Likelihood Bounds for Constrained Estimation with Uncertainty

Sikandar Samar
Stanford University
sikandar@stanford.edu

Dimitry Gorinevsky
Honeywell Labs
dimitry.gorinevsky@honeywell.com

Stephen Boyd
Stanford University
boyd@stanford.edu

***Abstract*— This paper addresses the problem of finding bounds on the optimal maximum a posteriori (or maximum likelihood) estimate in a linear model under the presence of model uncertainty. We introduce the novel concepts of *at least as likely as the maximum a posteriori* (ALAMAP) estimate, or *at least as likely as the maximum likelihood* (ALAML) estimate. The concept is formulated in terms of a convex optimization problem. We specifically make use of second order conic programming techniques to compute the likelihood bounds in an efficient manner. The procedure of computing the bounds is illustrated by examples in state estimation (smoothing/filtering) and in system identification.**

I. Introduction

Many estimation problems in decision and control are ill-conditioned. These include state estimation or trending, identification of the system model parameters, and others. Reliable estimation in the presence of noise, uncertainty and ill-conditioning can be achieved by using a prior knowledge of the unknown state. The *maximum a posteriori* (MAP) estimate is based on this concept of using an a priori known probabilistic distributions of the unknown state. Another method of reliable computation of the estimates is through robust estimation, which makes explicit use of the model uncertainty, see [1], [2], [3], [4]. Tikhonov's regularization is another well known technique to overcome ill-conditioning in the problem data, see [5], [6]. The MAP and robust estimation techniques essentially provide a systematic way of choosing the regularization for ill-conditioned estimation problems.

While providing a regularized solution (no bad behavior of the inverse along small singular values) it is not clear how reliable is the estimate. The solutions may be very sensitive to noise and perturbations in the data matrices. What are the confidence bounds for such a *maximum likelihood* (ML) solution? The answer is practically important to understand how far the real parameters can be from the ML estimates. We provide a novel approach of computing bounds for the ML solution in the presence of data perturbations. The main contribution of this work is to introduce the concept of at least as likely as the maximum likelihood (ALAML) solution. The proposed approach is applicable to systems where constrained linear state estimation is required in the presence of bounded data uncertainty. A large class of estimation problems result in a quadratic programming formulation, and can be efficiently solved using numerical optimization to compute the estimates.

The solution approach to the ALAML relies on constrained convex optimization based estimation. It provides upper and lower likelihood bounds for the ML estimate under uncertainty instead of computing the robust optimal solution. Constrained estimation using numerical optimization has been studied extensively, see [7], [8], [9]. However most previous work in determining bounds for the ML solution is done in the unconstrained least squares framework, where ellipsoidal sets of all possible states consistent with the given measurements are found. The work presented in this paper is similar in spirit but the implementation in the constrained framework requires additional tools that are provided by second order conic programming (SOCP) techniques.

The paper is organized as follows; Section II explains the problem statement of finding likelihood bounds under uncertainty. The problem is formulated mathematically in Section III, where we also propose a solution involving second order conic constraints. The concept of at least as likely as the maximum a posteriori estimate (ALAMAP) is explained using a simple univariate example in Section IV. Section V shows the application of the concept to monotonic trending using constrained state estimation. The proposed concept is applied to finite impulse response (FIR) model identification in Section VI. Some concluding remarks are given in Section VII.

II. Technical problem statement

In this work, we deal with linear state estimation problems in the presence of sensor noise and data uncertainty. We consider both the constrained and the unconstrained formulations. The objective is to find meaningful confidence bounds for the unknown state given some observed parameters. The fundamental assumption in this paper is the linear dependence of the observed parameters on the unknown state. This data model can be conveniently expressed in the form

$$y = Ax + e, \tag{1}$$

where $y \in \mathbf{R}^m$ is the vector of observations, $x \in \mathbf{R}^n$ is the vector of unknown state vector, and $A \in \mathbf{R}^{m\times n}$ is the known data matrix that defines the linear mapping between the unknown state and the observed vector. $e \in R^m$ is the vector of noise sequence. The noise term is added in the above data model to account for any modeling errors and sensor noise.

