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 $\ell_1$ or $\ell_2$ norm residual and one based on minimization of the level within which the $\ell_1$ or $\ell_2$ 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 $\ell_2$ norm case and upper-bounds on the optimal solutions in the $\ell_1$ case.

Keywords: Uncertain least-squares, random uncertainty, robust convex optimization, value at risk, $\ell_1$ 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 $\ell_p$ 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 $\delta$ of random uncertain parameters, that is $A \triangleq A(\delta)$ and $b \triangleq b(\delta)$. Due to 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 $\ell_2$ norm of the residual is minimized.

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 $\ell_2$ and $\ell_1$ 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 $\epsilon$.

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 $\ell_p$ norm of vector $x$; $\|X\|_F$ denotes the Frobenius norm of matrix $X$, that is $\|X\|_F = \sqrt{\text{Tr}(X^\top X)}$, where Tr is the trace operator. The notation $\delta \sim (\bar{\delta}, D)$ means that $\delta$ is a random vector with expected value $E\{\delta\} = \bar{\delta}$ and covariance matrix $\text{var}\{\delta\} = D$. 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

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

where

$$\begin{align*}
\|A(\delta) - A_0\|_2^2 & \leq \sum_{i=1}^q \delta_i \|A_i - A_0\|_2^2, \\
\|b(\delta) - b_0\|_2 & \leq \sum_{i=1}^q \delta_i \|b_i - b_0\|_2.
\end{align*}$$

For a given $\delta$, the parameter estimation problem with ambiguity on the uncertainty is to find $\hat{\theta}_\delta$ such that the $\ell_1$ norm or $\ell_2$ norm of the residual $\|A(\delta)\hat{\theta}_\delta - b(\delta)\|$ is minimized. For given $\epsilon \in (0, 1)$, we denote by $\hat{\theta}_\epsilon$ the solution to the following optimization problem:

$$\begin{align*}
\min_{\hat{\theta}} & \quad \|A(\hat{\theta}) - b\|_1, \\
\text{subject to} & \quad \Pr\{|\|A(\hat{\theta}) - b\|_1\| > \epsilon \|A_0 - b_0\|_1\} \leq \epsilon.
\end{align*}$$

This problem can be solved efficiently using convex programming techniques.

Relations between the min-max approach and regularization techniques are also discussed in [9] and in [13]. Generalizations of this approach to $\ell_1$ and $\ell_\infty$ norms are proposed in [11].
where \( \delta = [\delta_1 \cdots \delta_q]^T \) 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:

\[
f_p(\theta, \delta) \triangleq \| A(\delta)\theta - b(\delta) \|_p
\]

\[
= \| [A_1 \theta - b_1] \cdots [A_q \theta - b_q] \| + (A_0 \theta - b_0)\|_p = \| L(\theta) z \|_p,
\]

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

\[
L(\theta) = [L^{(d)}(\theta) \quad L^{(1)}(\theta)],
\]

with

\[
L^{(d)}(\theta) = [(A_1 \theta - b_1) \cdots (A_q \theta - b_q)] \in \mathbb{R}^{m,q},
\]

\[
L^{(1)}(\theta) = A_0 \theta - b_0 \in \mathbb{R}^m.
\]

In the following we assume that \( \mathbb{E}\{\delta\} = 0 \) and \( \text{var}\{\delta\} = I_q \).

This can be done without loss of generality, since data can always be pre-processed so as to comply with this assumption as detailed in the following remark.

