INTERNATIONAL JOURNAL OF
Robust and Nonlinear Control

A Convex Optimization Approach to Compute Trapping Regions for Lossless Quadratic Systems

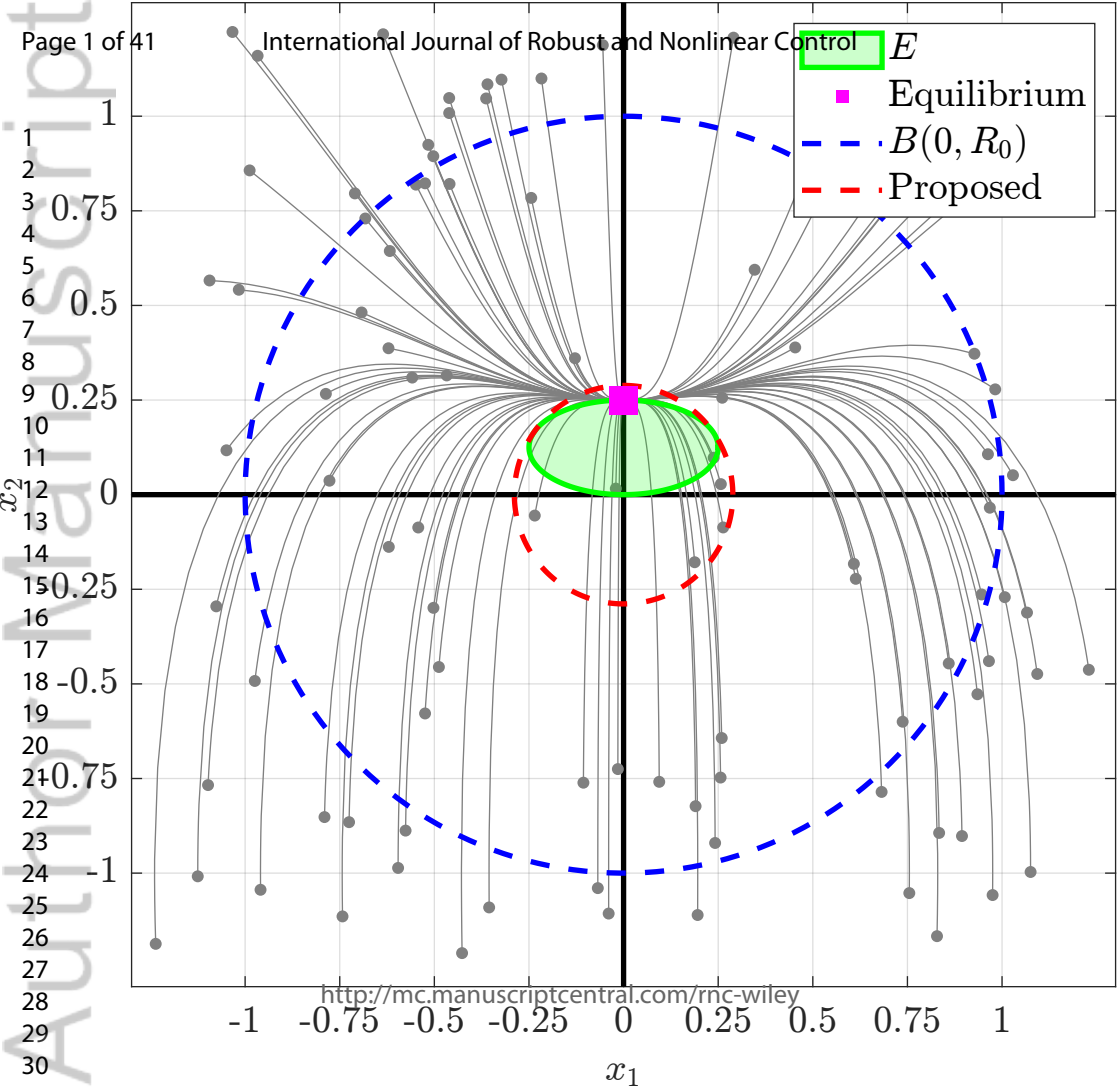
Journal:	*International Journal of Robust and Nonlinear Control*
Manuscript ID	RNC-24-0069.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	24-Sep-2024
Complete List of Authors:	Liao, Shih-Chi; University of Michigan College of Engineering Heide, A. Leonid; University of Minnesota Twin Cities Department of Aerospace Engineering and Mechanics Hemati, Maziar S.; University of Minnesota Twin Cities Department of Aerospace Engineering and Mechanics Seiler, Peter; University of Michigan College of Engineering
Keywords:	Boundedness, Trapping Region, Lyapunov Method, Semidefinite Programming
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Code_TrappingSDP.zip	

SCHOLARONE™
Manuscripts

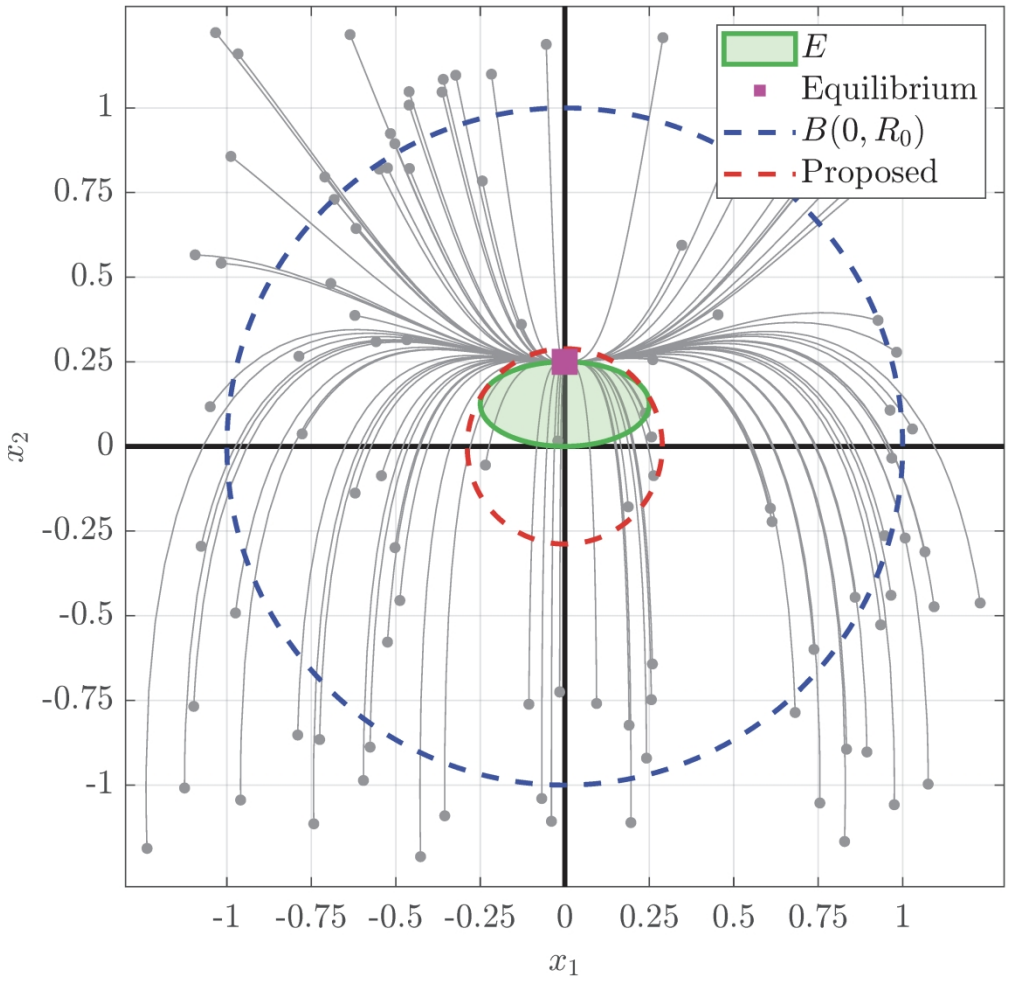
This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/rnc.7807

This article is protected by copyright. All rights reserved.

