Proceedings of Machine Learning Research vol 242:286–298, 2024 6th Annual Conference on Learning for Dynamics and Control

Uncertainty Quantification of Set-Membership Estimation in Control and Perception: Revisiting the Minimum Enclosing Ellipsoid

Yukai Tang TANGYK20@MAILS.TSINGHUA.EDU.CN
Department of Electrical Engineering, Tsinghua University, China

Jean-Bernard Lasserre LASSERRE@LAAS.FR
LAAS-CNRS and Toulouse School of Economics (TSE), Toulouse, France

Heng Yang HANKYANG@SEAS.HARVARD.EDU
School of Engineering and Applied Sciences, Harvard University, USA

Editors: A. Abate, M. Cannon, K. Margellos, A. Papachristodoulou

Abstract

Set-membership estimation (SME) outputs a set estimator that guarantees to cover the groundtruth. Such sets are, however, defined by (many) abstract (and potentially nonconvex) constraints and therefore difficult to manipulate. We present tractable algorithms to compute simple and tight over-approximations of SME in the form of *minimum enclosing ellipsoids* (MEE). We first introduce the *hierarchy of enclosing ellipsoids* proposed by Nie and Demmel (2005), based on *sums-of-squares relaxations*, that asymptotically converge to the MEE of a basic semialgebraic set. This framework, however, struggles in modern control and perception problems due to computational challenges. We contribute three computational enhancements to make this framework practical, namely constraints pruning, generalized relaxed Chebyshev center, and handling non-Euclidean geometry. We showcase numerical examples on system identification and object pose estimation.

Keywords: Set-Membership Estimation, Minimum Enclosing Ellipsoid, Semidefinite Relaxations

1. Introduction

Model estimation and learning from measurements is a central task in numerous disciplines (Stengel, 1994; Barfoot, 2017). Let $\theta \in \Theta \subseteq \mathbb{R}^n$ be an unknown model, and $\mathbb{Z} = \{z_i\}_{i=1}^N \in \mathcal{Z}^N$ be a set of N measurements such that, if z_i is a perfect (noise-free) measurement, it holds $r(\theta, z_i) = 0, i = 1, \ldots, N$, with $r : \Theta \times \mathcal{Z} \to \mathbb{R}^m$ a given *residual* function that is typically designed from first principles describing how z_i is generated from θ (see Examples 1-2 below).

Maximum Likelihood Estimation. The most popular approach to estimate θ from $\mathbb{Z}$ is *maximum likelihood estimation* (MLE). When the measurements are noisy, the residual is adjusted to

$$r(\theta, z_i) = \epsilon_i, \quad i = 1, \ldots, N, \tag{1}$$

with $\epsilon_i \in \mathbb{R}^m$ some small noise. The most common distributional assumption is that $\epsilon_i \sim \mathcal{N}(0, \Sigma)$ follows a Gaussian distribution, leading to the MLE estimator that is the solution of a *nonlinear least squares* problem (Nocedal and Wright, 1999; Pineda et al., 2022)[1]

$$\theta^\star_{\text{MLE}} \in \arg\min_{\theta \in \Theta} \sum_{i=1}^N \|r(\theta, z_i)\|^2_{\Sigma^{-1}}. \tag{2}$$

When the measurement set $\mathbb{Z}$ contains *outliers*, the least squares objective in (2) can be replaced by a *robust loss* (Huber, 2004; Antonante et al., 2021). Despite significant interests and progress in the

1. Or maximum *a posteriori* estimation (MAP) if there is prior knowledge about the distribution of θ.

literature, two shortcomings of the MLE framework exist. On one hand, the Gaussian assumption is questionable. In fact, in Tang et al. (2023) we show noises generated by modern neural networks on a computer vision example deviate far from a Gaussian distribution.[2] On the other hand, in safety-critical applications, *provably correct uncertainty quantification* of a given estimator is desired (*e.g.,* how close is $\theta^\star_{\text{MLE}}$ to the groundtruth). The uncertainty of the MLE estimator (2), however, is nontrivial to quantify due to the potential nonlinearity in the residual function r.

Set-Membership Estimation. An alternative framework, known as *set-membership estimation* (SME, or unknown-but-bounded estimation) (Milanese and Vicino, 1991), seeks to resolve the two shortcomings of MLE. Instead of making a distributional assumption on the measurement noise (1), SME only requires the noise to be bounded

$$\|\epsilon_i\| = \|r(\theta, z_i)\| \leq \beta_i, \quad i = 1, \dots, N, \tag{3}$$

where $\|\cdot\|$ indicates the ℓ_2 vector norm (one can choose ℓ_1 or ℓ_∞ norm and our algorithm would still apply). Characterizing the bound of the noise is easier than characterizing the distribution of the noise and can be conveniently done using, *e.g.,* conformal prediction with a calibration dataset (Angelopoulos and Bates, 2021). Given (3), SME returns a *set estimation* of the model

$$\mathcal{S} = \{\theta \in \Theta \mid \|r(\theta, z_i)\| \leq \beta_i, i = 1, \dots, N\}, \tag{SME}$$

i.e., $\mathcal{S}$ contains all models *compatible* with the measurements $\mathbb{Z}$ under assumption (3). Clearly, the groundtruth must belong to $\mathcal{S}$, and the "size" of $\mathcal{S}$ informs the uncertainty of the estimated model.

Challenges. It is almost trivial to write down the set $\mathcal{S}$ as in (SME), which, however, turns into a highly nontrivial object to manipulate. The reasons are (a) each of the constraints "$\|r(\theta, z_i)\| \leq \beta_i$" may be a nonconvex constraint, and/or (b) the number of constraints N may be very large. We use two examples in control and perception, respectively, to illustrate the challenges.