The known data matrix A is in general allowed to be time varying. The linear data model in (1) is useful for a wide class of linear state estimation problems. It appears in many practical applications related to data trending and system identification. One case of interest is the trending of fault parameters, where the unknown state x represents a fault condition in a physical system. In the fault estimation setting, the observed quantity y is referred to as the vector of residuals, and the linear map A is the matrix of fault signatures. It describes a linear dependence of the unknown faults on the observed vector of residuals. For implementation of the linear data model in practical trending applications, we refer the reader to [10].

To obtain a statistically optimal estimate of the unknown parameter vector x, such as the ML estimate or the MAP estimate, we make use of the concept of conditional probability. The MAP estimate of the unknown vector x given y is

$$\text{MAP} := \arg\max_{x} p_{x|y}; \tag{2}$$

where $p_{x|y}$ denotes the conditional probability density of x given the observed vector y. For $p_{x|y} \neq 0$, we define the loss index J as the negative log-likelihood of the conditional probability density, i.e.,

$$J(x) \;:=\; -\log p_{x|y} \tag{3}$$
$$\;=\; -\log p_{y|x} - \log p_x + c, \tag{4}$$

where (4) follows from (3) by a direct application of the Bayes' rule. The constant $c = \log p_y$ has no role in determining the MAP estimate, which is obtained by minimizing the loss index J. The first term in the loss index, $-\log p_{y|x}$, depends on the noise sequence e in the data model (1). We assume for now that the noise is uncorrelated gaussian with zero mean and covariance Q, i.e.,

$$e \sim N(0, Q). \tag{5}$$

For this gaussian distribution of the sensor noise e, we can substitute for the conditional probability density $-\log p_{y|x}$. The MAP estimate is then obtained by minimizing the loss index

$$J(x) = [Ax - y]^T Q^{-1} [Ax - y] - \log p_x \tag{6}$$

Some other noise distributions that result in a convex loss index can be handed accordingly. However, for the scope of this work we limit our focus to the gaussian noise sequence.

It is evident from the loss index (6) that the MAP estimate depends on the assumption we make about the probabilistic distribution of the unknown state vector x. The MAP estimate reduces to the ML estimate if the initial condition covariance is assumed very large. This is the case when no a priori information is available about initial condition distribution. We deal with a more broad problem formulation where any a priori knowledge about the unknown state can be incorporated in the form of constraints. In general, the minimization of the loss index (6) may be subject to constraints on the unknown state, which we represent as

$$x \in \mathcal{C}, \tag{7}$$

where we require $\mathcal{C}$ to be a convex set. This results in a constrained quadratic programming (QP) problem. For the examples in this paper, we are only interested in linear state constraints for which $\mathcal{C}$ is a polyhedral set. The penalty terms in the loss index are determined by $-\log p_x$, e.g., an assumed gaussian distribution of the unknown state results in a quadratic penalty term which is explained in Section V.

The formulation in (6) assumes that the linear map is known precisely over the entire interval of interest. Assume that the resulting minimization yields the optimal estimate x^* and the corresponding minimum value of the loss index J^*. In most practical applications, there is always some uncertainty in the assumed linear map A. This uncertainty can be due to a variety of reasons. The most common cause of uncertainty is that all physical systems are inherently nonlinear. The linearized model (1) is used only as an approximation of the actual system. It is therefore natural to deal with the imprecise nature of the linearization by considering uncertainty in the linear map A. Our formulation also accounts for any uncertainty in the observed vector y due to possible sensor limitations.

Assuming that the matrix A and the vector y are not known precisely, we introduce uncertainty parameters ΔA and Δy for the data matrix and the observed vector respectively. The uncertainty is considered to be norm bounded. We assume ΔA to be of the same dimension as A. The individual columns of the matrix ΔA account for the uncertainty in the corresponding columns of the matrix A. For the problem with model uncertainty

$$y \;\longrightarrow\; y + \Delta y, \tag{8}$$
$$A \;\longrightarrow\; A + \Delta A. \tag{9}$$

