Systems & Control Letters 58 (2009) 39–46

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Parameter estimation with expected and residual-at-risk criteria

Giuseppe Calafiore [a,*], Ufuk Topcu [b], Laurent El Ghaoui [c]

[a] Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy

[b] Department of Mechanical Engineering, University of California, Berkeley, CA, 94720, USA

[c] Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, 94720, USA

ARTICLE INFO

Article history:
Received 11 February 2008
Received in revised form
23 July 2008
Accepted 23 July 2008
Available online 16 August 2008

Keywords:
Uncertain least-squares
Random uncertainty
Robust convex optimization
Value at risk
ℓ_1 norm approximation

ABSTRACT

In this paper we study a class of uncertain linear estimation problems in which the data are affected by random uncertainty. 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.

1. Introduction

To introduce the problem treated in this paper, let us first consider a standard parameter estimation problem, where an unknown parameter $\theta \in \mathbb{R}^n$ is to be determined, so as 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 arises, 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 [5,7]. For $p = 1$, the solution to the norm minimization problem can be efficiently computed via linear programming, [3, Section 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 [8], 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 [8] and in [12]. Generalizations of this approach to ℓ_1 and ℓ_∞ norms are proposed in [10].

In the case when the uncertainty is assumed to be random with a 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 [4,10].

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.

* Corresponding author.
E-mail addresses: giuseppe.calafiore@polito.it (G. Calafiore), utopcu@jagger.me.berkeley.edu (U. Topcu), elghaoui@eecs.berkeley.edu (L. El Ghaoui).

0167-6911/$ – see front matter
doi:10.1016/j.sysconle.2008.07.007

The rest of the paper is organized as follows: Section 2 sets up the problem and gives some preliminary results. Section 3 is devoted to the estimation scheme with worst-case expected residual criterion, while Section 4 contains the results for the residual-at-risk criterion. Numerical examples are presented in Section 5 and conclusions are drawn in Section 6. Appendices A and B contain some of the technical derivations.

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 they can easily 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).

2. Problem setup and preliminaries

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

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

where $\delta = [\delta_1 \ \cdots \ \delta_q]^\top$ is a vector of 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 (which is a function of both θ and δ):

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

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

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

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 $L^{(\delta)}(\theta) = [(A_1\theta - b_1)\ \cdots\ (A_q\theta - b_q)]Q$, $L^{(1)}(\theta) = [(A_1\theta - b_1)\ \cdots\ (A_q\theta - b_q)]\hat{\delta} + (A_0\theta - b_0)$. $\triangleleft$

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) and (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) [6,11], 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

$$\min_{\theta\in\mathbb{R}^n,\gamma\geq 0} \quad \gamma$$

$$\text{subject to:} \quad \sup_{\delta\sim(0,I)} \mathrm{Prob}\{\delta : \|L(\theta)z\|_p \geq \gamma\} \leq \epsilon, \quad z^\top \doteq [\delta^\top\ 1],$$

where $p \in \{1, 2\}$, $L(\theta)$ is given in (3) and (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 Problems 1 and 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.

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*

$$E_{\mathrm{wc}} \doteq \sup_{x\sim(\hat{x},\Gamma)} \mathrm{E}\{\phi(x)\}$$

$$P_{\mathrm{wc}} \doteq \sup_{x\sim(\hat{x},\Gamma)} \mathrm{Prob}\{x \in S\}$$

$$Q \doteq \begin{bmatrix} \Gamma + \hat{x}\hat{x}^\top & \hat{x} \\ \hat{x}^\top & 1 \end{bmatrix}.$$

Then,

$$E_{\mathrm{wc}} = \inf_{M=M^\top} \mathrm{Tr}\, QM \quad \textit{subject to: } z^\top Mz \geq \phi(x), \quad \forall x \in \mathbb{R}^n \tag{6}$$

and

$$P_{\mathrm{wc}} = \inf_{M\succeq 0} \mathrm{Tr}\, QM \quad \textit{subject to: } z^\top Mz \geq 1, \quad \forall x \in S. \quad \triangleleft \tag{7}$$

A proof of Lemma 1 is essentially based on the results in [2], and it is given in Appendix A.

