Robust Maximum Likelihood Estimation in the Linear Model ⋆

Giuseppe Calafiore [1]

Dipartimento di Automatica e Informatica – Politecnico di Torino, Cso Duca degli Abruzzi 24, 10129 Torino, Italy.

Laurent El Ghaoui

Electrical Engineering and Computer Sciences Department – University of California at Berkeley.

Abstract

This paper addresses the problem of maximum likelihood parameter estimation in linear models affected by gaussian noise, whose mean and covariance matrix are uncertain. The proposed estimate maximizes a lower bound on the worst-case (with respect to the uncertainty) likelihood of the measured sample, and is computed solving a semidefinite optimization problem (SDP). The problem of linear robust estimation is also studied in the paper, and the the statistical and optimality properties of the resulting linear estimator are discussed.

Key words: Robust estimation, Distributional robustness, Least squares, Convex optimization, Linear Matrix Inequalities.

1 Introduction

The problem of estimating parameters from noisy observed data has a long history in engineering and experimental science in general. When the observations and the unknown parameters are related by a linear model, and a stochastic setting is assumed, then the application of the Maximum Likelihood (ML) principle (see for instance the monograph [3]) leads to the well-known Least

⋆ This work was supported in part by Italy CNR funds.

[1] Author to whom correspondence should be sent. Tel.: +39-011-564.7066; Fax: +39-011-564.7066; E-mail: calafiore@polito.it

Squares (LS) parameter estimate. However, the well-established ML principle assumes that the true parametric model for the data is exactly known, a seldom verified assumption in practice, where models only approximate reality; [18], Ch. 5. This paper introduces a family of estimators that are based on a robust version of the ML principle, where uncertainty in the underlying statistical model is explicitly taken into account. In particular, we will study estimators that maximize a lower bound on the worst-case (with respect to model uncertainty) value of the likelihood function. Next, we will analyze the case of linear robust estimation, and discuss the bias, variance and optimality properties of the resulting estimator.

The undertaken minimax approach to robustness is in the spirit of the *distributional robustness* approach discussed in [16] for parametrized families of distributions. In our case, the minimax is performed with respect to unknown-but-bounded parameters appearing in the underlying statistical model. The techniques introduced in this paper may also be viewed as the stochastic counterpart of the deterministic robust estimation methods that appeared recently in [7,10]. In particular, the model uncertainty will here be represented using the Linear Fractional Transformation (LFT) formalism [10], which allows to treat cases where the regression matrix has a particular form, such as Toeplitz or Vandermonde, and where the uncertainty affects the data in a structured way. Robust estimation trades accuracy, which is best achieved using standard techniques as LS or TLS (Total Least Squares, [22]), with robustness, i.e. insensitvity with respect to parameters variations. In this latter context, links between robust estimation, sensitivity, and regularization techniques, such as Tikhonov regularization [21], may be found in [4,9,10] and references therein. Here, we will treat a mixed-uncertainty problem, where the regression matrix is affected by deterministic, structured and norm-bounded uncertainty, while the measure is affected by gaussian noise whose covariance matrix is also uncertain. In this setting, we will compute a robust (with respect to the deterministic model uncertainty) estimate via semidefinite programming (SDP).

The main focus of this paper is on introducing a theoretical framework in which robust estimation problems may be solved efficiently (i.e. in polynomial-time) using available numerical tools for semidefinite programming; see e.g. [13]. Since the problem we treat is an extension of the classical stochastic least squares framework, we believe that there are countless diverse areas of possible application. Therefore, the subject is here treated in its generality, without reference to any specific area of application. An example of application of the introduced theory to the estimation of dynamic parameters of a robot manipulator from real experimental data is instead presented in Section 4.

For a square matrix X, $X \succ 0$ (resp. $X \succeq 0$) means X is symmetric, and positive-definite (resp. semidefinite). $\lambda_{max}(X)$, where $X = X^T$, denotes the maximum eigenvalue of X. $\|X\|$ denotes the operator (maximum singular value) norm of X. For $P \in \mathbf{R}^{n\times n}$, with $P \succ 0$, and $x \in \mathbf{R}^n$, the notation $\mathbf{x} \sim N(\bar{x}, P)$ means that $\mathbf{x}$ is a gaussian random vector with expected value $\bar{x}$ and covariance matrix P.

2 Problem Statement

We consider the problem of estimating a parameter from noisy observations that are related to the unknown parameter by a linear statistical model. To set up the problem, we shall take the Bayesian point of view, and assume an a-priori normal distribution on the unknown parameter $\mathbf{x} \in \mathbb{R}^n$, i.e. $\mathbf{x} \sim N(\bar{x}, P(\Delta_p))$, where $\bar{x} \in \mathbb{R}^n$ is the expected value of $\mathbf{x}$, and the a-priori covariance $P(\Delta_p) \in \mathbb{R}^{n,n}$ depends on a matrix Δ_p of uncertain parameters, as it will be discussed in detail in Section 2.1. Similarly, the observations vector $\mathbf{y} \in \mathbb{R}^m$ is assumed to be independent of $\mathbf{x}$, and with normal distribution $\mathbf{y} \sim N(\bar{y}, D(\Delta_d))$, with $\bar{y} \in \mathbb{R}^m$, and $D(\Delta_d) \in \mathbb{R}^{m,m}$. The linear statistical model assumes further that the expected values of $\mathbf{x}$ and $\mathbf{y}$ are related by a linear relation which, in our case, is also uncertain

$$\bar{y} = C(\Delta_c)\bar{x}.$$

Given some a-priori estimate x_s of $\mathbf{x}$, and given the vector of measurements y_s, we seek an estimate of $\bar{x}$ that maximizes a lower bound on the worst-case (with respect to the uncertainty) a-posteriori probability of the observed event. When no deterministic uncertainty is present on the model, this is the celebrated maximum likelihood (ML) approach to parameter estimation, which enjoys special properties such as efficiency and unbiasedness, see for instance [3,15,19]. For the important special case of linear estimation, we will discuss in Section 3.2 how these properties extend to the robust estimator, and how the resulting estimate is related to the minimum a-posteriori variance estimator.

