Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Parameter Estimation with Expected and Residual-at-Risk Criteria

Giuseppe Calafiore, Ufuk Topcu, and Laurent El Ghaoui

Abstract— **We study a class of uncertain linear estimation problems in which the data are affected by random uncertainty. In this setting, we consider two estimation criteria, one based on minimization of the expected ℓ_1 or ℓ_2 norm residual and one based on minimization of the level within which the ℓ_1 or ℓ_2 norm residual is guaranteed to lie with an a-priori fixed probability (residual at risk). The random uncertainty affecting the data is characterized by means of its first two statistical moments, and the above criteria are intended in a worst-case probabilistic sense, that is worst-case expectations and probabilities over all possible distribution having the specified moments are considered. The ensuing estimation problems can be solved efficiently via convex programming, yielding exact solutions in the ℓ_2 norm case and upper-bounds on the optimal solutions in the ℓ_1 case.**

Keywords: **Uncertain least-squares, random uncertainty, robust convex optimization, value at risk, ℓ_1 norm approximation.**

I. Introduction

To introduce the problem treated in this paper, let us consider a standard parameter estimation problem where an unknown parameter $\theta \in \mathbb{R}^n$ is to be determined so to minimize a norm residual of the form $\|A\theta - b\|_p$, where $A \in \mathbb{R}^{m,n}$ is a given regression matrix, $b \in \mathbb{R}^m$ is a measurement vector, and $\|\cdot\|_p$ denotes the ℓ_p norm. In this setting, the most relevant and widely studied case arise of course for $p = 2$, where the problem reduces to classical least-squares. The case of $p = 1$ also has important applications due to its resilience to outliers and to the property of producing "sparse" solutions, see for instance [6], [8]. For $p = 1$, the solution to the norm minimization problem can be efficiently computed via linear programming, [3, §6.2].

In this paper we are concerned with an extension of this basic setup that arises in realistic cases where the problem data A, b are imprecisely known. Specifically, we consider the situation where the entries of A, b depend affinely on a vector δ of random uncertain parameters, that is $A \doteq A(\delta)$ and $b \doteq b(\delta)$. Due its practical significance, the parameter estimation problem in the presence of uncertainty in the data has attracted much attention in the literature. When the uncertainty is modeled as unknown-but-bounded, a min-max approach is followed in [9], where the maximum over the uncertainty of the ℓ_2 norm of the residual is minimized. Relations between the min-max approach and regularization techniques are also discussed in [9] and in [13]. Generalizations of this approach to ℓ_1 and ℓ_∞ norms are proposed in [11].

In the case when the uncertainty is assumed to be random with given distribution, a classical stochastic optimization approach is often followed, whereby a θ is sought that minimizes the expectation of the ℓ_p norm of the residual with respect to the uncertainty. This formulation leads in general to numerically "hard" problem instances, that can be solved approximately by means of stochastic approximation methods, see, e.g., [4]. In the special case where the squared Euclidean norm is considered, instead, the expected value minimization problem actually reduces to a standard least-squares problem, which has a closed-form solution, see [4], [11].

In this paper we consider the uncertainty to be random and we develop our results in a "statistical ambiguity" setting, in which the probability distribution of the uncertainty is only known to belong to a given family of distributions. Specifically, we consider the family of all distributions on the uncertainty having a given mean and covariance, and seek results that are guaranteed irrespective of the actual distribution within this class. We address both the ℓ_2 and ℓ_1 cases, under two different estimation criteria: the first criterion aims at minimizing the worst-case expected residual, whereas the second one is directly tailored to control residual tail probabilities. That is, for given risk $\epsilon \in (0, 1)$, we minimize the residual level such that the probability of residual falling above this level is no larger than ϵ.

A journal version of this paper will be available in [5].

Notation. The identity matrix in $\mathbb{R}^{n,n}$ and the zero matrix in $\mathbb{R}^{n,n}$ are denoted as I_n and 0_n, respectively (subscripts may be omitted when dimensions can be inferred from context). $\|x\|_p$ denotes the standard ℓ_p norm of vector x; $\|X\|_F$ denotes the Frobenius norm of matrix X, that is $\|X\|_F = \sqrt{\mathrm{Tr}\ X^\top X}$, where Tr is the trace operator. The notation $\delta \sim (\hat{\delta}, D)$ means that δ is a random vector with expected value $\mathrm{E}\{\delta\} = \hat{\delta}$ and covariance matrix $\mathrm{var}\{\delta\} \doteq \mathrm{E}\left\{(\delta - \hat{\delta})(\delta - \hat{\delta})^\top\right\} = D$. The notation $X \succ 0$ (resp. $X \succeq 0$) indicates that matrix X is symmetric and positive definite (resp. semi-definite).

II. Problem setup and preliminaries

Let $A(\delta) \in \mathbb{R}^{m,n}$, $b(\delta) \in \mathbb{R}^m$ be such that

$$[A(\delta)\ \ b(\delta)] \doteq [A_0\ \ b_0] + \sum_{i=1}^{q} \delta_i [A_i\ \ b_i], \tag{1}$$

G. Calafiore acknowledges support from the MACP4Log grant (RU/02/26) of Regione Piemonte, Italy. L. El Ghaoui wishes to acknowledge support from the National Science Foundation (grant NSF 0625409), the Coleman Fung Risk Management Center at UC Berkeley, and Google.