Author Manuscript

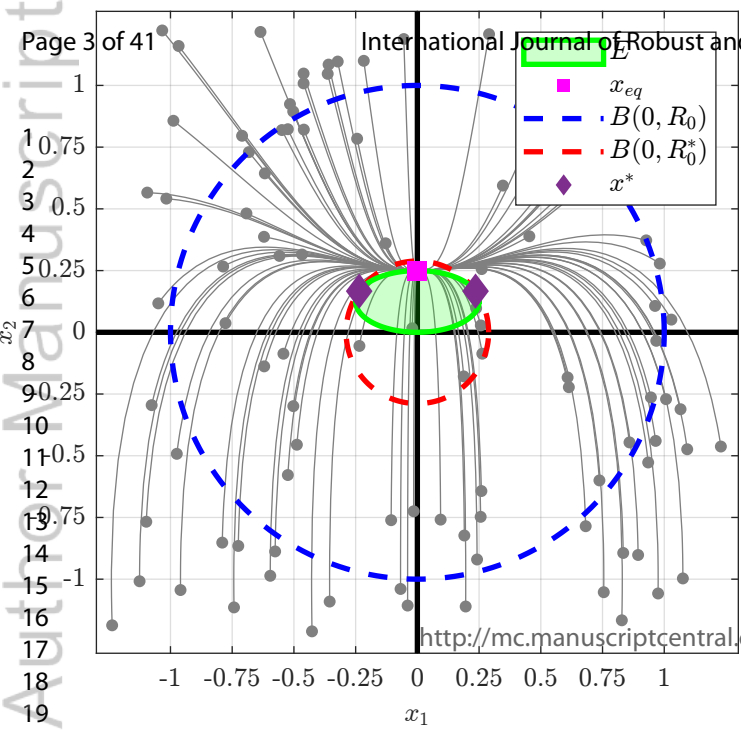


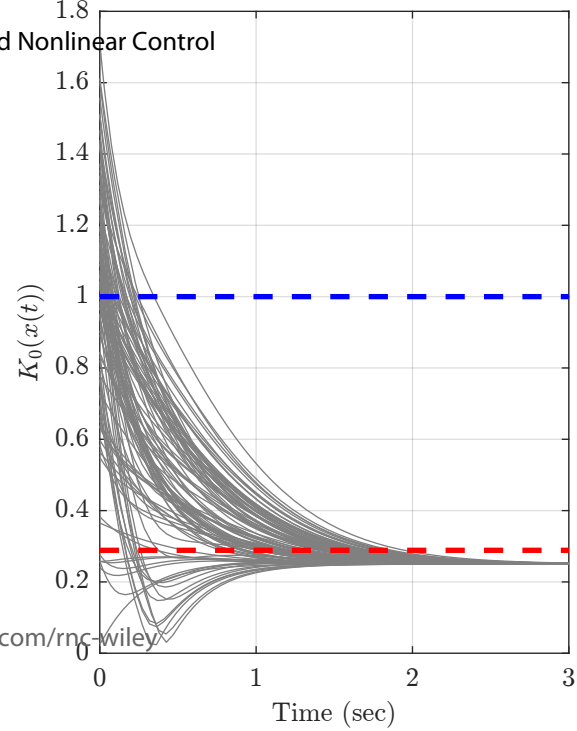
Author Manuscript



376x365mm (236 x 236 DPI)

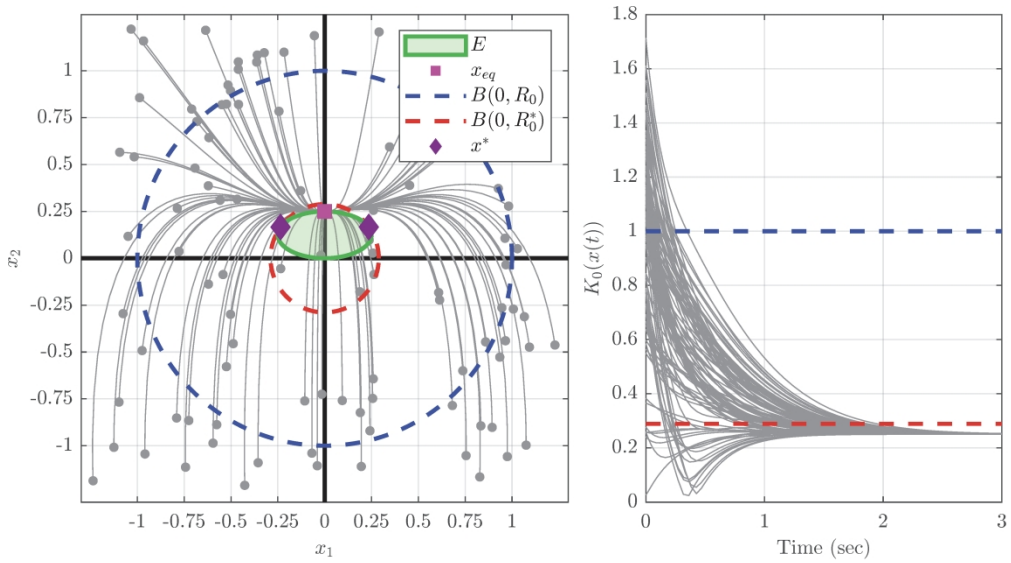
Author Manuscript





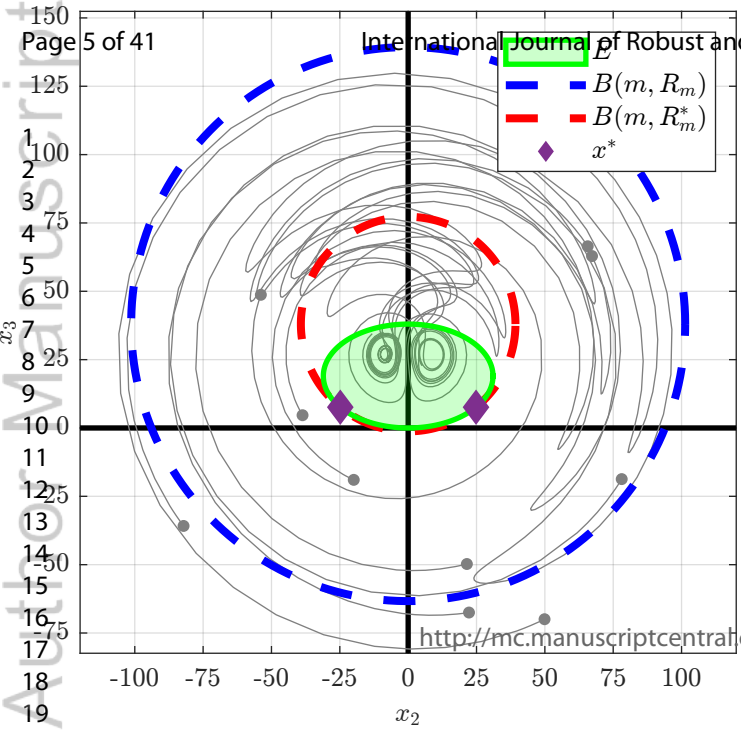
http://mc.manuscriptcentral.com/rnc-wiley

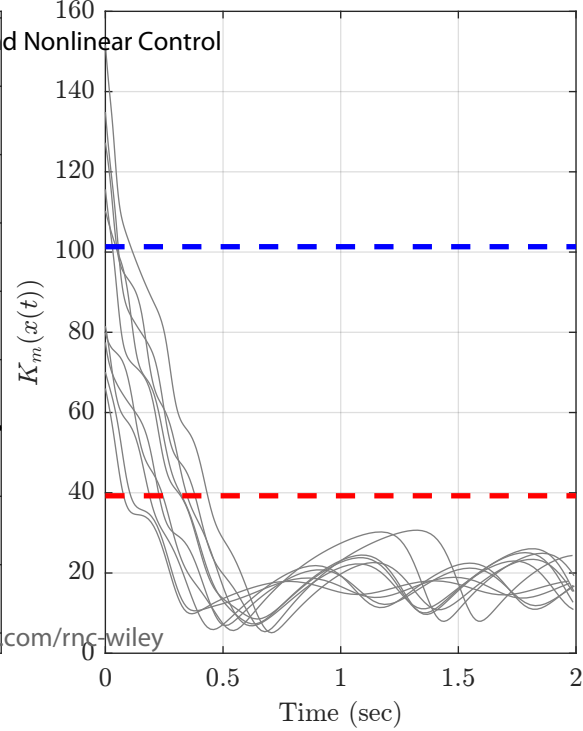
Author Manuscript



443x245mm (236 x 236 DPI)