To cast our problem in a ML setting, the log-likelihood function $\mathcal{L}$ is defined as the logarithm of the a-posteriori joint probability density of $\mathbf{x}, \mathbf{y}$

$$\mathcal{L}(\bar{x}, \Delta | x_s, y_s) = \log\left(f_x(x_s) f_y(y_s)\right),$$

where f_x, f_y are the probability density functions of $\mathbf{x}, \mathbf{y}$, respectively. Since

$\mathbf{x}, \mathbf{y}$ are independent gaussian vectors, maximizing the log-likelihood is equivalent to minimizing the following function

$$\ell(\bar{x}, \Delta) = (x_s - \bar{x})^T P^{-1}(\Delta_p)(x_s - \bar{x}) + (y_s - C(\Delta_c)\bar{x})^T D^{-1}(\Delta_d)(y_s - C(\Delta_c)\bar{x}),$$

where Δ is the total uncertainty matrix, containing the blocks $\Delta_p, \Delta_d, \Delta_c$. We notice that, for fixed Δ, computing the ML estimate reduces to solving the following standard norm minimization problem

$$\hat{x}_{ML}(\Delta) = \arg\min_{\bar{x}} \|F(\Delta)\bar{x} - g(\Delta)\|^2,$$

where

$$F(\Delta) = \begin{bmatrix} D^{-1/2}(\Delta_d)C(\Delta_c) \\ P^{-1/2}(\Delta_p) \end{bmatrix}; \quad g(\Delta) = \begin{bmatrix} D^{-1/2}(\Delta_d)y_s \\ P^{-1/2}(\Delta_p)x_s \end{bmatrix}. \tag{1}$$

If now Δ is allowed to vary in a given norm-bounded set, as precised in the next section, we define the worst-case maximum likelihood (WCML) estimate $\hat{x}_{WCML}$ as

$$\hat{x}_{WCML} = \arg\min_{\bar{x}} \max_{\Delta} \|F(\Delta)\bar{x} - g(\Delta)\|^2. \tag{2}$$

The WCML estimate provides therefore a guaranteed level of the likelihood function, for any possible value of the uncertainty. In the next section, we detail the uncertainty model used troughout the paper, and state a fundamental technical lemma.

2.1 LFT Uncertainty Models

We shall consider matrices subject to structured uncertainty in the so called linear-fractional (LFT) form

$$M(\Delta) = M + L\Delta(I - H\Delta)^{-1}R, \tag{3}$$

where M, L, H, R are constant matrices, while the uncertainty matrix Δ belongs to the set $\mathbf{\Delta}_1$, where $\mathbf{\Delta}_1 = \{\Delta \in \mathbf{\Delta} : \|\Delta\| \leq 1\}$, and $\mathbf{\Delta}$ is a linear subspace. The norm used is the spectral (maximum singular value) norm. The subspace $\mathbf{\Delta}$, referred to as the *structure subspace* in the sequel, defines the structure of the perturbation, which is otherwise only bounded in norm.

Together, the matrices M, L, H, R and the subspace $\mathbf{\Delta}$, constitute a *linear-fractional representation* of an uncertain model. We will make from now on the standard assumption that all LFT models are *well-posed* over $\mathbf{\Delta}_1$, meaning that $\det(I - H\Delta) \neq 0$, for all $\Delta \in \mathbf{\Delta}_1$, see [14]. We also introduce the following linear subspace $\mathcal{B}(\mathbf{\Delta})$, referred to as the *scaling subspace*

$$\mathcal{B}(\mathbf{\Delta}) = \left\{(S, T, G) \mid S\Delta = \Delta T, \ G\Delta = -\Delta^T G^T \text{ for every } \Delta \in \mathbf{\Delta}\right\}.$$

LFT models of uncertainty are general and now widely used in robust control [14,25] (especially in conjunction with SDP techniques, see for instance [1]), in identification [23], and filtering [12,24]. This uncertainty framework includes the case when parameters perturb each coefficient of the data matrices in a (polynomial or) rational manner, as stated in the representation lemma in [8].

2.2 Robustness Lemma

The main results in this paper rely on the following lemma, which provides a sufficent condition for a linear matrix inequality (LMI, [5]) to hold for any allowed value of the uncertainty. Given a LFT model (3) in the matrix variable $\Delta \in \mathbb{R}^{p,q}$, and given a real symmetric matrix W, we seek a sufficient LMI condition ensuring

$$\det(I - H\Delta) \neq 0 \tag{4}$$

and

$$\begin{bmatrix} M^T(\Delta) \ I \end{bmatrix} W \begin{bmatrix} M^T(\Delta) \ I \end{bmatrix}^T \succ 0 \tag{5}$$

for all $\Delta \in \mathbf{\Delta}_1$. This is expressed in the following lemma.