G. Calafiore is with the Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy. e-mail: giuseppe.calafiore@polito.it

U. Topcu is with the Department of Mechanical Engineering, University of California, Berkeley, CA, 94720 USA. e-mail: utopcu@jagger.me.berkeley.edu

L. El Ghaoui is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, 94720, USA. e-mail: elghaoui@eecs.berkeley.edu.

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

where $\delta = [\delta_1 \ \cdots \ \delta_q]^\top$ is a vector random uncertainties, $[A_0 \ \ b_0]$ represents the "nominal" data, and $[A_i \ \ b_i]$ are the matrices of coefficients for the uncertain part of the data. Let $\theta \in \mathbb{R}^n$ be a parameter to be estimated, and consider the following norm residual:

$$\begin{aligned} f_p(\theta, \delta) &\doteq \|A(\delta)\theta - b(\delta)\|_p \\ &= \ \| [(A_1\theta - b_1) \ \cdots \ (A_q\theta - b_q)] \, \delta + \\ &+ \ \ (A_0\theta - b_0)\|_p \doteq \|L(\theta) z\|_p, \end{aligned} \tag{2}$$

where we defined $z \doteq [\delta^\top \ 1]^\top$, and $L(\theta) \in \mathbb{R}^{m,q+1}$ is partitioned as

$$L(\theta) \doteq [L^{(\delta)}(\theta) \ \ L^{(1)}(\theta)], \tag{3}$$

with

$$\begin{array}{l} L^{(\delta)}(\theta) \doteq [(A_1\theta - b_1) \ \cdots \ (A_q\theta - b_q)] \in \mathbb{R}^{m,q}, \\ L^{(1)}(\theta) \doteq A_0\theta - b_0 \in \mathbb{R}^m. \end{array} \tag{4}$$

In the following we assume that $\mathrm{E}\{\delta\} = 0$ and $\mathrm{var}\{\delta\} = I_q$. This can be done without loss of generality, since data can always be pre-processed so to comply with this assumption, as detailed in the following remark.

Remark 1 (Preprocessing the data): Suppose that the uncertainty δ is such that $\mathrm{E}\{\delta\} = \hat{\delta}$ and $\mathrm{var}\{\delta\} = D \succeq 0$, and let $D = QQ^\top$ be a full-rank factorization of D. Then, we may write $\delta = Q\nu + \hat{\delta}$, with $\mathrm{E}\{\nu\} = 0$, $\mathrm{var}\{\nu\} = I_q$, and redefine the problem in terms of uncertainty $\nu \sim (0, I)$, with $L^{(\delta)}(\theta) = [(A_1\theta - b_1) \ \cdots \ (A_q\theta - b_q)]Q$, $L^{(1)}(\theta) = [(A_1\theta - b_1) \ \cdots \ (A_q\theta - b_q)]\hat{\delta} + (A_0\theta - b_0)$. ◁

We next state the two estimation criteria and the ensuing problems that are tackled in this paper.

Problem 1: (Worst-case expected residual minimization) Determine $\theta \in \mathbb{R}^n$ that minimizes $\sup_{\delta \sim (0,I)} \mathrm{E}\{f_p(\theta, \delta)\}$, that is solve

$$\min_{\theta \in \mathbb{R}^n} \sup_{\delta \sim (0,I)} \mathrm{E}\{\|L(\theta)z\|_p\}, \quad z^\top \doteq [\delta^\top \ 1], \tag{5}$$

where $p \in \{1, 2\}$, $L(\theta)$ is given in (3), (4), and the supremum is taken with respect to all possible probability distributions having the specified moments (zero mean and unit covariance).

In some applications, such as in financial Value-at-Risk (V@R) [7], [12], one is interested in guaranteeing that the residual remains "small" in "most" of the cases, that is one seeks θ such that the corresponding residual is small with high probability. An expected residual criterion such as the one considered in Problem 1 is not suitable for this purpose, since it concentrates on the average case, neglecting the tails of the residual distribution. The second criterion that we consider is hence focused on controlling the risk of having residuals above some level $\gamma \geq 0$, where risk is expressed as the probability $\mathrm{Prob}\{\delta : \ f(\theta, \delta) \geq \gamma\}$. Formally, we state the following second problem.

Problem 2: (Guaranteed residual-at-risk minimization) Fix a risk level $\epsilon \in (0, 1)$. Determine $\theta \in \mathbb{R}^n$ such that a residual level γ is minimized while guaranteeing that $\mathrm{Prob}\{\delta : \ f_p(\theta, \delta) \geq \gamma\} \leq \epsilon$. That is, solve

$$\begin{array}{l} \min_{\theta \in \mathbb{R}^n, \gamma \geq 0} \ \gamma \text{ subject to:} \\ \sup_{\delta \sim (0,I)} \ \mathrm{Prob}\{\delta : \|L(\theta)z\|_p \geq \gamma\} \leq \epsilon, \end{array}$$

where $z^\top \doteq [\delta^\top \ 1]$, $p \in \{1, 2\}$, $L(\theta)$ is given in (3), (4), and the supremum is taken with respect to all possible probability distributions having the specified moments (zero mean and unit covariance).

A key preliminary result opening the way for the solution of Problem 1 and Problem 2 is stated in the next lemma. This lemma is a powerful consequence of convex duality, and provides a general result for computing the supremum of expectations and probabilities over all distributions possessing a given mean and covariance matrix, see Section 16.4 in [2].