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. In this respect, the following equivalence holds:

$$\sup_{x\sim(\hat{x},\Gamma)} \mathrm{E}\{\phi(x)\} \leq \gamma \tag{8}$$

$$\Updownarrow$$

$$\exists M = M^\top : \mathrm{Tr}\, QM \leq \gamma,\ z^\top Mz \geq \phi(x), \quad \forall x \in \mathbb{R}^n. \tag{9}$$

To verify this fact, consider first the $\Downarrow$ direction: if (8) holds, we let M be the solution that achieves the optimum in (6), and we have that (9) holds. On the other hand, if (9) holds for some $M = \bar{M}$, then $\sup_{x\sim(\hat{x},\Gamma)} \mathrm{E}\{\phi(x)\} = \inf \mathrm{Tr}\, QM \leq \mathrm{Tr}\, Q\bar{M} \leq \gamma$, which concludes proves the statement. By an analogous reasoning, we can verify that

$$\sup_{x\sim(\hat{x},\Gamma)} \text{Prob}\,\{x\in S\}\le\epsilon$$
$$\Updownarrow$$
$$\exists M=M^\top : M\succeq 0,\quad \text{Tr}\,QM\le\epsilon,\quad z^\top Mz\ge 1,\quad \forall x\in S.\quad \triangleleft$$

3. 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

$$\psi_p(\theta)\doteq \sup_{\delta\sim(0,I)} \mathrm{E}\{\|L(\theta)z\|_p\},\quad \text{with}\ z^\top\doteq[\delta^\top\ 1],\tag{10}$$

$$r\doteq[0\cdots 0\ 1/2]^\top\in\mathbb{R}^{q+1},\tag{11}$$

where $L(\theta)\in\mathbb{R}^{m,q+1}$ is an affine function of parameter θ, given in (3) and (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*

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

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

Proof. From Lemma 1 we have that

$$\psi_p(\theta)=\inf_{M=M^\top}\ \text{Tr}\,M$$
$$\text{subject to: } z^\top Mz\ge\|L(\theta)z\|_p,\quad\forall\delta\in\mathbb{R}^q.$$

Since

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

it follows that $z^\top Mz\ge\|L(\theta)z\|_p$ holds for all δ if and only if

$$z^\top Mz\ge u^\top L(\theta)z,\quad \forall\delta\in\mathbb{R}^q \text{ and } \forall\, u\in\mathbb{R}^m:\ \|u\|_{p^*}\le 1.$$

Now, since $z^\top r = 1/2$, we write $u^\top L(\theta)z = z^\top(ru^\top L(\theta)+L^\top(\theta)ur^\top)z$, whereby the above condition is satisfied if and only if

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

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 (10). Then, the following holds for the worst-case expected residuals in the ℓ_1- and ℓ_2-norm cases.*

1. *Case $p = 1$: Define*

$$\overline{\psi}_1(\theta)\doteq\sum_{i=1}^m\left\|L_i^\top(\theta)\right\|_2,\tag{12}$$

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

$$\frac{2}{\pi}\overline{\psi}_1(\theta)\le\psi_1(\theta)\le\overline{\psi}_1(\theta).\tag{13}$$

2. *Case $p = 2$:*

$$\psi_2(\theta)=\sqrt{\text{Tr}\,L^\top(\theta)L(\theta)}=\|L(\theta)\|_F.\quad\triangleleft\tag{14}$$

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

$$\psi_1(\theta)=\inf_{M=M^\top}\ \text{Tr}\,M\tag{15}$$
$$\text{subject to: } M-L^\top(\theta)ur^\top-ru^\top L(\theta)\succeq 0,$$
$$\forall u:\ \|u\|_\infty\le 1.$$

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

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

where

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

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 (15) takes the form (23) considered in Theorem 4 in Appendix B. Consider thus the following relaxation of problem (15):

$$\varphi\doteq\inf_{M=M^\top,X_i=X_i^\top}\ \text{Tr}\,M\tag{16}$$
$$\text{subject to: }\quad -X_i+C_i\preceq 0,\ -X_i-C_i\preceq 0,\quad i=1,\ldots,m,$$
$$\sum_{i=1}^m X_i-M\preceq 0,$$

where we clearly have $\psi_1\le\varphi$. The dual of problem (16) can be written as

$$\varphi^D=\sup_{\Lambda_i,\Gamma_i}\sum_{i=1}^m\text{Tr}\,((\Lambda_i-\Gamma_i)C_i)\tag{17}$$
$$\text{subject to: }\quad \Lambda_i+\Gamma_i=I_{q+1},$$
$$\Gamma_i\succeq 0,\ \Lambda_i\succeq 0,\quad i=1,\ldots,m.$$

Since the problem in (16) is convex and Slater conditions are satisfied, $\varphi=\varphi^D$. Next we show that φ^D equals $\overline{\psi}_1$ given in (12). To this end, observe that (17) 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 $C_i=V_i\Theta_iV_i^\top$, 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 Appendix B, we have $\overline{\psi}_1=\varphi=\varphi^D$ and $\psi_1\le\overline{\psi}_1$.