Lemma 1 *The conditions (4), (5) hold for all $\Delta \in \mathbf{\Delta}_1$, if there exist a triple $(S, T, G) \in \mathcal{B}(\mathbf{\Delta})$ such that*

$$S \succ 0, T \succ 0, \textit{ and} \tag{6}$$

$$\begin{bmatrix} M & L \\ I & 0 \end{bmatrix}^T W \begin{bmatrix} M & L \\ I & 0 \end{bmatrix} - \begin{bmatrix} R & H \\ 0 & I \end{bmatrix}^T \begin{bmatrix} T & G \\ G^T & -S \end{bmatrix} \begin{bmatrix} R & H \\ 0 & I \end{bmatrix} \succ 0. \tag{7}$$

The above conditions are also necessary when $\mathbf{\Delta} = \mathbb{R}^{p,q}$, i.e. in the case of unstructured *perturbation.*

The proof of the above lemma is reported in Appendix A.

Remark 2 *The above lemma provides in general a sufficient condition only. A discussion on the tightness of this condition and its approximation error is out of the scope of this paper; general results and a further discussion may be found in [2,11].*

3 Robust Estimation

In order to compute a robust estimate, we first set up the complete uncertainty model for (2) in LFT form. Let $C(\Delta_c)$ be given in the LFT form as

$$C(\Delta_c) = C + L_c\Delta_c(I - H_c\Delta_c)^{-1}R_c,$$

and let

$$\begin{aligned} D^{-1/2}(\Delta_d) &= D^{-1/2} + L_d\Delta_d(I - H_d\Delta_d)^{-1}R_d, \\ P^{-1/2}(\Delta_p) &= P^{-1/2} + L_p\Delta_p(I - H_p\Delta_p)^{-1}R_p, \end{aligned}$$

be the LFT representation of the Cholesky factors of $D(\Delta_p)$ and $P(\Delta_p)$, respectively. Then, using the common rules for LFT operation (see for instance [25]) we obtain an LFT representation of

$$\begin{bmatrix} F(\Delta) \; g(\Delta) \end{bmatrix} = \begin{bmatrix} F \; g \end{bmatrix} + L\Delta(I - H\Delta)^{-1}\begin{bmatrix} R_F \; R_g \end{bmatrix},$$

where $F(\Delta)$, $g(\Delta)$ are given in (1), Δ is a structured matrix containing the (possibly repeated) blocks $\Delta_c, \Delta_d, \Delta_p$ on the diagonal, and

$$F = \begin{bmatrix} D^{-1/2}C \\ P^{-1/2} \end{bmatrix}; \quad g = \begin{bmatrix} D^{-1/2}y_s \\ P^{-1/2}x_s \end{bmatrix}. \tag{8}$$

The WCML estimation problem (2) may then be cast in the form

$$\begin{aligned} \eta^2_{WCML} = \arg\min\nolimits_{x,\eta}\; &\eta^2 \text{ subject to} \\ &\|F(\Delta)x - g(\Delta)\|^2 < \eta^2, \;\; \forall\Delta \in \mathbf{\Delta}_1, \end{aligned}$$

which may in turn be rewritten as a robust semidefinite optimization problem (SDP, see [11]) as

$$\eta^2_{WCML} = \arg\min_{x,\eta} \eta^2 \text{ subject to} \quad \begin{bmatrix} I & F(\Delta)x - g(\Delta) \\ (F(\Delta)x - g(\Delta))^T & \eta^2 \end{bmatrix} \succ 0, \quad \forall \Delta \in \mathbf{\Delta}_1. \tag{9}$$

3.1 Robust ML Estimation

The WCML estimate $\hat{x}_{WCML}$ is defined as the value of x at the optimum of problem (9). However, the solution of the above problem is in general numerically hard to compute, [2,11]. In order to obtain a computable solution, we shall apply the robustness lemma to the robust LMI constraint (9). In this way, we obtain a convex inner approximation of the feasible set, and the minimization of η^2 subject to this new constraint will provide an upper bound on the optimal objective of (9). The so-obtained solution $\hat{x}_{RML}$ will be called a robust maximum likelihood estimate (RML). This is summarized in the following theorem.

Theorem 3 *The robust maximum likelihood estimate $\hat{x}_{RML}$ is obtained solving the SDP*

$$\eta^2_{RML} = \arg \min_{x,S,G,T,\eta} \quad \eta^2 \quad \textit{subject to} \tag{10}$$

$$(S,G,T) \in \mathcal{B}(\mathbf{\Delta}), S \succ 0, T \succ 0, \tag{11}$$

$$\left[\begin{array}{c|c} \Theta(S,G,T) & \begin{matrix} Fx - g \\ R_F x - R_g \end{matrix} \\ \hline \begin{bmatrix} (Fx-g)^T & (R_F x - R_g)^T \end{bmatrix} & \eta^2 \end{array}\right] \succ 0, \tag{12}$$

where

$$\Theta(S,G,T) = \begin{bmatrix} I - LTL^T & -L(TH^T + G) \\ -(HT + G^T)L^T & S - HTH^T - HG - G^T H^T \end{bmatrix}. \tag{13}$$

The optimal upper bound η_{RML} is exact (i.e. $\eta_{RML} = \eta_{WCML}$) when $\mathbf{\Delta} = \mathbb{R}^{p,q}$.

PROOF. The result in the theorem follows immediately from the application of Lemma 1 to the LMI constraint in (9). In particular, the result is obtained setting

$$M(\Delta) \doteq \begin{bmatrix} F^T & 0 \\ g^T & 0 \end{bmatrix} + \begin{bmatrix} R_F^T \\ R_g^T \end{bmatrix} \Delta (I - H^T \Delta)^{-1} [L^T \quad 0];$$

$$W \doteq \left[\begin{array}{c|cc} 0 & 0 & \tilde{x} \\ \hline 0 & I & 0 \\ \tilde{x}^T & 0 & \eta^2 \end{array}\right]; \quad \tilde{x} = \begin{bmatrix} x \\ -1 \end{bmatrix}.$$

□

Remark 4 *When there is no model uncertainty, we can set* $L = 0, H = 0, R_F = 0, R_g = 0$*. In this case, the results of the previous theorem reduce to the standard LS estimate, and the robust estimate is consistent with the idealized (uncertainty free) model.*

While the result of the previous theorem is useful to obtain a numerical estimate of the parameters, due to complicated non-linear dependence of the estimate on the data x_s, y_s, it is ackward to study further the statistical properties of the resulting estimator. To pursue this study, in the next section we will cosider the additional constraint that the estimator should be *linear* in the data.