Lemma 1: Let $S \subseteq \mathbb{R}^n$ be a measurable set (not necessarily convex), and $\phi : \mathbb{R}^n \to \mathbb{R}$ a measurable function. Let $z^\top = [x^\top \ 1]$, and define

$$\begin{aligned} E_{\rm wc} &\doteq \sup_{x \sim (\hat{x}, \Gamma)} \ \mathrm{E}\{\phi(x)\} \\ P_{\rm wc} &\doteq \sup_{x \sim (\hat{x}, \Gamma)} \ \mathrm{Prob}\{x \in S\} \\ Q &\doteq \left[\begin{array}{cc} \Gamma + \hat{x}\hat{x}^\top & \hat{x} \\ \hat{x}^\top & 1 \end{array} \right]. \end{aligned}$$

Then,

$$\begin{array}{rl} E_{\rm wc} = & \inf_{M = M^\top} \ \mathrm{Tr}\, QM \text{ subject to:} \\ & z^\top M z \geq \phi(x), \quad \forall x \in \mathbb{R}^n \end{array}$$

and

$$\begin{array}{rl} P_{\rm wc} = & \inf_{M \succeq 0} \ \mathrm{Tr}\, QM \text{ subject to:} \\ & z^\top M z \geq 1, \quad \forall x \in S. \end{array}$$

◁

A proof of Lemma 1 can be found in [5].

Remark 2: Lemma 1 provides a result for computing worst-case expectations and probabilities. However in many cases of interest we shall need to impose constraints on these quantities in order to eventually optimize them with respect to some other design variables. It is however a simple matter to verify that the following equivalences hold:

$$\begin{array}{c} \sup_{x \sim (\hat{x}, \Gamma)} \ \mathrm{E}\{\phi(x)\} \leq \gamma \\ \Updownarrow \\ \exists M = M^\top : \ \mathrm{Tr}\, QM \leq \gamma, \\ z^\top M z \geq \phi(x), \ \forall x \in \mathbb{R}^n, \end{array}$$

and

$$\begin{array}{c} \sup_{x \sim (\hat{x}, \Gamma)} \ \mathrm{Prob}\{x \in S\} \leq \epsilon \\ \Updownarrow \\ \exists M = M^\top \succeq 0 : \ \mathrm{Tr}\, QM \leq \epsilon, \\ z^\top M z \geq 1, \ \forall x \in S. \end{array}$$

◁

III. Worst-case expected residual minimization

In this section we focus on Problem 1 and provide an efficiently computable exact solution for the case $p = 2$, and efficiently computable upper and lower bounds on the solution for the case $p = 1$. Define

$$\psi_p(\theta) \doteq \sup_{\delta \sim (0,I)} \mathrm{E}\{\|L(\theta)z\|_p\}, \tag{6}$$

$$\text{with } z^\top \doteq [\delta^\top \ 1], \ r \doteq [0 \cdots 0 \, 1/2]^\top \in \mathbb{R}^{q+1}, \tag{7}$$

where $L(\theta) \in \mathbb{R}^{m,q+1}$ is an affine function of parameter θ, given in (3), (4). We have the following preliminary lemma.

Lemma 2: For given $\theta \in \mathbb{R}^n$, the worst-case residual expectation $\psi_p(\theta)$ is given by

$$\begin{array}{l}\psi_p(\theta) = \inf_{M=M^\top} \text{ Tr } M \quad \text{subject to:}\\ M - ru^\top L(\theta) - L(\theta)^\top ur^\top \succeq 0,\\ \quad \forall u \in \mathbb{R}^m : \|u\|_{p^*} \leq 1,\end{array}$$

where $\|u\|_{p^*}$ is the dual ℓ_p norm. ◁

Proof. From Lemma 1 we have that

$$\begin{array}{rl}\psi_p(\theta) = & \inf_{M=M^\top} \text{ Tr } M \text{ subject to:}\\ & z^\top Mz \geq \|L(\theta)z\|_p, \forall \delta \in \mathbb{R}^q.\end{array}$$

Since

$$\|L(\theta)z\|_p = \sup_{\|u\|_{p^*}\leq 1} u^\top L(\theta)z,$$

it follows that $z^\top Mz \geq \|L(\theta)z\|_p$ holds for all δ if and only if $z^\top Mz \geq u^\top L(\theta)z$ holds $\forall \delta \in \mathbb{R}^q$ and $\forall\, u \in \mathbb{R}^m$: $\|u\|_{p^*} \leq 1$. Now, since $z^\top r = 1/2$, we write $u^\top L(\theta)z = z^\top(ru^\top L(\theta) + L(\theta)^\top ur^\top)z$, whereby the above condition is satisfied if and only if

$$M - ru^\top L(\theta) + L(\theta)^\top ur^\top \succeq 0, \quad \forall\, u : \|u\|_{p^*} \leq 1,$$

which concludes the proof. □

We are now in position to state the following key theorem.

Theorem 1: Let $\theta \in \mathbb{R}^n$ be given, and let $\psi_p(\theta)$ be defined as in (6). Then, the following holds for the worst-case expected residuals in the ℓ_1- and ℓ_2-norm cases.