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

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 \( \sup_{\delta \sim (0, I)} \mathbb{E}\{f_p(\theta, \delta)\} \), that is solve

\[
\min_{\theta \in \mathbb{R}^n} \sup_{\delta \sim (0, I)} \mathbb{E}\{\|L(\theta) z\|_p\}, \quad z^T = [\delta^T \ 1],
\]

where \( z^T = [\delta^T \ 1], p \in \{1, 2\}, L(\theta) \) 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 \( z^T = [x^T \ 1] \), and define

\[
E_{wc} \triangleq \sup_{x \sim (\hat{x}, I)} \mathbb{E}\{\phi(x)\}
\]

\[
P_{wc} \triangleq \sup_{x \sim (\hat{x}, I)} \text{Prob}\{x \in S\}
\]

Then,

\[
E_{wc} = \inf_{M = M^T} \text{Tr} \ QM \text{ subject to: } z^T M z \geq \phi(x), \forall x \in \mathbb{R}^n
\]

and

\[
P_{wc} = \inf_{M \succeq 0} \text{Tr} \ QM \text{ subject to: } z^T M z \geq 1, \forall x \in S.
\]

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:

\[
\sup_{x \sim (\hat{x}, I)} \mathbb{E}\{\phi(x)\} \leq \gamma \uparrow \exists M = M^T : \text{Tr} \ QM \leq \gamma, \quad z^T M z \geq \phi(x), \forall x \in \mathbb{R}^n,
\]

and

\[
\sup_{x \sim (\hat{x}, I)} \text{Prob}\{x \in S\} \leq \epsilon \downarrow \exists M = M^T \succeq 0 : \text{Tr} \ QM \leq \epsilon, \quad z^T M z \geq 1, \forall x \in S.
\]

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

\[
\psi_p(\theta) \triangleq \sup_{\delta \sim (0, I)} \mathbb{E}\{\|L(\theta) z\|_p\}, \quad \text{with } z^T = [\delta^T \ 1], r \equiv [0 \cdots 0 \ 1/2]^T \in \mathbb{R}^{q+1},
\]

where \( g = \theta^T \) is the cost function to be minimized subject to the constraints (5).
where $L(\theta) \in \mathbb{R}^{m \times n + 1}$ is an affine function of parameter $\theta$, 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

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

M - ru^\top L(\theta) - L(\theta)^\top ur^\top \succeq 0,

\forall u \in \mathbb{R}^m : \|u\|_{p^*} \leq 1,
$$

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

Proof. From Lemma 1 we have that

$$
\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}^p.
$$

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 $\delta$ if and only if $z^\top Mz \geq u^\top L(\theta)z$ holds for $\delta \in \mathbb{R}^p$ 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

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

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 $\ell_1$- and $\ell_2$-norm cases.

1) Case $p = 1$: Define

$$
\overline{\psi}_1(\theta) := \sum_{i=1}^m \|L_i(\theta)^\top\|_2, \quad (8)
$$

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

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

2) Case $p = 2$: \[ \psi_2(\theta) = \sqrt{\text{Tr } L(\theta)^\top L(\theta)} = \|L(\theta)\|_F. \] (10)

Proof. (Case $p = 1$) The dual $\ell_1$ norm is the $\ell_\infty$ norm, hence applying Lemma 2 we have

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

M - L(\theta)^\top ur^\top - ru^\top L(\theta) \succeq 0, \forall u : \|u\|_\infty \leq 1. \quad (11)
$$

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

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

where

$$
C_i \doteq ru^\top L_i + L_i r^\top \begin{bmatrix} 0 & \frac{1}{2} L_i^{(4)} \top \frac{1}{2} L_i^{(4)} \top \\ \frac{1}{2} L_i^{(4)} \top \frac{1}{2} L_i^{(4)} \top \\ \frac{1}{2} L_i^{(4)} \top \frac{1}{2} L_i^{(4)} \top \\ \end{bmatrix},
$$

where $L_i$ is partitioned according to (4) as $L_i = [L_i^{(1)} \; L_i^{(1)}]$, with $L_i^{(1)} \in \mathbb{R}^{d \times d}$, and $L_i^{(1)} \in \mathbb{R}$. The characteristic polynomial of $C_i$ is $p_i(s) = s^{r-1}(s^2 - L_i^{(1)} s - \|L_i^{(1)}\|_2^2/4)$, hence $C_i$ has $q - 1$ null eigenvalues, and two non-zero eigenvalues at $\eta_{i,1} = (L_i^{(1)} + \|L_i^{(1)}\|_2)/2 > 0$, $\eta_{i,2} = (L_i^{(1)} - \|L_i^{(1)}\|_2)/2 < 0$. 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):

$$
\varphi \doteq \inf_{M = M^\top, x_i = x_i^\top} \text{Tr } M \text{ subject to: }