3.2 Robust Linear Estimation

The goal of this section is to compute a robust estimate which is linear in the observations. This is done in order to recover some of the nice features related to linear estimators, and to allow for further analysis of the bias and variance characteristics of the estimate. To this end, let K be an unknown gain matrix, and let $x = Kz$, with $z \doteq [y_s^T \quad x_s^T]^T$, $K \doteq [K_y \quad K_x]$, and $A \doteq [F^T \quad R_F^T]^T$, $h \doteq [g^T \quad R_g^T]^T \doteq \mathcal{G}z$, where $\mathcal{G}$ is some given matrix that can be deduced from (1), (8). Then, the main result on the optimal robust linear estimate (RLE) is provided by the following theorem.

Theorem 5 *Let*

$$\nu_{RLE}^2 = \arg \min_{K,S,G,T,\nu} \quad \nu^2 \quad \text{subject to} \tag{14}$$

$$(S, G, T) \in \mathcal{B}(\mathbf{\Delta}), S \succ 0, T \succ 0, \tag{15}$$

$$\left[\begin{array}{c|c} \Theta(S,G,T) & AK-\mathcal{G} \\ \hline (AK-\mathcal{G})^T & \nu^2 I \end{array}\right] \succ 0, \quad (16)$$

and let K_{RLE} be the value of K at the optimum of (14), then

$$\hat{x}_{RLE} = K_{RLE} z \doteq K_x x_s + K_y y_s$$

is a robust linear estimate (RLE) guaranteeing that $\|F(\Delta)\hat{x} - g(\Delta)\| \leq \eta_{RLE}$ for all admissible values of the uncertainty, where

$$\eta_{RLE} = \|z\| \nu_{RLE}.$$

Thus, η_{RLE} is a minimized upper bound on η_{RML} (i.e. $\eta_{RLE} \geq \eta_{RML}$), and the optimal gain K_{opt} is independent of the observations z.

PROOF. We start from the result of Theorem 3, assume that $z \neq 0$ (the case $z = 0$ may be trivially considered aside) and introduce a new variable K such that $x = Kz$. We now rewrite problem (10) in the equivalent form

$$\eta^2_{RML} = \arg \min_{K,S,G,T,\eta} \quad \eta^2 \text{ subject to} \quad (17)$$

$$(S, G, T) \in \mathcal{B}(\mathbf{\Delta}), S \succeq 0, T \succeq 0, \quad (18)$$

$$\left[\begin{array}{c|c} \Theta(S,G,T) & (AK-\mathcal{G})z \\ \hline z^T(AK-\mathcal{G})^T & \eta^2 \end{array}\right] \succ 0. \quad (19)$$

The previous is only a restatement of (10), and the resulting estimate is not yet linear in the observations, as the optimal gain K will depend on z. However, we now show that condition (19) is satisfied whenever condition (16) is satisfied. This is because, taking Schur complements, (19) is equivalent to

$$\Theta \succ 0, \quad (20)$$

$$\eta^2 > z^T(AK-\mathcal{G})^T \Theta^{-1} (AK-\mathcal{G}) z. \quad (21)$$

Since

$$z^T(AK-\mathcal{G})^T \Theta^{-1}(AK-\mathcal{G})z \leq \|z\|^2 \lambda_{max}((AK-\mathcal{G})^T \Theta^{-1}(AK-\mathcal{G})),$$

then (21) is implied by $\eta^2 > \|z\|^2 \lambda_{max}((AK-\mathcal{G})^T\Theta^{-1}(AK-\mathcal{G}))$. Introducing the new variable $\nu = \eta/\|z\|$, this latter condition, together with (20), may be restated in the form of (16), applying again the Schur complement rule. As the initial constraint has been replaced by a more stringent one, it immediately follows that all solutions to (14) will be feasible for (10), therefore $\hat{x}$ will be a robust estimate, and η_{RLE} an upper bound on η_{RML}. Minimizing over ν (i.e. solving problem (14)) amounts to finding the best possible upper bound on η_{RML}, based on the premise that the estimate is linear in the samples. □

3.3 Bias and Variance of the Robust Linear Estimate

In this section, we examine the bias and variance characteristics of the linear robust estimator, and present a result for the computation of a robust linear unbiased estimator.

The estimation bias is defined as

$$b \doteq E\{\hat{x} - \bar{x}\},$$

where $\bar{x}$ is the (unknown) expected value of $\mathbf{x}$, and $\hat{x}$ is a linear estimate in the form

$$\hat{x} = K_x x_s + K_y y_s. \tag{22}$$

We then have that

$$b = b(\Delta_c) = E\left(K_x x_s + K_y y_s - \bar{x}\right) = B(\Delta_c)\bar{x},$$

where

$$B(\Delta_c) = B + K_y L_c \Delta_c (I - H_c \Delta_c)^{-1} R_c; \quad B \doteq K_x - I + K_y C, \tag{23}$$

therefore, the bias is a linear function of the unknown mean $\bar{x}$, with uncertain coefficients.