We consider the two uncertainty parameters to be norm bounded, i.e.,

$$|\Delta y| \;\leq\; r_y, \tag{10}$$
$$|\Delta A| \;\leq\; r_A, \tag{11}$$

where $r_A := max_{j=1,\ldots,n} \sum_{i=1}^{m} |\Delta a_{ij}|$. The choice of norm for the uncertainty bounds on ΔA and Δy depends on the specific nature of the problem. Typically we are interested in the ℓ_∞ norm bound but in some applications the ℓ_1 or ℓ_2 norm may be more suitable. The question of how far the solution of the problem with uncertainty can be off the nominal solution x^* is addressed in this paper. We consider a novel *at least as likely as the MAP* (ALAMAP) or *at least as likely as the ML* (ALAML) setting. The choice depends on whether we are interested in the MAP estimate, or we have no a prior knowledge about the distribution p_x and only require the ML estimate of the unknown state. In the problem with model uncertainty, the loss index (6) becomes a function of the parameters ΔA and Δy. We consider a set W such that

$$W = \{(x, \Delta A, \Delta y) : J(x; \Delta A, \Delta y) \leq J^*\}, \tag{12}$$

where J^* is obtained from the optimal solution of the nominal problem (6). We call it the *ALAMAP* or the *ALAML set*. The set contains all the possible uncertainty parameter values that yield an estimate of the unknown state which is at least as likely as the optimal estimate x^* for the nominal model (6). For practical purposes we are interested in the likelihood bounds or extreme points (worst case solution) for a given uncertainty. We follow the approach of the next section to compute these bounds.

III. Solution Approach

We now mathematically formulate the problem of finding the likelihood bounds for the unknown state given some uncertainty in the problem data. Introducing the uncertainty parameters ΔA and Δy in the loss index (6), we get

$$\begin{aligned} J(x, \Delta A, \Delta y) &= [(A + \Delta A)x - y + \Delta y]^T Q^{-1} \\ &\quad [(A + \Delta A)x - y + \Delta y] - \log p_x, \end{aligned} \tag{13}$$

Define a new uncertainty variable $z \in \mathbf{R}^m$ as

$$z := \Delta A x - \Delta y. \tag{14}$$

Substituting z in (6) yields

$$\begin{aligned} J(x; z) &= [Ax + z - y]^T Q^{-1} \\ &\quad [Ax + z - y] - \log p_x, \end{aligned} \tag{15}$$

To obtain the likelihood bounds in the presence of uncertainty, the ALAMAP estimate problem can be formulated as

$$\min_x c^T x, \tag{16}$$

subject to the following constraints

$$J(x; z) \leq J^* \tag{17}$$

$$|z| \leq r_A|x| + r_y. \tag{18}$$

To obtain the upper bound we simply solve $\min_x -c^T x$. The row vector $c^T \in \mathbf{R}^{1\times n}$ is used to pick point wise the components of x, and can be thought of as a unit vector along the corresponding x. The minimization is also subject to any original problem constraints on the unknown state, as given in (7). To ensure the convexity of the problem, we replace the decision variable x in constraint (18) by the optimal x^* obtained from the nominal problem, i.e.,

$$|z| \leq r_A|x^*| + r_y \tag{19}$$

The above simplification will yield an approximate solution of the ALAMAP problem. The approximation can be improved by replacing x^* with the solution of the minimization problem (16) in successive iterations.

We now show that the ALAMAP estimation is a convex optimization problem. The objective (16) is linear. The constraints on the state in (7) are assumed convex. The uncertainty bound constraint in (19) is linear. If we can cast (17) as a convex constraint then the likelihood bounds can be easily obtained by solving a linear objective subject to the convex constraints. We now show that (17) can be formulated as a second order conic constraint. Second order conic programming (SOCP) problems are well known in optimization theory. For a detailed description of the SOCP formulation, see [11]. Rewrite the first term in the loss index (15) for the problem with uncertainty in terms of a new matrix $P \in \mathbf{R}^{m\times(n+m+1)}$ and a vector $v \in \mathbf{R}^{n+m+1}$, where

$$P := \begin{bmatrix} A & I & -y \end{bmatrix}, \tag{20}$$

$$v := \begin{bmatrix} x \\ z \\ 1 \end{bmatrix}, \tag{21}$$

where $I \in \mathbf{R}^{m\times m}$ is the identity matrix. For notational simplicity, assume unit covariance of the noise term e, i.e., $Q = 1$. The constraint (17) can now be conveniently written as

$$J(x; z) = \|Pv\|^2 - \log p_x \leq J^* \tag{22}$$

The probabilistic distribution of the unknown state determines the second term in the above expression. In most cases an assumed distribution of x results in either a linear or a quadratic penalty term in the loss index. We define the function $f : \mathbf{R}^n \to \mathbf{R}$ as