Example 1 (System Identification (Kosut et al., 1992; Li et al., 2023)) *Consider a discrete-time dynamical system with state $x \in \mathbb{R}^{n_x}$, control $u \in \mathbb{R}^{n_u}$, and unknown system parameters $\theta \in \mathbb{R}^n$*

$$x_{k+1} = \phi_0(x_k, u_k) + \sum_{i=1}^{n} \theta_i \phi_i(x_k, u_k) + \epsilon_k = \Phi(x_k, u_k)\tilde{\theta} + \epsilon_k, \quad k = 0, \dots, N-1, \tag{4}$$

where $\Phi(x_k, u_k) = [\phi_0(x_k, u_k), \dots, \phi_n(x_k, u_k)] \in \mathbb{R}^{n_x \times (n+1)}$ is a (nonlinear) activation function, $\epsilon_k \in \mathbb{R}^{n_x}, k = 0, \dots, N-1$ are unknown noise vectors assumed to satisfy (3),[3] and $\tilde{\theta} = [1, \theta^\mathsf{T}]^\mathsf{T}$. Given a system trajectory $\{x_0, u_0, \dots, x_{N-1}, u_{N-1}, x_N\}$, the SME of the parameters θ is

$$\mathcal{S} = \left\{\theta \in \mathbb{R}^n \mid \|x_{k+1} - \Phi(x_k, u_k)\tilde{\theta}\| \leq \beta_k, k = 0, \dots, N-1\right\}. \tag{5}$$

Example 2 (Object Pose Estimation (Hartley and Zisserman, 2003; Yang and Pavone, 2023)) *Consider a 3D point cloud $\{Y_i\}_{i=1}^N$ and a camera, at an unknown rotation and translation $\theta = (R, t) \in \text{SE}(3)$,[4] observing the point cloud as a set of 2D image keypoints*

$$y_i = \Pi(RY_i + t) + \epsilon_i, \quad i = 1, \dots, N, \tag{6}$$

2. One can replace the Gaussian assumption with more sophisticated distributional assumptions, but the resulting optimization often becomes more difficult to solve.
3. Note that our algorithm can allow θ to appear nonlinearly in the dynamics, but it is sufficient to restrict to dynamics linear in θ as in (4) for system identification tasks in many real applications.
4. $\text{SE}(3) := \text{SO}(3) \times \mathbb{R}^3$ with $\text{SO}(3) := \{R \in \mathbb{R}^{3\times 3} \mid R^\mathsf{T}R = RR^\mathsf{T} = \mathbf{I}, \det R = +1\}$.

where $\Pi : \mathbb{R}^3 \to \mathbb{R}^2, \Pi(v) = [v_1/v_3, v_2/v_3]^\mathsf{T}$ *denotes the projection of a 3D point onto the 2D image plane and* $\epsilon_i \in \mathbb{R}^2$ *denotes measurement noise satisfying* (3)*. Given pairs of 3D-2D correspondences* $\{y_1, Y_1, \ldots, y_N, Y_N\}$*, the SME of the camera pose* θ *is*

$$\mathcal{S} = \{\theta \in \mathrm{SE}(3) \mid \|y_i - \Pi(RY_i + t)\| \leq \beta_i, i = 1, \ldots, N\}\,. \tag{7}$$

The SME (5) is convex –defined by quadratic inequality constraints– but N can be large given a long trajectory. The SME (7), shown by (Yang and Pavone, 2023, Proposition 3) to be defined by N quadratic inequalities and N linear inequalities, is unfortunately nonconvex despite that $N < 10$.

Contributions. We propose tractable algorithms based on semidefinite programming (SDP) to simplify the set-membership estimator (SME) while maintaining tight uncertainty quantification. Specifically, we seek to find the *minimum enclosing ellipsoid* (MEE) of (SME), *i.e.,* the ellipsoid that contains $\mathcal{S}$ with minimum volume. Such an ellipsoid allows us to (i) use the center of the ellipsoid as a point estimator, (ii) provide a (minimum) worst-case error bound for the point estimator, and (iii) generate samples in $\mathcal{S}$ via straightforward rejection sampling (*i.e.,* sample inside the ellipsoid and accept the sample if inside $\mathcal{S}$).[5] Our algorithms are based on the *sums-of-squares* (SOS) relaxation framework proposed in Nie and Demmel (2005), revisited in Kojima and Yamashita (2013), for computing *a hierarchy of enclosing ellipsoids* that *asymptotically* converge to the MEE of a *basic semialgebraic set*, *i.e.,* a set defined by finitely many polynomial (in-)equalities, to which both (5) and (7) belong.[6] When the SME is convex, together with the celebrated *Löwner-John's ellipsoid theorem* (Henk, 2012), we give an algorithm that can provide a *certificate* of convergence when the MEE has been attained by the SOS hierarchy. We show this vanilla algorithm already outperforms the confidence set of least-squares estimation (Abbasi-Yadkori and Szepesvári, 2011) in a simple instance of Example 1. Unfortunately, applying this algorithm to Example 1 with large N and Example 2 encounters three challenges. (C1) A long system trajectory (*e.g.,* $N = 1000$) in Example 1 renders high-order SOS relaxations overly expensive. We therefore introduce a *preprocessing* algorithm to prune redundant constraints in the set (5) (*e.g.,* over 900 constraints are deemed redundant). (C2) We empirically found for (SME) sets that are nonconvex or defined by many constraints, the SDP solver would encounter serious numerical issues and simply fail. To circumvent this, we propose a two-step approach, where step one draws random samples from (SME) to approximate the *shape matrix* of the enclosing ellipsoid, and step two minimizes the size of the ellipsoid with a fixed shape. The second step, coincidentally, becomes a strict generalization of the *relaxed Chebyshev center* method proposed in Eldar et al. (2008). (C3) The last challenge is to handle the non-Euclidean geometry when enclosing an (SME) of $\mathrm{SO}(3)$. By leveraging the special geometry of *unit quaternions*, we show a reduction of computing the MEE on a Riemannian manifold to computing the MEE in Euclidean space. With these computational enhancements, we conduct experiments on system identification (including those that are hard to learn (Tsiamis and Pappas, 2021)) and object pose estimation that were not possible to perform in existing literatures.

Paper Organization. We first present the SOS-based MEE framework in Section 2. We describe in Section 3 our computational enhancements. We present numerical experiments in Section 4 and conclude in Section 5.

Related works, proofs, extra experiments, and preliminaries are presented in (Tang et al., 2023).

5. Clearly, this would also allow us to obtain an estimate of the volume of $\mathcal{S}$ by counting the acceptance rate.
6. Surprisingly, the SOS-based MEE framework seems unexploited in the context of Examples 1-2.

