On the Convergence Rates of Set Membership Estimation of Linear Systems with Disturbances Bounded by General Convex Sets (Supplementary Material)

Haonan Xu and Yingying Li

Abstract— **This paper studies the uncertainty set estimation of system parameters of linear dynamical systems with bounded disturbances, which is motivated by robust (adaptive) constrained control. Departing from the confidence bounds of least square estimation from the *machine-learning literature*, this paper focuses on a method commonly used in (robust constrained) *control literature*: set membership estimation (SME). SME tends to enjoy better empirical performance than LSE's confidence bounds when the system disturbances are bounded. However, the theoretical guarantees of SME are not fully addressed even for i.i.d. bounded disturbances. In the literature, SME's convergence has been proved for general convex supports of the disturbances [1], but SME's convergence rate assumes a special type of disturbance support: l_∞ ball [2]. The main contribution of this paper is relaxing the assumption on the disturbance support and establishing the convergence rates of SME for general convex supports, which closes the gap on the applicability of the convergence and convergence rates results. Numerical experiments on SME and LSE's confidence bounds are also provided for different disturbance supports. (The additional content in this supplementary manuscript is in Appendix E.)**

I. INTRODUCTION

Recent years have witnessed significant interests and progresses on the non-asymptotic analysis of system identification of linear dynamical systems, e.g.

$$x_{t+1} = A^* x_t + B^* u_t + w_t, \qquad (1)$$

by leveraging statistical learning tools [3], [4], [5], [6], [7], [8]. For example, least-square estimation (LSE) is a popular method to estimate the system parameters and has been shown to achieve the optimal convergence rate for Gaussian noises w_t [5], [6].

However, for most safety-critical applications, it is also crucial to characterize the uncertainties of the system parameter estimation and satisfy safety requirements (e.g. constraint satisfaction, stability, etc.) despite the uncertainties [9], [10]. One possible way to achieve this is by estimating the uncertainty sets of the system parameters and designing robustly safe controllers to satisfy the safety requirements for any possible system parameters in the uncertainty sets [10], [11]. It is well-known that the size of the uncertainty sets heavily affects the robust (constrained) control performance since a large uncertainty set will result in over-conservative controllers, thus generating worse performance; while an uncertainty set that fails to contain the true parameters may lead to unsafe control behaviors [10], [12]. Therefore, it is vital to reduce the size of the uncertainty sets as much as possible while still containing the true parameters.

In the literature, there are two major methodologies for the uncertainty set estimation for linear systems.

1) Confidence-bound-based approaches, such as the confidence bounds for LSE [8], [5], which rely on the statistical properties of w_t. This methodology has been very popular in the machine-learning community and recent learning-based control literature [13], [14].
2) Set-membership-based approaches, which are based on bounded disturbances but do not require any statistical properties on w_t to construct an uncertainty set containing the true parameters.This methodology has a long history in the control literature [15], [16], [17], [18], [19], [20] and has been a popular method in the literature on robust adaptive control with constraints [1], [16], [13], [21], [22], [23], [24], [25], [26], [27].

Set-membership estimation (SME) is known for generating valid uncertainty sets without statistical assumptions on w_t. But even when w_t enjoys some statistical properties, such as i.i.d., SME still tends to outperform the confidence bounds of LSE empirically [20], [2]. The promising empirical performance of SME motivates more theoretical analysis of SME's convergence rates.

Perhaps surprisingly, despite the long history of SME in the control literature, the theoretical convergence rate analysis of SME have been limited, especially on dynamical systems. Most of the existing convergence rate analysis considers stochastic linear regression problems: $y_t = \theta^* z_t + \epsilon_t$, where ϵ_t are i.i.d. and x_t are independent of history $\epsilon_0, \ldots, \epsilon_{t-1}$ [28], [29], [30], [31], [32].[1] However, these convergence rate results do not directly apply to linear dynamical systems because our (x_t, u_t) depends on the history $w_0, \ldots, w_{t-1}$ through (1). This issue has been overlooked in the large body of literature on the control design based on SME until 2019, when [34] identifies this issue and provides the first convergence results of SME when w_t is tightly bounded by a convex compact set $\mathbb{W}$. Later, [2]

Haonan Xu and Yingying Li are with the Department of Industrial and Enterprise Systems Engineering and the Coordinated Science Lab, University of Illinois Urbana-Champaign, IL, USA (e-mail: haonan9, yl101@illinois.edu).

[1] There are also theoretical analysis of SME under deterministic w_t (e.g. [33]), but SME does not converge for general deterministic sequences of w_t, so these bounds are usually converging to a neighborhood around the true parameters. It is an interesting future work direction to study what conditions on the deterministic sequence of w_t can guarantee the convergence of SME.

established the convergence rate of SME for a specific set $\mathbb{W} = \{w : \|w\|_\infty \leq w_{\max}\}$, which is an l_∞ ball, and $\mathbb{W}$ is assumed to be a tight bound on the true support of w_t on all directions.

The assumption in [2] on the l_∞ ball is rather restrictive and one can easily find applications that do not satisfy this assumption. For example, consider voltage control in distribution power systems, where the disturbances on each substation/nodes are the uncontrollable power injections, which equals power generation at this substation minus the power consumption at this substation [35]. Some substations may have renewable energy generators, such as solar panels and wind turbines, so the disturbances on these substations can be positive, while the substations without generators will only have negative disturbances. Further, different substations usually have very different magnitudes of power consumption, e.g. a high-density residential area may consume much more power than a low-density residential area. Therefore, the true support of the disturbances on all substations are mostly likely an irregular hyper-rectangle that does not center at 0, which does not satisfy the assumption in [2].

Therefore, it is important to analyze the convergence rate of SME under more general $\mathbb{W}$ and bridges the gap between the assumptions that guarantees the convergence of SME and the ones that guarantees a convergence rate of SME.

Our contributions. This paper tackles this gap by providing convergence rates of SME for general convex and compact $\mathbb{W}$. To achieve this, we construct a unifying assumption that bridges the assumptions in [34] and [2]. Our convergence bounds are instance-specific, i.e., depending on the probability distributions of w_t, which is similar to the previous convergence rate analysis of SME in both the linear dynamical system and linear regression. To provide more insight, we explain our convergence rates in several different distributions of w_t and discusses how the shape of $\mathbb{W}$ affects the convergence rates. Numerical experiments are also provided that compares SME and LSE's confidence bounds.

Notations. Let $\mathbb{N}$ be the set of all positive integers and $\mathbb{N}_0$ be the set of all non-negative integers. For two matrices $M_1 \in \mathbb{R}^{k\times t}, M_2 \in \mathbb{R}^{k\times r}$, let $\begin{pmatrix} M_1 & M_2 \end{pmatrix}$ denote the concatenated matrix, and the same applies to vector concatenation. Let $||\cdot||_F$ denote the Frobenius norm of a matrix. $\forall\, p \in \mathbb{N}$, denote $||\cdot||_p$ the p-norm of a vector. Specially, $||\cdot||_\infty$ is the vector's infinity norm. For events A and B, let $A \wedge B$ denote the event where both A and B hold, and let $A^{\complement}$ denote the complement of A. In $\mathbb{R}^n$, the closed 2-norm ball with radius $r > 0$ and center x is denoted $\mathbb{B}_r(x)$; and the 2-norm sphere with with radius $r > 0$ and center x is denoted $\mathbb{S}_r(x)$. For two sets C and D, we denote $C \subseteq D$ if $\forall\, c \in C$, $c \in D$. The tilde big-O notation $\tilde{O}(h(n))$, denotes asymptotic bounds ignoring logarithmic terms. In a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we say $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$ is a filtration if $\forall\, i \in \mathbb{N}$, $\mathcal{F}_i$ is a sub-σ-algebra of $\mathcal{A}$ and $\forall\, i \leq j$ one has $\mathcal{F}_i \subseteq \mathcal{F}_j$. If a statement holds with probability 1, we say it holds almost surely, denoted $a.s.$. Denote the σ-algebra generated by a collection of random variables $\sigma\{\cdots\}$. The interior of a set E is denoted $\mathring{E} := \{e \in E \ : \ \exists\, \epsilon > 0 \ s.t.\ \mathbb{B}_\epsilon(e) \subseteq E\}$. $F \succ 0$ denotes that a matrix F is positive definite.