$$f(x) := -\log p_x \tag{23}$$

where $f(x)$ is either a quadratic or a linear function of the unknown state x. In its most general form, an SOCP constraint is expressed as

$$\|Pv + b\| \leq a^T v + d, \tag{24}$$

where $b \in \mathbf{R}^m$, $a \in \mathbf{R}^n$, and $d \in \mathbf{R}$ can be chosen according to the problem at hand. It is straight forward to see that the constraints (19) and (22) can be easily incorporated as one SOCP constraint of the form (24). Most state constraints (7) are also captured by this SOCP formulation. The problem

The main contribution of this work is to introduce the novel concept of ALAMAP or ALAML estimate and to cast the problem of finding the likelihood bounds as a minimization of a linear objective subject to convex (second order conic) constraints. The convex optimization framework allows for efficient computations of the bounds in a reasonable amount of time. We now illustrate the concept by some examples in the following sections.

IV. Description of ALAMAP Estimate

We now explain the concept of finding the estimate that is at least as likely as the MAP estimate using a one dimensional example. In this univariate case, y, A, x, and

e are all scalars. The assumed values for these scalars can be thought of as the values of the quantities at a particular instant during an interval of interest. The values chosen for the simulation are; observed parameter $y = 10$, $A = 1$, unit noise covariance for the noise e, i.e., $Q = 1$, The unknown state x is assumed gaussian with zero mean and covariance r. This implies that the term $-\log p_x$ in the loss index is just a quadratic penalty rx^2, where r is covariance of x. We choose $r = 10$ for the MAP estimate in this simulation. The nominal loss index for this case without uncertainty is plotted using (6) and is shown in Fig. 1.

We now introduce uncertainty in this problem setup. We assume the uncertainty ΔA to be bounded by $r_A = 0.1$, i.e., at most than 10% uncertainty in the given A. For this example $\Delta y = 0$ and z is simply ΔAx. Limiting our

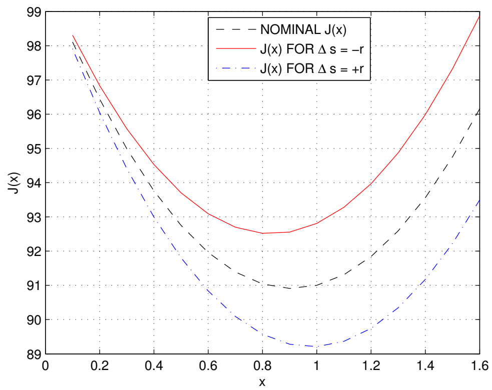

Fig. 1. Nominal loss index J(x) with positive and negative uncertainty

attention to the worst case uncertainty, we first substitute for $\Delta A = r_A = 0.1$, and then $\Delta A = -r_A = -0.1$ in the uncertainty loss index (15). The two corresponding parabolas are shown in Fig. 1 along with the loss index for the nominal case.

The optimal value for the nominal loss index naturally occurs at the vortex of the parabola. In this example x^* is 0.9 and the corresponding optimal value of loss index is $J^*(x^*) = 90.9$. The introduction of negative uncertainty, $\Delta A = -0.1$, shifts the parabola upwards. This results in a higher value of optimal loss index, i.e., $J^*(x, z = -0.1x) > J^*$. In this case the constraint in (17) is not satisfied and our set W in (12) of uncertainty parameters that yield an estimate at least as likely as the nominal MAP estimate is empty. On the other hand, when we consider the positive worst case uncertainty $\Delta A = 0.1$, the parabola $J(x, z)$ is shifted below the nominal parabola $J(x)$. In this case, we have a range of solutions that are at least as likely as the nominal MAP solution x^*, and satisfy the constraint in (17). It is therefore meaningful to find the likelihood bounds for this point estimate by performing the minimization in (16).

