removed by a median filter (controller gains).

considered in a different paper.

References

[1] Abraham, B. and A. Chuang (1989). Outlier detection and time series modeling. *Technometrics*, 31(2):241-248.

[2] Åström, K. and B. Wittenmark (1984). *Computer Controlled Systems*. Prentice-Hall.

[3] Chang, I., G.Tiao and C. Chen (1988). EStimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193-204.

[4] Coito, F., J. M. Lemos, R. N. Silva, E. Mosca (1997). “Adaptive control of a solar energy plant: Exploiting accessible disturbances”, *Int. J. Adapt. Control Signal Process.*, **11**, 327-342.

[5] Girosi, F. (1991). *Models of Noise and Robust Estimates*. MIT, Artificial Intelligence Laboratory Memo No. 1287.

[6] Huber, P. J. (1981). *Robust Statistics*, Wiley.

[7] Kalata, P. R. (1984). The tracking index: A generalized parameter for $\alpha - \beta$ and $\alpha - \beta - \gamma$ target trackers. *IEEE Trans. Aerospace and Electronic Syst.*, AES-20(2):174-182.

[8] Ljung, G. M. (1993). On outlier detection in time series. *J. Royal Stat. Soc.*, series B, 55:559-567.

[9] Marques, J. S. and J. M. Lemos (2001). Optimal and Suboptimal Shape Tracking Based on Multiple Switched Dynamic Models. *Image and Vision Computing*, 19(8):539-550.

[10] E. Mosca. G. Zappa. and J.M. Lemos (1989). ”Robustness of multipredictor adaptive regulators: MUSMAR” *Automatica*, 25:521 - 529.

[11] Meer, P., D. Mintz and A. Rosenfeld (1991). Robust Regression Methods for Computer Vision: A review. *Int. J. of Computer Vision*, 6(1):59-70.

[12] Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*, Wiley Interscience.

[13] T. O. Santos, R. B. Caetano, J. M. Lemos and F. J. Coito (2000). ”Multipredictive Adaptive Control of Arc Welding Trailing Centerline Temperature”. *IEEE Trans. Control Syst. Technology*, 8(1):159-169.

[14] Silva, R. N., P. O. Shirley, J. M. Lemos and A. C. Gonçalves (2000). Adaptive regulation of super-heated steam temperature: a case study in an industrial boiler. *Control Eng. Practice*, 8:1405-1415.

[15] Smith, A. and M. West (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics* 39:867-878.

[16] Taplin, R. and A. Raftery (1994). Analysis of agricultural field trials in the presence of outliers and fertility jumps. *Biometrics* 50:764-781.

[17] Tiku, M. L., W. Y. Tan and N. Balakrishnan (1986). *Robust Inference*, Marcel Dekker.

[18] Van Trees, H. L. (1967). *Detection, Estimation, and Modulation Theory, Part I*. Wiley.