II. Problem Formulation and Preliminaries

A. Problem Formulation

In this paper, we consider a discrete-time linear dynamical system with additive process noises:

$$x_{t+1} = A^* x_t + B^* u_t + w_t, \quad t \geq 0 \tag{2}$$

where $x_t \in \mathbb{R}^{n_x}$, $u_t \in \mathbb{R}^{n_u}$, and $w_t \in \mathbb{R}^{n_x}$ respectively denote the state, the control input, and the process noise at time $t \in \mathbb{N}_0$, and the system parameters (A^*, B^*) are unknown. For ease of notations, we denote $\theta^* = \begin{pmatrix} A^* & B^* \end{pmatrix} \in \mathbb{R}^{n_x\times(n_x+n_u)}$, $z_t = \begin{pmatrix} x_t^\top & u_t^\top \end{pmatrix}^\top \in \mathbb{R}^{n_z}$, and $n_z = n_x + n_u$. Accordingly, the system (2) can be written as

$$x_{t+1} = \theta^* z_t + w_t. \tag{3}$$

This paper focuses on the estimation of the uncertainty set of the unknown parameter θ^*. In particular, given a sequence of data $\{x_t, u_t, x_{t+1}\}_{t=0}^{T-1}$ in horizon T for $T \geq 1$, we aim to construct an uncertainty set Θ_T that contains the true parameter θ^* and is as small as possible.

The uncertainty set of the system parameters are commonly used in robust constrained control, such as robust model predictive control [36], robust control barrier functions [21], and robust system level synthesis [37], which is a popular methodology for safety-critical applications, such as power systems [38], robotics [39], etc. The robust constrained control usually aims to achieve robust constraint satisfaction for any possible θ in the uncertainty set Θ_T as well as for any other system uncertainties in the corresponding uncertainty sets. In this way, the constraints are guaranteed to be satisfied for the true θ^* since $\theta^* \in \Theta_T$. Naturally, the size of the uncertainty set Θ_T will heavily influence the conservativeness of the robust constraints, thus having a significant impact on the control performance. Therefore, it is crucial to reduce the size of Θ_T as much as possible while still guaranteeing $\theta^* \in \Theta_T$.

Formally, we quantify the size of the uncertainty set by its diameter as defined below.

Definition 1 (Diameter of a set of matrices)**.** *For a set of matrices $\Theta \in \mathbb{R}^{n_x\times n_z}$, we define its diameter to be diam$(\Theta) := \sup_{\theta_1,\theta_2\in\Theta} ||\theta_1 - \theta_2||_F$.*

B. Preliminaries on Set Membership Estimation

This paper focuses on a specific uncertainty-set-estimation method: set membership estimation (SME). SME enjoys a long history in the control literature and has been widely adopted in robust adaptive control algorithm design [40].[2]

SME considers bounded noises w_t, and leverages the boundedness of the noises to construct the uncertainty sets.

[2]There is also vast literature using SME for state estimation in the output-feedback systems [41], but this paper only studies the state-feedback case and consider system-parameter identification.

In particular, suppose the support of w_t, denoted by $\mathbb{W}$, is known, then SME constructs the uncertainty set:

$$\Theta_T = \left\{ \hat{\theta} \in \mathbb{R}^{n_x \times n_z} \; : \; \forall\, 0 \le t \le T-1, x_{t+1} - \hat{\theta} z_t \in \mathbb{W} \right\}$$

The idea behind this algorithm is quite intuitive and straightforward. This algorithm has a long history and has been introduced in multiple pieces of literature [42], [43]. It is trivial to validate that $\theta^* \in \Theta_T$ for any $T \ge 0$ because $x_{t+1} - \theta^* z_t = w_t \in \mathbb{W}$.

Existing literature has shown promising empirical performance of SME compared with the confidence-bound-based uncertainty set estimation, e.g. least square estimator's confidence bounds [20], [2]. This has attracted interests on the theoretical analysis of SME's convergence behaviors. For example, [34] established the convergence of SME for general convex set $\mathbb{W}$, while [2] provided the convergence rate for a special $\mathbb{W}$: $\mathbb{W} = \{w : \|w\|_\infty \le w_{\max}\}$. This paper tries to extend the results in [2] by considering more general $\mathbb{W}$ than l_∞ ball. In the next section, we will discuss the assumptions needed by our theoretical analysis and discuss how we relax the assumption on $\mathbb{W}$.

C. Assumptions and Some Discussions on $\mathbb{W}$

First, we list the assumptions consistent with [2]. Then, we will discuss how we relax the assumption on $\mathbb{W}$.

1) Standard Assumptions from the literature [2].:

Assumption 1 (i.i.d. noise)**.** *The additive noise $\{w_t\}_t$ are identically and independently sampled from a noise set $\mathbb{W}$ with $\mathbb{E}(w_t) = \vec{\mu}_w$ and Cov $(w_t) = \Sigma_w \succ \mathbf{0}$.*

Assumption 2 (bounded z_t and BMSB condition [6])**.** *$\exists\, b_z$ such that $\forall\, t \ge 0$, $||z_t||_2 \overset{a.s.}{\le} b_z$. Meanwhile, given the filtration $\{\mathcal{F}_t\}_{t=1}^{T-1}$, where $\mathcal{F}_t := \sigma\{w_0, \cdots, w_{t-1}, z_0, \cdots, z_t\}$. $\exists\, p_z \in (0,1]$, $\sigma_z > 0$ such that $\forall\, \lambda \in \{\lambda \in \mathbb{R}^{n_z} \; : \; ||\lambda||_2 = 1\}$:*

$$\forall\, t \ge 0, \; \mathbb{P}\left(|\lambda^\top z_{t+1}| \ge \sigma_z \;\; | \;\; \mathcal{F}_t\right) \ge p_z$$

The boundedness condition can be naturally satisfied when applying SME to robust constrained control, where the controllers are guaranteed to satisfy safety constraints (which are usually bounded) for any uncertain parameters in some priorly known initial uncertainty set. Further, the bounded z_t condition can also be achieved by applying bounded-input-bounded-state controllers since our disturbances are bounded.

Regarding the BMSB (block-martingale small-ball) condition as defined in [6], this could be achieved by adding an i.i.d. noise with positive definite covariance to the control policies as shown in [4].

2) Discussions on the shape of $\mathbb{W}$ and the tight bound assumptions: Our analysis is based on the assumption that $\mathbb{W}$ is a tight bound for the support of w_t. The existing literature [2] presents analysis when $\mathbb{W}$ is an ∞-norm ball. In realistic applications, this assumption may not be sufficient to model problems with asymmetric constraints on the noise. For example, we consider a power supply system where each utility has a stochastic demand in its respective interval. In this paper, we extend the theory in [2] to scenarios where $\mathbb{W}$ is a general polytope. Particularly, we describe $\mathbb{W}$ by linear constraints. Each linear constraint, in the form $v^\top w \ge h(v)$ for some **normal vector** v and $h(v) \in \mathbb{R}$, encodes a supporting hyperplane of $\mathbb{W}$.

Definition 2 (normal vector and supporting hyperplane)**.** *For the sake of convenience, we denote V the set of all normal vectors of $\mathbb{W}$. Formally, we have:$V := \{v \in \mathbb{S}_1(0) \; : \; \forall\, w^0 \in \arg\min_{\tilde{w} \in \mathbb{W}} v^\top \tilde{w}, \forall\, w \in \mathbb{W}, \; v^\top (w - w^0) \ge 0\}$.*

Accordingly, the map $h : \mathbb{S}_1(0) \to \mathbb{R}$ is defined by $h(v) = \min_{\tilde{w} \in \mathbb{W}} v^\top \tilde{w}$. The half space corresponding to $(v, h(v))$ is denoted $H(v) := \{w \in \mathbb{R}^{n_x} \; : \; v^\top w \ge h(v)\}$.

When $\mathbb{W}$ is compact and convex, we have the following Proposition

Proposition 1 (properties of V and $H(v)$)**.** *For V and $H(v)$, we have: a) $V = \mathbb{S}_1(0)$; b) $\cap_{v \in \mathbb{S}_1(0)} H(v) = \mathbb{W}$.*

The proof of Proposition can be found in the Appendix. Despite the fact that $\cap_{v \in \mathbb{S}_1(0)} H(v) = \mathbb{W}$, we do not necessarily need every single linear constraints to determine $\mathbb{W}$. For example, it takes as few as $2n_x$ linear constraints to define an n_x-dimensional, ∞-norm ball. However, for $\mathbb{W}$ in the form of a 2-norm ball, we will need all vectors in $\mathbb{S}_1(0)$ to represent it. All the extra linear inequalities (if any) resembles the redundant constraints in real-world implementations of general convex optimization problems. Notice that redundant linear constraints have no effect on the algorithm outputs. Therefore, we will utilize $\tilde{V} \subseteq V$, the non-redundant set of normal vectors in our analysis.