For the case of positive worst case uncertainty $z = r_A|x|$, when our feasible set W is not empty, it is useful to get an idea about the convexity of the constraint by plotting the function $J(x, z)$. The loss index $J(x, z)$ in (15) is plotted in the (x, z) plane. The result is shown in Fig. 2. The important

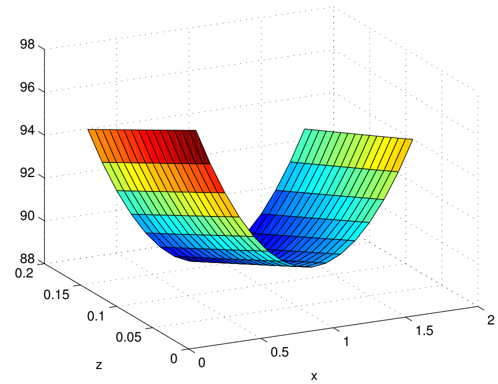

Fig. 2. Quadratic J(x,z) surface in the (x,z) plane

thing to note about this region is that it is convex (quadratic). This makes the constraint minimization (16) of finding the point wise confidence bounds for the estimate a convex optimization problem which is computationally feasible.

The problem of finding the estimate that is at least as likely as the nominal MAP or ML estimate can be easily explained by Fig 3. We find the level sets of $J(x, z)$ that satisfy $J(x, z) \leq J^* = 90.9$. A contour that satisfies the equality is shown in Fig 3. The uncertainty bound $z = r_A x$ is superimposed on the contour $J(x, z) = J^*$ in Fig. 3. The uncertainty bound is satisfied by all the points below the constraint line $z = rx$. As can be seen, there is a range of solutions (x, z) that satisfy the constraint (17) and the set W is not empty. There is a range of possible values for which the estimate is at least as likely as the nominal MAP or ML estimate and the extreme points that give the likelihood bounds can be computed using the minimization in (16).

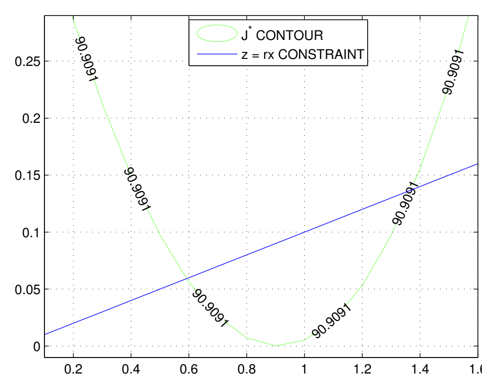

Fig. 3. Feasible quadratic domain in (x,z) plane

V. Application to monotonic trending

We now apply the concept of ALAML estimate to monotonic trending. Monotonic trends are a priori known to increase (or decrease) with time. This prior knowledge appears in the form of constraints on the unknown state. Monotonic trends are particularly common in a fault estimation setting. The unknown trend may denote a gradually accumulating fault state such as mechanical damage during the course of a flight. For details about monotonic trending reference to fault estimation, see [10], [12]. This example illustrates the application of the concept of likelihood bounds to the case of constrained state estimation, i.e., where the nominal MAP or ML estimate cannot be obtained using least squares formulation.

We consider a univariate formulation. Assume that the time interval of interest is $t = \{1, \dots, N\}$ for which we have the known data sequence

$$y = \{y(1), \dots, y(N)\}. \tag{25}$$

We take A to be an identity matrix of size N. The noise distribution e is assumed to be gaussian with covariance $Q = 1$. The sequence y (observed raw data) is generated by adding random noise to an underlying monotonic trend as shown in Fig. 4. There are $N = 25$ time samples for this simulation.

The monotonic trending problem is to estimate the unknown state x, where

$$x = \{x(1), \dots, x(N)\}, \tag{26}$$

given the observed sequence y and the data matrix A. Each diagonal entry of A represents the linear relationship between the unknown state $x(t)$ and the observed parameter $y(t)$ at a particular instant. For a linear time invariant state estimation problem, all the entries of the diagonal of A are the same. In this example we consider a one sided gaussian distribution for the unknown state x. If the covariance of the state is r, then the penalty term in the loss index (6) reduces to

$$-\log p_x = \sum_{t=2}^{N} r[x(t) - x(t-1)]^2, \tag{27}$$

We choose $r = 1$ in this example. The initial state covariance is assumed large, i.e., no prior knowledge is available about the state $x(0)$. As a result we are interested in computing the ML estimate. The loss index (6) is minimized subject to the monotonic state constraints

$$x(t+1) \geq x(t). \tag{28}$$

This is a linearly constrained quadratic programming (QP) problem which yields the ML estimate for the nominal uncertain model. The nominal ML estimate is shown in Fig. 4. We introduce uncertainty parameters ΔA and Δy to compute the likelihood bounds. The bounds are chosen to reflect a 10% uncertainty in the measured y and the known map A, i.e., worst case $\Delta A = 0.1A$ and the worst case $\Delta y = 0.1y$. The likelihood bounds obtained by solving the minimization in (16) subject to the constraints (17), (19) and the state constraints (28) are shown in Fig. 4. The likelihood bounds can be made tight or loose depending upon