Notice that the robust estimate will be in general affected by bias. A condition for having robustly zero bias is of course given by $B(\Delta_c) = 0$, that is

$$\begin{aligned} B &= 0, \\ K_y L_c &= 0. \end{aligned} \tag{24}$$

The first condition requires in particular that $K_x = I - K_y C$, which means that the estimate should be in the classical "innovations" form

$$\hat{x} = x_s + K_y(y_s - C x_s).$$

The second condition requires orthogonality between the gain K_y and the matrix L_c describing the uncertainty on the regression matrix $C(\Delta_c)$. In particular, when there is no uncertainty on C (but we still allow for uncertainty in the covariance matrices $D(\Delta_d), P(\Delta_p)$), we have $L_c = 0$, and we can have unbiased estimates, provided that $B = 0$, i.e. $K_x = I - K_y C$. Further, notice that both conditions (24) impose linear constraints on the gain K, which may be easily added (i.e. the resulting problem is still an SDP) to the constraints of problem (14), in order to compute linear robust unbiased estimates. Notice also that the additional constraints on the gain will not destroy feasibility, but simply decrease the level of the achievable robust ML performance.

Our result on robust linear unbiased estimation (RLUE) is summarized in the following theorem.

Theorem 6 *Assume no uncertainty acts on the regression matrix C, i.e. $L_c = 0$. Let*

$$\hat{x}_{RLUE} = x_s + K_{RLUE}(y_s - Cx_s),$$

where K_{RLUE} is the value of K_y at the optimum of

$$\nu^2_{RLUE} = \arg \min_{K_y, S, G, T, \nu} \quad \nu^2 \quad \textit{subject to} \tag{25}$$

$$(S, G, T) \in \mathcal{B}(\mathbf{\Delta}), S \succ 0, T \succ 0, \tag{26}$$

$$\left[\begin{array}{c|c} \Theta(S,G,T) & A[K_y \;\; I - K_y C] - \mathcal{G} \\ \hline (A[K_y \;\; I - K_y C] - \mathcal{G})^T & \nu^2 I \end{array}\right] \succ 0. \tag{27}$$

Then, $\hat{x}_{RLUE}$ is a robust linear unbiased estimate (RLUE) guaranteeing that $\|F(\Delta)\hat{x} - g(\Delta)\| \leq \eta_{RLUE}$ for all admissible values of the uncertainty, where

$$\eta_{RLUE} = \|z\| \nu_{RLUE}.$$

η_{RLUE} is the best possible upper bound on η_{RLE}, based on the premise that the robust estimate must be linear and unbiased, therefore $\eta_{RLUE} \geq \eta_{RLE} \geq \eta_{RML}$.

PROOF. The proof follows from the previous discussion.

3.3.1 Covariance of the Robust Linear Estimate

In this section, we discuss the a-posteriori covariance properties of the estimator in the parameter space. For a generic linear estimate in the form (22), the a-posteriori covariance matrix is defined as $R \doteq E\{(\hat{x} - \bar{x})(\hat{x} - \bar{x})^T\}$. A standard manipulation then yields

$$R(\Delta) = K_x P(\Delta_p) K_x^T + K_y D(\Delta_d) K_y^T + B(\Delta_c)\bar{x}\bar{x}^T B^T(\Delta_c), \tag{28}$$

where $B(\Delta_c)$ is defined as in (23).

We notice that (28) depends on the unknown mean $\bar{x}$, therefore an empirical estimate of the covariance may be obtained substituting the estimated value $\hat{x}$ in the place of the unknown $\bar{x}$. However, the significance of the covariance matrix in case of biased estimate may be questionable. A more interesting result is obtained in the case of unbiased estimates (obtained by means of Theorem 6, when C is exactly known), where $R(\Delta)$ reduces to

$$R(\Delta) = K_x P(\Delta_p) K_x^T + K_y D(\Delta_d) K_y^T. \tag{29}$$

Remark 7 *Notice that, using standard rules for operations with LFT's, one can determine an LFT representation for $R(\Delta)$, and use it in a recursive estimation framework, collecting a new observation y_s, and setting $x_s \leftarrow \hat{x}_{RLUE}$, $P(\Delta_c) \leftarrow R(\Delta)$, etc. Further study is however needed to analyze the behaviour of the recursive RLUE.*

It is important at this point to make some observations. It is well-known that, when the linear model is perfectly known, the application of the ML principle provides estimates which are unbiased and efficient, in the sense that the a-posteriori covariance in parameter space reaches the Cramér-Rao lower bound, see for instance [18], Sec. 6.4. In our context of robust estimation, we saw that unbiased estimates can be obtained (Theorem 6) only if no uncertainty is acting on the regression matrix C, while allowing for uncertainty in $P(\Delta_p)$ and $D(\Delta_d)$. When this is not the case, estimation bias will be unavoidable due to imperfect knowledge of the linear relation $\bar{y} = C(\Delta_c)\bar{x}$ between the mean values of $\mathbf{x}, \mathbf{y}$.

As for the a-posteriori parameter covariance, one may ask how the robust maximization of the likelihood function is related to robust minimization of the a-posteriori covariance. The answer to this question is that our robust ML linear estimator is also the one that minimizes an upper bound on the worst-case a-posteriori covariance, therefore the estimate provided by Theorem 6 is also a minimum variance unbiased estimate (MVUE), in the robust sense explained above. The reason for this resides in a deep and general duality result between the maximization of the log-likelihood function and the minimization (in matrix sense) of the a-posteriori covariance. For linear models, both problems have an equivalent formulation, provided that suitable dual bases are chosen to write the optimization problem; for a thorough discussion of this issue, the reader is referred to [17], Ch. 15. Finally, we remark that the robust estimation framework proposed in this paper has similarities with the minimax approach to ML estimates and minimax variance estimates discussed in [16], Ch. 4, where robustness issued are considered with respect to parametrized families of distributions.