We do not limit the number of linear constraints used to define $\mathbb{W}$, and allow its continuation towards infinity. In the case where we allow infinitely many linear constraints, $\mathbb{W}$ is generalized to any convex, compact set.

Now, we are ready to present the tight bound assumption on $\mathbb{W}$, which is similar to the assumptions in [2], [34].

Assumption 3 (Tight bound)**.** *Let $\mathbb{W}$ be compact, convex, and have a non-empty interior. And $\forall\, \epsilon > 0$, $\exists\, q_w(\epsilon) > 0$ such that $\forall\, t \ge 0$, $\forall\, v \in V$: $\mathbb{P}\left(v^\top w_t - h(v) < \epsilon\right) \ge q_w(\epsilon)$.*

The intuitive explanation of Assumption 3 is that for every "thin" slice of $\mathbb{W}$ near its boundary, w_t has a non-vanishing probability to visit that area. Figure 1 demonstrates the underlined "thin slice" in 2-dimensional $\mathbb{W}$:

Remark 1. *The tight bound assumption guarantees that the noise has a non-vanishing probability to be arbitrarily close to the boundary of $\mathbb{W}$. Though this assumption can be restrictive, it is a common assumption for the convergence analysis of SME in the literature [2], [34], [31]. The absence of this assumption leads the SME to be only able to converge to a neighborhood of θ^*. Meanwhile, some pieces of recent literature (e.g. [44], [2]) tries to design SME-based algorithms that do not require a tight bound on w_t. It is our ongoing work to establish convergence rates of those algorithms.*

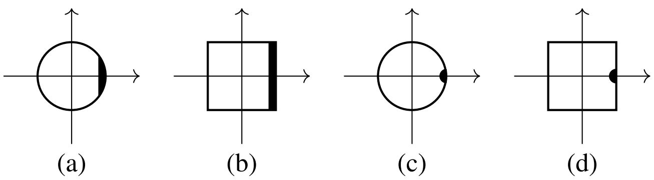

Fig. 1: This figure visualizes the definition of $q_w(\epsilon)$ in both this paper and in the literature [2] and [34]. In particular, (a) and (b) are the "thin" slices we consider in this paper. In the literature [34], nevertheless, the tight bound assumption is based on the ϵ-neighborhoods of any boundary point of $\mathbb{W}$ as is indicated in (c) and (d). In other words, we consider a looser assumption on the noise's performance near the boundary of $\mathbb{W}$. Besides, for l_∞ balls illustrated in (b), our assumption is the same as the assumption in [2], so the disturbances considered in [2] can be viewed as a special case of the disturbances considered in this paper.

III. Theoretical Derivation & Analysis

In this section, we first propose a non-asymptotic bound for the SME's diameter under the most general circumstance where $\mathbb{W}$ is an arbitrary compact and convex set. For the elegance of the analysis, we investigate into the non-redundant normal vector set $\tilde{V}$ instead of $\mathbb{S}_1(0)$. Notice that since $\cap_{v\in\mathbb{S}_1(0)} H(v)$ and $\cap_{v\in\tilde{V}} H(v)$ represent exactly the same set $\mathbb{W}$, the difference only lies within the analysis but not in the outputs of the set membership algorithm. Before stating the theorem, we need a constant ξ that encodes some geometric properties of the vectors in $\tilde{V}$.

Definition 3. *Let $\tilde{V} \subseteq \mathbb{S}_1(0)$ be a set of normal vectors of $\mathbb{W}$ such that $\cap_{v\in\tilde{V}} H(v) = \mathbb{W}$, define the **projection constant** of $\tilde{V}$ by $\xi = \inf_{c\in\mathbb{S}_1(0)} \max_{v\in\bar{\tilde{V}}} v^\top c$ where $\bar{\tilde{V}}$ denotes the closure of $\tilde{V}$.*

Lemma 1. *For any $\tilde{V}$, its projection constant ξ is strictly positive.*

The proof of the Lemma can be found in the appendix. More discussion about the property of ξ is also presented after the main theorem is stated and proven. With $\xi > 0$ defined, we hereby state the theorem for general convex and compact noise set $\mathbb{W}$:

A. The Main Theorem & Its Proof

Theorem 1. *$\forall T > m > 0, \delta > 0$, one has*

$$\mathbb{P}\left(diam(\Theta_T) > \delta\right) \leq \underbrace{\frac{T}{m}\tilde{O}(n_z^{5/2}) a_2^{n_z} \exp(-a_3 m)}_{\text{Term 1}} + \underbrace{\tilde{O}\left((n_x n_z)^{5/2}\right) a_4^{n_x n_z} [1 - q_w(\frac{a_1 \delta \xi}{4})]^{\lceil (T-1)/m \rceil}}_{\text{Term 2}} \tag{4}$$

Here $a_1 = \frac{1}{4}\sigma_z p_z$, $a_2 = \max\{1, \frac{64 b_z^2}{\sigma^2 p_z^2}\}$, $a_3 = \frac{1}{8}p_z^2$, $a_4 = \max\{1, \frac{4b_z}{a_1 \xi}\}$.

With this theorem, we can estimate the asymptotic error bound given $\mathbb{W}$ and the distribution of w_t. We propose the following corollaries as examples, to help the readers understand the intuition of the rate underlined in Theorem 1:

Corollary 1. *If $q_w(\epsilon) = O(\epsilon)$, given $m \geq O(n_z + \log T - \log \epsilon)$, then with probability no less than $1 - 2\epsilon$, one has*

$$diam(\Theta_T) \leq \tilde{O}\left(\frac{n_x n_z}{T\xi}\right)$$

Proof of this Corollary is in the appendix.

Notice that Corollary 1 gives a $\tilde{O}(1/T)$ convergence rate if $q_w(\epsilon) = O(\epsilon)$, which is consistent with the convergence rate in [2] for $\mathbb{W}$ being an l_∞ ball. We will show in the following examples that $q_w(\epsilon) = O(\epsilon)$ can be not only satisfied by the l_∞ ball, but also other polytopes. Further, Corollary 1 indicates a better convergence rate in terms of T compared with the convergence rate of LSE, which is $O(1/\sqrt{T})$. This is likely because SME takes advantage of the additional assumption on bounded w_t, while LSE allows unbounded sub-Gaussian w_t.

Example 1 ($O(\epsilon)$ tight-bounds)**.** *We consider w_t uniformly, identically and independently sampled from (A weighted ∞-norm ball) $\mathbb{W} = \{w \in \mathbb{R}^{n_x} : \max_{i\in[n_x]} \left\{\frac{1}{a_i}|w_i|\right\} \leq 1\}$ for positive constants $a_1, \cdots, a_{n_x}$.*[3] *The $q_w(\epsilon)$ corresponding to this $\mathbb{W}$ is $O(\epsilon)$.*

Example 2 (A weighted 1-norm ball)**.** *We consider w_t uniformly sampled from $\mathbb{W} = \{w \in \mathbb{R}^{n_x} : \sum_{i=1}^{n_x} \frac{1}{a_i}|w_i| \leq 1\}$ for positive constants $a_1, \cdots, a_{n_x}$.*[4] *The volume of $\mathbb{W}$ is bounded by $[\frac{(2a)^{n_x}}{n_x!}, \frac{(2A)^{n_x}}{n_x!}]$, where $a = \min_{i\in[n_x]} a_i$ and $A = \max_{i\in[n_x]} a_i$. In this case $q_w(\epsilon) = O(\epsilon)$.*

Proofs of these two examples are available in the appendix.

B. Discussion: How ξ Influences the Error Bounds

Though well-defined, the projection constant ξ lacks intuitive illustrations. In this subsection, we propose several examples demonstrating its properties. For the most trivial case, $\mathbb{W} = \{w \in \mathbb{R}^{n_x} : ||w||_\infty \leq 1\}$ is the unit ∞-norm ball. In this case $\tilde{V} = \{\pm e_i : i \in [n_x]\}$, where $\{e_i\}_{i=1}^{n_x}$ is the canonical basis in $\mathbb{R}^{n_x}$. In this case, $v^\top c$ is optimized when c points at each of the vertices of $\mathbb{W}$, that is: $c \in \{\frac{1}{n_x}\}^{n_x}$ and it follows that $\xi = \frac{1}{\sqrt{n_x}}$. If we plug in this ξ value into the bound in (4), we will get the same result as in [2].