2. Minimum Enclosing Ellipsoid by SOS Relaxations

Consider a basic semialgebraic set defined by a finite number of polynomial constraints

$$\mathcal{S} = \{\theta \in \mathbb{R}^n \mid g_i(\theta) \geq 0, i = 1, \ldots, l_g, h_j(\theta) = 0, j = 1, \ldots, l_h\} \tag{8}$$

with $g_i, h_j \in \mathbb{R}[\theta]$ real polynomials in θ. We use the same notation $\mathcal{S}$ here as in (SME), (5), (7) because the membership sets we consider are all basic semialgebraic sets. Let $\xi = (\theta_{i_1}, \ldots, \theta_{i_d}) = P\theta$ be a d-dimensional subvector of θ (with P a selection matrix), $\mathcal{S}_\xi := \{P\theta \mid \theta \in \mathcal{S}\}$ be the restriction of $\mathcal{S}$ on ξ, and consider a d-dimensional ellipsoid

$$\mathcal{E} = \{\xi \in \mathbb{R}^d \mid 1 - (\xi - \mu)^T E(\xi - \mu) \geq 0\}, \tag{9}$$

with $\mu \in \mathbb{R}^d$ the center and $E \in \mathbb{S}^d_{++}$ the shape matrix (E is positive definite). We want to find the ellipsoid $\mathcal{E}$ with minimum volume that encloses $\mathcal{S}_\xi$

$$\max_{\mu \in \mathbb{R}^d, E \in \mathbb{S}^d_{++}} \{\log\det E \mid \xi \in \mathcal{E}, \forall \xi \in \mathcal{S}_\xi\}, \tag{MEE}$$

where $\det E$ is inversely proportional to the volume of $\mathcal{E}$ (*i.e.,* (MEE) minimizes the volume of $\mathcal{E}$). We remark that considering an ellipsoid in the subvector ξ is general and convenient as (i) $P = I$ recovers the n-dimensional ellipsoid, (ii) in Example 2 it is desired to enclose the rotation R and translation t separately because R lives in $\mathrm{SO}(3)$ and t lives in $\mathbb{R}^3$, and (iii) having $\xi = \theta_i \in \mathbb{R}$ for some dimension i allows the ellipsoid $\mathcal{E}_i$ to be a line segment that encloses θ_i and hence $\mathcal{E}_1 \times \ldots \times \mathcal{E}_n$ forms an enclosing box of $\mathcal{S}$.

Problem (MEE) is generally intractable even when $\mathcal{S}$ is convex.[7] However, denoting $e(\xi) := 1 - (\xi - \mu)^T E(\xi - \mu)$ in (9) as the polynomial in ξ, we observe the constraint in (MEE) simply asks $e(\xi)$ to be *nonnegative on the set* $\mathcal{S}$. This observation allows a hierarchy of convex relaxations of (MEE) based on sums-of-squares (SOS) programming (Lasserre, 2001; Blekherman et al., 2012).

Theorem 1 (MEE Approximation by SOS Programming) *Assume $\mathcal{S}$ is Archimedean,[8] consider the SOS program with an integer κ such that $2\kappa \geq \max\{2, \{\deg(g_i)\}_{i=1}^{l_g}, \{\deg(h_j)\}_{j=1}^{l_h}\}$*

$$\max_{E, b, c, \sigma_i, \lambda_j} \quad \log\det E \tag{10a}$$

$$\textit{subject to} \quad 1 - (\xi^\mathsf{T} E \xi + 2b^\mathsf{T}\xi + c) = \textstyle\sum_{i=0}^{l_g} \sigma_i(\theta) g_i(\theta) + \sum_{j=1}^{l_h} \lambda_j(\theta) h_j(\theta) \tag{10b}$$

$$\sigma_i(\theta) \in \mathrm{SOS}[\theta], \;\; \deg(\sigma_i g_i) \leq 2\kappa, \quad i = 0, \ldots, l_g \tag{10c}$$

$$\lambda_j(\theta) \in \mathbb{R}[\theta], \;\; \deg(\lambda_j h_j) \leq 2\kappa, \quad j = 1, \ldots, l_h \tag{10d}$$

$$\begin{bmatrix} E & b \\ b^\mathsf{T} & c \end{bmatrix} \succeq 0 \tag{10e}$$

where $g_0(\theta) := 1$, $\mathrm{SOS}[\theta]$ is the set of SOS polynomials in θ, and $\deg(\cdot)$ denotes the degree of a polynomial. Let $(E_\star, b_\star, c_\star)$ be an optimal solution of (10), then,

(i) for any κ, we have $\xi^\mathsf{T} E_\star \xi + 2b_\star^\mathsf{T}\xi + c_\star = (\xi - \mu_\star)^\mathsf{T} E_\star (\xi - \mu_\star)$ with $\mu_\star = -E_\star^{-1} b_\star$, and the ellipsoid $\mathcal{E}_\kappa = \{\xi \in \mathbb{R}^d \mid (\xi - \mu_\star)^\mathsf{T} E_\star (\xi - \mu_\star) \leq 1\}$ encloses $\mathcal{S}_\xi$;

7. When $\mathcal{S}$ is a set of points, then (MEE) is easy to solve (Gärtner, 1999; Magnani et al., 2005; Moshtagh et al., 2005).
8. The definition of Archimedeanness is given in (Blekherman et al., 2012, Def. 3.137). One can make the Archimedean condition trivially hold by adding a constraint $M_\theta - \theta^\mathsf{T}\theta \geq 0$ to (8), which is easy for Examples 1-2.

(ii) $\mathrm{Vol}(\mathcal{E}_\kappa)$ *decreases as* κ *increases, and* $\mathcal{E}_\kappa$ *tends to the solution of* (MEE) *as* $\kappa \to \infty$.

Although it seems that Equation (10b) is hard to satisfy because the LHS only contains a subset of variables comparing to the RHS, we claim that it is not the case. This is because the inequalities $\{g_i\}$ and equalities $\{h_j\}$ contain all variables. Thus, monomial cancellation is easy to happen. Theorem 1 is inspired by Nie and Demmel (2005); Magnani et al. (2005). Problem (10) is convex and can be readily implemented by YALMIP (Lofberg, 2004) and solved by MOSEK (ApS, 2019). Its intuition is simple: (10b)-(10d) ensures $1-(\xi^\mathsf{T} E\xi+2b^\mathsf{T}\xi+c)$ is nonnegative on $\mathcal{S}$ and hence every feasible solution is an enclosing ellipsoid, (10e) uses a lifting technique to convexify the bilinearity of μ and E in the original ellipsoid parametrization (9), and the objective (10a) seeks to minimize the volume. The convergence of the hierarchy follows from Putinar's Positivstellensatz (Putinar, 1993).

Certifying Convergence. Unlike applying SOS relaxations to polynomial optimization, where a certificate of convergence is known (Henrion and Lasserre, 2005), detecting the convergence of (10) is in general difficult (Lasserre, 2015). However, when the set $\mathcal{S}$ is convex, it is possible to leverage Löwner-John's ellipsoid theorem (Xie, Miaolan, 2016) to derive a simple convergence certificate. We present such a certificate and a numerical example in Tang et al. (2023).