4 Example

In this section, we report a result of the application of the presented methodology to the experimental estimation of dynamic parameters of a SCARA two-link IMI manipulator available at the Politecnico di Torino Robotics Lab. Details related to the experimental setup, manipulator model and data treatment are discussed in [6]. The goal is to estimate eight dynamic and friction parameters of the manipulator from noisy joint torque data. Denoting by $q^T = [q_1, q_2]$ the joint positions, and by $\tau^T = [\tau_1, \tau_2]$ the measured joint torques, the following manipulator model was developed for identification

$$\tau = C_\delta(q, \dot{q}, \ddot{q})\theta + d,$$

where $\theta \in \mathbb{R}^8$ is the vector of identifiable parameters, $d \in \mathbb{R}^2$ is a zero mean gaussian noise vector, and $C(q, \dot{q}, \ddot{q}) \in \mathbb{R}^{2,8}$ is the nominal regression matrix, which is a non-linear function of $q, \dot{q}, \ddot{q}$. From data acquired over repetitions of a given reference trajectory, we estimated the measurement covariance as $\sigma_1^2 = 5.23$ N^{2}m^2, $\sigma_2^2 = 0.075$ N^{2}m^2, therefore $D = \mathbf{diag}\,(\sigma_1^2, \sigma_2^2)$. Similarly, we determined confidence bounds for the angular positions measurements and velocity and acceleration data, obtaining $r_{q1} = 0.053$ rad, $r_{q2} = 0.056$ rad, $r_{\dot{q}1} = 0.082$ rad/s, $r_{\dot{q}2} = 0.085$ rad/s, $r_{\ddot{q}1} = 0.589$ rad/s^2, $r_{\ddot{q}2} = 1.648$ rad/s^2. Each position, velocity and acceleration data is therefore assumed to be of the form $q_1(\delta) = q_1 + r_{q1}\delta_1$, $q_2(\delta) = q_2 + r_{q2}\delta_2$, $\dot{q}_1(\delta) = \dot{q}_1 + r_{\dot{q}1}\delta_3$, etc..., with $\delta^T = [\delta_1, \delta_2, \ldots, \delta_6]$, $\|\delta\|_\infty \leq 1$. No prior information on θ is assumed, therefore $P \to \infty$.

To take into account the structured uncertainty entering the regression matrix C, we developed a linearized model for the uncertainty in the LFT form. For given data $q, \dot{q}, \ddot{q}$, the regression matrix is first expressed as

$$C_\delta = C + \sum_{i=1}^{6} \delta_i C_i,$$

and then rewritten in LFT format as $C_\delta = C + L\Delta R$, where $C_i = L_i R_i$ is a full-rank factorization of C_i, $r_i = \mathbf{Rank}\, C_i$, and

$$L = [L_1 \ \cdots \ L_6]; \quad R = \begin{bmatrix} R_1 \\ \vdots \\ R_6 \end{bmatrix}; \quad \Delta = \mathbf{diag}\,(\delta_1 I_{r_1}, \ldots, \delta_6 I_{r_6}).$$

For the actual estimation of θ, we collected data at six time instants on the reference trajectory, and then stacked the relative torque measurements and

LFT regression models. This gave an augmented regression model

$$\tilde{\tau} = (\tilde{C} + \tilde{L}\tilde{\Delta}\tilde{R})\theta + \tilde{d},$$

where $\tilde{\tau} \in \mathbb{R}^{12}$, $\tilde{C} \in \mathbb{R}^{12,8}$, $\tilde{\Delta} = \mathbf{diag}\,(\delta_1 I_{r_1}, \ldots, \delta_{36} I_{r_{36}})$. The robust estimate was then computed solving (10) by means of a MATLAB code based on the LMITOOL SDP solver; [13]

$$\hat{\theta}_{RML} = [3.9689,\ \ 0.0027,\ \ 0.0270,\ \ -0.0000,\ \ -0.9811,\ \ 12.2065,\ \ 1.3067,\ \ 0.8017].$$

The least squares estimate obtained from the same torque measurements, using $\tilde{C}$ as regression matrix is

$$\hat{\theta}_{LS} = [4.1593,\ \ 0.5295,\ \ -0.1146,\ \ -0.1889,\ \ -1.6205,\ \ 14.5631,\ \ 1.1408,\ \ 1.2199].$$

It is worth to remind that *on the estimation data set* the LS estimate yields a smaller residual than the robust estimate, indeed we have $\epsilon_{LS} \doteq \|D^{-1/2}(\tilde{\tau} - \tilde{C}\hat{\theta}_{LS})\| = 3.24$ and $\epsilon_{RML} \doteq \|D^{-1/2}(\tilde{\tau} - \tilde{C}\hat{\theta}_{RML})\| = 3.54$. However, the quality of the two parameter estimates should be compared on data sets different from the one used for the estimation (validation data sets). We therefore computed the residuals ϵ_{LS} and ϵ_{RML} using data collected over several different trajectories. The result are shown in Figure 1, where each point in the plot represents the estimation residual computed for a group of six time samples.