For the lower bound on ψ_1 in (13), assume that the problem in (17) is not feasible. Then, for $M\succeq 0$, we have that

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

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

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

and

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

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

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

Applying the LMI robustness lemma (Lemma 3.1 of [9]), we have that the previous semi-infinite problem is equivalent to the following SDP

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

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*

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

For $p = 1$, it holds that

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

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

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

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 $\text{Tr}\, L^\top(\theta)L(\theta)$. ◁

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 [4,10]. This might not be obvious, since in general $\text{E}\{\|\cdot\|^2\} \neq (\text{E}\{\|\cdot\|\})^2$. ◁

4. Guaranteed residual-at-risk minimization

4.1. The ℓ_2-norm case

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

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

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

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

where the constraint can be written equivalently as

$$\begin{aligned}& z^\top \left(M - \text{diag}(0_q, 1)\right) z \geq 0, \\ & \forall\delta : z^\top \left(L^\top(\theta)L(\theta) - \text{diag}(0_q, \gamma^2)\right) z \geq 0.\end{aligned}$$

Applying the lossless S-procedure, the condition above is in turn equivalent to the existence of $\tau \geq 0$ such that $\left(M - \text{diag}(0_q, 1)\right) \succeq \tau\left(L^\top(\theta)L(\theta) - \text{diag}(0_q, \gamma^2)\right)$, therefore we obtain

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

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

$$\begin{bmatrix} M - \text{diag}(0_q, 1 - \tau\gamma^2) & \tau L^\top(\theta) \\ \tau L(\theta) & \tau I_m \end{bmatrix} \succeq 0. \quad (18)$$

We now notice, by the reasoning in Remark 2, that the condition $P_{\text{wc}_2}(\theta) \leq \epsilon$ with $\epsilon \in (0, 1)$ is equivalent to the conditions

$$\exists\tau \geq 0, M \succeq 0 \text{ such that: Tr } M \leq \epsilon, \text{ and (18) holds.}$$

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 $P_{\text{wc}_2}(\theta) \leq \epsilon$ 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:*

$$\inf_{\tau>0, M\succ 0, \theta\in\mathbb{R}^n, \gamma^2>0} \gamma^2, \quad (19)$$

$$\begin{aligned}\textit{subject to:} \quad & \text{Tr } M \leq \tau\epsilon \\ & \begin{bmatrix} M - \text{diag}(0_q, \tau - \gamma^2) & L^\top(\theta) \\ L(\theta) & I_m \end{bmatrix} \succeq 0. \quad ◁\end{aligned}$$

Remark 4. The optimization problem in Theorem 2 is a standard and efficiently solvable SDP with $N = (q + 1)(q + 2)/2 + n + 2$ variables (the entries of symmetric matrix M, the entries of θ, and τ, γ^2). The exact numerical complexity of its solution depends on the specific algorithm employed, and it is guaranteed to grow at most as a polynomial of the problem size, see, e.g., [13,14]. ◁

4.2. 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

$P_{\mathrm{wc}_1}(\theta) \leq \epsilon$, where $P_{\mathrm{wc}_1}(\theta)$ is the worst-case ℓ_1-norm residual tail probability

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

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*

$$\|v\|_1 = \frac{1}{2}\inf_{D\in\mathcal{D}} \sum_{i=1}^{m}\left(\frac{v_i^2}{d_i} + d_i\right) = \frac{1}{2}\inf_{D\in\mathcal{D}}\left(v^\top D^{-1} v + \text{Tr}\, D\right), \quad (20)$$