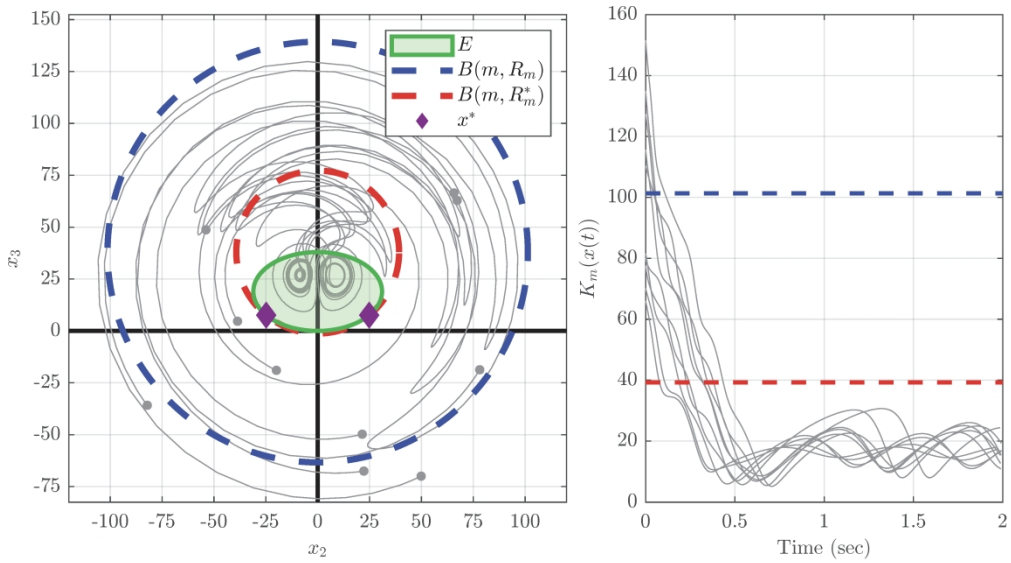
Author Manuscript





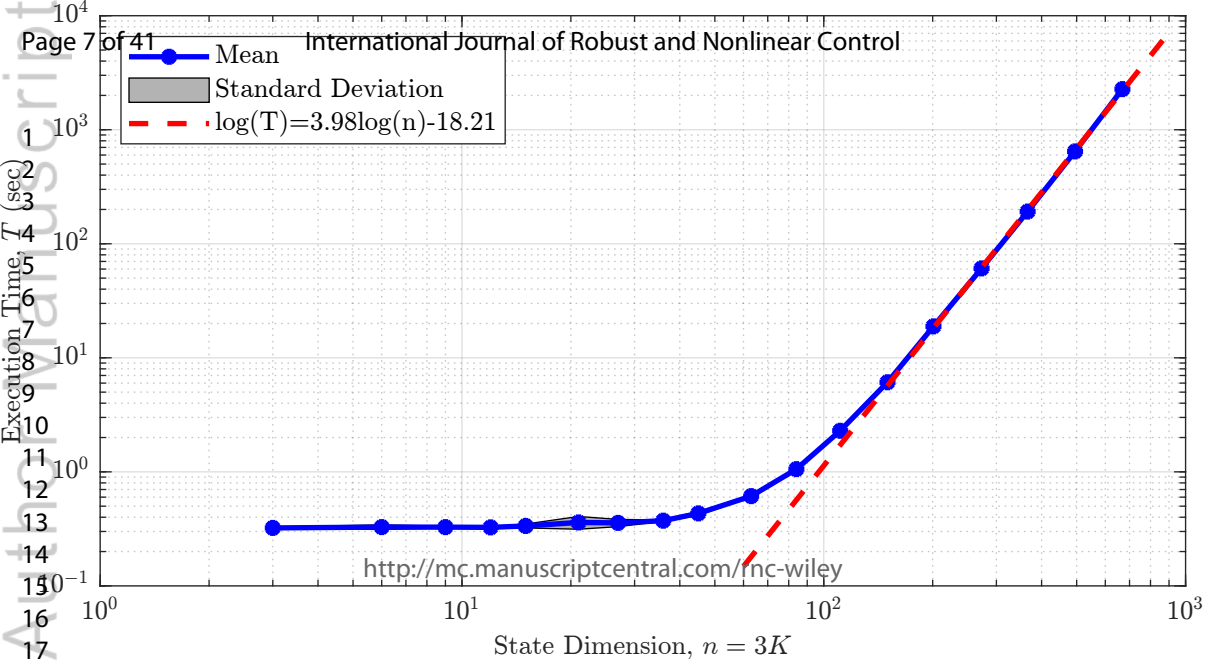
http://mc.manuscriptcentral.com/rnc-wiley

Author Manuscript



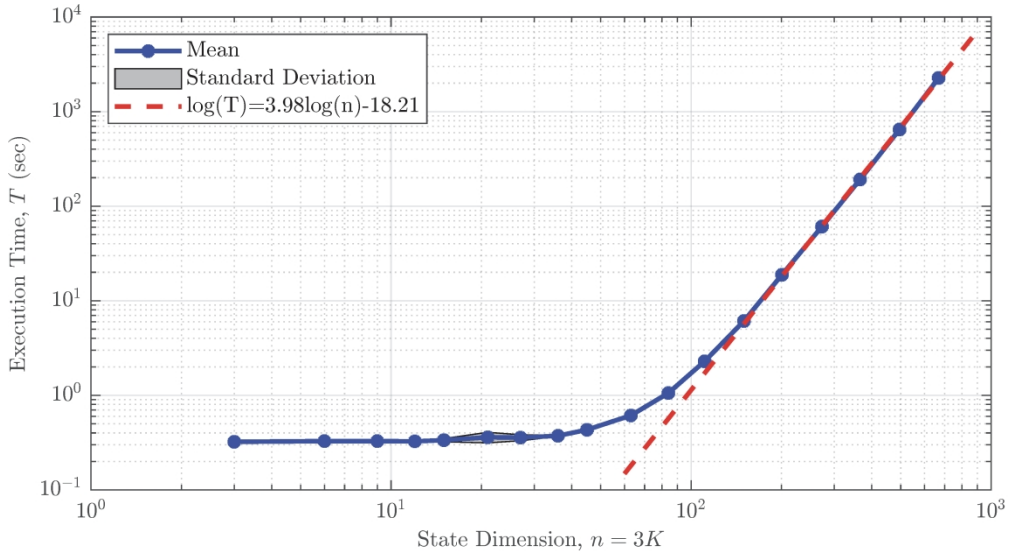
443x245mm (236 x 236 DPI)

Author Manuscript



http://mc.manuscriptcentral.com/rnc-wiley

Manuscript



405x222mm (236 x 236 DPI)

Author Manuscript

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

A Convex Optimization Approach to Compute Trapping Regions for Lossless Quadratic Systems

Shih-Chi Liao[1] | **A. Leonid Heide**[2] | **Maziar S. Hemati**[2] | **Peter J. Seiler**[1]

[1]Electircal and Computer Engineering, University of Michigan, Ann Arbor, Michigan, USA

[2]Aerospace Engineering and Mechanics, University of Minnesota, Minnesota, USA

Correspondence
Shih-Chi Liao, Electrical and Computer Engineering, University of Michigan, Ann Arbor.
Email: shihchil@umich.edu

Funding Information
This research was supported by the This material is based upon work supported by the Army Research Office under grant number W911NF-20-1-0156 and the Air Force Office of Scientific Research under grant number FA9550-21-1-0434.

Abstract

Quadratic systems with lossless quadratic terms arise in many applications, including models of atmosphere and incompressible fluid flows. Such systems have a trapping region if all trajectories eventually converge to and stay within a bounded set. Conditions for the existence and characterization of trapping regions have been established in prior work for boundedness analysis. However, prior solutions have used non-convex optimization methods, resulting in conservative estimates. In this paper, we build on this prior work and provide a convex semidefinite programming condition for the existence of a trapping region. The condition allows for precise verification or falsification of the existence of a trapping region. If a trapping region exists, then we provide a second semidefinite program to compute the least conservative radius of the spherical trapping region. Two low-dimensional systems are provided as examples to illustrate the results. A third high-dimensional example is also included to demonstrate that the computation required for the analysis can be scaled to systems of up to $\sim O(100)$ states. The proposed method provides a precise and computationally efficient numerical approach for computing trapping regions. We anticipate this work will benefit future studies on modeling and control of lossless quadratic dynamical systems.