1) Case $p = 1$: Define

$$\overline{\psi}_1(\theta) \doteq \sum_{i=1}^{m} \left\|L_i(\theta)^\top\right\|_2, \tag{8}$$

where $L_i(\theta)^\top$ denotes the i-th row of $L(\theta)$. Then,

$$\frac{2}{\pi}\overline{\psi}_1(\theta) \leq \psi_1(\theta) \leq \overline{\psi}_1(\theta). \tag{9}$$

2) Case $p = 2$:

$$\psi_2(\theta) = \sqrt{\text{Tr } L(\theta)^\top L(\theta)} = \|L(\theta)\|_F. \tag{10}$$

◁

Proof. (Case $p = 1$) The dual ℓ_1 norm is the ℓ_∞ norm, hence applying Lemma 2 we have

$$\psi_1(\theta) = \inf_{M=M^\top} \text{ Tr } M \text{ subject to:} \tag{11}$$

$$M - L(\theta)^\top ur^\top - ru^\top L(\theta) \succeq 0,\ \forall u : \|u\|_\infty \leq 1. \tag{12}$$

For ease of notation, we drop the dependence on θ in the following derivation. Note that

$$L^\top ur^\top + ru^\top L = \sum_{i=1}^{m} u_i C_i,$$

where

$$C_i \doteq rL_i^\top + L_i r^\top = \begin{bmatrix} 0_q & \frac{1}{2}L_i^{(\delta)} \\ \frac{1}{2}L_i^{(\delta)\top} & L_i^{(1)} \end{bmatrix},$$

where $L_i^\top$ is partitioned according to (4) as $L_i^\top = [L_i^{(\delta)\top}\ \ L_i^{(1)}]$, with $L_i^{(\delta)\top} \in \mathbb{R}^{1,q}$, and $L_i^{(1)} \in \mathbb{R}$. The characteristic polynomial of C_i is $p_i(s) = s^{q-1}(s^2 - L_i^{(1)}s - \|L_i^{(\delta)}\|_2^2/4)$, hence C_i has $q-1$ null eigenvalues, and two non-zero eigenvalues at $\eta_{i,1} = (L_i^{(1)} + \|L_i\|_2)/2 > 0$, $\eta_{i,2} = (L_i^{(1)} - \|L_i\|_2)/2 < 0$. Since C_i is rank two, the constraint in problem (11) takes the form (19) considered in Theorem 4 in the Appendix. Consider thus the following relaxation of problem (11):

$$\begin{array}{r}\varphi \doteq \inf_{M=M^\top, X_i=X_i^\top} \text{Tr}\, M \text{ subject to:}\\ -X_i + C_i \preceq 0,\ \ -X_i - C_i \preceq 0,\ i = 1,\ldots,m,\\ \sum_{i=1}^{m} X_i - M \preceq 0,\end{array}$$

where we clearly have $\psi_1 \leq \varphi$. The dual of problem (13) can be written as

$$\begin{array}{rcl}\varphi^D = & \sup_{\Lambda_i,\Gamma_i} \sum_{i=1}^{m} \text{Tr}\,((\Lambda_i - \Gamma_i)C_i) & (13)\\ \text{subject to:} & \Lambda_i + \Gamma_i = I_{q+1}, & \\ & \Gamma_i \succeq 0,\ \Lambda_i \succeq 0,\ i = 1,\ldots,m. & \end{array}$$

Since the problem in (13) is convex and Slater conditions are satisfied, $\varphi = \varphi^D$. Next we show that φ^D equals $\overline{\psi}_1$ given in (8). To this end, observe that (13) is decoupled in the Γ_i, Λ_i variables and, for each i, the subproblem amounts to determining $\sup_{0\preceq\Gamma_i\preceq I} \text{ Tr}\,(I - 2\Gamma_i)C_i$. By diagonalizing C_i as $C_i = V_i\Theta_i V_i^\top$, with $\Theta_i = \text{diag}(0,\ldots,0,\eta_{i,1},\eta_{i,2})$, each subproblem is reformulated as $\sup_{0\preceq\tilde{\Gamma}_i\preceq I} \text{ Tr}\, C_i - 2\text{Tr}\,\Theta_i\tilde{\Gamma}_i$, where it immediately follows that the optimal solution is $\tilde{\Gamma}_i = \text{diag}(0,\ldots,0,0,1)$, hence the supremum is $(\eta_{i,1} + \eta_{i,2}) - 2\eta_{i,2} = \eta_{i,1} - \eta_{i,2} = |\eta_{i,1}| + |\eta_{i,2}| = \|\text{eig}(C_i)\|_1$, where $\text{eig}(\cdot)$ denotes the vector of the eigenvalues of its argument. Now, we have $\|\text{eig}(C_i)\|_1 = \|L_i^\top\|_2$, then $\varphi^D = \sum_{i=1}^{m} \|L_i^\top\|_2$, and by the first conclusion in Theorem 4 in the Appendix, we have $\overline{\psi}_1 = \varphi = \varphi^D$ and $\psi_1 \leq \overline{\psi}_1$.

For the lower bound on ψ_1 in (9), assume that the problem in (13) is not feasible. Then, for $M \succeq 0$, we have that

$$\begin{array}{l}\left\{M :\ \text{Tr}\, M = \varphi^D\right\} \bigcap\\ \quad \{M :\ X_i \succeq \pm C_i,\ \ \sum_{i=1}^{n} X_i \preceq M\} = \emptyset.\end{array}$$

This last emptiness statement, coupled with the fact that, for $i = 1,\ldots,n$, C_i is of rank two, implies, by the second conclusion in Theorem 4, that

$$\begin{array}{l}\left\{M :\ \text{Tr}\, M = \varphi^D\right\} \bigcap\\ \{M :\ M \succeq \sum_{i=1}^{n} u_i C_i,\ \forall u :\ |u_i| \leq \pi/2\} = \emptyset\end{array}$$

and

$$\begin{array}{l}\left\{\tilde{M} :\ \text{Tr}\, \tilde{M} = \frac{\varphi^D}{\pi/2}\right\} \bigcap\\ \left\{\tilde{M} :\ \tilde{M} \succeq \sum_{i=1}^{n} \tilde{u}_i C_i,\ \forall \tilde{u} :\ |\tilde{u}_i| \leq 1\right\} = \emptyset.\end{array}$$

Consequently, we have $\psi_1 \geq \frac{\varphi^D}{\pi/2} = \frac{\overline{\psi}_1}{\pi/2}$, which concludes the proof of the $p = 1$ case.