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:*

$$\inf_{\tau>0, M\succeq 0, D\in\mathcal{D}, \theta\in\mathbb{R}^n, \gamma\geq 0} \gamma \quad (21)$$

subject to:

$$\text{Tr}\, M \leq \tau\epsilon$$

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

with $J \doteq \text{diag}(0_q, 1)$. 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

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

For ease of notation we drop the dependence on θ in the following derivation. Using (20) we have that, for any $D \in \mathcal{D}$,

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

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

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

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

$$P_{\mathrm{wc}_1} = \sup_{\delta\sim(0,I)} \text{Prob}\,\{\delta \in S\} \leq \inf_{D\in\mathcal{D}} \sup_{\delta\sim(0,I)} \text{Prob}\,\{\delta \in S(D)\} \doteq \bar{P}_{\mathrm{wc}1}.$$

Note that, for fixed $D \in \mathcal{D}$, we can compute $P_{\mathrm{wc}1}(D) \doteq \sup_{\delta\sim(0,I)}$ Prob $\{\delta \in S(D)\}$ from its equivalent dual:

$$P_{\mathrm{wc}1}(D) = \inf_{M\succeq 0} \text{Tr}\, M : z^\top M z \geq 1, \quad \forall \delta \in S(D)$$
$$= \inf_{M\succeq 0} \text{Tr}\, M : z^\top M z \geq 1, \quad \forall\delta : z^\top L^\top D^{-1} L z + \text{Tr}\, D \geq 2\gamma$$
$$\times\ [\text{applying the lossless S-procedure}]$$
$$= \inf_{M\succeq 0, \tau>0} \text{Tr}\, M : M \succeq \tau L^\top D^{-1} L + (1 - 2\tau\gamma + \tau \text{Tr}\, D)J,$$

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

$$\bar{P}_{\mathrm{wc}1} = \inf_{M\succeq 0, \tau>0, D\in\mathcal{D}} \text{Tr}\, M : M \succeq \tau L^\top D^{-1} L + (1 - 2\tau\gamma + \tau \text{Tr}\, D)J$$
$$\times\ [\text{by change of variable } \tau D \to D$$
$$\times\ (\text{whence } \tau \text{Tr}\, D \to \text{Tr}\, D,\ \tau D^{-1} \to \tau^2 D^{-1})]$$
$$= \inf_{M\succeq 0, \tau>0, D\in\mathcal{D}} \text{Tr}\, M : M \succeq \tau^2 L^\top D^{-1} L + (1 - 2\tau\gamma + \text{Tr}\, D)J$$
$$= \inf_{M\succeq 0, \tau>0, D\in\mathcal{D}} \text{Tr}\, M : \begin{bmatrix} M - (1 - 2\tau\gamma + \text{Tr}\, D)J & \tau L^\top \\ \tau L & D \end{bmatrix} \succeq 0.$$

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

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

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

$$\text{Tr}\, M \leq \tau\epsilon, \quad \text{and} \quad \begin{bmatrix} M - (\tau - 2\gamma + \text{Tr}\, D)J & L^\top(\theta) \\ L(\theta) & D \end{bmatrix} \succeq 0. \quad (22)$$

Notice that, since $L(\theta)$ is affine in θ, condition (22) 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. □

5. Numerical examples

5.1. Worst-case expected residual minimization

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

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

with $A_0 = \begin{bmatrix} 3 & 1 & 4 \\ 0 & 1 & 1 \\ -2 & 5 & 3 \\ 1 & 4 & 5.2 \end{bmatrix}$ $A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, 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

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

with nominal residuals of 1.7838 and 1.8235, respectively.

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

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

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 Fig. 1. We observe that both ℓ_1 and ℓ_2 expected residuals tend to a constant value for large ρ.

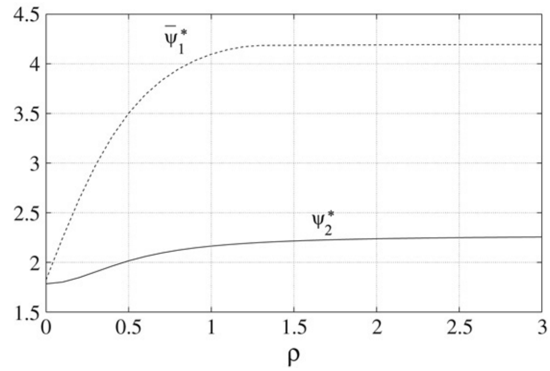

Fig. 1. Plot of ψ_2^* (solid) and $\bar{\psi}_1^*$ (dashed) as a function of perturbation level ρ.