3. Computational Enhancement

The SOS-based algorithm (10) works very well on simple examples. In this section, we describe three challenges of applying (10) and present three techniques to enhance its performance.

3.1. Pruning Redundant Constraints

The first challenge arises when the number of inequalities l_g is very large, in which case the convex optimization (10) has l_g+1 positive semidefinite (PSD) variables (whose sizes grow quickly with the relaxation order κ). In system identification (Example 1), a large $l_g = N$ is common as practitioners often collect a long system trajectory to accurately identify the system.

Nevertheless, since the dimension n of the parameter θ is usually much smaller than the trajectory length N, one would expect many constraints in (5) to be *redundant*. The difficultly lies in how to *identify and prune* the redundant constraints. When all the constraints are linear, algorithms from linear programming can identify redundancy (Caron et al., 1989; Telgen, 1983; Paulraj et al., 2010; Cotorruelo et al., 2020). To handle the quadratic constraints in (5), we propose Algorithm 1.

Algorithm 1: Prune Redundant Constraints in (5)

1 **Input:** Constraint set $\{g_k(\theta)\}_{k\in[N-1]}$ with $g_k(\theta) := \beta_k^2 - \|x_{k+1} - \Phi(x_k,u_k)\tilde{\theta}\|^2$ (*cf.* (5))
2 **Output:** Pruned set of constraints $\{g_k(\theta)\}_{k\in\mathcal{I}}$
3 Initialize $\mathcal{I} = [N-1]$
4 **for** $k \leftarrow 0$ *to* $N-1$ **do**
5 Solve $\underline{g}_k \le g_k^\star = \min_\theta \{g_k(\theta) \mid g_i(\theta) \ge 0, i \in \mathcal{I}\backslash\{k\}\}$ ◁ First-order moment-SOS hierarchy
6 **if** $\underline{g}_k^\star \ge 0$ **then** $\mathcal{I} \leftarrow \mathcal{I}\backslash\{k\}$;
7 **end**
8 **Return:** $\mathcal{I}$

Intuitively, line 5 of Algorithm 1 seeks to minimize $g_k(\theta)$ when the rest of the constraints hold. If $g_k^\star \ge 0$, then $g_k(\theta)$ is redundant as it is implied by the rest of the constraints. The optimization in line 5, however, is nonconvex because $g_k(\theta)$ is a concave polynomial. Therefore, we use the first-order moment-SOS hierarchy to obtain a lower bound $\underline{g}_k \le g_k^\star$: if $\underline{g}_k \ge 0$, then $g_k^\star \ge 0$ must

hold and $g_k(\theta)$ is deemed redundant. In practice, the first-order moment-SOS hierarchy is very efficient and easily scales to $N = 1000$. In a pendulum system identification experiment presented in Section 4.1.4, we show Algorithm 1 effectively prunes over 90% constraints.

3.2. Generalized Relaxed Chebyshev Center

The second challenge comes from the "$\log\det E$" objective in (10). Maximizing the "$\log\det$" of E is convex (Vandenberghe et al., 1998), yet it cannot be written as a standard linear SDP.[9] Consequently, it is often replaced by the geometric mean (Lofberg, 2004) and modelled with an exponential cone constraint (ApS, 2019), causing serious numerical issues in our experiments.[10] Even after pruning redundant constraints using Algorithm 1, we found problem (10) still difficult to solve. Similar observations have been reported in (Lasserre, 2023, page 950).

This motivates solving the following *generalized Chebyshev center* (GCC) problem

$$\eta^\star = \min_{\mu\in\mathbb{R}^d} \max_{\xi\in\mathcal{S}_\xi} \quad \|\mu-\xi\|_Q^2 \tag{GCC}$$

with a given $Q \succ 0$ (recall $\xi = P\theta$ is a subvector of θ). When $Q = I$, (GCC) reduces to the usual Chebyshev center problem (Milanese and Vicino, 1991; Eldar et al., 2008). In (GCC), given any μ, the inner "$\max_\xi$" computes the maximum Q-weighted distance from μ to the set $\mathcal{S}_\xi$, denoted as $\eta_\mu^\star$. Therefore, the ellipsoid $\mathcal{E}_{Q,\mu} = \{\xi \mid \|\xi-\mu\|_Q^2 \leq \eta_\mu^\star\}$ must enclose the set $\mathcal{S}_\xi$. Via the outer "$\min_\mu$", (GCC) seeks the μ such that $\mathcal{E}_{Q,\mu}$ is the smallest enclosing ellipsoid with a given shape matrix Q. We found (GCC) to be much easier to solve than (10) due to removing the "$\log\det E$" objective.

Estimating Q from Samples. Intuitively, (GCC) may be more conservative than (MEE) because (GCC) assumes Q is given while (MEE) optimizes the shape matrix. However, if we can uniformly draw samples $(\xi^1,\ldots,\xi^{N_s})$ from $\mathcal{S}_\xi$, then a good estimate of Q can be obtained as

$$Q = \frac{1}{N_s}\sum_{i=1}^{N_s}(\xi^i-\bar{\xi})(\xi^i-\bar{\xi})^\mathsf{T}, \quad \bar{\xi} = \frac{1}{N_s}\sum_{i=1}^{N_s}\xi^i. \tag{11}$$

We can use three different algorithms to uniformly sample $\mathcal{S}_\xi$. (i) When $\mathcal{S}$ is convex, we can use the *hit-and-run* algorithm (Bélisle et al., 1993), where each iteration involves solving a convex optimization. (ii) For Example 2, Yang and Pavone (2023) proposed a RANSAG algorithm, where each iteration involves solving a geometry problem. (iii) We can first solve (GCC) with $Q = I$ to obtain an enclosing ball, then perform *rejection sampling* inside the enclosing ball. We use (iii) in our experiments as it is general and does not require convexity.

Convex Relaxations for (GCC). Problem (GCC) is still nonconvex, but we can design a hierarchy of SDP relaxations that asymptotically converges to $\eta^\star$.