KEYWORDS
Boundedness, Trapping Region, Lyapunov Method, Semidefinite Programming

1 | INTRODUCTION

Many physical phenomena, including those of the atmosphere[1] and incompressible fluid flows[2], can be modeled by nonlinear dynamics with lossless quadratic nonlinearities. Here "lossless" means the quadratic term neither introduces nor dissipates energy. Such models should exhibit long-term boundedness as this is a feature of the corresponding physical phenomena. Similarly, closed-loop control of lossless quadratic systems has been studied in the literature[3]. In such cases, the boundedness can be used to maintain the closed-loop system in a certain operation condition[4]. Hence, methods to verify boundedness are required for this class of models. One straightforward approach is establishing an equilibrium's global stability, entailing all trajectories to converge and be bounded. However, a system can have an unstable equilibrium yet still be bounded, e.g., the Lorenz system[1]. Thus, the typical Lyapunov stability analysis is often too strict for such systems, and a dedicated boundedness analysis is required.

A trapping region is a bounded set that all trajectories eventually converge into and stay within. This notion has been proposed in the literature as an approach to analyze boundedness[1,5]. Lorenz[1] proposed an energy function as a certificate for the existence of a trapping region to study the boundedness of the chaotic three-state system. Schlegel and Noack[5] formalized this idea with a necessary and sufficient condition for the existence of a spherical trapping region. They applied the condition to demonstrate the boundedness of several reduced-order models of incompressible fluid flows with $\sim O(10)$ state variables. Furthermore, these trapping region conditions have been exploited in physics-informed system identification algorithms[6,7] to obtain long-term bounded models of fluid flows with $\sim O(10)$ state variables from data.

While trapping regions provide a certificate of boundedness, verifying the existence of a trapping region is non-trivial. Several approaches are discussed in Section 7 of Schlegel and Noack[5]. However, they are either limited by the state dimension of systems (as they involve an algebraic condition on the eigenvalues of the system matrices) or involve non-convex optimization

Author Manuscript
For Review

relying on a good initial seed for the algorithm. The approaches offer no guarantees despite using physical understanding and simulations to aid the analysis. Thus the approaches may fail to find a trapping region even if one exists. Furthermore, even if the existence of a trapping region is verified, Schlegel and Noack provide a conservative radius estimation of the spherical trapping region. These limitations restrict the application of the boundedness analysis using trapping regions.

This paper provides three contributions to the trapping region analysis. First, a convex semidefinite programming (SDP) condition is proposed as the necessary and sufficient condition for the system to have a trapping region. This convex condition provides theoretical guarantees and corresponding efficient numerical solutions. Hence, readily available SDP solvers can be used to verify or falsify the existence of a trapping region. Second, a quadratically constrained quadratic program (QCQP) is derived to compute the least conservative radius of the spherical trapping region if one exists. This QCQP can be solved exactly by its SDP dual problem. Also, the solution of the dual SDP characterizes the set of points with instantaneously non-decreasing energy on the boundary of the trapping region. This set of points can be useful for further study via simulations. Lastly, the proposed SDP-based trapping region analysis is demonstrated by three examples. The first two examples are low-dimensional systems to illustrate the concepts and the analysis process. The last example highlights that the proposed method can scale to $\sim O(100)$ states. The SDP-based method can easily verify the existence of the trapping region and provide the least conservative estimate for the trapping region. This convex formulation can potentially benefit the relevant modeling[6,7] and control[3] approaches for lossless quadratic systems.

The rest of the paper is structured in four sections. Section 2 formulates the problem of this study and presents relevant prior work. Section 3 presents the proposed SDP-based trapping region analysis, including verification, computing the least conservative radius, and characterizing critical points. Section 4 provides numerical examples to illustrate the process and computational load of the analysis. Lastly, Section 5 summarizes the paper and discusses possible future directions.

2 | PRELIMINARIES

2.1 | Lossless Quadratic Systems

Let $c \in \mathbb{R}^n$, $L \in \mathbb{R}^{n\times n}$, and symmetric matrices $\{Q^{(i)}\}_{i=1}^n \subset \mathbb{R}^{n\times n}$ be given. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a quadratic function with i^{th} entry defined by $f_i(x) = x^\top Q^{(i)} x$ for $i = 1 \dots n$. Given this notation, define the following n-state nonlinear system:

$$\dot{x}(t) = c + Lx(t) + f(x(t)), \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector at time t. The time dependency of the state vector $x(t)$ is occasionally omitted to simplify the notation. Furthermore, we assume the quadratic nonlinearity is lossless: $x^\top f(x) = 0$ for all $x \in \mathbb{R}^n$. The lossless condition is:

$$0 = x^\top f(x) = \sum_{i=1}^{n} x_i \left(x^\top Q^{(i)} x\right) = \sum_{i,j,k=1}^{n} Q^{(i)}_{jk} x_i x_j x_k \qquad \forall x \in \mathbb{R}^n. \tag{2}$$

Note that every term in the product $x^\top f(x)$ is cubic in the entries of x. The lossless condition holds if and only if the coefficient of each term $x_i x_j x_k$ is zero. Accounting for the fact that $x_i x_j x_k = x_j x_k x_i = x_k x_i x_j$, the nonlinearity $f(x)$ is lossless if and only if the symmetric matrices $\{Q^{(i)}\}_{i=1}^n$ satisfy the following condition:

$$Q^{(i)}_{jk} + Q^{(j)}_{ik} + Q^{(k)}_{ij} = 0, \qquad \forall i, j, k \in \{1, \dots, n\}. \tag{3}$$

The nonlinear term is called "lossless" as it does not change the energy over time. Specifically, assume the energy in state x is $K_0(x) = \frac{1}{2}\|x\|_2^2$. Then the time derivative of the energy along trajectories of (1) is

$$\frac{d}{dt} K_0(x(t)) = x(t)^\top \left(c + Lx(t) + f(x(t))\right) \tag{4}$$

$$= x(t)^\top \left(c + Lx(t)\right). \tag{5}$$

The nonlinear term does not affect the time derivative of energy due to the lossless property. This property is used in subsequent sections to develop sufficient conditions for the boundedness of (1).

Manuscript
For Peer Review

One important feature of a quadratic system is that the nonlinearities are preserved under coordinate translation. Specifically, define $y := x - m$, where $m \in \mathbb{R}^n$ is a constant translation. The dynamics after translation are:

$$\dot{y}(t) = d(m) + A(m)\, y(t) + f(y(t)), \tag{6}$$

where $d(m) := c + Lm + f(m)$ and

$$A(m) := L + 2 \begin{bmatrix} m^\top Q^{(1)} \\ \vdots \\ m^\top Q^{(n)} \end{bmatrix}. \tag{7}$$

The quadratic nonlinearity in (6) is exactly the same as in (1). Hence the nonlinearity in (6) is also lossless. The symmetric part of $A(m)$ is given by $A_s(m) := \frac{1}{2}(A(m) + A(m)^\top)$. This symmetric part plays an important role in the boundedness analysis introduced later. It can be expressed simply as:

$$A_s(m) = L_s - \sum_{k=1}^{n} m_k Q^{(k)}, \tag{8}$$

where $L_s = \frac{1}{2}(L + L^\top)$ is the symmetric part of L and m_k is the kth element of the coordinate shift m. Equation 8 is due to the lossless property of the nonlinearity (3):

$$\begin{bmatrix} m^\top Q^{(1)} \\ \vdots \\ m^\top Q^{(n)} \end{bmatrix}_{i,j} + \begin{bmatrix} Q^{(1)} m & \dots & Q^{(n)} m \end{bmatrix}_{i,j} = \sum_{k=1}^{n} m_k Q^{(i)}_{jk} + \sum_{k=1}^{n} m_k Q^{(j)}_{ik} = -\sum_{k=1}^{n} m_k Q^{(k)}_{ij}. \tag{9}$$

2.2 | Boundedness of a System

Let $\phi(t; x_0, t_0)$ denote the solution of (1) at time $t \geq t_0$ starting from the initial condition $x_0 \in \mathbb{R}^n$ at time t_0. As the stability of a nonlinear system is hard to analyze in general, the boundedness of solutions $\phi(t; x_0, t_0)$ serves as an alternative notion to characterize the behavior of a system. The boundedness of the system can be defined as in Chapter 4.8 of Khalil[4].