5.2. Guaranteed residual at risk minimization

As a first example of guaranteed residual at risk minimization, we consider again the variable perturbation level problem of the previous section. Here, we fix the risk level to $\epsilon = 0.1$ and solve repeatedly problems (19) and (21) for increasing values of ρ. A plot of the resulting optimal residuals at risk as a function of ρ is shown in Fig. 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.

As a further example, we consider a system identification problem where one seeks to estimate the impulse response of a discrete-time linear FIR system from its input/output measurements. Let the system be described by the convolution

$$y_k = \sum_{j=1}^{N} h_j \cdot u_{k-\tau+1}, \quad k = 1, 2, \ldots,$$

where u_i is the input, y_i is the output, and $h = [h_1 \cdots h_N]$ is the impulse response to be estimated. If N input/output measurements are collected in vectors $u^\top \doteq [u_1\, u_2 \cdots u_N]$, $y^\top \doteq [y_1\, y_2 \cdots y_N]$, then the impulse response h can be computed by solving the system of linear equations $Uh = y$, where U is a lower-triangular Toeplitz matrix having u in the first column. In practice, however, both u and y might be affected by errors, that is $U(\delta_u) = U + \sum_{i=1}^{N} \delta_{ui} U_i$, $y(\delta_y) = y + \sum_{i=1}^{N} \delta_{yi} e_i$ where e_i is the i-th column of the identity matrix in $\mathbb{R}^N$, and U_i is a lower-triangular Toeplitz matrix having e_i in the first column. These uncertain data are easily rewritten in the form (1) by setting $\theta = h$, $q = 2N$, $A_0 = U$, $b_0 = y$, and, for $i = 1, \ldots, 2N$,

$$A_i = \begin{cases} U_i, & \text{if } i \le N \\ 0_N, & \text{otherwise}, \end{cases} \qquad b_i = \begin{cases} 0, & \text{if } i \le N \\ e_{i-N}, & \text{otherwise}, \end{cases}$$

$$\delta_i = \begin{cases} \delta_{ui}, & \text{if } i \le N \\ \delta_{y,i-N}, & \text{otherwise}. \end{cases}$$

We considered the same data used in a similar example in [8], that is $u^\top = [1\ 2\ 3]$, $y^\top = [4\ 5\ 6]$, and assumed that the input/output measurements are affected by i.i.d. errors with zero mean and unit variance. The nominal system $Uh = y$ has minimum-norm solution $h_{\text{nom}} = [4\ {-3}\ 0]^\top$. First, we solved the ℓ_2 residual-at-risk problem in (19), setting risk level $\epsilon = 0.1$. The optimal solution yielded an optimal worst-case residual at risk $\gamma = 11.09$, with corresponding parameter

$$h_{\ell_2,\text{rar}} = [0.7555\ 0.4293\ 0.1236]^\top.$$

This means that, no matter what the distribution on the uncertainty is (as long has it has the assumed mean and covariance), the probability of having a residual larger than $\gamma = 11.09$ is smaller than $\epsilon = 0.1$.

We next performed an a-posteriori Monte-Carlo test of this solution against the nominal one h_{nom}, for two specific noise distributions, namely the uniform and the Normal distribution. The empirical cumulative distribution functions of the resulting norm residuals are shown in Fig. 3. Some relevant statistics are also shown in Table 1.

We next solved the same problem using an ℓ_1-norm residual criterion. In this case, solution of problem (21) yielded an optimal upper bound on ℓ_1 residual at risk $\gamma = 19.2$, with corresponding parameter

$$h_{\ell_1,\text{rar}} = [0.7690\ 0.4464\ 0.1344]^\top.$$

Table 1
ℓ_2-norm residual statistics from a-posteriori test on nominal and residual-at-risk solutions, with uniform and Normal distribution on the noise

	Uniform		Normal	
	h_{nom}	$h_{\ell_2,\text{rar}}$	h_{nom}	$h_{\ell_2,\text{rar}}$
min	0.19	0.49	0.19	0.15
max	19.40	10.79	28.98	11.75
mean	7.75	5.52	7.50	5.51
median	7.50	5.53	7.01	5.49
std	3.02	1.44	3.58	1.43
rar@$\epsilon = 0.1$	11.85	7.39	12.39	7.36