-x_i + C_i \succeq 0, \quad -x_i - C_i \succeq 0, \quad i = 1, \ldots, m,

\sum_{i=1}^m x_i - M \succeq 0,
$$

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

$$
\varphi^D = \sup_{\Lambda_1, \Gamma_i} \sum_{i=1}^m \text{Tr } (\Lambda_i - \Gamma_i C_i) \quad (13)
$$

subject to:

$$\Lambda_1 + \Gamma_i = I_{q+1},
$$

$$\Gamma_i \succeq 0, \quad \Lambda_1 \succeq 0, \quad i = 1, \ldots, m.
$$

Since the problem in (13) is convex and Slater conditions are satisfied, $\varphi = \varphi^D$. Next we show that $\varphi^D$ equals $\overline{\psi}_1$ given in (8). To this end, observe that (13) is decoupled in the $\Gamma_i, \Lambda_i$ variables and, for each $i$, the subproblem amounts to determining $\sup_{0 \leq \psi_i \leq \chi_i} \text{Tr } (I - 2\Gamma_i) C_i$. By diagonalizing $C_i$ as $C_i = V_i \Theta_i \psi_i \psi_i^\top$, with $\Theta_i = \text{diag}(0, \ldots, 0, \eta_{i,1}, \eta_{i,2})$, each subproblem is reformulated as $\sup_{0 \leq \psi_i \leq \chi_i} \text{Tr } C_i - 2\Theta_i \Gamma_i$, where it immediately follows that the optimal solution is $\Gamma_i = \text{diag}(0, \ldots, 0, 0, 1)$, hence the supremum is $(\eta_{i,1} + \eta_{i,2}) - 2\eta_{i,2} = \eta_{i,1} - \eta_{i,2} = \eta_{i,1} + |\eta_{i,1} - |\eta_{i,1}||$, where $\text{eig}(\cdot)$ denotes the vector of the eigenvalues of its argument. Now, we have $\|\text{eig}(C_i)\|_1 = \|L_i\|_2$, 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 $\psi_1$ in (9), assume that the problem in (13) is not feasible. Then, for $M \succeq 0$, we have that

$$
\{ M : \text{Tr } M = \varphi^D \} \cap \{ M : X_i \succeq \pm C_i, \sum_{i=1}^m X_i \preceq M \} = \emptyset.
$$

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

$$
\{ M : \text{Tr } M = \varphi^D \} \cap \{ M : M \succeq \sum_{i=1}^m u_i C_i, \forall u : |u_i| \leq \pi/2 \} = \emptyset
$$

and

$$
\{ M : \text{Tr } M = \varphi^D \} \cap \{ M : M \succeq \sum_{i=1}^m \tilde{u}_i C_i, \forall \tilde{u} : |\tilde{u}_i| \leq 1 \} = \emptyset.
$$