Theorem 2 (Generalized Relaxed Chebyshev Center) *Let $\mathcal{S}$ in (8) be Archimedean, κ be an integer such that $2\kappa \geq \max\{2, \{\deg(g_i)\}_{i=1}^{l_g}, \{\deg(h_j)\}_{j=1}^{l_h}\}$, and $s_n(d) := \binom{n+d}{d}$, consider the following convex quadratic SDP*

$$-\eta_\kappa^\star = \min_{z\in\mathbb{R}^{s_n(2\kappa)}} \quad L_z(\theta)^\mathsf{T}P^\mathsf{T}QPL_z(\theta) - \langle C, M_\kappa(z)\rangle \tag{12a}$$

$$\text{subject to} \quad M_\kappa(z) \succeq 0, \quad M_0(z) = 1 \tag{12b}$$

$$M_{\kappa-\lceil\deg(g_i)/2\rceil}(g_i z) \succeq 0, \quad i = 1,\ldots,l_g \tag{12c}$$

$$L_z(h_j[\theta]_{2\kappa-\deg(h_j)}) = 0, \quad j = 1,\ldots,l_h \tag{12d}$$

9. SDPT3 (Toh et al., 1999) natively supports $\log\det$ but empirically we found it performs worse than MOSEK.
10. Particularly, MOSEK returns a solution with large duality gap.

where $z \in \mathbb{R}^{s_n(2\kappa)}$ is the pseudomoment vector in θ of degree up to 2κ, $L_z(u)$ and $M_\kappa(uz)$ are both linear functions of z given coefficients of a certain polynomial $u(\theta)$, and C is a constant matrix whose expression is given in Tang et al. (2023). Let $z_{\star,\kappa}$ be an optimal solution to (12) and $\mu_{\star,\kappa} = L_{z_{\star,\kappa}}(\theta)$, then,

(i) for any κ, the ellipsoid $\mathcal{E}_{Q,\kappa} := \{\|\xi - \mu_{\star,\kappa}\|_Q^2 \leq \eta_\kappa^\star\}$ encloses $\mathcal{S}_\xi$;

(ii) $\eta_\kappa^\star \geq \eta^\star$ for any κ and $\eta_\kappa^\star$ converges to $\eta^\star$ as $\kappa \to \infty$.

The proof of Theorem (2), together with a lifting technique to write (12) as a standard linear SDP, is given in Tang et al. (2023). The basic strategy is to first apply the moment-SOS hierarchy, with order κ, to relax the inner "$\max_{\xi \in \mathcal{S}_\xi}$" (a polynomial optimization) as a convex SDP whose decison variable is $z \in \mathbb{R}^{s_n(2\kappa)}$, the pseudomoment vector constrained by (12b)-(12d). Then, by invoking Sion's minimax theorem (Sion, 1958), one can switch "$\min_\mu$" and "$\max_z$", after which the "$\min_\mu$" admits a closed-form solution $\mu = L_z(\theta)$ and leads to a single-level optimization with cost (12a). It is worth noting that letting $\kappa = 1$ (and $Q = I$) in (12) recovers the relaxed Chebyshev center in Eldar et al. (2008) for bounded error estimation. For this reason, we call $\mu_{\star,\kappa}$ the *generalized relaxed Chebyshev center* of order κ (GRCC-κ). As we will see in Section 4.1.2, the enclosing balls at GRCC-κ with $\kappa \geq 2$ is significantly smaller than those of the RCC.

3.3. Handling the non-Euclidean Geometry of $\mathrm{SO}(3)$

With the (GCC) formulation, we can approximate the minimum ellipsoid that encloses $\mathcal{S}_t$, the translation part of the SME in Example 2. The last challenge lies in enclosing $\mathcal{S}_R$, the rotation part of (7). Since the concept of an ellipsoid is not well defined on a Riemannian manifold such as $\mathrm{SO}(3)$, it is natural to seek an enclosing *geodesic ball*. This is closely related to the problem of finding the *Riemannian minimax center* of $\mathcal{S}_R$, defined as (Arnaudon and Nielsen, 2013)

$$\mu_{R,\star} = \operatorname*{arg\,min}_{\mu \in \mathrm{SO}(3)} \max_{R \in \mathcal{S}_R} d_{\mathrm{SO}(3)}(\mu, R), \tag{13}$$

where, given two rotations $\mu, R \in \mathrm{SO}(3)$, $d_{\mathrm{SO}(3)}(\mu, R)$ denotes the geodesic distance on $\mathrm{SO}(3)$ defined by $d_{\mathrm{SO}(3)}(\mu, R) = \arccos((\mathrm{tr}\left(\mu^\mathsf{T} R\right) - 1)/2)$. Clearly, if we can find $\mu_{R,\star}$, then we have found the best point estimate from which the worst-case error bound is the smallest.

Unfortunately, problem (13) cannot be directly solved using Theorem 2 because (i) $\mu \in \mathrm{SO}(3)$ is constrained (instead of in (GCC) $\mu \in \mathbb{R}^d$ is unconstrained), and (ii) the objective "$d_{\mathrm{SO}(3)}(\mu, R)$" is not a polynomial. The next result resolves these issues by leveraging unit quaternions.

Theorem 3 (Minimum Enclosing Ball for Rotations) *Suppose there exists a 3D rotation $\bar{R}$ such that $d_{\mathrm{SO}(3)}(\bar{R}, R) \leq \frac{\pi}{2}, \forall R \in \mathcal{S}_R$, and let $\bar{q} \in S^3 := \{q \in \mathbb{R}^4 \mid \|q\| = 1\}$ (or $-\bar{q}$) be any one of the two unit quaternions associated with $\bar{R}$. Consider the following Chebyshev center problem in $\mathbb{R}^4$*