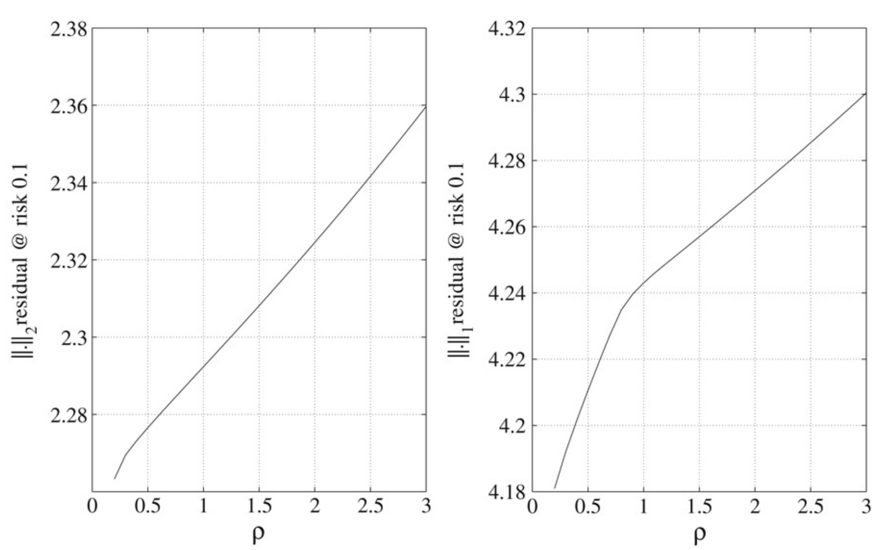

Fig. 2. Worst-case ℓ_2 and ℓ_1 residuals at risk as a function of perturbation level ρ.

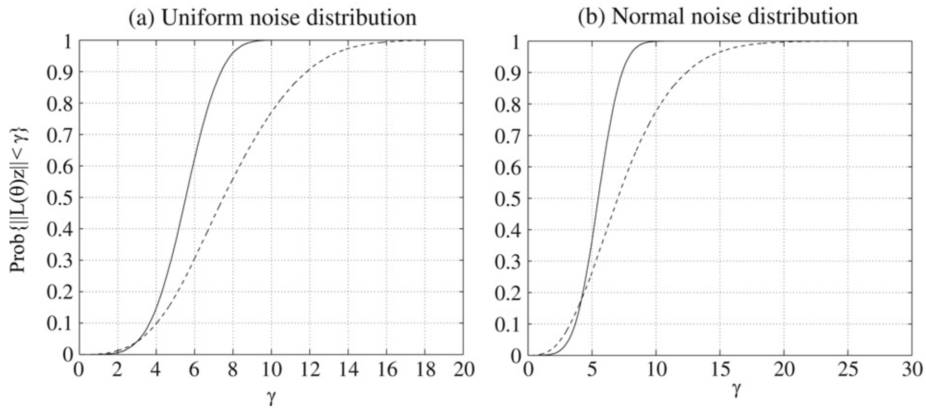

Fig. 3. Empirical ℓ_2-norm residual cumulative distribution for the $h_{\ell_2,\mathrm{rar}}$ solution (solid) and the nominal h_{nom} solution (dashed). Simulation in case (a) assumes uniform noise distribution; case (b) assumes Normal distribution.