Consequently, we have $\psi_1 \geq \varphi^D = \overline{\psi}_1 = \overline{\varphi} = \overline{\psi}_1$, which concludes the proof of the $p = 1$ case.
(Case \( p = 2 \)) The dual \( \ell_2 \) norm is the \( \ell_2 \) norm itself, hence applying Lemma 2 we have
\[
\psi_2 = \inf_{M = M^T} \text{Tr} \ M \text{ subject to: } M - ru^T L - L^T ur^T \geq 0, \quad \forall u : \|u\|_2 \leq 1.
\]
Applying the LMI robustness lemma (Lemma 3.1 of [10]), we have that the previous semi-infinite problem is equivalent to the following SDP
\[
\psi_2(\theta) = \inf_{M = M^T, \tau > 0} \text{Tr} \ M \text{ subject to: } \begin{bmatrix} M - \tau rr^T & L \\ L^T & \tau I_m \end{bmatrix} \succeq 0.
\]
By the Schur complement rule, the latter constraint is equivalent to \( \tau > 0 \) and \( M \succeq \frac{1}{\tau} (L^T L) + \tau rr^T \). Thus, the infimum of \( \text{Tr} \ M \) is achieved for \( M = \frac{1}{\tau} (L^T L) + \tau rr^T \) and, since \( rr^T = \text{diag}(0_q, 1/4) \), the infimum of \( \text{Tr} \ M \) over \( \tau > 0 \) is achieved for \( \tau = 2\sqrt{\text{Tr} L^T L} \). From this, it follows that \( \psi_2 = \sqrt{\text{Tr} L^T L} \), 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 \( \theta \), 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 \( \theta \). This is formalized in the following corollary, whose simple proof is omitted.

**Corollary 1:** (Worst-case expected residual minimization)
Let
\[
\psi_p^* \overset{\Delta}{=} \min_{\theta \in \mathbb{R}^n} \sup_{\delta \sim (0, I)} \mathbb{E} \{ \|L(\theta) z\|_p \}, \quad z^T \overset{\Delta}{=} [\delta^T 1].
\]
For \( p = 1 \), it holds that
\[
2^{-1} \psi_1^* \leq \psi_1^* \leq \psi_1^*,
\]
where \( \psi_1^* \) is computed by solving the following second-order-cone (SOCP) program:
\[
\psi_1^* = \min_{\theta \in \mathbb{R}^n} \sum_{i=1}^m \|L_i(\theta)^T\|_2.
\]
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 \( \text{Tr} L^T(\theta)L(\theta) \).

**Remark 3:** Notice that in the specific case of \( \delta \sim (0, I) \) we have that \( \psi_2^* = \text{Tr} L^T(\theta)L(\theta) = \sum_{i=0}^q \|A_i \theta - b_i\|_2^2 \), 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 \( \ell_2 \)-norm minimization problem discussed for instance in [4], [11]. This might not be obvious, since in general \( \mathbb{E} \{ \|\cdot\|_2^2 \} \neq (\mathbb{E} \{ \|\cdot\| \})^2 \).

### IV. Guaranteed Residual-at-Risk Minimization

#### A. The \( \ell_2 \)-norm case
Assume first \( \theta \in \mathbb{R}^n \) is fixed, and consider the problem of computing
\[
P_{\text{wec}}(\theta) = \sup_{\delta \sim (0, I)} \text{Prob} \{ \delta : \|L(\theta) z\|_2 \geq \gamma \}
\]
\[
= \sup_{\delta \sim (0, I)} \text{Prob} \{ \delta : \|L(\theta) z\|_2^2 \geq \gamma^2 \},
\]
where \( z^T \overset{\Delta}{=} [\delta^T 1] \). By Lemma 1, this probability corresponds to the optimal value of the optimization problem
\[
P_{\text{wec}}(\theta) = \inf_{M \succeq 0} \text{Tr} \ M \text{ subject to: }
\]
\[
z^T M z \geq 1, \quad \forall \delta : \|L(\theta) z\|_2^2 \geq \gamma^2,
\]
where the constraint can be written equivalently as \( z^T (M - \text{diag}(0_q, 1)) z \geq 0, \quad \forall \delta; \)
\[
z^T (L(\theta)^T L(\theta) - \text{diag}(0_q, \gamma^2)) z \geq 0.
\]
Applying the lossless S-procedure, the condition above is in turn equivalent to the existence of \( \tau \geq 0 \) such that \( (M - \text{diag}(0_q, 1)) \succeq \tau (L(\theta)^T L(\theta) - \text{diag}(0_q, \gamma^2)) \), therefore we obtain
\[
P_{\text{wec}}(\theta) = \inf_{M \succeq 0, \tau > 0} \text{Tr} \ M \text{ subject to: }
\]
\[
M \succeq \tau L(\theta)^T L(\theta) + \text{diag}(0_q, 1 - \tau \gamma^2),
\]
where the latter expression can be further elaborated using the Schur complement formula into
\[
\begin{bmatrix} M - \text{diag}(0_q, 1 - \tau \gamma^2) & \tau L(\theta)^T \\ \tau L(\theta) & \tau I_m \end{bmatrix} \succeq 0.
\]
We now notice, by the reasoning in Remark 2, that the condition \( P_{\text{wec}}(\theta) \leq \epsilon \) with \( \epsilon \in (0, 1) \) is equivalent to the conditions: \( \exists \tau \geq 0, \ M \geq 0 \) such that \( \text{Tr} M \leq \epsilon \) and (14) holds. Dividing both conditions by \( \tau > 0 \) and then renaming variables so that \( M/\tau \rightarrow M, 1/\tau \rightarrow \tau \), we have that a parameter \( \theta \) that minimizes the residual-at-risk level \( \gamma \) while satisfying the condition \( P_{\text{wec}}(\theta) \leq \epsilon \) can be computed by solving a convex semidefinite optimization problem (SDP) as formalized in the next theorem.