(Case $p=2$) The dual ℓ_2 norm is the ℓ_2 norm itself, hence applying Lemma 2 we have

$$\psi_2 = \inf_{M=M^\top} \text{Tr } M \text{ subject to:}$$
$$M - ru^\top L - L^\top ur^\top \succeq 0, \quad \forall u : \|u\|_2 \le 1.$$

Applying the LMI robustness lemma (Lemma 3.1 of [10]), we have that the previous semi-infinite problem is equivalent to the following SDP

$$\psi_2(\theta) = \inf_{M=M^\top, \tau>0} \text{Tr } M \text{ subject to:}$$
$$\begin{bmatrix} M - \tau rr^\top & L^\top \\ L & \tau I_m \end{bmatrix} \succeq 0.$$

By the Schur complement rule, the latter constraint is equivalent to $\tau > 0$ and $M \succeq \frac{1}{\tau}(L^\top L) + \tau rr^\top$. Thus, the infimum of Tr M is achieved for $M = \frac{1}{\tau}(L^\top L) + \tau rr^\top$ and, since $rr^\top = \text{diag}(0_q, 1/4)$, the infimum of Tr M over $\tau > 0$ is achieved for $\tau = 2\sqrt{\text{Tr } L^\top L}$. From this, it follows that $\psi_2 = \sqrt{\text{Tr } L^\top L}$, thus concluding the proof. □

Starting from the results in Theorem 1, it is easy to observe that we can further minimize the residuals over the parameter θ, in order to find a solution to Problem 1. Convexity of the ensuing minimization problem is a consequence of the fact that $L(\theta)$ is an affine function of θ. This is formalized in the following corollary, whose simple proof is omitted.

Corollary 1: (Worst-case expected residual minimization) Let

$$\psi_p^* \doteq \min_{\theta \in \mathbb{R}^n} \sup_{\delta \sim (0,I)} \mathrm{E}\{\|L(\theta)z\|_p\}, \quad z^\top \doteq [\delta^\top \; 1].$$

For $p = 1$, it holds that

$$\frac{2}{\pi}\overline{\psi}_1^* \le \psi_1^* \le \overline{\psi}_1^*,$$

where $\overline{\psi}_1^*$ is computed by solving the following second-order-cone (SOCP) program:

$$\overline{\psi}_1^* = \min_{\theta \in \mathbb{R}^n} \sum_{i=1}^m \|L_i(\theta)^\top\|_2.$$

For $p = 2$, it holds that

$$\psi_2^* = \min_{\theta \in \mathbb{R}^n} \|L(\theta)\|_F,$$

where a minimizer for this problem can be computed via convex quadratic programming, by minimizing Tr $L^\top(\theta)L(\theta)$. ◁

Remark 3: Notice that in the specific case of $\delta \sim (0, I)$ we have that $\psi_2^2 = \text{Tr } L^\top(\theta)L(\theta) = \sum_{i=0}^q \|A_i\theta - b_i\|_2^2$, hence the minimizer can in this case be determined by standard Least-Squares solution method. Interestingly, this solution coincides with the solution of the expected *squared* ℓ_2-norm minimization problem discussed for instance in [4], [11]. This might not be obvious, since in general $\mathrm{E}\{\|\cdot\|^2\} \neq (\mathrm{E}\{\|\cdot\|\})^2$. ◁

IV. Guaranteed residual-at-risk minimization

A. The ℓ_2-norm case

Assume first $\theta \in \mathbb{R}^n$ is fixed, and consider the problem of computing

$$\begin{aligned} P_{\mathrm{wc}_2}(\theta) &= \sup_{\delta \sim (0,I)} \text{Prob}\{\delta : \|L(\theta)z\|_2 \ge \gamma\} \\ &= \sup_{\delta \sim (0,I)} \text{Prob}\{\delta : \|L(\theta)z\|_2^2 \ge \gamma^2\}, \end{aligned}$$

where $z^\top \doteq [\delta^\top \; 1]$. By Lemma 1, this probability corresponds to the optimal value of the optimization problem

$$P_{\mathrm{wc}_2}(\theta) = \inf_{M \succeq 0} \text{Tr } M \text{ subject to:}$$
$$z^\top M z \ge 1, \quad \forall \delta : \|L(\theta)z\|_2^2 \ge \gamma^2,$$

where the constraint can be written equivalently as $z^\top (M - \text{diag}(0_q, 1)) z \ge 0$, $\forall \delta$: $z^\top \left(L(\theta)^\top L(\theta) - \text{diag}(0_q, \gamma^2)\right) z \ge 0$. Applying the lossless S-procedure, the condition above is in turn equivalent to the existence of $\tau \ge 0$ such that $(M - \text{diag}(0_q, 1)) \succeq \tau\left(L(\theta)^\top L(\theta) - \text{diag}(0_q, \gamma^2)\right)$, therefore we obtain

$$P_{\mathrm{wc}_2}(\theta) = \inf_{M \succeq 0, \tau > 0} \text{Tr } M \text{ subject to:}$$
$$M \succeq \tau L(\theta)^\top L(\theta) + \text{diag}(0_q, 1 - \tau\gamma^2),$$

where the latter expression can be further elaborated using the Schur complement formula into

$$\begin{bmatrix} M - \text{diag}(0_q, 1 - \tau\gamma^2) & \tau L(\theta)^\top \\ \tau L(\theta) & \tau I_m \end{bmatrix} \succeq 0. \tag{14}$$

We now notice, by the reasoning in Remark 2, that the condition $P_{\mathrm{wc}_2}(\theta) \le \epsilon$ with $\epsilon \in (0, 1)$ is equivalent to the conditions: $\exists \tau \ge 0$, $M \succeq 0$ such that Tr $M \le \epsilon$ and (14) holds. Dividing both conditions by $\tau > 0$ and then renaming variables so that $M/\tau \to M$, $1/\tau \to \tau$, we have that a parameter θ that minimizes the residual-at-risk level γ while satisfying the condition $P_{\mathrm{wc}_2}(\theta) \le \epsilon$ can be computed by solving a convex semidefinite optimization problem (SDP) as formalized in the next theorem.