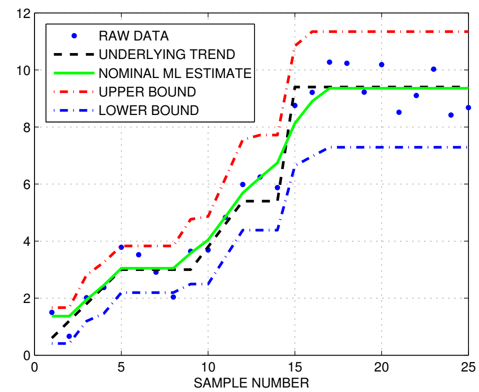

Fig. 4. Likelihood bounds for constrained state estimation

the magnitude of the allowed uncertainty. Fig. 5 shows a comparison of the likelihood bounds for uncertainty of 10% and 20%. The bounds for 10% uncertainty are much tighter as expected. The uncertainty bounds r_A and r_y can thus be used as tuning parameters to obtain different likelihood bounds.

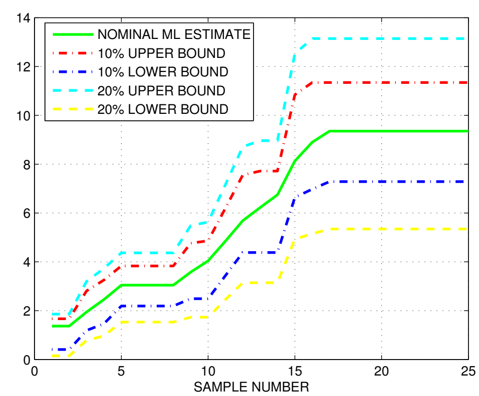

Fig. 5. Likelihood bounds for two uncertainty levels

VI. Application to System Identification

We now apply the concept of ALAML to estimate a moving average (MA) or finite impulse response (FIR) model. We measure input $u(t)$ and output $y(t)$ for $t = 0, \dots, N$ of unknown system. The system identification problem deals

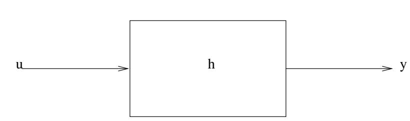

Fig. 6. System identification schematic

with finding a reasonable model for a system based on measured input output data u, y. We illustrate the ALAML concept by an example of scalar input u and output y. The vector case is handled readily.

Consider a moving average model with n delays

$$y(t) = h_0 u(t) + h_1 u(t-1) + \cdots + h_n u(t-n) \qquad (29)$$

where $h_0, \ldots, h_n \in \mathbf{R}$. We can write the model or predicted output as

$$\begin{bmatrix} y(n) \\ y(n+1) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} u(n) & \cdots & u(0) \\ u(n+1) & \cdots & u(1) \\ \vdots & \vdots & \vdots \\ u(N) & \cdots & u(N-n) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_n \end{bmatrix} + e \qquad (30)$$

The above model is in the standard linear state estimation form (1). These models arise in a variety of different applications. The objective is to find the likelihood bounds for the FIR kernel h. The input and output pair used for simulation is shown in Fig. 7, where $N = 12$. Assuming a unit covariance