Another simple example is when $\mathbb{W}$ is a regular polytope centered at the origin. The following figures are examples in $\mathbb{R}^2$. Here $c^\top v$ is optimized when c is colinear with the angle bisector between any two "adjacent" normal vectors. Namely, for an N-gon, we always have $\xi = \cos\left(\frac{\pi}{N}\right)$. Notice that the error bound is not necessarily improved with a larger ξ, as we notice that the scale of $q_w(\cdot)$ also changes as

[3]Even though we only consider uniform distributions, but the bounds also holds for truncated Gaussian distributions (see [2]).

[4]Even though we only consider uniform distributions, but the bounds also holds for truncated Gaussian distributions (see [2]).

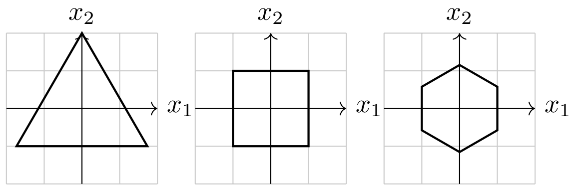

Fig. 2: Examples of polytopal $\mathbb{W}$ in $\mathbb{R}^2$

we consider $\mathbb{W}$ with more linear constraints. Considering the 2-dimensional polygons' example, we denote $2b_N > 0$ the edge length of a regular polygon with N edges and with all its vertices on $\mathbb{S}_1(0)$. We then have, for w_t uniformly, identically and independently distributed on $\mathbb{W}$, $\forall \epsilon > 0$: $q_w(\epsilon) = \frac{\epsilon(2b_N + O(\epsilon))}{S_N}$, where S_N is the area of the regular N-gon inscribed in the unit circle. We have $S_N \in [\frac{3\sqrt{3}}{4}, \pi]$. Notice that $\xi = \cos\left(\frac{\pi}{N}\right) = \sqrt{1 - b_N^2}$ and consider $q_w\left(\frac{a_1\delta\xi}{4}\right) = \frac{a_1\delta\xi(2\sqrt{1-\xi^2}+O(\xi\delta))}{4S_N}$ It follows that

$$q_w\left(\frac{a_1\delta\xi}{4}\right) = \begin{cases} O(\delta), & \text{if } \xi\delta \leq 2\sqrt{1-\xi^2} \\ O(\delta^2), & \text{otherwise.} \end{cases}$$

The example above indicates that the asymptotic estimation of q_w with respect to δ is not "continuous". Likewise, it explains why the theoretical error bound performs worse when ξ goes to 1. Namely, $\exists T_0 = T_0(N) \in \mathbb{N}$ such that

$$\delta \sim \begin{cases} \tilde{O}\left(\sqrt{\frac{n_x n_z}{T}}\right) & \text{if } T \leq T_0 \\ \tilde{O}\left(\frac{n_x n_z}{T}\right) & \text{if } T > T_0 \end{cases}$$

Meanwhile, T_0 is monotonically increasing with N, and is unbounded. That is, with $N \to +\infty$, one has this threshold $T_0 \to +\infty$. For general n_x, we have the following Corollary.

Corollary 2. *If* $\mathbb{W} = \{w \in \mathbb{R}^{n_x} : ||w||_2 \leq 1\}$ *and* $m \geq O(n_z + \log T - \log \epsilon)$*, then with probability no less than* $1 - 2\epsilon$*, one has diam*$(\Theta_T) \leq \tilde{O}\left(\left(\frac{n_x n_z}{T}\right)^{1/n_x}\right)$.

The proof of this Corollary is in the appendix.

Although theoretical derivations show that the error bound for $\mathbb{W}$ in the form of a 2-norm ball is much worse compared to the LSE's bound $O(1/\sqrt{T})$ for large n_x, yet in numerical experiments no significant distinction in these two cases' performance is observed (see Figure 4 for $n_x = 10$). This may suggest the poor dependence on T is a proof-artifact. Improving this convergence rate is our ongoing work.

IV. Proof Sketch of Theorem 1

The proof of Theorem 1 is inspired by the proof in [2]. Notice that the estimation error is denoted $\gamma := \hat{\theta} - \theta^*$. $\forall t \geq 0$, one has $x_{t+1} - \hat{\theta} z_t = w_t - (\hat{\theta} - \theta^*) z_t$. We define the error set by

$$\Gamma_T := \cap_{t=0}^{T-1} \{\gamma \ : \ w_t - \gamma z_t \in \mathbb{W}\}$$

Notice that Γ_T is attained by translating Θ_T by $-\theta^*$, which is an isometry. It follows that $\text{diam}(\Gamma_T) = \text{diam}(\Theta_T)$. In the rest of this paper, we will focus on $\text{diam}(\Gamma_T)$. Next, we define two events as follows:

Definition 4. *Define the event in which there exists an error estimator with large diameter by*

$$\mathcal{E}_1 := \left\{\exists\, \gamma \in \Gamma_T \ s.t.\ ||\gamma||_F \geq \frac{\delta}{2}\right\}$$

Define the event in which we have persistence excitation by

$$\mathcal{E}_2 := \left\{\frac{1}{m}\sum_{s=1}^{m} z_{km+s} z_{km+s}^\top \succeq a_1^2 I_{n_z},\ \forall\, 0 \leq k \leq \frac{T}{m} - 1\right\}$$

With the above two events defined, we can divide the event $\{\text{diam}(\Gamma_T) > \delta\}$ by the following.

$$\begin{aligned}\mathbb{P}(\text{diam}(\Gamma_T) > \delta) &\leq \mathbb{P}(\mathcal{E}_1) = \mathbb{P}\left((\mathcal{E}_1 \cap \mathcal{E}_2) \sqcup (\mathcal{E}_1 \cap \mathcal{E}_2^C)\right) \\ &\leq \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) + \mathbb{P}\left(\mathcal{E}_2^C\right)\end{aligned}$$

Notice that the bound on $\mathbb{P}\left(\mathcal{E}_2^C\right)$ follows directly from Lemma 1 in [2] as stated below.

Lemma 2 (Lemma 1 in [2])**.**

$$\mathbb{P}\left(\mathcal{E}_2^C\right) \leq \frac{T}{m}\tilde{O}(n_z^{5/2}) a_2^{n_z} \exp(-a_3 m)$$

The following Lemma proposes an upper bound for the other term.

Lemma 3.

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \leq \tilde{O}\left((n_x n_z)^{5/2}\right) a_4^{n_x n_z} [1 - q_w(\frac{a_1\delta\xi}{4})]^{\lceil (T-1)/m \rceil}$$

A complete proof of Lemma 3 can be found in the Appendix (part E). The basic idea of this proof is analogical to its ∞-norm constrained counterpart (Lemma 2 in [2]). The key is to design a sequence of events whose probability is easy to represent. Then we use these new events to cover each of the $\mathcal{E}_{1,i} \cap \mathcal{E}_2$ and derive the upper bound. Specially, instead of tracking the greatest ∞-norm of $\{w_t - \gamma z_t\}_{t=km+1}^{(k+1)m}$ and bounding it by $w_{\max}$, we now track the minimal projection denoted $v_{i,km+L_{i,k}}^\top(\gamma z_{km+L_{i,k}})$, and bound it below by $h(v_{i,km+L_{i,k}})$. Notice that in [2] the bound with $w_{\max}$ is in both directions due to the symmetry. In this paper's case, however, we only have one side of such inequalities. Without the introduction of the projection constant ξ, it is hard to derive the upper bound. Another noticeable difference is: we now compare the projection gap $v_{i,km+L_{i,k}} w_{km+L_{i,k}} - h(v_{i,km+L_{i,k}})$ with $\frac{a_1\delta\xi}{4}$. The projection constant ξ exists in the new derivation because of the properties of the normal vector v.

Consequently, Theorem 1 is directly implied by combining Lemma 2 and Lemma 3.

V. Numerical Experiments

In this section, we implement several numerical experiments of the set membership algorithm. We propose the descending curves of the uncertainty set's diameter as we feed more data to the model. We also present a comparison between the SME diameter and that of a least square estimator's 95%-confidence region.