Theorem 2 (ℓ_2 residual-at-risk estimation): A solution of Problem 2 in the ℓ_2 case can be found by solving the following SDP:

$$\inf_{\tau>0, M \succ 0, \theta \in \mathbb{R}^n, \gamma^2 > 0} \gamma^2, \quad \text{subject to:} \tag{15}$$
$$\text{Tr } M \le \tau\epsilon$$
$$\begin{bmatrix} M - \text{diag}(0_q, \tau - \gamma^2) & L^\top(\theta) \\ L(\theta) & I_m \end{bmatrix} \succeq 0.$$

◁

B. The ℓ_1-norm case

We next consider the problem of determining $\theta \in \mathbb{R}^n$ such that the residual-at-risk level γ is minimized while guaranteeing that $P_{\mathrm{wc}_1}(\theta) \le \epsilon$, where $P_{\mathrm{wc}_1}(\theta)$ is the worst-case ℓ_1-norm residual tail probability

$$P_{\mathrm{wc}_1}(\theta) = \sup_{\delta \sim (0,I)} \text{Prob}\{\delta : \|L(\theta)z\|_1 \ge \gamma\},$$

and $\epsilon \in (0,1)$ is the a-priori fixed risk level. To this end, define

$$\mathcal{D} \doteq \{D \in \mathbb{R}^{m,m} : D \text{ diagonal}, D \succ 0\}$$

and consider the following proposition (whose statement may be easily proven by taking the gradient with respect to D and setting it to zero).

Proposition 1: For any $v \in \mathbb{R}^m$, it holds that

$$\|v\|_1 = \frac{1}{2} \inf_{D\in\mathcal{D}} \sum_{i=1}^{m} \left(\frac{v_i^2}{d_i} + d_i\right) = \frac{1}{2} \inf_{D\in\mathcal{D}} \left(v^\top D^{-1} v + \mathrm{Tr}\, D\right), \quad (16)$$

where d_i is the i-th diagonal entry of D. ◁

The following key theorem holds.

Theorem 3 (ℓ_1 residual-at-risk estimation): Consider the following SDP:

$$\inf_{\tau>0, M\succeq 0, D\in\mathcal{D}, \theta\in\mathbb{R}^n, \gamma\geq 0} \gamma, \quad \text{subject to:} \quad (17)$$

$$\mathrm{Tr}\ M \leq \tau\epsilon$$

$$\begin{bmatrix} M - (\tau - 2\gamma + \mathrm{Tr}\, D)J & L^\top(\theta) \\ L(\theta) & D \end{bmatrix} \succeq 0,$$

with $J \doteq \mathrm{diag}(0_q, 1)$. The optimal value of this SDP provides an upper bound for Problem 2 in the ℓ_1 case, that is an upper bound on the minimum level γ for which there exist θ such that $P_{\mathrm{wc}_1}(\theta) \leq \epsilon$. ◁

Proof. Define

$$S \doteq \{\delta : \|L(\theta)z\|_1 \geq \gamma\}$$
$$S(D) \doteq \{\delta : z^\top L(\theta)^\top D^{-1} L(\theta) z + \mathrm{Tr}\, D \geq 2\gamma\},$$

with $D \in \mathcal{D}$. For ease of notation we drop the dependence on θ in the following derivation. Using (16) we have that, for any $D \in \mathcal{D}$,

$$2\|Lz\|_1 \leq z^\top L^\top D^{-1} L z + \mathrm{Tr}\, D,$$

hence $\delta \in S$ implies $\delta \in S(D)$, thus $S \subseteq S(D)$, for any $D \in \mathcal{D}$. This in turn implies that

$$\mathrm{Prob}\{\delta \in S\} \leq \mathrm{Prob}\{\delta \in S(D)\} \leq \sup_{\delta\sim(0,I)} \mathrm{Prob}\{\delta \in S(D)\}$$

for any probability measure and any $D \in \mathcal{D}$, and therefore

$$P_{\mathrm{wc}_1} = \sup_{\delta\sim(0,I)} \mathrm{Prob}\{\delta \in S\} \leq \inf_{D\in\mathcal{D}} \sup_{\delta\sim(0,I)} \mathrm{Prob}\{\delta \in S(D)\} \doteq \bar{P}_{\mathrm{wc1}}.$$

Note that, for fixed $D \in \mathcal{D}$, we can compute $P_{\mathrm{wc1}}(D) \doteq \sup_{\delta\sim(0,I)} \mathrm{Prob}\{\delta \in S(D)\}$ from its equivalent dual:

$$P_{\mathrm{wc1}}(D) = \inf_{M\succeq 0} \mathrm{Tr}\, M : z^\top M z \geq 1, \forall \delta \in S(D)$$
$$= \inf_{M\succeq 0} \mathrm{Tr}\, M : z^\top M z \geq 1, \quad \forall\delta : z^\top L^\top D^{-1} L z + \mathrm{Tr}\, D \geq 2\gamma$$