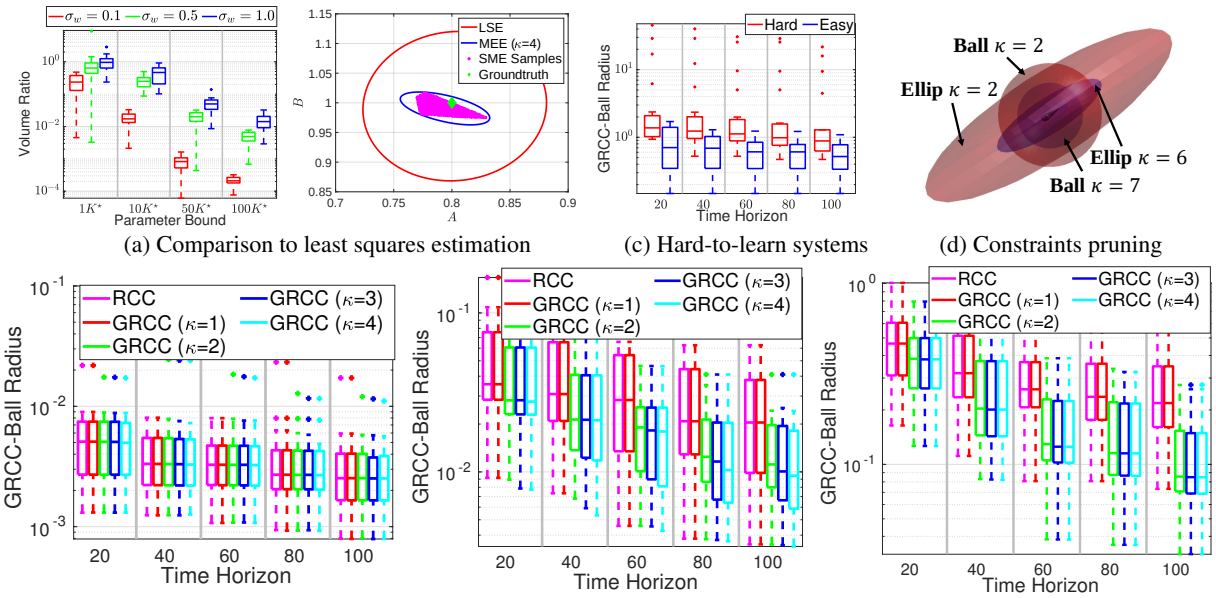
$$\mu_\star = \operatorname*{arg\,min}_{\mu \in \mathbb{R}^4} \max_{q \in \mathbb{R}^4} \left\{ \|\mu - q\|^2 \middle| \; q \in S^3, \mathcal{R}(q) \in \mathcal{S}_R, \bar{q}^\mathsf{T} q \geq 0 \right\}, \tag{14}$$

where $\mathcal{R} : S^3 \to \mathrm{SO}(3)$ maps a unit quaternion to a 3D rotation matrix (whose expression is given in Tang et al. (2023)). Then the projection of $\mu_\star$ onto S^3 is the solution to (13), i.e., $\mu_{R,\star} = \mu_\star / \|\mu_\star\|$.

Theorem 3 states that we can solve the original Riemannian minimax problem (13) by solving (14), which is easily verified to be an instance of (GCC) and hence can be approximated by the SDP relaxations in Theorem 2. Theorem 3 requires the existence of $\bar{R}$ such that all rotations in $\mathcal{S}_R$ are reasonably close to $\bar{R}$ and similar assumptions are needed in Arnaudon and Nielsen (2013). In other words, $\mathcal{S}_R$ needs to have relatively small uncertainty. This assumption can be numerically verified by first using the RANSAG algorithm in Yang and Pavone (2023) to compute $\bar{R}$ and then compute the maximum distance between $\bar{R}$ and $\mathcal{S}_R$, which we found empirically to be below $\frac{\pi}{2}$ for a majority of the test problems.

For a summary of the whole workflow, see Tang et al. (2023).

4. Experiments



(a) Comparison to least squares estimation (c) Hard-to-learn systems (d) Constraints pruning

(b) Identification of random linear systems, from left to right: $\beta = 0.01$, $\beta = 0.1$, $\beta = 1.0$

Figure 1: Experimental results on system identification (Example 1).

4.1. System Identification (Example 1)

4.1.1. Comparison with Least Squares Estimation

Consider a one-dimensional linear system $x_{k+1} = A_\star x_k + B_\star u_k + w_k$ with $A_\star = 0.8, B_\star = 1.0$, and $w_k \sim \mathcal{N}(0, \sigma_w^2)$.[11] Suppose we collected a system trajectory $x_0, u_0, \ldots, x_N$, we wish to find a set $\mathcal{S}$ such that we can guarantee $\mathbb{P}[(A_\star, B_\star) \in \mathcal{S}] \geq 80\%$. (Abbasi-Yadkori and Szepesvári, 2011, Theorem 1) provides an ellipse centered at the least-squares estimator (LSE) that satisfies the probabilistic coverage guarantee, which requires knowing a bound such that $A_\star^2 + B_\star^2 \leq K^2$. Let $K_\star = \sqrt{A_\star^2 + B_\star^2}$ be the best bound.

We can get much smaller coverage sets via the SME (5) and its enclosing ellipsoids without knowledge on $K_\star$. Let $\beta > 0$ be the smallest number such that $\mathbb{P}[w_k^2 \leq \beta] \geq (80\%)^{1/N}$, implying $\mathbb{P}[w_0^2 \leq \beta, \ldots, w_{N-1}^2 \leq \beta] \geq 80\%$ (assume w_k's are independent). Consequently, the SME (5) covers $(A_\star, B_\star)$ with probability at least 80%. Let $\mathcal{E}_4$ be the enclosing ellipse of (5) computed via Theorem 1 using $\kappa = 4$ and $\mathcal{E}_{\text{LSE}}$ be the confidence ellipse from Abbasi-Yadkori and Szepesvári

11. This is the same setup as in Li et al. (2023).

(2011). We compute $\mathrm{Vol}(\mathcal{E}_2)/\mathrm{Vol}(\mathcal{E}_{\mathrm{LSE}})$ as the volume ratio, which indicates SME finds a smaller enclosing ellipse when it is below 1.

Fig. 1(a) boxplots the volume ratio (in log scale) under different choices of σ_w and K with fixed $N = 50$ (we perform 20 random experiments in each setup). We observe that LSE outperforms SME only when (i) LSE has access to a precise upper bound $K_\star$ (which we think is often unrealistic) and (ii) the noise w_k is large. In all other cases, the enclosing ellipse of SME is orders of magnitude smaller than that of LSE. A sample visualization is provided in Fig. 1(a). The parameters are $K = K^*, \sigma_w = 0.1$ and the number of sampled points is 5000.

However, it is worth noting that in order to apply SME, we need to know the noise bound (or a high probability bound) β, while LSE does not neccessarily require such bounds or distributional assumptions unless one wants to estimate the uncertainty.