To assess the performance of SME, we consider a system of $n_x = 10$-dimensions. The ground-truth parameters A^*, B^* are randomly generated and rescaled to be stable. We consider w_t uniformly sampled from the noise set $\mathbb{W}$ in two forms: $\mathbb{W}_1 = \{w \in \mathbb{R}^{10} : \max_{i \in [n_x]} \left\{ \frac{1}{a_i} |w_i| \right\} \leq 1\}$, a weighted ∞-norm ball/hyper-rectangle, and $\mathbb{W}_2 = \{w \in \mathbb{R}^{10} : ||w||_2 \leq 1\}$, a 2-norm ball. For each of $\mathbb{W}_1, \mathbb{W}_2$, we generate a sequence of data $\{x_t\}_{t=0}^{T}$ with i.i.d. controller $u \in \mathbb{R}^{10}$ that follows the same distribution as w_t does. For each case, we plot the descending curve of diam(Θ_T) along with the estimated error bound for the LSE 95%-confidence region's diameter. We adapt to the error bound proposed by [45] for the LSE 95%-confidence region. For the upper bound $S \geq ||\theta^*||_F$, we take the exact lower bound (i.e. we set $S = ||\theta^*||_F$). And for the proxy variance, we refer to [46] and use the lower bound $L \geq \text{Var}(w_i)$. For $\mathbb{W}_1$, $L = \sqrt{\frac{10}{12}}$; for $\mathbb{W}_2$, $L = \frac{\max_{i \in [n_x]} \{a_i\}}{\sqrt{3}}$. The $(1 - \delta)$ value is set to be 0.95 for a bound of the 95%-confidence region. It is worth mentioning that our choices of parameters are carefully chosen to improve the performance of the LSE bounds, for example, the LSE's confidence bound increases with S and L so we choose the smallest possible values for S and L. We also tune the parameters of λ for better performance of LSE. Meanwhile, we apply entry-wise outer approximation for the set membership estimation for ease of computation. Therefore, the real SME diameters can be less than the plotted values.

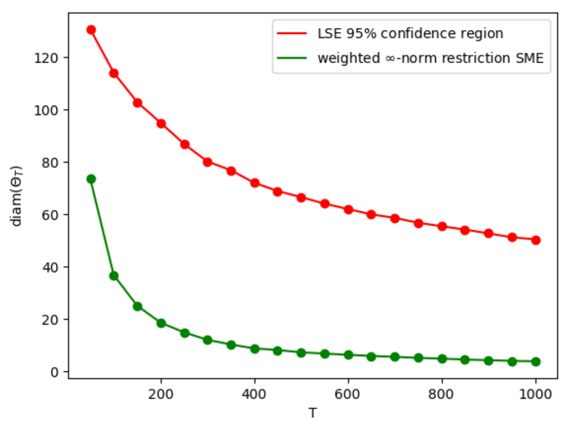

Fig. 3: Convergence rates of uncertainty sets of SME and LSE's confidence bounds when $\mathbb{W} = \mathbb{W}_1$ is a weighted l_∞ ball (hyper-rectangle).

Figure 3 plots the diameter of the uncertainty sets generated by SME and LSE's 95% confidence bounds when the disturbance w_t is uniformly distributed on $\mathbb{W}_1$. It can be observed that SME significantly outperforms LSE's confidence bounds, which is consistent with our theoretical insights in Corollary 1.

Figure 4 plots the diameter of the uncertainty sets when w_t is uniformly distributed on $\mathbb{W}_2$. Notice that the gaps between SME and LSE's confidence bounds are much smaller. However, the rates of convergence are rather similar. This is quite interesting because our theoretical convergence rate for the l_2 ball is much worse than the theoretical bound of LSE. This mismatch between theory and simulation motivates more efforts on improving the convergence rates of SME, which is our ongoing work.

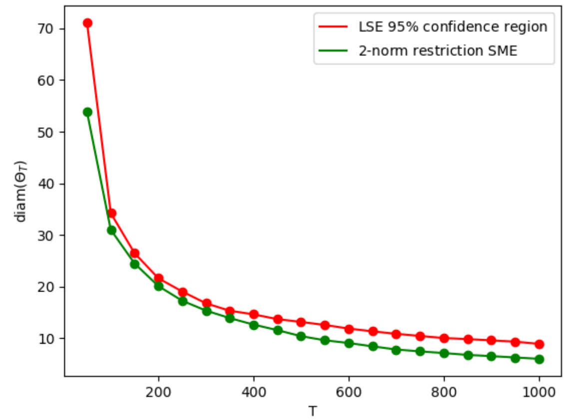

Fig. 4: Convergence rates of uncertainty sets of SME and LSE's confidence bounds when $\mathbb{W} = \mathbb{W}_2$ is a 2-norm ball.

More simulation can be found in our supplementary [47].

VI. Conclusions and Future Directions

In this paper, we propose a non-asymptotic analysis for the diameter of the set membership estimator using data generated from an unknown linear system whose noise is i.i.d. generated from a convex and compact set $\mathbb{W}$. We also provide error bound estimations on the SME and numerical experiments comparing the performance of SME with that of LSE. The bounds are based on our current best knowledge in this field and we expect to improve this project in the future. We are interested in the following directions: 1) improving the convergence rate on 2-norm ball; 2) seeking for a non-stochastic representation for the error bound; 3) analyzing the computational complexity of set membership; 4) investigating SME's performance when the tight-bound assumption is absent; 5) assessing set membership's performance with robust adaptive model predictive control/distributionally robust control; 6) estimating the volume of the SME instead of diameter; 7) discovering the fundamental limit of this algorithm; 8) adapting the theory to non-linear systems, 9) comparing with the credible regions of Bayesian approaches, such as Gaussian processes and Thompson sampling, etc.

Appendix

A. Proof of Proposition 1

a) Consider any $s \in \mathbb{S}_1(0)$. $\forall w^0 \in \arg\min_{\tilde{w} \in \mathbb{W}} s^\top \tilde{w}$ and $\forall w \in \mathbb{W}$, one has $s^\top (w - w^0) = s^\top w - \min_{\tilde{w} \in \mathbb{W}} s^\top \tilde{w} \geq 0$. Therefore, $\mathbb{S}_1(0) \subseteq V$. The other side of inclusion is trivial.

b) To show this, we prove by contradiction. Suppose there exists some $w^* \in \cap_{v \in \mathbb{S}_1(0)}$ such that $w^* \in \mathbb{W}^{\complement}$. By the strong convexity of the 2-norm, $\exists! w^0 \in \partial \mathbb{W}$ such that $||w^* - w^0||_2 = \min_{\tilde{w} \in \partial \mathbb{W}} ||w^* - \tilde{w}||_2 > 0$. Consider

$v^* := -\frac{w^*-w^0}{||w^*-w^0||_2}$. Since $\mathbb{W}$ is convex. By the supporting hyperplane theorem, $w^0 \in \arg\max_{\tilde{w}\in\mathbb{W}}(v^*)^\top \tilde{w}$. It follows that $(v^*)^\top w^* = (v^*)^\top w^0 - ||w^* - w^0||_2 < \min_{\tilde{w}\in\mathbb{W}}(v^*)^\top \tilde{w}$ and we have a contradiction against the assumption that $w^* \in \cap_{v\in\mathbb{S}_1(0)} H(v)$. The other side of implication is trivial.

B. Proof of Lemma 1

Since $\bar{\tilde{V}}$ is compact and the inner product map $v \to v^\top c$ is continuous, then $\max_{v\in\bar{\tilde{V}}} v^\top c$ is well-defined. Next we proof by contradiction: suppose that $\inf_{c\in\mathbb{S}_1(0)} \max_{v\in\bar{\tilde{V}}} v^\top c < 0$. Namely, $\forall\, \epsilon > 0$, $\forall\, n \in \mathbb{N}$, $\exists\, c^{(n)} \in \mathbb{S}_1(0)$ s.t. $\max_{v\in\bar{\tilde{V}}} v^\top c^{(n)} \leq \frac{\epsilon}{n}$. Consider $w_0 \in \mathbb{W}$ such that $\forall\, v \in \tilde{V}$, $v^\top w_0 > h(v) + \epsilon$, we then have $v^\top(w_0 - nc^{(n)}) \geq h(v)$. Notice that $c^{(n)} \in \mathbb{S}_1(0)$. Then $\{w_0 - nc^{(n)}\}_{n\geq 0} \subseteq \mathbb{W}$. This cannot hold since we assumed that $\mathbb{W}$ is compact. □

C. Proofs of Corollaries 1 & 2

a) Corollary 2: In this case we have $\tilde{V} = \mathbb{S}_1(0)$, and it follows that $\xi = 1$. We first claim that Term $1 \leq \epsilon$. To show this, notice that

$$\begin{aligned} m &\geq O(n_z + \log T - \log \epsilon) \\ &= \frac{1}{a_3}\left[O((\log a_2)n_z + \frac{5}{2}\log n_z + O(\log T) - O(\log \epsilon)\right] \end{aligned}$$