[applying the lossless S-procedure]

$$= \inf_{M\succeq 0, \tau>0} \mathrm{Tr}\, M : \quad M \succeq \tau L^\top D^{-1} L + (1 - 2\tau\gamma + \tau \mathrm{Tr}\, D)J,$$

where $J = \mathrm{diag}(0_q, 1)$. Hence, $\bar{P}_{\mathrm{wc1}}$ is obtained by minimizing $P_{\mathrm{wc1}}(D)$ over $D \in \mathcal{D}$, which results in

$$\bar{P}_{\mathrm{wc1}} = \inf_{M\succeq 0, \tau>0, D\in\mathcal{D}} \mathrm{Tr}\, M : \quad M \succeq \tau L^\top D^{-1} L + (1 - 2\tau\gamma + \tau \mathrm{Tr}\, D)J$$

[by change of variable $\tau D \to D$]

$$= \inf_{M\succeq 0, \tau>0, D\in\mathcal{D}} \mathrm{Tr}\, M : \quad M \succeq \tau^2 L^\top D^{-1} L + (1 - 2\tau\gamma + \mathrm{Tr}\, D)J$$
$$= \inf_{M\succeq 0, \tau>0, D\in\mathcal{D}} \mathrm{Tr}\, M : \quad \begin{bmatrix} M - (1 - 2\tau\gamma + \mathrm{Tr}\, D)J & \tau L^\top \\ \tau L & D \end{bmatrix} \succeq 0.$$

Now, from the reasoning in Remark 2, we have that (re-introducing the dependence on θ in the notation) $\bar{P}_{\mathrm{wc}_1}(\theta) \leq \epsilon$ if and only if there exist $M \succeq 0$, $\tau > 0$ and $D \in \mathcal{D}$ such that $\mathrm{Tr}\ M \leq \epsilon$ and

$$\begin{bmatrix} M - (1 - 2\tau\gamma + \mathrm{Tr}\, D)J & \tau L(\theta)^\top \\ \tau L(\theta) & D \end{bmatrix} \succeq 0.$$

Dividing both conditions by $\tau > 0$ and then renaming the variables as $M/\tau \to M$, $D/\tau \to D$, $1/\tau \to \tau$, these conditions become $\mathrm{Tr}\ M \leq \tau\epsilon$, and

$$\begin{bmatrix} M - (\tau - 2\gamma + \mathrm{Tr}\, D)J & L^\top(\theta) \\ L(\theta) & D \end{bmatrix} \succeq 0. \quad (18)$$

Notice that, since $L(\theta)$ is affine in θ , condition (18) is an LMI in $M, D, \theta, \tau, \gamma$. We can thus minimize the residual level γ subject to the condition $\bar{P}_{\mathrm{wc}_1}(\theta) \leq \epsilon$ (which implies $P_{\mathrm{wc}_1}(\theta) \leq \epsilon$), and this results in the statement of the theorem. □

V. NUMERICAL EXAMPLE

As a numerical example, we used data from a test appeared in [4]. Let

$$A(\delta) = A_0 + \sum_{i=1}^{3} \delta_i A_i, \quad b^T = \begin{bmatrix} 0 & 2 & 1 & 3 \end{bmatrix}, \quad \text{with}$$

$$A_0 = \begin{bmatrix} 3 & 1 & 4 \\ 0 & 1 & 1 \\ -2 & 5 & 3 \\ 1 & 4 & 5.2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and let δ_i be independent random perturbations with zero mean and standard deviations $\sigma_1 = 0.067$, $\sigma_2 = 0.1$, $\sigma_3 = 0.2$. The standard ℓ_2 and ℓ_1 solutions (obtained neglecting the uncertainty terms, i.e. setting $A(\delta) = A_0$) result to be

$$\theta_{\mathrm{nom2}} = \begin{bmatrix} -10 \\ -9.728 \\ 9.983 \end{bmatrix}, \quad \theta_{\mathrm{nom1}} = \begin{bmatrix} -11.8235 \\ -11.5882 \\ 11.7647 \end{bmatrix},$$

with nominal residuals of 1.7838 and 1.8235, respectively.

Expected residual minimization. Applying Theorem 1, the minimal worst-case expected ℓ_2 residual resulted to be

$\psi_2^* = 2.164$, whereas the minimal upper bound on worst-case expected ℓ_1 residual resulted to be $\bar{\psi}_1^* = 4.097$. The corresponding parameter estimates are

$$\theta_{\text{ewc2}} = \begin{bmatrix} -2.3504 \\ -2.0747 \\ 2.4800 \end{bmatrix}, \quad \theta_{\text{ewc1}} = \begin{bmatrix} -2.8337 \\ -2.5252 \\ 2.9047 \end{bmatrix}.$$

We next analyzed numerically how the worst-case expected residuals increase with the level of perturbation. To this end, we consider the previous data with standard deviations on the perturbation depending on a parameter $\rho \geq 0$: $\sigma_1 = \rho \cdot 0.067$, $\sigma_2 = \rho \cdot 0.1$, $\sigma_3 = \rho \cdot 0.2$. A plot of the worst-case expected residuals as a function of ρ is shown in Figure 1. We observe that both ℓ_1 and ℓ_2 expected residuals tend to a constant value for large ρ.