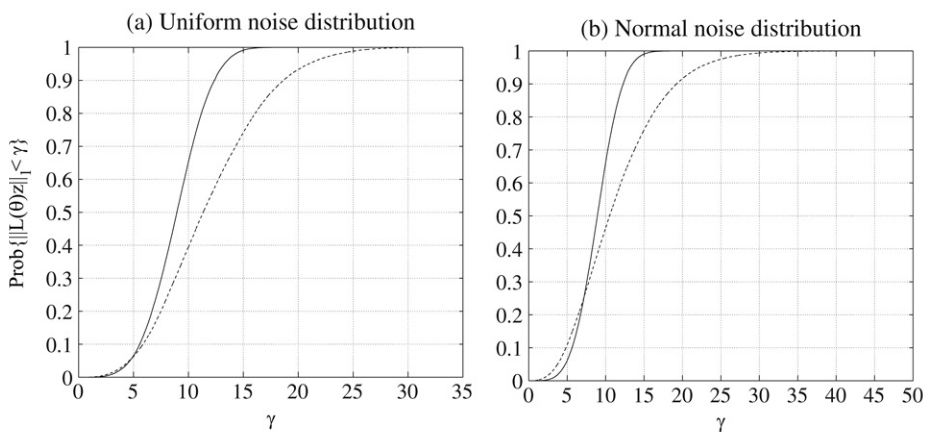

Fig. 4. Empirical ℓ_1-norm residual cumulative distribution for the $h_{\ell_1,\mathrm{rar}}$ solution (solid) and the nominal h_{nom} solution (dashed). Simulation in case (a) assumes uniform noise distribution; case (b) assumes Normal distribution.

Table 2
ℓ_1-norm residual statistics from a-posteriori test on nominal and residual-at-risk solutions, with uniform and Normal distribution on the noise

	Uniform		Normal	
	h_{nom}	$h_{\ell_1,\mathrm{rar}}$	h_{nom}	$h_{\ell_1,\mathrm{rar}}$
min	0.38	0.81	0.11	0.57
max	32.12	18.39	47.93	19.63
mean	11.89	8.95	11.42	8.94
median	11.34	8.95	10.52	8.93
std	5.01	2.61	5.76	2.56
rar@$\epsilon = 0.1$	18.66	12.37	19.22	12.29

An a-posteriori Monte-Carlo test then produced the residual distributions shown in Fig. 4. Residual statistics are reported in Table 2.

6. 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 A. Proof of Lemma 1

We start by recalling a preliminary result. Let $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, be functions whose expectations q_i are given and finite, and consider the following problem (P) and its dual (D):

$$(P): \quad Z^P \doteq \sup_x \mathrm{E}\{\phi(x)\}$$
$$\text{subject to: } \mathrm{E}\{f_i(x)\} = q_i, \quad i = 0, 1, \ldots, m;$$
$$(D): \quad Z^D \doteq \inf_y \mathrm{E}\{y^\top f(x)\} = \inf y^\top q$$
$$\text{subject to: } y^\top f(x) \geq \phi(x), \quad \forall x \in \mathbb{R}^n,$$

where $f_0(x) = 1$ and $q_0 = 1$ correspond to the implied probability-mass constraint, and the infimum and the supremum are taken with respect to all probability distributions satisfying the specified moment constraints. Then, strong duality holds: $Z^P = Z^D$, hence $\sup_x \mathrm{E}\{\phi(x)\}$ can be computed by solving the dual problem (D); see for instance Section 16.4 of [2].

Now, the primal problem that we wish to solve in Lemma 1 for computing E_{wc} is

$$(P): Z^P = E_{\rm wc} = \sup_x \mathrm{E}\{\phi(x)\}$$

$$\text{subject to:}\quad \mathrm{E}\{x\} = \hat{x}$$
$$\mathrm{E}\{xx^\top\} = \Gamma + \hat{x}\hat{x}^\top,$$

where $\Gamma \succ 0$ is the covariance matrix of x, and the functions f_i are x_k, $k = 1, \ldots, n$, and $x_k x_j$, $1 \le k \le j \le n$. Then, the dual problem is

$$(D):\quad Z^D = \inf_{y_0\in\mathbb{R}, y\in\mathbb{R}^n, Y=Y^\top\in\mathbb{R}^{n,n}} y_0 + y^\top \hat{x} + \mathrm{Tr}\,(\Gamma + \hat{x}\hat{x}^\top)Y$$
$$\text{subject to: } y_0 + y^\top x + \mathrm{Tr}\; xx^\top Y \ge \phi(x), \quad \forall x \in \mathbb{R}^n,$$

where the dual variable y_0 is associated with the implicit probability-mass constraint. Defining

$$M = \begin{bmatrix} Y & \frac{1}{2}y \\ \frac{1}{2}y^\top & y_0 \end{bmatrix}, \qquad Q = \begin{bmatrix} \Gamma + \hat{x}\hat{x}^\top & \hat{x} \\ \hat{x}^\top & 1 \end{bmatrix}, \qquad z = \begin{bmatrix} x \\ 1 \end{bmatrix},$$

this latter problem writes

$$(D):\quad Z^D = \inf_{M=M^\top\in\mathbb{R}^{n+1}} \mathrm{Tr}\; QM$$
$$\text{subject to: } z^\top M z \ge \phi(x), \quad \forall x \in \mathbb{R}^n,$$

which is (6), thus concluding the first part of the proof.