It follows that $\exp(-a_3 m) \leq \frac{a_2^{-n_z} n_z^{-5/2}\epsilon}{T}$ And therefore, Term $1 \leq \epsilon/m \leq \epsilon$ Next we let Term $2 = \epsilon$ Denoting the volume of a spherical cap with height ϵ in an n_x-dimensional unit ball to be $V_\epsilon^{n_x}$, we notice that $V_\epsilon^{n_x} \geq \frac{\pi^{n_x/2}\epsilon^{n_x}}{2\Gamma(n_x/2+1)}$. For a uniform distribution, the probability that w_t falls in this cap is $\frac{V_\epsilon^{n_x}}{V_{n_x}}$ where V_{n_x} is the volume of the n_x-dimensional unit ball. And $V_n = \frac{\pi^{\frac{n_x}{2}}}{\Gamma(\frac{n_x}{2}+1)}$. We then have: $\frac{V_\epsilon^{n_x}}{V_{n_x}} \geq O\left(\epsilon^{n_x}\right)$ We can then take $q_w(\epsilon) = O\left(\epsilon^{n_x}\right)$. It follows that

$$\begin{aligned} \delta^{n_x} &= O((\frac{4}{a_1})^{n_x})\left\{1 - [\epsilon\tilde{O}((n_x n_z)^{5/2})a_4^{-n_x n_z}]^{m/T}\right\} \\ &\leq O((\frac{4}{a_1})^{n_x}\frac{m}{T}\tilde{O}(n_x n_z) \\ &= \tilde{O}\left(\left(\frac{4}{a_1}\right)^{n_x}\cdot\frac{n_x n_z}{T}\right) \end{aligned}$$

The inequality comes from $\forall\, x,\ x - 1 \geq \log x$. Therefore, $\delta \leq \tilde{O}\left(\left(\frac{n_x n_z}{T}\right)^{1/n_x}\right)$.

b) Corollary 1: The proof is similar to that for Corollary 2. Since Term 1 in Theorem 4 is the same as that in Theorem 1, we can show that Term $1 \leq \epsilon$. Now, letting Term $2 = \epsilon$, one has $O\left(\frac{a_1\delta\xi}{4}\right) = 1 - \left[\epsilon\tilde{O}\left((n_x n_z)^{-5/2}\right)a_4^{-n_x n_z}\right]^{m/T}$. Using the $x - 1 \geq \log x$ trick again, we have $\delta \leq \tilde{O}\left(\frac{n_x n_z}{T\xi}\right)$.

D. Proofs of Examples 1 & 2

a) Example 1: Alternatively, we can write $\mathbb{W} = [-a_1, a_1] \times \cdots \times [-a_{n_x}, a_{n_x}]$. A valid set of normal vectors V can be $V = \{\pm e_i\}$, where e_i is the i-th vector in the canonical basis of $\mathbb{R}^n$. Without loss of generality, we consider $w_t \in \cdot_\epsilon = [a_1 - \epsilon, a_1] \times [-a_2, a_2] \times [-a_{n_x}, a_{n_x}]$. Namely, when w_t is sampled in a thin slice near a certain facet. Then $\mathbb{P}(w_t \in \Delta_\epsilon) = \frac{\text{Volume of } \Delta_\epsilon}{\text{Volume of } \mathbb{W}} = \frac{2^{n_x-1}\epsilon\prod_{i=2}^{n_x} a_i}{2^{n_x}\prod_{i=1}^{n_x} a_i} = \frac{\epsilon}{2a_1} = O(\epsilon)$.

b) Example 2: The volume of the ball $\hat{\mathbb{W}} = \{w \in \mathbb{R}^{n_x} : \sum_{i=1}^{n_x}\frac{1}{a_i}|w_i| \leq 1 - \epsilon\}$ is accordingly bounded by $[(1-\epsilon)^{n_x}\frac{(2a)^{n_x}}{n_x!}, (1-\epsilon)^{n_x}\frac{(2A)^{n_x}}{n_x!}]$. It follows that $2^{n_x} q_w(\epsilon) \approx \frac{\text{Volume of } \mathbb{W}\backslash\hat{\mathbb{W}}}{\text{Volume of } \mathbb{W}} \in \left[[1-(1-\epsilon)^{n_x}]\left(\frac{a}{A}\right)^{n_x}, [1-(1-\epsilon)^{n_x}]\left(\frac{A}{a}\right)^{n_x}\right]$. It follows that $\left(\frac{a}{2A}\right)^{n_x} \leq \frac{q_w(\epsilon)}{1-(1-\epsilon)^{n_x}} \leq \left(\frac{A}{2a}\right)^{n_x}$ Notice that $1-(1-\epsilon)^{n_x} \sim O(\epsilon)$. Thus, $q_w(\epsilon)$ is also $O(\epsilon)$.

E. Complete Proof of Lemma 3

For the $(n_x \times n_z)$-dimensional unit Frobenius-norm sphere $\mathbb{S}_1^F(0) := \{\gamma \in \mathbb{R}^{n_x\times n_z} : ||\gamma||_F = 1\}$, we consider covering it with smaller balls with radius $\epsilon_\gamma = \frac{1}{a_4}$, and denote the corresponding ϵ_γ-net to be $\mathcal{M} := \{\gamma_i\}_{i=1}^{v_\gamma}$. Here v_γ is the number of small ϵ_γ-balls required to cover the sphere so that $\forall\,\gamma \in \mathbb{S}_1^F(0)$, $\exists\gamma_i \in \mathcal{M}$ such that $||\gamma_i - \gamma||_F \leq 2\epsilon_\gamma$. For the theory of covering number, we refer the readers to [48], [49] and Appendix D.1 in [2]. On the other hand, we define the stopping time $L_{i,k} := \min\{m+1, \min\{l \geq 1 : ||\gamma_i z_{km+l}||_2 \geq a_1\}\}$. Since z_t is $\mathcal{F}_t$-measurable and $\{L_{i,k} = \ell\} \in \mathcal{F}_{km+\ell}$, then $v_{i,t}$ is $\mathcal{F}_t$-measurable, and $L_{i,k}$ a stopping time with respect to the original filtration. $\forall\, t \geq 0$, define the adapted process $\{v_{i,t}\}_{t\geq 0}$: $v_{i,t} := \arg\max_{v\in cl(\tilde{V})} v^\top(\gamma_i z_t)$. Here $cl(\tilde{V})$ denotes the closure of $\tilde{V}$. Notice that since $cl(\tilde{V})$ is compact, then the maximum is well-defined. However, the maximizer may not be unique. We can take arbitrary one of them. Accordingly, we define $\mathcal{E}_{1,i} = \left\{\exists\,\gamma \in \Gamma_T :\ v_{i,km+L_{i,k}}^\top(\gamma z_{km+L_{i,k}}) \geq \frac{a_1\delta\xi}{4},\ \forall\, k\right\}$. We want to cover the event $\mathcal{E}_1 \cap \mathcal{E}_2$ with these events. We have the following technical Lemma:

Lemma 4.

$$\mathbb{P}\left(\mathcal{E}_1 \cap \mathcal{E}_2\right) \leq \sum_{i=1}^{v_\gamma}\mathbb{P}\left(\mathcal{E}_{1,i} \cap \mathcal{E}_2\right)$$

Proof of Lemma 4. By $\mathcal{E}_2$, choosing arbitrary $\gamma_i \in \mathcal{M}$, we have: $\frac{1}{m}\sum_{s=1}^m ||\gamma_i z_{km+s}||_2^2 \geq a_1^2$. By pigeonhole principle: $\max_{s\in\{1,\cdots,m\}} ||\gamma_i z_{km+s}||_2 \geq a_1$. It follows that $L_{i,k} \leq m$ and $||\gamma_i z_{km+L_{i,k}}||_2 \geq a_1$.