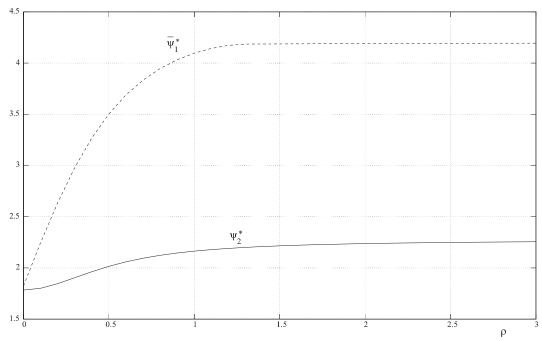

Fig. 1. Plot of ψ_2^* (solid) and $\bar{\psi}_1^*$ (dashed) as a function of perturbation level ρ.

Residual at risk minimization. Consider again the variable perturbation level problem of the previous paragraph. Now, we fix the risk level to $\epsilon = 0.1$ and solve repeatedly problems (15) and (17) for increasing values of ρ. A plot of the resulting optimal residuals at risk as a function of ρ is shown in Figure 2. These residuals grow with the covariance level ρ, as it might be expected since increasing the covariance increases the tails of the residual distribution.

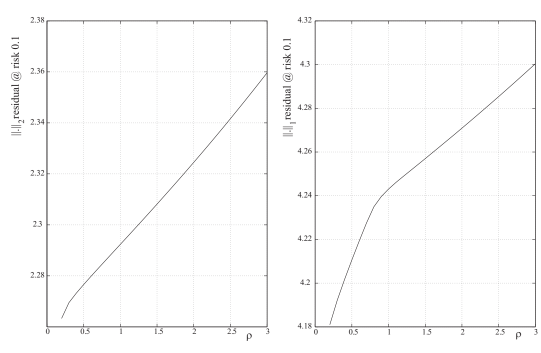

Fig. 2. Worst-case ℓ_2 and ℓ_1 residuals at risk as a function of perturbation level ρ.

VI. Conclusions

In this paper we discussed two criteria for linear parameter estimation in presence of random uncertain data, under both ℓ_2 and ℓ_1 norm residuals. The first criterion is a worst-case residual expectation and leads to exact and efficiently computable solutions for the ℓ_2 norm case. For the ℓ_1 norm, we can efficiently compute upper and lower bounds on the optimal solution, by means of convex second order cone programming. The second criterion considered in the paper is the worst-case residual for a given risk level ϵ. With this criterion, an exact solution for the ℓ_2 norm case can be computed by a solving a convex semi-definite optimization problem, and an analogous computational effort is required for computing an upper bound on the optimal solution in the ℓ_1 norm case. The estimation setup proposed in the paper is "distribution free," in the sense that only information about the mean and covariance of the random uncertainty need be available to the user: the results are guaranteed irrespective of the actual shape of uncertainty distribution.

Appendix

Theorem 4 (Matrix cube relaxation; [1]): Let B^0, B^1, ..., $B^L \in \mathbf{R}^{n \times n}$ be symmetric and $B^1, \ldots, B^L$ be of rank two. Let the problem P_ρ be defined as:

$$\mathrm{P}_\rho: \quad \text{Is } B^0 + \sum_{i=1}^{L} u_i B^i \succeq 0, \ \forall u : \|u\|_\infty \leq \rho? \qquad (19)$$

and the problem $\mathrm{P}_{\text{relax}}$ be defined as:

$\mathrm{P}_{\text{relax}}$: Do there exist symmetric matrices $X_1, \ldots, X_L \in \mathbf{R}^{n \times n}$ satisfying

$$X_i \succeq \pm\rho B^i, \ i = 1, \ldots, L,$$
$$\sum_{i=1}^{L} X_i \preceq B^0?$$

Then, the following statements hold:

1) If $\mathrm{P}_{\text{relax}}$ is feasible, then P_ρ is feasible.
2) If $\mathrm{P}_{\text{relax}}$ is not feasible, then $\mathrm{P}_{\frac{\pi}{2}\rho}$ is not feasible.

References

[1] A. Ben-Tal and A. Nemirovski. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization*, 12(3):811–833, 2002.

[2] D. Bertsimas and J. Sethuraman. Moment problems and semidefinite optimization. In *Handbook of semidefinite programming*, pages 469–509. Kluwer Acad. Publ., Boston, MA, 2000.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, 2004. ISBN: 0-521-83378-7.

[4] G. Calafiore and F. Dabbene. Near optimal solutions to least-squares problems with stochastic uncertainty. *Systems and Control Letters*, 54(12):1219–1232, December 2005.

[5] G. Calafiore, U. Topcu, and L. El Ghaoui. Parameter estimation with expected and residual-at-risk criteria. *Systems and Control Letters*, to appear.

[6] C. Dossal. A necessary and sufficient condition for exact recovery by ℓ_1 minimization. *http://hal.archives-ouvertes.fr/hal-00164738/fr/*, 2008.

[7] D. Duffie and J. Pan. An overview of value at risk. *Journal of Derivatives*, pages 7–49, Spring 1997.

[8] M.P. Friedlander and P. Tseng. Exact regularization of convex programs. *SIAM J. on Optimization*, 18(4):1326–1350, 2007.

[9] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[10] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM J. Optimization*, 9(1):33–52, 1998.

[11] H.A. Hindi and S.P. 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.

[12] R.T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking and Finance*, 26:1443–1471, 2002.

[13] A.H. Sayed, V.H. Nascimento, and F.A.M. Cipparrone. A regularized robust design criterion for uncertain data. *SIAM J. on Matrix Analysis and Applications*, 23(4):1120–1142, 2002.