The result in (7) can then be obtained by specializing (6) to the case when $\phi(x) = \mathbb{I}_S(x)$, where $\mathbb{I}_S$ is the indicator function of set S, since $\mathrm{Prob}\{x \in S\} = \mathrm{E}\{\mathbb{I}_S(x)\}$. We thus have that $Z^P = Z^D$ for

$$(P):\quad Z^P = P_{\rm wc} = \sup_x \mathrm{Prob}\{x \in S\}$$
$$\text{subject to:}\quad \mathrm{E}\{x\} = \hat{x}, \quad \mathrm{E}\{xx^\top\} = \Gamma + \hat{x}\hat{x}^\top,$$
$$(D):\quad Z^D = \inf_{M=M^\top\in\mathbb{R}^{n+1}} \mathrm{Tr}\; QM$$
$$\text{subject to: } z^\top M z \ge \mathbb{I}_S(x), \quad \forall x \in \mathbb{R}^n.$$

The constraint $z^\top M z \ge \mathbb{I}_S(x) \forall x \in \mathbb{R}^n$ can be rewritten as $z^\top M z \ge 1\ \forall x \in S$, $z^\top M z \ge 0\ \forall x \in \mathbb{R}^n$, and this latter constraint is equivalent to requiring $M \succeq 0$, which explains (7) and concludes the proof. □

Appendix B. Matrix cube theorem

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

$$\mathrm{P}_\rho : \textit{Is } B^0 + \sum_{i=1}^{L} u_i B^i \succeq 0, \quad \forall u : \|u\|_\infty \le \rho? \tag{23}$$

and the problem P_{relax} be defined as:

P_{relax}*: Do there exist symmetric matrices $X_1, \ldots, X_L \in \mathbf{R}^{n\times n}$ satisfying*

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

Then, the following statements hold:

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

References

[1] A. Ben-Tal, A. Nemirovski, On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty, SIAM Journal on Optimization 12 (3) (2002) 811–833.
[2] D. Bertsimas, J. Sethuraman, Moment problems and semidefinite optimization, in: Handbook of Semidefinite Programming, Kluwer Acad. Publ., Boston, MA, 2000, pp. 469–509.
[3] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, New York, NY, ISBN: 0-521-83378-7, 2004.
[4] G. Calafiore, F. Dabbene, Near optimal solutions to least-squares problems with stochastic uncertainty, Systems and Control Letters 54 (12) (2005) 1219–1232.
[5] C. Dossal, A necessary and sufficient condition for exact recovery by ℓ_1 minimization, 2008. http://hal.archives-ouvertes.fr/hal-00164738/fr/.
[6] D. Duffie, J. Pan, An overview of value at risk, Journal of Derivatives (1997) 7–49.
[7] M.P. Friedlander, P. Tseng, Exact regularization of convex programs, SIAM Journal on Optimization 18 (4) (2007) 1326–1350.
[8] L. El Ghaoui, H. Lebret, Robust solutions to least-squares problems with uncertain data, SIAM Journal on Matrix Analysis and Applications 18 (4) (1997) 1035–1064.
[9] L. El Ghaoui, F. Oustry, H. Lebret, Robust solutions to uncertain semidefinite programs, SIAM Journal Optimization 9 (1) (1998) 33–52.
[10] H.A. Hindi, S.P. Boyd, Robust solutions to l_1, l_2, and l_∞ uncertain linear approximation problems using convex optimization, in: Proc. American Control Conf., vol. 6, 1998, pp. 3487–3491.
[11] R.T. Rockafellar, S. Uryasev, Conditional value-at-risk for general loss distributions, Journal of Banking and Finance 26 (2002) 1443–1471.
[12] A.H. Sayed, V.H. Nascimento, F.A.M. Cipparrone, A regularized robust design criterion for uncertain data, SIAM Journal on Matrix Analysis and Applications 23 (4) (2002) 1120–1142.
[13] M. Todd, Semidefinite optimization, Acta Numerica 10 (2001) 515–560.
[14] L. Vandenberghe, S. Boyd, Semidefinite programming, SIAM Review 38 (1) (1996) 49–95.