By the ball-covering theory, there $\exists\gamma_i \in \mathcal{M}$ such that $||\gamma - \gamma_i||_F \leq 2\epsilon_\gamma$. Intuitively, we can check the projection of γz on $v_{i,km+L_{i,k}}$. Namely, for any $\gamma \in \bar{\mathbb{S}}_1(\mathbf{0})$:

$$\begin{aligned} v_{i,km+L_{i,k}}^\top(\gamma z_{km+L_{i,k}}) &= v_{i,km+L_{i,k}}^\top(\gamma_i z_{km+L_{i,k}}) \\ &\quad - v_{i,km+L_{i,k}}^\top((\gamma_i - \gamma)z_{km+L_{i,k}}) \\ &\geq a_1\xi - ||(\gamma_i - \gamma)z_{km+L_{i,k}}||_2 \\ &\geq a_1\xi - \underbrace{||\gamma_i - \gamma||_2}_{\leq 2\epsilon_\gamma}\underbrace{||z_{km+L_{i,k}}||_2}_{\leq b_z} \\ &\geq a_1\xi - 2\epsilon_\gamma b_z \geq \frac{a_1\xi}{2} \end{aligned}$$

Notice that $\forall\gamma \in \mathbb{R}^{n_x\times n_z}$, $v_{i,km+L_{i,k}}^\top(\gamma z_{km+L_{i,k}}) \geq \frac{a_1||\gamma||_F\xi}{2}$. Given $\mathcal{E}_1$, we have: $v_{i,km+L_{i,k}}^\top(\gamma z_{km+L_{i,k}}) \geq \frac{a_1\xi}{2}\cdot||\gamma||_F > \frac{a_1\delta\xi}{4}$. That is, $\mathcal{E}_1\cap\mathcal{E}_2 \subseteq \cup_{i=1}^{v_\gamma}(\mathcal{E}_{1,i}\cap\mathcal{E}_2)$, from which we can deduce Lemma 4. □

However, it is still difficult to deal with $\mathcal{E}_{1,i}$. We will need a further step of covering. We know that $\gamma := \hat{\theta}-\theta^*$. Thus, $w_t - \gamma z_t = x_{t+1} - \hat{\theta}z_t \in \mathbb{W}$. It follows that:

$$\forall\, v\in\tilde{V},\ v^\top(w_t-\gamma z_t) \geq h(v)$$

This is also true at the stopping time $L_{i,k}$. That is:

$$\begin{aligned}&v_{i,km+L_{i,k}}^\top w_{km+L_{i,k}} - h(v_{i,km+L_{i,k}})\\ \geq& v_{i,km+L_{i,k}}^\top(\gamma z_{km+L_{i,k}})\end{aligned}$$

$\forall\, k\geq 0$, define

$$\begin{aligned}G_{i,k} =\{&\left(v_{i,km+L_{i,k}}^\top w_{km+L_{i,k}} - h(v_{i,km+L_{i,k}}) \geq \frac{a_1\delta\xi}{4}\right)\\ &\wedge\left(\frac{1}{m}\sum_{s=1}^m z_{km+s}z_{km+s}^\top \succeq a_1^2 I_{n_z}\right)\}\end{aligned}$$

For $\{G_{i,k}\}_k$, we have the following Lemma:

Lemma 5.

$$\mathbb{P}(\mathcal{E}_{1,i}\cap\mathcal{E}_2) = \mathbb{P}(G_{i,0})\prod_{t=1}^{\lceil T/m\rceil-1}\mathbb{P}\left(G_{i,t}\mid\cap_{k=0}^{t-1}G_{i,k}\right)$$

Proof of Lemma 5. By $\mathcal{E}_{1,i}\cap\mathcal{E}_2$ holds, $\exists\,\gamma\in\Gamma_T$ such that

$$v_{i,km+L_{i,k}}^\top(\gamma z_{km+L_{i,k}}) \geq \frac{a_1\delta\xi}{4}$$

which implies:

$$v_{i,km+L_{i,k}}^\top w_{km+L_{i,k}} - h(v_{i,km+L_{i,k}}) \geq \frac{a_1\delta\xi}{4}$$

which is $G_{i,k}$ exactly. Therefore, $\mathcal{E}_{1,i}\cap\mathcal{E}_2 \subseteq \cap_{k=1}^{T/m}G_{i,k}$. Consequently,

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{1,i}\cap\mathcal{E}_2) &\leq \mathbb{P}\left(\cap_{k=1}^{\lceil T/m\rceil-1}G_{i,k}\right)\\ &= \mathbb{P}(G_{i,0})\prod_{t=1}^{\lceil T/m\rceil-1}\mathbb{P}\left(G_{i,t}\mid\cap_{k=0}^{t-1}G_{i,k}\right)\end{aligned}$$

□

With Lemma 5, we can proceed to find a bound for the probability of $\cap G_{i,k}$. For the sake of convenience, we now denote the events:

$$\begin{aligned}&\mathcal{A}(i,k,m,L_{i,k}) := \{(L_{i,k}\leq m)\wedge\\ &\left(v_{i,km+L_{i,k}}^\top w_{km+L_{i,k}} - h(v_{i,km+L_{i,k}}) \geq \frac{a_1\delta\xi}{4}\right)\}\end{aligned}$$

Consider an arbitrary factor of the consecutive product

$$\begin{aligned}&\mathbb{P}\left(G_{i,k}\mid\cap_{s=0}^{k-1}G_{i,s}\right) = \mathbb{P}\left(\mathcal{A}(i,k,m,L_{i,k})\mid\cap_{s=0}^{k-1}G_{i,s}\right)\\ &=\sum_{l=1}^m\mathbb{P}\left(\mathcal{A}(i,k,m,l),\ L_{i,k}=l\mid\cap_{s=0}^{k-1}G_{i,s}\right)\\ &=\sum_{l=1}^m\left[\mathbb{P}\left(\mathcal{A}(i,k,m,l)\mid\cap_{s=0}^{k-1}G_{i,s},L_{i,k}=l\right)\right.\\ &\quad\left.\mathbb{P}\left(L_{i,k}=l\mid\cap_{s=0}^{k-1}G_{i,s}\right)\right]\end{aligned}$$

The second equation comes from the law of total probability, the third is deduced by the Bayes' law. Let $\{v_m\}_{m=0}^{km+l}$ be a sequence of realizations of w_t. Notice that $\forall\, l\in\{1,\cdots,m\}$. Taking integrals on the realizations, we have:

$$\begin{aligned}&\mathbb{P}\left(\mathcal{A}(i,k,m,l)\mid\cap_{s=0}^{k-1}G_{i,s},L_{i,k}=l\right)\\ &=\int_{v_{0:km+l}}\mathbb{P}\left(\mathcal{A}(i,k,m,l)\mid\cap_{s=0}^{k-1}G_{i,s},L_{i,k}=l\right)dv_{0:km+l}\\ &=\int_{v_{0:km+l}}\left[\mathbb{P}\left(\mathcal{A}(i,k,m,l)\mid w_{0:km+l}=v_{0:km+l}\right)\right.\\ &\quad\left.\mathbb{P}\left(w_{0:km+l}=v_{0:km+l}\mid L_{i,k}=l,\ \cap_{s=0}^{k-1}G_{i,s}\right)\right]dv_{0:km+l}\\ &\leq\sup\left[\mathbb{P}\left(\mathcal{A}(i,k,m,l)\mid w_{0:km+l}=v_{0:km+l}\right)\right]\end{aligned}$$

Notice that by the assumptions:

$$v_{i,km+l}^\top w_{km+l} - h_{i,km+l} \geq \frac{a_1\delta\xi}{4}$$

And it follows that:

$$\mathbb{P}\left(\mathcal{A}(i,k,m,l)\mid v_{0:km+l}\right) \leq 1-q_w\left(\frac{a_1\delta\xi}{4}\right)$$

which leads to

$$\mathbb{P}\left(G_{i,k}\mid\cap_{l=0}^{k-1}G_{i,l}\right)\leq 1-q_w\left(\frac{a_1\delta\xi}{4}\right)$$

With an upper bound found for all components, we have:

$$\begin{aligned}\mathbb{P}(\mathcal{E}_1\cap\mathcal{E}_2) &= \sum_{i=1}^{v_\gamma}\mathbb{P}(\mathcal{E}_{1,i}\cap\mathcal{E}_2) \leq \sum_{i=1}^{v_\gamma}\mathbb{P}\left(\mathbb{P}(\cap_{k=0}^{T/m-1}G_{i,k})\right)\\ &= \mathbb{P}(G_{i,0})\prod_{t=1}^{\lceil T/m\rceil-1}\mathbb{P}\left(G_{i,t}\mid\cap_{k=0}^{t-1}G_{i,k}\right)\\ &\leq\sum_{i=1}^{v_\gamma}\left[1-q_w\left(\frac{a_1\delta\xi}{4}\right)\right]^{\lceil(T-1)/m\rceil}\\ &\leq\tilde{O}((n_xn_z)^{5/2})a_4^{n_xn_z}\left[1-q_w\left(\frac{a_1\delta\xi}{4}\right)\right]^{\lceil(T-1)/m\rceil}\end{aligned}$$

This finishes the proof of Lemma 3.

References

[1] X. Lu and M. Cannon, "Robust adaptive model predictive control with persistent excitation conditions," *Automatica*, vol. 152, p. 110959, 2023.

[2] Y. Li, J. Yu, L. Conger, and A. Wierman, "Learning the uncertainty sets for control dynamics via set membership: A non-asymptotic analysis," *arXiv preprint arXiv:2309.14648*, 2023.