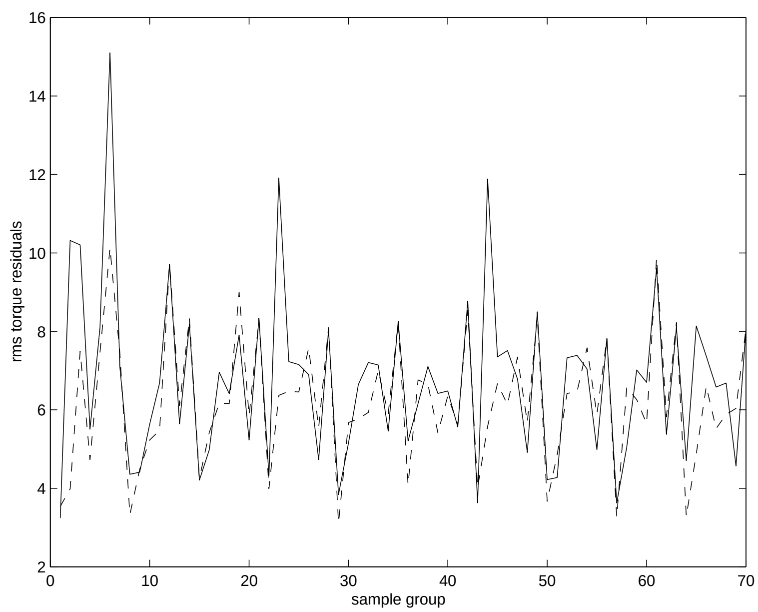

Fig. 1. Torque estimation residuals (rms values) computed for 70 groups of 6 time samples over various manipulator trajectories. LS estimate: solid line; Robust estimate: dotted line.

Considering the data of Figure 1, the average residual resulted to be 6.2 for the robust estimate and 6.76 for the LS estimate. More interestingly, if we compare the peak values of the residuals we obtain 10.1 for the robust estimate and 15.1 for the LS estimate, which is about 50% worse. Also, we notice that the robust estimate is more “regular” than the LS estimate; regularity may be measured by the variance of the residuals, which is 2.65 for the robust estimate and 4.52 for the LS estimate, which is again about 70% worst than the robust estimate.

5 Conclusions

In this paper, we have shown that the maximum likelihood estimation problem with uncertainty in the regression matrix and in the observations covariance can be solved in a worst-case setting using convex programming. This implies that in practice these problems can be solved efficently in polynomial time using available software [13].

The paper also presents specialized results for robust linear estimation and unbiased robust linear estimation. In particular, this latter estimator recovers most of the nice features of standard ML estimators, and seems to be suitable for implementation in a recursive estimation framework.

Robust esitmation has been applied to an experimental problem of manipulator parameters identification, which has inherent uncertainty in the regression matrix. The reported results show that a consistent improvement can be obtained over standard estimation methods.

A Appendix

Proof of Lemma 1

We first observe that the lower-right block of the matrix in (7) is $H^T T H + H^T G + G^T H - S$, therefore the condition (7) implies well-posedness of the LFT (3), see for instance [14] for a proof. Now, if the LFT for $M(\Delta)$ is well-posed, then condition (5) is satisfied if and only if

$$f_0(u,p) := \begin{bmatrix} u \\ p \end{bmatrix}^T \begin{bmatrix} M & L \\ I & 0 \end{bmatrix}^T W \begin{bmatrix} M & L \\ I & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} > 0$$

for all u, p such that $p = \Delta(Ru + Hp)$ for some $\Delta \in \mathbf{\Delta}_1$. Let then $q = Ru + Hp$, and $(S, T, G) \in \mathcal{B}(\mathbf{\Delta})$, with $T \succ 0$. The condition $p = \Delta q$ for some $\Delta \in \mathbf{\Delta}_1$ implies that

$$q^T G p = q^T G \Delta q = 0,$$

by skew-symmetry of $G\Delta$. In addition we have

$$\begin{aligned} q^T T q - p^T S p &= q^T (T - \Delta^T S \Delta) q \\ &= q^T T^{1/2} (I - T^{-1/2} \Delta^T \Delta T^{1/2}) T^{1/2} q \geq 0. \end{aligned}$$

In the above, we have used the fact that $S\Delta = \Delta T$, and that the matrix $T^{-1/2}\Delta^T \Delta T^{1/2}$ is actually symmetric, and has eigenvalues less or equal to one. We conclude that

$$\begin{bmatrix} u \\ p \end{bmatrix}^T \begin{bmatrix} R & H \\ 0 & I \end{bmatrix}^T \begin{bmatrix} T & G \\ G^T & -S \end{bmatrix} \begin{bmatrix} R & H \\ 0 & I \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} \geq 0$$

for every u, p such that $p = \Delta(Ru + Hp)$ for some $\Delta \in \mathbf{\Delta}_1$, and every triple $(S, T, G) \in \mathcal{B}(\mathbf{\Delta})$ with $T \succ 0$. Based on this fact, we obtain a sufficient condition for (5) to hold, i.e. that for every non-zero pair (u, p), we have

$$\begin{gathered} \begin{bmatrix} u \\ p \end{bmatrix}^T \begin{bmatrix} M & L \\ I & 0 \end{bmatrix}^T W \begin{bmatrix} M & L \\ I & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} \\ > \begin{bmatrix} u \\ p \end{bmatrix}^T \begin{bmatrix} R & H \\ 0 & I \end{bmatrix}^T \begin{bmatrix} T & G \\ G^T & -S \end{bmatrix} \begin{bmatrix} R & H \\ 0 & I \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} \end{gathered}$$

for some triple $(S, T, G) \in \mathcal{B}(\mathbf{\Delta})$, with $T \succ 0$. The above condition is exactly the one stated in the theorem.