4.1.2. Identification of Random Linear Systems

Consider a linear system $x_{k+1} = A_\star x_k + Bu_k + w_k, k = 0, \ldots, N-1$. We generate random $A_\star \in \mathbb{R}^{n_x \times n_x}, B \in \mathbb{R}^{n_x \times n_u}$ with each entry following a standard Gaussian distribution $\mathcal{N}(0,1)$, then truncate the singular values of $A_\star$ that are larger than 1 to 1. u_k follows $\mathcal{N}(0, I)$, and w_k is random inside a ball with radius β. We treat B as known. We compute enclosing balls of the SME (5) using two algorithms: the RCC algorithm in Eldar et al. (2008), and our GRCC algorithm presented in Theorem 2. Fig. 1(b) plots the radii of the enclosing balls with $N \in \{20, 40, 60, 80, 100\}$, $n_x = 2$, and $\beta \in \{0.01, 0.1, 1.0\}$ (each boxplot summarizes 20 random experiments). We observe that (i) GRCC ($\kappa = 1$) leads to exactly the same result as RCC, verifying that our GRCC algorithm recovers the RCC algorithm with $\kappa = 1$. (ii) The enclosing balls get much smaller with high-order relaxations (*i.e.,* larger κ). More results with $n_x = 3$ and $n_x = 4$, also the runtime for the $n_x = 2$ experiment are presented in Tang et al. (2023).

4.1.3. Linear Systems that Are Hard to Learn

Tsiamis and Pappas (2021) presented examples of linear systems that are hard to learn. We show that SME and its enclosing balls are *adaptive to the hardness of identification*, *i.e.,* one gets large enclosing balls when the system is hard to learn. Consider the system $x_{k+1} = Ax_k + Hw_k$ with

$$A = \begin{bmatrix} 0 & \theta_1 & 0 \\ 0 & 0 & \theta_2 \\ 0 & 0 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

where w_k is a random disturbance with noise bound 0.1. Consider (i) an easy-to-learn system with $\theta_2 = 1$, and (ii) a hard-to-learn system with $\theta_2 = 10^{-5}$. Let $\theta_1 = 1$. Fig. 1(c) boxplots the radii of the enclosing balls for the SME of (θ_1, θ_2) with $\kappa = 2$ and increasing N. Clearly, we observe that the radii for hard-to-learn systems are much larger than that for easy systems.

4.1.4. Constraints Pruning in A Long Trajectory

Consider the continuous-time dynamics of a simple pendulum

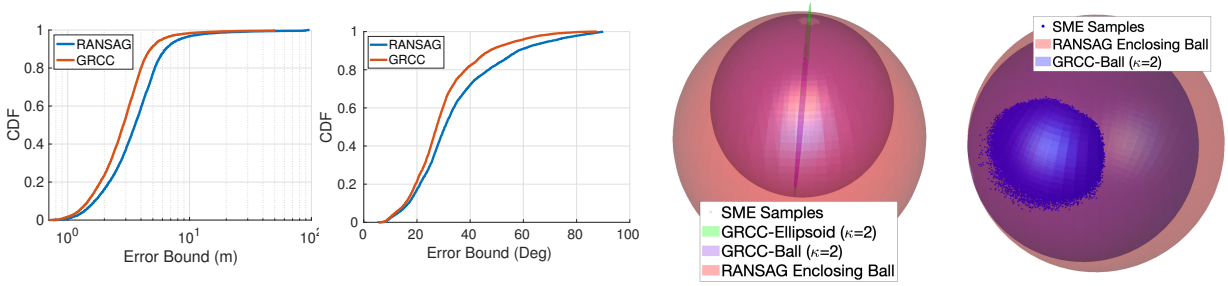
$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -\frac{\mathfrak{b}}{\mathfrak{m}\mathfrak{l}^2} x_2 + \frac{1}{\mathfrak{m}\mathfrak{l}^2} u - \frac{\mathfrak{g}}{\mathfrak{l}} \sin x_1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\frac{\mathfrak{b}}{\mathfrak{m}\mathfrak{l}^2} & \frac{1}{\mathfrak{m}\mathfrak{l}^2} & -\frac{\mathfrak{g}}{\mathfrak{l}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ u \\ \sin x_1 \end{bmatrix}$$

where $x = (x_1, x_2)$ is the state (angle and angular velocity), $\mathfrak{b}$ is the damping ratio, $\mathfrak{m}$ is the mass, $\mathfrak{l}$ is the length of the pole, and $\mathfrak{g}$ is the gravity constant. We wish to identify $\theta_1 = \frac{\mathfrak{b}}{\mathfrak{m}\mathfrak{l}^2}$, $\theta_2 = \frac{1}{\mathfrak{m}\mathfrak{l}^2}$

and $\theta_3 = \frac{\mathfrak{g}}{\mathfrak{l}}$. To do so, we discretize the dynamics using Euler method with $dt = 0.01$, add random disturbance w_k that has bounded norm 0.1, and collect a single trajectory of length $N = 1000$. Without constraint pruning, we can only run the GRCC algorithm with $\kappa = 3$. We then prune the constraint set using Algorithm 1, which leads to a much smaller set of 19 constraints (only 1.9% of the original number of constraints). We can then increase the relaxation order of GRCC and the resulting enclosing ellipsoid and ball become much smaller as visualized in Fig. 1(d). The volume of the enclosing ball with $\kappa = 7$ is only 1.08×10^{-6}, indicating the SME has almost converged to a single point. The time result is provided in Tang et al. (2023).

4.2. Object Pose Estimation (Example 2)

We follow the same procedure as Yang and Pavone (2023), *i.e.,* we use conformal prediction to calibrate the norm bounds β_i for the noise vectors ϵ_i (*cf.* (6)) generated by the pretrained neural network in Pavlakos et al. (2017) and form the SME (7). We then use the GRCC algorithm in Theorem 2 with $\kappa = 3$ to compute enclosing balls (for the rotation we apply GRCC in Theorem 3 to compute enclosing geodesic balls). We compare the radii of the enclosing balls obtained by GRCC with the radii of the enclosing balls obtained by RANSAG of Yang and Pavone (2023).[12] Fig. 2(a) shows the empirical cumulative distribution function (CDF) of the translation bounds and rotation bounds, respectively (there are 7035 translation problems and 6661 rotation problems). We observe that the error bounds obtained by GRCC are smaller than those obtained by RANSAG. Examples of enclosing balls and ellipsoids for the translation SME are shown in Fig. 2(b), where we observe that the enclosing ellipsoid precisely captures the shape of the SME. Fig. 2(b) also plots the enclosing balls of the rotation SME using stereographic projection.



(a) CDF plots. Left: translation, right: rotation

(b) Enclosing balls/ellipsoids. Left: translation, right: rotation.

Figure 2: Experimental results on object pose estimation (Example 2).

For extra experiments, we refer the readers to Tang et al. (2023).