Definition 1 (Boundedness)**.** The solutions of system (1) are globally uniformly ultimately bounded if there exists a scalar $\beta > 0$, independent of $t_0 \geq 0$, and a function $T : \mathbb{R} \to \mathbb{R}$ such that:

$$\|\phi(t; x_0, t_0)\|_2 \leq \beta \qquad \forall x_0 \in \mathbb{R}^n \text{ and } \forall t \geq t_0 + T(\|x_0\|_2). \tag{10}$$

In words, a system is globally uniformly ultimately bounded with bound $\beta > 0$ if the trajectory from any initial condition eventually converges to within a spherical ball centered at the origin and with radius β. Moreover, $T(\|x_0\|_2)$ is the time required to converge within the bound β. This time is independent of t_0 but depends on the norm of the initial condition. In this paper, only globally uniformly ultimate boundedness is investigated and we will simply refer to this as the "boundedness" in the remainder of the paper.

Note that the boundedness of the shifted system (6) is equivalent to the boundedness of the original system (1) as the coordinates only differ by translation. In addition, the triangle inequality implies $\|x(t)\|_2 \leq \|y(t)\|_2 + \|m\|_2$. Hence, if a trajectory in shifted coordinates is bounded by β, i.e. $\|y(t)\|_2 \leq \beta$, then the corresponding trajectory in the original coordinates is bounded by $\beta + \|m\|_2$. Hence, one can show the boundedness of the original system (1) by showing the boundedness of the shifted system (6).

2.3 | A Sufficient Condition for Boundedness

This section reviews a sufficient condition to certify the boundedness of quadratic systems (1). This section is mainly based on early results discussed by Lorenz[1], which were further formalized by Schlegel and Noack[5]. First, we introduce the notion of a monotonically attracting trapping region as defined in Schlegel and Noack[5].

Manuscript
For Peer Review

Definition 2 (Trapping region)**.** A trapping region $D \subseteq \mathbb{R}^n$ is a compact set that is forward invariant, i.e., if $x(t_0) \in D$ then $x(t) \in D$ for all $t \geq t_0$. In other words, once a trajectory enters a trapping region, it remains in the trapping region for all future time. A trapping region is termed globally monotonically attracting if an energy function is strictly monotonically decreasing along all trajectories starting from an arbitrary state outside of D.

The existence of a monotonically attracting trapping region for a system implies that the system is bounded, as all trajectories will enter the compact set D. We will focus on globally monotonically attracting trapping regions and will refer to this simply as a "trapping region." Furthermore, we will only consider spherical regions as candidates of a trapping region in this paper.

Section 2 of Lorenz's seminal 1963 paper[1] provides a condition for the existence of a trapping region for the dynamical system (1). The condition uses the energy and states that a trapping region exists if $L_s < 0$. A more precise characterization for a trapping region is presented in Schlegel and Noack[5]. These results are summarized in the following lemma with the definition of a closed-ball with center ν and radius R given by $B(\nu, R) := \{x \in \mathbb{R}^n \,:\, \|x - \nu\|_2 \leq R\}$.

Lemma 1. *If $L_s \prec 0$, a trapping region exists, and the system* (1) *is bounded. Moreover, if λ_1 is the largest (least negative) eigenvalue of L_s and $R_0 := \frac{\|c\|_2}{|\lambda_1|}$, then $B(0, R_0)$ is a trapping region.*

Proof. The energy was previously defined in Section 2.1 as $K_0 := \frac{1}{2}\|x\|_2^2$. The derivative of this energy function along trajectories of the system (1) is:

$$\frac{d}{dt}K_0\left(x(t)\right) := \frac{1}{2}\left(x(t)^\top \dot{x}(t) + \dot{x}(t)^\top x(t)\right) = c^\top x(t) + x(t)^\top L_s x(t). \tag{11}$$

As noted previously, the derivative of $K_0(x(t))$ does not depend on the quadratic dynamics f as the nonlinearity is lossless.

Assume $L_s \prec 0$. The energy evolution (11) is dominated by the quadratic term and is negative if $\|x\|_2$ is sufficiently large. The energy can increase only when it is close to the origin. Therefore, a trapping region around the origin exists as all trajectories converge toward the origin or stay close to the origin. In summary, if Ls is negative definite, a trapping region exists, and the system is bounded.

Next, the time derivative of K_0 is bounded using the Cauchy-Schwartz inequality and the largest (least negative) eigenvalue λ_1:

$$\frac{d}{dt}K_0(x(t)) \leq \|c\|_2\,\|x(t)\|_2 + \lambda_1\|x(t)\|_2^2 = \left(\|c\|_2 + \lambda_1\|x(t)\|_2\right)\cdot\|x(t)\|_2. \tag{12}$$

Thus, the derivative of K_0 is negative if $\|x(t)\|_2 > \frac{\|c\|_2}{|\lambda_1|}$. Hence $B(0, R_0)$ is a trapping region for the system (1) in original coordinates with the radius $R_0 := \frac{\|c\|_2}{|\lambda_1|}$. □

Lemma 1 provides a sufficient condition to check the boundedness of the system (1). The condition can be easily verified by eigenvalue computation. Note that the condition is independent of the equilibrium point and depends only on c and L_s. This allows analyzing the boundedness of system (1) even when it has multiple equilibrium points or chaotic behavior. Furthermore, if a trapping region exists, all equilibrium points would be inside the region as the energy evolution at an equilibrium is always zero.

Nonetheless, the radius bound in Lemma (1) is conservative in general and is attained only when the vector c aligns exactly with the eigenvector corresponding to λ_1. This is due to the derivation using Cauchy-Schwartz inequality in (12). To demonstrate the conservatism, consider the set of states where the energy is (instantaneously) non-decreasing, $E := \left\{x(t) \in \mathbb{R}^n \,:\, \frac{d}{dt}K_0(x(t)) \geq 0\right\}$. We can characterize E using completion of squares to rewrite (11) as follows:

$$\frac{d}{dt}K_0(x(t)) = \left(x(t) + \frac{1}{2}L_s^{-1}c\right)^\top L_s \left(x(t) + \frac{1}{2}L_s^{-1}c\right) - \frac{1}{4}c^\top L_s^{-1}c. \tag{13}$$

The boundary of E corresponds to the states where $\frac{d}{dt}K_0(x(t)) = 0$. It follows from (13) that this boundary is an ellipsoid.† The point $x(t) = -\frac{1}{2}L_s^{-1}c \in E$ is the ellipsoid center and is the point of maximum energy growth: $\frac{d}{dt}K_0(x(t)) = -\frac{1}{4}c^\top L_s^{-1}c > 0$ (assuming $c \neq 0$). By Lemma 1, $B(0, R_0)$ is a trapping region, where the energy is strictly decreasing outside this ball. Hence $E \subseteq B(0, R_0)$.

† The matrix L_s is symmetric and negative definite. Thus, it has an eigenvalue decomposition $L_s = U\Lambda U^\top$ where U is orthogonal and $\Lambda < 0$ is a diagonal matrix with the eigenvalues $\lambda_n \leq \cdots \leq \lambda_1 < 0$ along the diagonal. Define the rotated coordinates $z(t) := U^\top x(t)$. The condition $\frac{d}{dt}K_0(x(t)) = 0$ can be written in these rotated coordinates as $(z(t) - v)^\top \Lambda\,(z(t) - v) = \frac{1}{4}c^\top L_s^{-1}c$ where $v := -\frac{1}{2}U^\top L_s^{-1}c \in \mathbb{R}^n$ is the ellipsoid center in the rotated coordinates. This can be further simplified to $\sum_{i=1}^n \frac{1}{\alpha_i^2}(z_i - v_i)^2 = 1$ where $\alpha_i := \frac{1}{2}\sqrt{\frac{c^\top L_s^{-1}c}{\lambda_i}}$ are the semi-axis lengths.