It remains to prove that our condition is also necessary in the unstructured case, $\mathbf{\Delta} = \mathbb{R}^{p,q}$. In this case, the set $\mathcal{B}(\mathbf{\Delta})$ reduces to the set of triples (S, T, G), with $S = \tau I_p$, $T = \tau I_q$, $\tau \in \mathbb{R}$, and $G = 0$. First, we note that the well-posedness sufficient condition $H^T T H + H^T G + G^T H - S \succ 0$ for some $(S, T, G) \in \mathcal{B}(\mathbf{\Delta})$ is equivalent to $\|H\| < 1$, which is the exact well-posedness condition. Second, we note that for every u, p, we have $p = \Delta(Ru + Hp)$ for some Δ, $\|\Delta\| \leq 1$ if and only if

$$f_1(u,p) := \begin{bmatrix} u \\ p \end{bmatrix}^T \begin{bmatrix} R & H \\ 0 & I \end{bmatrix}^T \begin{bmatrix} I_q & 0 \\ 0 & -I_p \end{bmatrix} \begin{bmatrix} R & H \\ 0 & I \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} \geq 0.$$

We note that the above inequality is strictly feasible, that is there exists a pair (u_0, p_0) such that $f_1(u_0, p_0) > 0$ (since $\|H\| < 1$, it suffices to choose $u_0 = 0$ and $p_0 \neq 0$). In this case, the $\mathcal{S}$-procedure [5] provides a necessary and sufficient condition for the quadratic constraint $f_0(u, p) > 0$ to hold for every non-zero pair (u, p) such that $f_1(u, p) \geq 0$. This condition is that there exists a scalar $\tau \geq 0$ such that, for every non-zero (u, p), we have $f_0(u, p) > \tau f_1(u, p)$. This is exactly the condition of the theorem in the unstructured case, written with $S = \tau I_p$, $T = \tau I_q$, and $G = 0$. □

References

[1] T. Asai, S. Hara, and T. Iwasaki, "Simultaneous modelling and synthesis for robust control by LFT scaling." Proc. of *13th IFAC World Congress,* vol.G, pp.309-14, San Francisco, CA, USA, July 1996.

[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovskii, "Robust Semidefinite Programming." In R. Saigal, L. Vandenberghe, H. Wolkowicz, editors, *Handbook of Semidefinite Programming,* Kluwer Academic Publishers, Waterloo, Canada, 2000.

[3] J.O. Berger, R. Wolpert, *The Likelihood Principle.* Hayward, CA: Institute of Mathematical Statistics, 1988.

[4] A. Björk, "Component-wise perturbation analysis and error bounds for linear least squares solutions." *BIT,* vol.31, pp.238-244, 1991.

[5] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory.* Studies in Applied Mathematics, Philadelphia: SIAM, 1994.

[6] G. Calafiore, M. Indri, "Experiment design for robot dynamic calibration." Proc. of the *1998 IEEE International Conference on Robotics and Automation*, Leuven, May 16-21, 1998.

[7] S. Chandrasekaran, G. Golub, M. Gu, and A.H. Sayed, "Parameter estimation in the presence of bounded data uncertainties." *SIAM J. Matrix Anal. Appl.,* vol.19, no.1, pp.235-252, Jan. 1998.

[8] S. Dussy and L. El Ghaoui, "Measurement-scheduled control for the RTAC problem." *Int. J. Robust and Nonlinear Control*, vol.8, no.(4-5), pp.377-400, 1998.

[9] L. Elden, "Perturbation theory for the least-squares problem with linear equality constraints." *BIT,* vol.24, pp.472-476, 1985.

[10] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data." *SIAM J. Matrix Anal. Appl.,* vol.18, no.4, pp.1035-1064, 1997.

[11] L. El Ghaoui, F. Oustry, and H. Lebret, "Robust Solutions to Uncertain Semidefinite Programs." *SIAM J. Optimization,* vol.9, no.1, pp.33-52, 1998.

[12] L. El Ghaoui and G. Calafiore, "Deterministic state prediction under structured uncertainty." In *Proc. of the American Control Conference,* San Diego, California, 1999.

[13] L. El Ghaoui and J.-L. Commeau. `lmitool` *version* 2.0, January 1999. Available via `http://www.ensta.fr/~gropco`.

[14] M.K.H. Fan, A.L. Tits, and J.C. Doyle, "Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics." *IEEE Trans. Aut. Control,* vol.36, no.1, pp.25-38, Jan. 1991.

[15] G.C. Goodwin and R.L. Payne, *Dynamic System Identification: experiment design and data analysis.* New York: Academic Press, 1977.

[16] P.J. Huber, *Robust Statistics.* New York: Wiley, 1981.

[17] T. Kailath, A.H. Sayed, B. Hassibi, *Linear Estimation.* Englewood Cliffs, NJ: Prentice Hall, 2000.

[18] K. Knight, *Mathematical Statistics.* New York: Chapman & Hall/CRC, 2000.

[19] L. Ljung, *System identification: theory for the user.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[20] R.E. Skelton, T. Iwasaki, and K. Grigoriadis, *A Unified Algebraic Approach to Linear Control Design.* London: Taylor & Francis, 1998.

[21] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems.* New York: Wiley, 1977.

[22] S. Van Huffel and J. Vandewalle, *The total least squares problem: Computational aspects and analysis.* Philadelphia: SIAM, 1991.

[23] G. Wolodkin, S. Rangan, and K. Poolla, "An LFT approach to parameter estimation." In Proc. of the *American Control Conference,* Albuquerque, New Mexico, 1997.

[24] L. Xie, Y.C. Soh, C.E. de Souza, "Robust Kalman filtering for uncertain discrete-time systems," *IEEE Trans. on Autom. Control.* vol.39, no.6, pp.1310-1314, June 1994.

[25] K. Zhou, J.C. Doyle, and K. Glover, *Robust and Optimal Control.* Upper Saddle River: Prentice-Hall, 1996.