5. Conclusions

We introduced a suite of computational algorithms based on semidefinite programming relaxations to compute minimum enclosing ellipsoids of set-membership estimation in system identification and object pose estimation. Three computational enhancements are highlighted, namely constraints pruning, generalized relaxed Chebyshev center, and handling non-Euclidean geometry. These algorithms are still limited to small- and medium-sized problems (though these problems are already interesting) due to computational challenges in semidefinite programming. Multiple future research directions are possible, *e.g.,* applying SME to system identification with partial observations, extending SME on object pose estimation to more perception problems, and integrating SME with adaptive control and reinforcement learning.

12. RANSAG first estimates an average rotation and translation, then uses SDP relaxations to compute the inner "max" problem in (GCC). In other words, RANSAG does not seek to find a better estimate with smaller error bounds.

References

Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26. JMLR Workshop and Conference Proceedings, 2011.

Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

Pasquale Antonante, Vasileios Tzoumas, Heng Yang, and Luca Carlone. Outlier-robust estimation: Hardness, minimally tuned algorithms, and applications. *IEEE Transactions on Robotics*, 38(1): 281–301, 2021.

Mosek ApS. Mosek optimization toolbox for matlab. *User's Guide and Reference Manual, Version*, 4:1, 2019.

Marc Arnaudon and Frank Nielsen. On approximating the riemannian 1-center. *Computational Geometry*, 46(1):93–104, 2013.

Timothy D Barfoot. *State estimation for robotics*. Cambridge University Press, 2017.

Claude JP Bélisle, H Edwin Romeijn, and Robert L Smith. Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266, 1993.

Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012.

RJ Caron, JF McDonald, and CM Ponic. A degenerate extreme point strategy for the classification of linear constraints as redundant or necessary. *Journal of Optimization Theory and Applications*, 62(2):225–237, 1989.

Andres Cotorruelo, Ilya Kolmanovsky, Daniel R Ramírez, Daniel Limon, and Emanuele Garone. Elimination of redundant polynomial constraints and its use in constrained control. *arXiv preprint arXiv:2006.14957*, 2020.

Yonina C Eldar, Amir Beck, and Marc Teboulle. A minimax chebyshev estimator for bounded error estimation. *IEEE transactions on signal processing*, 56(4):1388–1397, 2008.

Bernd Gärtner. Fast and robust smallest enclosing balls. In *European symposium on algorithms*, pages 325–338. Springer, 1999.

Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

Martin Henk. löwner-john ellipsoids. *Documenta Math*, 95:106, 2012.

Didier Henrion and Jean-Bernard Lasserre. Detecting global optimality and extracting solutions in gloptipoly. In *Positive polynomials in control*, pages 293–310. Springer, 2005.

Peter J Huber. *Robust statistics*, volume 523. John Wiley & Sons, 2004.

Masakazu Kojima and Makoto Yamashita. Enclosing ellipsoids and elliptic cylinders of semialgebraic sets and their application to error bounds in polynomial optimization. *Mathematical Programming*, 138(1-2):333–364, 2013.

Robert L Kosut, Ming K Lau, and Stephen P Boyd. Set-membership identification of systems with parametric and nonparametric uncertainty. *IEEE Transactions on Automatic Control*, 37(7): 929–941, 1992.

Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on optimization*, 11(3):796–817, 2001.

Jean B Lasserre. A generalization of löwner-john's ellipsoid theorem. *Mathematical Programming*, 152:559–591, 2015.

Jean B Lasserre. Pell's equation, sum-of-squares and equilibrium measures on a compact set. *Comptes Rendus. Mathématique*, 361(G5):935–952, 2023.

Yingying Li, Jing Yu, Lauren Conger, and Adam Wierman. Learning the uncertainty sets for control dynamics via set membership: A non-asymptotic analysis. *arXiv preprint arXiv:2309.14648*, 2023.

Johan Lofberg. Yalmip: A toolbox for modeling and optimization in matlab. In *2004 IEEE international conference on robotics and automation*, pages 284–289. IEEE, 2004.

Alessandro Magnani, Sanjay Lall, and Stephen Boyd. Tractable fitting with convex polynomials via sum-of-squares. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 1672–1677. IEEE, 2005.

Mario Milanese and Antonio Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991.

Nima Moshtagh et al. Minimum volume enclosing ellipsoid. *Convex optimization*, 111(January): 1–9, 2005.

Jiawang Nie and James W Demmel. Minimum ellipsoid bounds for solutions of polynomial systems via sum of squares. *Journal of Global Optimization*, 33(4):511–525, 2005.

Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96:293–320, 2003.

Sumathi Paulraj, P Sumathi, et al. A comparative study of redundant constraints identification methods in linear programming problems. *Mathematical Problems in Engineering*, 2010, 2010.

Georgios Pavlakos, Xiaowei Zhou, Aaron Chan, Konstantinos G Derpanis, and Kostas Daniilidis. 6-dof object pose from semantic keypoints. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 2011–2018. IEEE, 2017.

Luis Pineda, Taosha Fan, Maurizio Monge, Shobha Venkataraman, Paloma Sodhi, Ricky TQ Chen, Joseph Ortiz, Daniel DeTone, Austin Wang, Stuart Anderson, et al. Theseus: A library for differentiable nonlinear optimization. *Advances in Neural Information Processing Systems*, 35: 3801–3818, 2022.

Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

Maurice Sion. On general minimax theorems. 1958.

Robert F Stengel. *Optimal control and estimation*. Courier Corporation, 1994.

Yukai Tang, Jean-Bernard Lasserre, and Heng Yang. Uncertainty quantification of set-membership estimation in control and perception: Revisiting the minimum enclosing ellipsoid, 2023.

Jan Telgen. Identifying redundant constraints and implicit equalities in systems of linear constraints. *Management Science*, 29(10):1209–1222, 1983.

Kim-Chuan Toh, Michael J Todd, and Reha H Tütüncü. Sdpt3a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999.

Anastasios Tsiamis and George J Pappas. Linear systems can be hard to learn. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 2903–2910. IEEE, 2021.

Lieven Vandenberghe, Stephen Boyd, and Shao-Po Wu. Determinant maximization with linear matrix inequality constraints. *SIAM journal on matrix analysis and applications*, 19(2):499–533, 1998.

Xie, Miaolan. Inner approximation of convex cones via primal-dual ellipsoidal norms. Master's thesis, 2016. URL http://hdl.handle.net/10012/10474.

Heng Yang and Marco Pavone. Object pose estimation with statistical guarantees: Conformal keypoint detection and geometric uncertainty propagation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8947–8958, 2023.