Author Manuscript For Peer Review

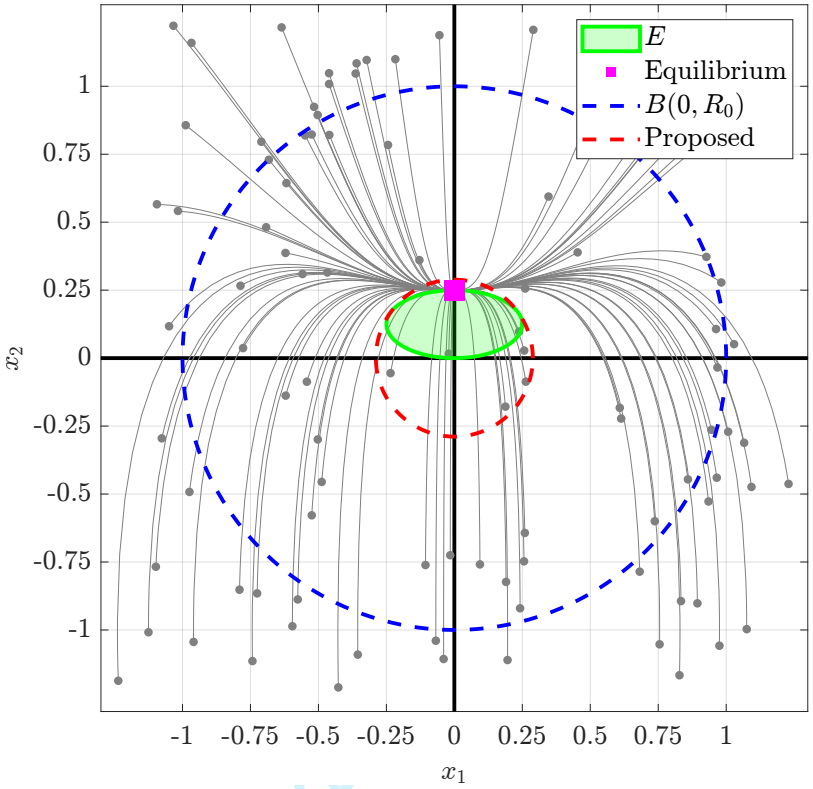


FIGURE 1 An illustration of trapping region for the two-state system (14) with kinetic energy $K_0(x) = \frac{1}{2}x^\top x$. All trajectories with random initial conditions (gray) converge to the equilibrium point at $(0, 0.25)$ (magenta square). The set of states with non-decreasing energy E (green solid ellipsoid) and a conservative trapping region $B(0, R_0)$ (blue-dashed circle) are shown. The tightest trapping region centered at the origin can be computed by our proposed method (Section 3.2) and is shown in the red-dashed circle.

However, $B(0, R_0)$ is not necessarily the tightest one that contains E as pointed out. In fact, any forward-invariant set containing E renders a trapping region as the energy is decreasing for any states outside of E. In Section 3.2, we will derive a convex optimization problem to reduce conservatism and find the tightest trapping region containing the non-decreasing energy set E.

The following two-state system with lossless quadratic nonlinearities serves as an example to illustrate the boundedness analysis using Lemma 1 and its conservatism:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix} x + \begin{bmatrix} -x_1 x_2 \\ x_1^2 \end{bmatrix}. \tag{14}$$

The trajectories (light gray lines) in Figure 1 are generated with random initial conditions and converge to the single stable equilibrium $(0, 0.25)$. The constant vector in the dynamics is $c = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ and the largest (least negative) eigenvalue of L_s is $\lambda_1 = -1$. Hence, $B(0, R_0)$ is a trapping region with $R_0 = \frac{\|c\|_2}{|\lambda_1|} = 1$ as derived above. This is shown as the blue-dashed circle in Figure 1. All trajectories eventually converge inside this ball as expected. The green ellipsoid in this figure corresponds to the non-decreasing energy states E. Based on (13), the boundary of this ellipse is defined by:

$$\frac{d}{dt}K_0(x(t)) = -x_1^2 - 4\left(x_2 - \frac{1}{8}\right)^2 + \frac{1}{16} = 0.$$

As shown in Figure 1, the circle $B(0, R_0)$ is a conservative estimate as any ball centered at the origin containing the ellipsoid E is a trapping region. Note that this conservatism is rooted in the derivation using Cauchy-Schwartz inequality in (12). We will derive in Section 3.2 a convex optimization to compute the tightest ball containing the ellipsoid E. This tightest (least conservative) ball is shown as the red-dashed circle in Figure 1.

Author Manuscript For Review

Author Manuscript

2.4 | Boundedness Using Coordinate Translations

The previous subsection presented a sufficient condition for boundedness using the energy centered at the state origin. Schlegel and Noack[5] further generalized the condition using coordinate translations. Consider the translated system (6) with $y(t) := x(t) - m$. The energy for the translated system is $K_m(y) := \frac{1}{2}\|y\|_2^2$ with corresponding time derivative:

$$\frac{d}{dt}K_m(y(t)) := \frac{1}{2}\left(y(t)^\top \dot{y}(t) + \dot{y}(t)^\top y(t)\right) = d(m)^\top y(t) + y(t)^\top A_s(m) y(t). \tag{15}$$

Using the same argument as in Section 2.3, if the matrix $A_s(m)$ is negative definite, then the shifted system (6) has a trapping region $B(0, R_m)$ in the shifted coordinate for some R_m. This implies the original system (1) is bounded by $\|m\|_2 + R_m$. Furthermore, a valid radius is $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$ where λ_1 is the largest (least negative) eigenvalue of $A_s(m)$. The following theorem in Schlegel and Noack[5] formalizes these discussions for monotonically attracting trapping regions.

Theorem 1 (Theorem 1 of Schlegel and Noack[5])**.** *The system* (1) *has a monotonically attracting trapping region* $B(m, R_m)$ *if and only if there exists an* $m \in \mathbb{R}^n$ *such that the real, symmetric matrix* $A_s(m)$ *is negative definite. If* $A_s(m)$ *has eigenvalues* $\lambda_n \leq \cdots \leq \lambda_1 < 0$*, then a trapping region is given by* $B(m, R_m)$ *with the radius* $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$.

A formal proof of this necessary and sufficient trapping region condition is given in Appendix B of Schlegel and Noack[5]. Note that the radius R_m given in Theorem 1 is different from the statement in Schlegel and Noack[5] as there is a minor typo that is also pointed out in Kaptanoglu et al[6]. Also, the proof in Schlegel and Noack[5] regarding the estimated radius R_m used a geometrical argument instead of the Cauchy-Schwartz inequality approach presented in (12). The two estimates are the same, and the Cauchy-Schwartz inequality is adopted in this paper for simplicity.

The key point of Theorem 1 is that the boundedness of the system (1) can be determined by an algebraic condition. In other words, if m exists such that $A_s(m)$ is negative definite, then the system has a trapping region $B(m, R_m)$. This corresponds to a bound R_m in the shifted coordinates and $\|m\|_2 + R_m$ in the original coordinates. This approach does not require solving the nonlinear differential equation for all possible initial conditions. Note that the converse is not true, i.e., a bounded system does not imply the existence of a trapping region. Consider a trivial system $\dot{x} = 0$. All solutions are stationary and thus bounded. However, this system does not have a monotonically attracting trapping region, i.e., one cannot find an m such that $A_s(m)$ is negative definite. If no trapping region exists, then Schlegel and Noack[5] point out that additional analysis of the linear and nonlinear terms is needed to determine whether the system is bounded.

Even though Theorem 1 provides an approach to verify the boundedness of the system, there are two remaining caveats. First, verifying this algebraic condition is non-trivial as it involves a search over $m \in \mathbb{R}^n$. An inf-sup problem is formulated in Schlegel and Noack[5] to minimize the largest eigenvalue of $A_s(m)$ and simulated annealing is proposed as a solution method. However, this formulation is a non-convex optimization, and hence there are no guarantees of convergence to the global optimum. Thus the test is inconclusive if the simulated annealing fails to find a feasible m with the largest eigenvalue of $A_s(m)$ being negative. In other words, a trapping region may exist even if the simulated annealing algorithm fails. The second caveat of Theorem 1 is that the estimated bound R_m is an approximate upper bound and is usually conservative. As mentioned in Section 2.3, the approximation is made by the analysis of eigenvalues of symmetric matrix $A_s(m)$ with the worst-case alignment of direction $d(m)$ in the Cauchy-Schwartz inequality (12). Hence, the upper bound approximation R_m is often conservative. We address both of these caveats in the next section by providing: (i) a convex optimization to compute the trapping region, and (ii) a tight bound on the trapping region radius.