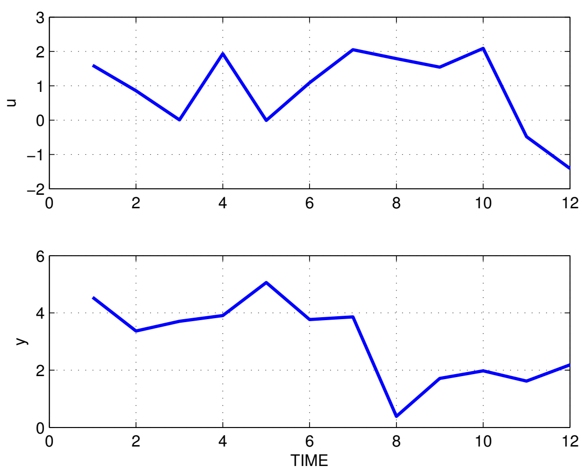

Fig. 7. Input and output for MA model

for the nose e, we are interested in estimating a model with $n = 10$ delays. The solution obtained through an over determined least squares (LS) problem is used to predict the nominal model output shown in Fig. 8. The likelihood bounds are obtained using a 10% uncertainty bound on the uncertainty in the vector y and the matrix of inputs u. The upper and lower bounds are shown in Fig. 8 along with the outputs from the actual MA model. In practice, the model order selection is an important consideration and may effect the likelihood bounds as well. However here we knew the actual model had n delays and so the issue of model order selection was not explored.

VII. Conclusion

In this paper we introduce the concept of ALAMAP (or ALAML) estimate in the presence of data uncertainty for constrained linear state estimation problems. The presented concepts are illustrated by application to FIR model identification and monotonic trending. The approach is based on convex optimization techniques and casts the problem in terms of minimization of a linear objective subject to second order cone constraints. The computed likelihood bounds can

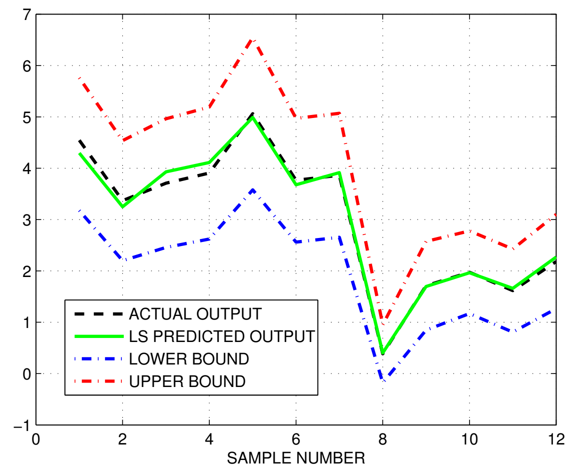

Fig. 8. Likelihood bounds for MA model predicted outputs

be tuned by varying the uncertainty bounds specified for a particular problem.

References

[1] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed. Parameter estimation in the presence of bounded data uncertainties. *SIAM Journal on Matrix Analysis and Applications*, 19(01):235–252, 1998.

[2] H. Hindi and S. Boyd. Robust solutions to ℓ_1, ℓ_2 and ℓ_∞ uncertain linear approximation problems using convex optimization. In *Proc. American Control Conf.*, volume 6, pages 3487–3491, 1998.

[3] L. El Ghaoui and G. Calafiore. Robust filtering for discrete-time systems with bounded noise and parametric uncertainty. *IEEE Transactions on Automatic Control*, 46(07):1084–1089, July 2001.

[4] G. Calafiore and L. El Ghaoui. Robust maximum likelihood estimation in the linear model. *Automatica*, 37(04):573–580, March 2001.

[5] A. N. Tikhonov and V. Y. Arsenin. *Solution of ill-posed problems*. John Wiley and Sons, 1977.

[6] L. Ljung. *System Identification - Theory For the User*. Prentice Hall, 1999.

[7] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, February 2003.

[8] G. C. Goodwin, M. M. Seron, and J. A. De Doná. *Constrained Control and Estimation–An Optimisation Approach*. Springer-Verlag, 2004.

[9] G. Ferrari-Trecate, D. Mignone, and M. Morari. Moving horizon estimation for hybrid systems. *IEEE Transactions on Automatic Control*, 47(10):1663–1676, 2002.

[10] D. Gorinevsky, S. Samar, J. Bain, and G. Aaseng. Integrated diagnostics of rocket flight control. In *Proceedings of the IEEE Aerospace Conference*, Big Sky, MT, March 2005.

[11] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Applications*, 284(1-3):193–228, November 1998.

[12] D. Gorinevsky. Monotonic regression filters for trending gradual deterioration faults. In *Proceedings of the American Control Conefrence*, Boston, MA, June 2004.