#### Theorem 2 (\( \ell_2 \)-residual-at-risk estimation): A solution of Problem 2 in the \( \ell_2 \) case can be found by solving the following SDP:
\[
\inf_{\tau > 0, M > 0, \theta \in \mathbb{R}^n, \gamma^2 > 0} \gamma^2, \quad \text{subject to: } \begin{bmatrix} M - \text{diag}(0_q, \tau - \gamma^2) & L(\theta)^T \\ L(\theta) & I_m \end{bmatrix} \succeq 0.
\]

#### B. The \( \ell_1 \)-norm case
We next consider the problem of determining \( \theta \in \mathbb{R}^n \) such that the residual-at-risk level \( \gamma \) is minimized while guaranteeing that \( P_{\text{wec}}(\theta) \leq \epsilon \), where \( P_{\text{wec}}(\theta) \) is the worst-case \( \ell_1 \)-norm residual tail probability
\[
P_{\text{wec}}(\theta) = \sup_{\delta \sim (0, I)} \text{Prob} \{ \delta : \|L(\theta) z\|_1 \geq \gamma \},
\]

669
and $\epsilon \in (0, 1)$ is the a-priori fixed risk level. To this end, define
\[ D = \{ D \in \mathbb{R}^{m,m} : D \text{ diagonal, } D > 0 \} \]
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
\[
\|v\|_1 = \frac{1}{2} \inf_{D \in D} \left( \sum_{i=1}^{m} \left( \frac{v_i^2}{d_i} + d_i \right) \right)
= \frac{1}{2} \inf_{D \in D} \left( v^\top D^{-1} v + \text{Tr } D \right),
\] (16)
where $d_i$ is the $i$-th diagonal entry of $D$.

The following key theorem holds.

**Theorem 3 (ℓ₁ residual-at-risk estimation):** Consider the following SDP:
\[
\begin{align*}
\inf_{\tau > 0, M \succeq 0, D \in D, \theta \in \mathbb{R}^n, \gamma \geq 0} \tau \epsilon, & \quad \text{subject to:} \\
& \text{Tr } M \leq \tau \epsilon \\
& \begin{bmatrix} M - (\tau - 2\gamma + \text{Tr } D) J & L(\theta)^\top D \cr L(\theta) & D \end{bmatrix} \succeq 0,
\end{align*}
\]
with $J \doteq \text{diag}(0, 1)$. The optimal value of this SDP provides an upper bound for Problem 2 in the ℓ₁ case, that is an upper bound on the minimum level $\gamma$ for which there exist $\theta$ such that $P_{\text{wc}1}(\theta) \leq \epsilon$.