3 | COMPUTATION OF TRAPPING REGIONS VIA CONVEX OPTIMIZATION

This section revisits the boundedness condition in Theorem 1. First, the condition is reformulated as a constraint in the form of a linear matrix inequality (LMI), whose feasible set is a convex set. The feasibility of this LMI condition can be tested numerically, and a feasible solution, if one exists, verifies the existence of a trapping region. This avoids the non-convex inf-sup formulation in Schlegel and Noack[5]. Second, a quadratically constrained quadratic program (QCQP) is derived to compute the exact radius of the bounded region without conservative approximation. Lastly, the set of furthest points with non-decreasing energy can be identified by the dual solution of the QCQP and shown to be a hypersphere.

For Peer Review

3.1 | Trapping Condition as a Convex Constraint

By Theorem 1, the system (1) has a trapping region if and only if there exists $m \in \mathbb{R}^n$ such that $A_s(m)$ is negative definite, i.e.,

$$A_s(m) = L_s - \sum_{i=1}^{N} m_i Q^{(i)} \prec 0. \tag{16}$$

Observe that $A_s(m)$ is an affine function of m. Hence, the negative definite constraint (16) is a linear matrix inequality (LMI) in the variable m[8,9]. LMIs have been widely studied in the optimization community and are known to have a convex feasible set, i.e., a convex combination of two solutions is also a solution (Chapter 2.3.2 of Boyd and Vandenberghe[8]). Hence, the set of $m \in \mathbb{R}^n$ that satisfies (16) is a convex set. LMI feasibility problems can be efficiently solved with many numerical optimization solvers, e.g., SeDuMi[10], MOSEK[11], and SDPT3[12]. There is also software that enables easy implementation and interface with these solvers, e.g., CVX[13]. Finally, LMI feasibility problems have useful duality results such as the theorem of alternatives (Chapter 5.8 of Boyd and Vandenberghe[8]). Thus, the numerical algorithm will return a coordinate shift $m \in \mathbb{R}^n$ satisfying (16) (and hence verifying the existence of a trapping region) or it will return a dual variable that certifies that no such feasible m exists.

The following lemma states the application of this LMI condition to verify the existence of a trapping region in system (1):

Lemma 2. *Consider system* (1) *and a semidefinite programming (SDP) to minimize the largest eigenvalue of* $A_s(m)$*:*

$$\begin{aligned} a^* = \min_{m \in \mathbb{R}^n, a \in \mathbb{R}} \quad & a \\ \text{s.t.} \quad & A_s(m) \preceq a I_n. \end{aligned} \tag{17}$$

A trapping region exists and the system (1) *is bounded if the optimal cost of the SDP* (17) *satisfies* $a^* < 0$*. If* $a^* \geq 0$*, there is no trapping region in the system* (1)*.*

Proof. The optimization (17) has a single LMI constraint with the linear objective given by the slack variable a. It falls into the category of SDPs (Chapter 4.6.2 of Boyd and Vandenberghe[8]). In addition, the SDP (17) is always feasible by choosing any $m \in \mathbb{R}^n$ and a sufficiently large value of a. Finally, there exists an m such that $A_s(m)$ is negative definite if the optimal cost $a^* < 0$. By Theorem 1, the system (1) has a trapping region and is bounded. If $a^* \geq 0$, there is no feasible m such that LMI (16) is satisfied. By Theorem (1), no trapping region exists in the system (1). □

SDPs are convex optimizations that can be efficiently solved to the global optimum (within numerical tolerances) using numerical solvers. This particular SDP (17) can be solved on a standard desktop for problems with dimensions up to $n \sim O(100)$. Specialized implementations can be made for larger problems. In contrast to the inf-sup formulation proposed in Schlegel and Noack[5], this SDP formulation is guaranteed to converge to the global optimum and provide a conclusive trapping region analysis.

3.2 | Computing the Tightest Trapping Region Radius with SDP

Assume the SDP (17) has an optimal val $a^* < 0$ so that there exists an m such that $A_s(m) \prec 0$. By Theorem 1, the system (1) has a monotonically attracting trapping region. Let λ_1 denote the largest (least negative) eigenvalue of $A_s(m)$. A conservative bound on the trapping region is given by $B(m, R_m)$ with the radius $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$. In this subsection, we utilize duality in optimization theory to compute the tightest (least conservative) radius R_m^* such that $B(m, R_m^*)$ is a trapping region for a certain choice of m. The discussion of choosing the best m to obtain the smallest R_m^* is briefly addressed at the end of the section.

If $\frac{d}{dt} K_m(y(t)) \geq 0$, then the energy is instantaneously non-decreasing. It follows from (15) that the set of all states where the energy is instantaneously non-decreasing (in shifted coordinates) is:

$$E := \left\{ y \in \mathbb{R}^n \,:\, d(m)^\top y + y^\top A_s(m) y \geq 0 \right\}. \tag{18}$$

Author Manuscript For Peer Review

The tightest (least conservative) radius R_m^* corresponds to the smallest ball (in shifted coordinates) that contains all states in E. Thus the smallest radius is given by:

$$\begin{aligned} R_m^* := \min_{R_m} \quad & R_m \\ \text{s.t.} \quad & E \subseteq B(0, R_m). \end{aligned}$$

Equivalently, we can find the state in E with the largest squared norm, i.e., maximize $\|y\|_2^2$ subject to $y \in E$. This equivalent formulation leads to the following optimization:

$$(R_m^*)^2 = \max_y \quad f_0(y) := y^\top y \tag{19a}$$

$$\text{s.t.} \quad f_1(y) := d(m)^\top y + y^\top A_s(m) y \geq 0. \tag{19b}$$

Note that (19) is a non-convex problem as we are maximizing a convex objective. Non-convex problems can, in general, have multiple local optima, and solving for global optima can be difficult.

In fact, (19) is a quadratically constrained quadratic program (QCQP) with only one quadratic constraint. This particular problem, while nonconvex, can be efficiently solved for the global optimum. The following lemma provides a convex SDP to compute the tightest radius R_m^* by solving QCQP (19):

Lemma 3. *Given system* (1) *and an m such that* $A_s(m) \prec 0$. *The following convex SDP exactly solves the QCQP* (19)*:*

$$(R_m^*)^2 = \min_{\lambda \geq 0, \gamma} \quad \gamma \tag{20a}$$

$$s.t. \quad f_2(\lambda, \gamma) := \begin{bmatrix} I + \lambda A_s(m) & \frac{\lambda}{2} d(m) \\ \frac{\lambda}{2} d(m)^\top & -\gamma \end{bmatrix} \preceq 0. \tag{20b}$$

Then, $B(m, R_m^*)$ *is the tightest (least conservative) trapping region of system* (1) *that centers at m.*

Proof. The SDP (20) is the Lagrange dual of the QCQP (19) (Appendix B of Boyd and Vandenberghe[8]). Two cases are discussed to show the optimal cost for this dual SDP (20) is identical to the optimal cost of the primal QCQP (19). First, if $d(m) \neq 0$, the strong duality holds between QCQP (19) and dual SDP (20). In particular, the point $y_0 = -\frac{1}{2} \frac{d(m)^\top d(m)}{d(m)^\top A_s(m) d(m)} d(m)$ is strictly feasible for the QCQP (19), i.e. $f_1(y_0) < 0$. Hence, the QCQP satisfies Slater's constraint qualification and the strong duality holds (Appendix B of Boyd and Vandenberghe[8]). Second, if $d(m) = 0$, then the constraint $f_1(y) \leq 0$ does not have a constant term and is not strictly feasible. The only feasible point of the QCQP (19) in this case is $y_0 = 0$, and therefore the optimal cost is $R_m^* = 0$. The dual (20) has the same optimal cost of 0 by the solution of $\gamma = 0$ with $\lambda > 0$ sufficiently large such that $I + \lambda A_s(m) \preceq 0$. This is possible since $A_s(m) \prec 0$. Thus, QCQP and dual SDP achieve the same cost.

Finally, $(R_m^*)^2$ is both the optimal cost of the primal problem (19) and the dual (20). The primal form implies that there exists a state y^* with $\|y^*\|_2 = R_m^*$ such that $f_1(y^*) = 0$ and it is contained in the trapping region. Any state y_0 such that $\|y_0\|_2 > R_m^*$ must have $f_1(y_0) < 0$. In other words, $\frac{d}{dt} K_m(y(t)) < 0$ for any state $y(t) = y_0$ with $\|y_0\|_2 > R_m^*$. Hence, the trapping region $B(m, r_m^*)$ of system (1) is the tightest (least conservative) trapping region centered at m.