[3] A. Sarker, P. Fisher, J. E. Gaudio, and A. M. Annaswamy, "Accurate parameter estimation for safety-critical systems with unmodeled dynamics," *Artificial Intelligence*, p. 103857, 2023.

[4] Y. Li, T. Zhang, S. Das, J. Shamma, and N. Li, "Non-asymptotic system identification for linear systems with nonlinear policies," *arXiv preprint arXiv:2306.10369*, 2023.

[5] M. Simchowitz and D. Foster, "Naive exploration is optimal for online lqr," in *International Conference on Machine Learning*. PMLR, 2020, pp. 8937–8948.

[6] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Conference On Learning Theory*. PMLR, 2018, pp. 439–473.

[7] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," *Advances in Neural Information Processing Systems*, vol. 31, pp. 4188–4197, 2018.

[8] Y. Abbasi-Yadkori and C. Szepesvári, "Regret bounds for the adaptive control of linear quadratic systems," in *Proceedings of the 24th Annual Conference on Learning Theory*. JMLR Workshop and Conference Proceedings, 2011, pp. 1–26.

[9] K. P. Wabersich, A. J. Taylor, J. J. Choi, K. Sreenath, C. J. Tomlin, A. D. Ames, and M. N. Zeilinger, "Data-driven safety filters: Hamilton-jacobi reachability, control barrier functions, and predictive methods for uncertain systems," *IEEE Control Systems Magazine*, vol. 43, no. 5, pp. 137–177, 2023.

[10] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, and A. P. Schoellig, "Safe learning in robotics: From learning-based control to safe reinforcement learning," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 411–444, 2022.

[11] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, "Learning-based model predictive control: Toward safe learning in control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 269–296, 2020.

[12] M. Petrik and R. H. Russel, "Beyond confidence regions: Tight bayesian ambiguity sets for robust mdps," *Advances in neural information processing systems*, vol. 32, 2019.

[13] S. Dean, S. Tu, N. Matni, and B. Recht, "Safely learning to control the constrained linear quadratic regulator," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 5582–5588.

[14] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *Foundations of Computational Mathematics*, pp. 1–47, 2019.

[15] M. Lauricella and L. Fagiano, "Set membership identification of linear systems with guaranteed simulation accuracy," *IEEE Transactions on Automatic Control*, vol. 65, no. 12, pp. 5189–5204, 2020.

[16] M. Lorenzen, M. Cannon, and F. Allgöwer, "Robust mpc with recursive model update," *Automatica*, vol. 103, pp. 461–471, 2019.

[17] M. M. Livstone and M. A. Dahleh, "Asymptotic properties of set membership identification algorithms," *Systems & control letters*, vol. 27, no. 3, pp. 145–155, 1996.

[18] E. Fogel and Y.-F. Huang, "On the value of information in system identification—bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229–238, 1982.

[19] D. P. Bertsekas, "Control of uncertain systems with a set-membership description of the uncertainty." Ph.D. dissertation, Massachusetts Institute of Technology, 1971.

[20] J. Yu, D. Ho, and A. Wierman, "Online adversarial stabilization of unknown networked systems," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 7, no. 1, pp. 1–43, 2023.

[21] B. T. Lopez, J.-J. E. Slotine, and J. P. How, "Robust adaptive control barrier functions: An adaptive and data-driven approach to safety," *IEEE Control Systems Letters*, vol. 5, no. 3, pp. 1031–1036, 2020.

[22] B. T. Lopez and J.-J. E. Slotine, "Unmatched control barrier functions: Certainty equivalence adaptive safety," in *2023 American Control Conference (ACC)*. IEEE, 2023, pp. 3662–3668.

[23] M. Bujarbaruah, X. Zhang, M. Tanaskovic, and F. Borrelli, "Adaptive stochastic mpc under time-varying uncertainty," *IEEE Transactions on Automatic Control*, vol. 66, no. 6, pp. 2840–2845, 2020.

[24] K. Zhang, Q. Sun, and Y. Shi, "Trajectory tracking control of autonomous ground vehicles using adaptive learning mpc," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 12, pp. 5554–5564, 2021.

[25] A. Parsi, A. Iannelli, M. Yin, M. Khosravi, and R. S. Smith, "Robust adaptive model predictive control with worst-case cost," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 4222–4227, 2020.

[26] A. Parsi, A. Iannelli, and R. S. Smith, "Active exploration in adaptive model predictive control," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 6186–6191.

[27] A. Sasfi, M. N. Zeilinger, and J. Köhler, "Robust adaptive mpc using control contraction metrics," *arXiv preprint arXiv:2209.11713*, 2022.

[28] E.-W. Bai, H. Cho, and R. Tempo, "Convergence properties of the membership set," *Automatica*, vol. 34, no. 10, pp. 1245–1249, 1998.

[29] H. Akçay, "The size of the membership-set in a probabilistic framework," *Automatica*, vol. 40, no. 2, pp. 253–260, 2004.

[30] W. Kitamura, Y. Fujisaki, and E.-W. Bai, "The size of the membership set in the presence of disturbance and parameter uncertainty," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 5698–5703.

[31] E.-W. Bai, R. Tempo, and H. Cho, "Membership set estimators: size, optimal inputs, complexity and relations with least squares," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 42, no. 5, pp. 266–277, 1995.

[32] J. Eising, S. Liu, S. Martínez, and J. Cortés, "Using data informativity for online stabilization of unknown switched linear systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 8–13.

[33] T. Sarkar, A. Rakhlin, and M. Dahleh, "Nonparametric system identification of stochastic switched linear systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 3623–3628.

[34] X. Lu, M. Cannon, and D. Koksal-Rivet, "Robust adaptive model predictive control: Performance and parameter estimation," *International Journal of Robust and Nonlinear Control*, 2019.

[35] J. Feng, Y. Shi, G. Qu, S. H. Low, A. Anandkumar, and A. Wierman, "Stability constrained reinforcement learning for decentralized real-time voltage control," *IEEE Transactions on Control of Network Systems*, 2023.

[36] V. Le, C. Stoica, D. Dumur, T. Alamo, and E. Camacho, "Robust tube-based constrained predictive control via zonotopic set-membership estimation," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 4580–4585.

[37] S. Chen, N. Matni, M. Morari, and V. M. Preciado, "System level synthesis-based robust model predictive control through convex inner approximation," *arXiv preprint arXiv:2111.05509*, 2021.

[38] B. Pal and B. Chaudhuri, *Robust control in power systems*. Springer Science & Business Media, 2006.

[39] H. Kress-Gazit, M. Lahijanian, and V. Raman, "Synthesis for robots: Guarantees and feedback for robot behavior," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 211–236, 2018.

[40] A. Tzes and K. Le, "Adaptive performance/robust control design using orthotopic set membership identification for low-dimensional systems," *International journal of adaptive control and signal processing*, vol. 11, no. 8, pp. 649–663, 1997.

[41] A. B. Kurzhanski and P. Varaiya, "Optimization of output feedback control under set-membership uncertainty," *Journal of optimization theory and applications*, vol. 151, pp. 11–32, 2011.

[42] G. Belforte, B. Bona, and V. Cerone, "Parameter estimation algorithms for a set-membership description of uncertainty," *Automatica*, vol. 26, no. 5, pp. 887–898, 1990.

[43] J. R. Deller, "Set membership identification in digital signal processing," *IEEE ASSP Magazine*, vol. 6, no. 4, pp. 4–20, 1989.

[44] D. Ho and J. C. Doyle, "Scalable robust adaptive control from the system level perspective," in *2019 American Control Conference (ACC)*, 2019, pp. 3683–3688.

[45] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Online least squares estimation with self-normalized processes: An application to bandit problems," *arXiv preprint arXiv:1102.2670*, 2011.

[46] J. Arbel, O. Marchal, and H. D. Nguyen, "On strict sub-gaussianity, optimal proxy variance and symmetry for bounded random variables," *ESAIM: Probability and Statistics*, vol. 24, pp. 39–55, 2020.

[47] H. Xu and Y. Li, *On the Convergence Rates of Set Membership Estimation of Linear Systems with Disturbances Bounded by General Convex Sets (online supplementary file)*, 2024, https://yingying.li/files/SMEconvex.pdf.

[48] C. A. Rogers, "Covering a sphere with spheres," *Mathematika*, vol. 10, no. 2, p. 157–164, 1963.

[49] J.-L. Verger-Gaugry, "Covering a ball with smaller equal balls in r n," *Discrete Computational Geometry - DCG*, vol. 33, pp. 143–155, 06 2005.