**Proof.** Define
\[ S \doteq \{ \delta : \| L(\theta) z \|_1 \geq \gamma \} \]
\[ S(D) \doteq \{ \delta : z^\top L(\theta)^\top D^{-1} L(\theta) z + \text{Tr } D \geq 2\gamma \}, \]
with $D \in D$. For ease of notation we drop the dependence on $\theta$ in the following derivation. Using (16) we have that, for any $D \in D$,
\[ 2\| L z \|_1 \leq z^\top L^\top D^{-1} L z + \text{Tr } D, \]
hence $\delta \in S$ implies $\delta \in S(D)$, thus $S \subseteq S(D)$, for any $D \in D$. This in turn implies that
\[ \text{Prob} \{ \delta \in S \} \leq \text{Prob} \{ \delta \in S(D) \} \]
for any probability measure and any $D \in D$, and therefore
\[ P_{\text{wc}1} = \sup_{\delta \sim (0, I)} \text{Prob} \{ \delta \in S \} \leq \inf_{D \in D} \sup_{\delta \sim (0, I)} \text{Prob} \{ \delta \in S(D) \} \doteq \tilde{P}_{\text{wc}1}. \]

Note that, for fixed $D \in D$, we can compute $P_{\text{wc}1}(D) = \sup_{\delta \sim (0, I)} \text{Prob} \{ \delta \in S(D) \}$ from its equivalent dual:
\[ P_{\text{wc}1}(D) = \inf_{M \succeq 0} \text{Tr } M : \quad z^\top M z \geq 1, \quad \forall \delta \in S(D) \]
\[ = \inf_{M \succeq 0} \text{Tr } M : \quad z^\top M z \geq 1, \quad z^\top L^\top D^{-1} L z + \text{Tr } D \geq 2\gamma \]
[applying the lossless S-procedure]
\[ = \inf_{M \succeq 0, \tau > 0} \text{Tr } M : \quad M \geq \tau L^\top D^{-1} L + (1 - 2\tau \gamma + \text{Tr } D) J, \]
where $J = \text{diag}(0, 1)$. Hence, $\tilde{P}_{\text{wc}1}$ is obtained by minimizing $P_{\text{wc}1}(D)$ over $D \in D$, which results in
\[ P_{\text{wc}1} = \inf_{M \succeq 0, \tau > 0, D \in D} \text{Tr } M : \quad M \geq \tau L^\top D^{-1} L + (1 - 2\tau \gamma + \text{Tr } D) J \]
[by change of variable $\tau D \rightarrow D$]
\[ = \inf_{M \succeq 0, \tau > 0, D \in D} \text{Tr } M : \quad M \geq \tau^2 L^\top D^{-1} L + (1 - 2\tau \gamma + \text{Tr } D) J \]
\[ = \inf_{M \succeq 0, \tau > 0, D \in D} \text{Tr } M : \quad M - (1 - 2\tau \gamma + \text{Tr } D) J \geq \frac{\tau \delta}{\tau L \delta} \geq 0. \]

Now, from the reasoning in Remark 2, we have that (re-introducing the dependence on $\theta$ in the notation) $\tilde{P}_{\text{wc}1}(\theta) \leq \epsilon$ if and only if there exist $M \succeq 0$, $\tau > 0$ and $D \in D$ such that $\text{Tr } M \leq \epsilon$ and
\[ \begin{bmatrix} M - (1 - 2\tau \gamma + \text{Tr } D) J & \tau L(\theta)^\top D \cr \tau L(\theta) & D \end{bmatrix} \succeq 0. \]
Dividing both conditions by $\tau > 0$ and then renaming the variables as $M/\tau \rightarrow M$, $D/\tau \rightarrow D$, $1/\tau \rightarrow \tau$, these conditions become $\text{Tr } M \leq \epsilon$, and
\[ \begin{bmatrix} M - (\tau - 2\gamma + \text{Tr } D) J & \tau L(\theta)^\top D \cr \tau L(\theta) & D \end{bmatrix} \succeq 0. \]
Notice that, since $L(\theta)$ is affine in $\theta$, condition (18) is an LMI in $M, D, \theta, \tau, \gamma$. We can thus minimize the residual level $\delta$ subject to the condition $P_{\text{wc}1}(\theta) \leq \epsilon$ (which implies $P_{\text{wc}1}(\theta) \leq \epsilon$), 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
\[ A(\delta) = A_0 + \sum_{i=1}^{3} \delta_i A_i, \quad b^T = [0 \ 2 \ 1 \ 3], \]
with
\[ A_0 = \begin{bmatrix} 3 & 1 & 4 \\ 0 & 1 & 1 \\ -2 & 5 & 3 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 4 & 5.2 \end{bmatrix}, \]
\[ A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \]
and let $\delta_i$ be independent random perturbations with zero mean and standard deviations $\sigma_1 = 0.067$, $\sigma_2 = 0.1$, $\sigma_3 = 0.2$. The standard $\ell_2$ and $\ell_1$ solutions (obtained neglecting the uncertainty terms, i.e. setting $A(\delta) = A_0$) result to be
\[ \theta_{\text{nom}2} = \begin{bmatrix} -10 \\ -9.728 \\ 9.983 \end{bmatrix}, \quad \theta_{\text{nom}1} = \begin{bmatrix} -11.8235 \\ -11.5882 \\ 11.7647 \end{bmatrix}, \]
with nominal residuals of 1.7838 and 1.8235, respectively.

**Expected residual minimization.** Applying Theorem 1, the minimal worst-case expected $\ell_2$ residual resulted to be

670
\( \psi_2^* = 2.164 \), whereas the minimal upper bound on worst-case expected \( \ell_1 \) residual resulted to be \( \psi_1^* = 4.097 \). 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 \( \rho \) is shown in Figure 1. We observe that both \( \ell_1 \) and \( \ell_2 \) expected residuals tend to a constant value for large \( \rho \).

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

\[\text{Fig. 1. Plot of } \psi_2^* \text{ (solid) and } \psi_1^* \text{ (dashed) as a function of perturbation level } \rho.\]

\[\text{Fig. 2. Worst-case } \ell_2 \text{ and } \ell_1 \text{ residuals at risk as a function of perturbation level } \rho.\]

\[\text{VI. CONCLUSIONS}\]

In this paper we discussed two criteria for linear parameter estimation in presence of random uncertain data, under both \( \ell_2 \) and \( \ell_1 \) norm residuals. The first criterion is a worst-case residual expectation and leads to exact and efficiently computable solutions for the \( \ell_2 \) norm case. For the \( \ell_1 \) 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 \( \epsilon \). With this criterion, an exact solution for the \( \ell_2 \) norm case can be computed by 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 \( \ell_1 \) 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.

\[\text{APPENDIX}\]

\[\text{Theorem 4 (Matrix cube relaxation; [1]): Let } B^0, B^1, \ldots, B^L \in \mathbb{R}^{n \times n} \text{ be symmetric and } B^1, \ldots, B^L \text{ be of rank two. Let the problem } P_\rho \text{ be defined as:}\]
\[P_\rho : \quad \begin{array}{l}
\text{Is } B^0 + \sum_{i=1}^L u_i B^i \succeq 0, \quad \forall u : \| u \|_{\infty} \leq \rho ?
\end{array}
\]

\[\text{and the problem } P_{\text{relax}} \text{ be defined as:}\]
\[P_{\text{relax}} : \quad \text{Do there exist symmetric matrices } X_1, \ldots, X_L \in \mathbb{R}^{n \times n} \text{ satisfying}\]
\[X_i \geq \pm \rho B^i, \quad i = 1, \ldots, L, \quad \sum_{i=1}^L X_i \leq B^0?\]

\[\text{Then, the following statements hold:}\]
\[\text{1) If } P_{\text{relax}} \text{ is feasible, then } P_\rho \text{ is feasible.}\]
\[\text{2) If } P_{\text{relax}} \text{ is not feasible, then } P_{\frac{2}{\rho}} \text{ is not feasible.}\]

\[\text{REFERENCES}\]