□

Lemma 3 provides a computationally feasible approach to solve for the global optimum. The dual SDP (20) can be solved by the numerical solvers mentioned above. Note that for the case $d(m) = 0$, the optimal cost $R_m^* = 0$ and the smallest trapping region is the origin in the shifted coordinates. Thus, $A_s(m) \prec 0$ and $d(m) = 0$ implies that the state m is a globally stable equilibrium point for the system (1).

In summary, we can characterize, with no conservatism, the smallest radius R_m^* such that the trapping region $B(m, R_m^*)$ contains all the states where the energy is instantaneously non-decreasing. All trajectories of the original system (1) will eventually converge to and be tightly bounded in the region $B(m, R_m^*)$. The global optimum R_m^* from the dual SDP (20) provides a tighter bound than the value $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$ given in Theorem 1. The discussion in this section applies to any m such that $A_s(m) \prec 0$. It can also be applied to the optimal m^* obtained from the SDP (17) assuming it has an optimal cost $a^* < 0$.

Note that one might potentially want to find the tightest trapping region among all possible coordinate shifts. This corresponds to jointly optimizing the problem (19) over (m, y) with the additional constraint (16) such that a trapping region exists. Recall that $d(m) = c + Lm + f(m)$. Hence, constraints of (19) have a term $f(m)^\top y$ that is cubic in the variables (m, y). The problem no longer

Author Manuscript For Peer Review

has the same strong duality results, in general. We are not aware of an efficient approach to jointly optimize the problem (19) over (m, y). Nonlinear optimization algorithms can be applied, but they are not guaranteed to converge to the global optima. Further investigation of solving this problem is left to future work.

3.3 | Recovering Furthest Points with Non-Decreasing Energy

In this subsection, we characterize the set of optimal solutions y^* of the QCQP (19). These correspond to states y^* with norm of R_m^* and non-decreasing energy. The case $d(m) = 0$ has the trivial optimal solution $y^* = 0$ as discussed in the previous subsection. This subsection focuses on the case of $d(m) \neq 0$.

Strong duality holds between the QCQP (19) and the dual SDP (20) when $d(m) \neq 0$. Hence any primal and dual optimal solutions must satisfy the first-order optimality conditions. Specifically, any primal and dual optimal variables $(y^*, \gamma^*, \lambda^*)$ must satisfy the following Karush–Kuhn–Tucker (KKT) conditions (Chapter 5.5.3 of Boyd and Vandenberghe[8]):

$$\textit{Stationary condition:} \qquad \nabla_y \left[f_0(y) + \lambda^* f_1(y)\right]\big|_{y=y^*} = 0 \tag{21a}$$

$$\textit{Primal feasibility:} \qquad f_1(y^*) \leq 0 \tag{21b}$$

$$\textit{Dual feasibility:} \qquad \lambda^* \geq 0, \quad f_2(\lambda^*, \gamma^*) \succeq 0 \tag{21c}$$

$$\textit{Complementary slackness:} \qquad \lambda^* f_1(y^*) = 0 \tag{21d}$$

The stationary condition (21a) for the given f_0 and f_1 is:

$$0 = \nabla_y \left[f_0(y) + \lambda^* f_1(y)\right]\big|_{y=y^*} = 2(I + \lambda^* A_s(m))y^* + \lambda^* d(m).$$

Let r denote the rank of $(I + \lambda^* A_s(m))$ and $V \in \mathbb{R}^{n\times(n-r)}$ be a matrix whose columns are orthonormal and span the null space of $(I + \lambda^* A_s(m))$. Then for any optimal point y^* there exists $c \in \mathbb{R}^{n-r}$ such that:

$$y^*(c) = -\frac{\lambda^*}{2}(I + \lambda^* A_s(m))^\dagger d(m) + Vc,$$

where $(\cdot)^\dagger$ is the pseudo-inverse of a matrix.

Next, observe that the dual constraint $f_2(\lambda^*, \gamma^*) \succeq 0$ in (21c) is infeasible when $\lambda^* = 0$. Hence, the dual constraints (21c) imply that $\lambda^* > 0$. The complementary slackness condition (21d) then further implies that $f_1(y^*(c)) = 0$. Note that $f_1(y^*(c)) = 0$ also implies the primal feasibility (21b). Therefore, the primal-dual point $(y^*(c), \lambda^*, \gamma^*)$ satisfies KKT conditions (21) for any $c \in \mathbb{R}^{n-r}$. The primal points $y^*(c)$ are only candidates for the optimal solution of the QCQP (19) as the KKT conditions are only necessary for optimality. The null space vector c^* yields an optimal point if $y^*(c^*)$ achieves the optimal cost:

$$(R_m^*)^2 = f_0(y^*(c^*)) = \frac{(\lambda^*)^2}{4} d(m)^\top \left((I + \lambda^* A_s(m))^\dagger\right)^2 d(m) + (c^*)^\top c^*. \tag{22}$$

Therefore, the set of optimal solutions $y^*(c^*)$ forms a $(n-r)$-dimensional hypersphere with radius of $\|c^*\|_2$:

$$\left\{ y^*(c^*) = -\frac{\lambda^*}{2}(I + \lambda^* A_s(m))^\dagger d(m) + Vc^* \;:\; \|c^*\|_2^2 = (R_m^*)^2 - \frac{(\lambda^*)^2}{4} d(m)^\top \left((I + \lambda^* A_s(m))^\dagger\right)^2 d(m) \right\}, \tag{23}$$

where the first term of $y^*(c^*)$ is the center of the hypersphere, columns of V are $(n-r)$ directions of the hypersphere in n-dimensional space. This $(n-r)$-dimensional hypersphere corresponds to the points y^* furthest from the origin of the shifted system (6) where the energy $K_m(y^*(c^*))$ is instantaneously non-decreasing.

3.4 | Comparison to Existing Results

The following theorem summarizes the proposed convex optimization method for computing a trapping region.

Theorem 2. *The system* (1) *has a monotonically attracting trapping region* $B(m, R_m^*)$ *if and only if the optimal cost of SDP* (17) *satisfies* $a^* < 0$. *Moreover, if* $a^* < 0$ *then there exists at least one* m *such that* $A_s(m) \prec 0$. *The tightest trapping region centered at any such* m *is given by* $B(m, R_m^*)$ *where* $(R_m^*)^2$ *is the optimal cost of SDP* (20).

Author Manuscript For Peer Review

Proof. The first statement follows from Lemma 2. The second statement follows from Lemma 3. □

The proposed method is to first solve the SDP (17) and, if $a* < 0$, then solve for R_m^* using the SDP in (20). As an aside, the SDP (20) also yields the optimal Lagrange multiplier λ^*. This can be used to compute the points on the boundary of $B(m, R_m^*)$ with non-decreasing energy. These correspond to the points $x^* = y^*(c) + m$ with $y*$ in the set (23).

As a comparison, this paper has three key differences to Theorem 1 (result in Schlegel and Noack[5]). First, this paper provides a computational approach to verify the existence of a trapping region with guarantees, while Schlegel and Noack's non-convex approach does not provide such guarantees. Second, the tightest trapping region is identified in this paper instead of an upper bounding radius in Schlegel and Noack. Third, this paper characterizes the set of points on the boundary of the tightest trapping region with non-decreasing energy, which is not discussed in Schlegel and Noack.

4 | NUMERICAL EXAMPLES

Here, we provide numerical examples to illustrate the proposed SDP-based analysis. First, the trapping region of the two-dimensional example system (14) is revisited. Second, the Lorenz attractor, which highlights the complex dynamics that can arise from models with lossless quadratic nonlinearities, is presented as a benchmark to compare our method against that given in Schlegel and Noack[5]. Lastly, a system with an adjustable number of states is used to study the scalability of the proposed method. All analyses are implemented in Matlab with CVX[13] and SDP solver MOSEK[11]. The examples are run on a standard desktop with an Intel(R) i7-9700 CPU @ 3.00GHz and 16 GB memory. The complete source code for the implementation and examples can be found on GitHub at https://github.com/SCLiao47/Boundedness_LosslessQuadSys.

4.1 | Bounded Two-Dimensional System

Recall the two-dimensional system previously introduced in Section 2.3:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix} x + \begin{bmatrix} -x_1 x_2 \\ x_1^2 \end{bmatrix}. \tag{24}$$

The linear symmetric part is $L_s = \left[\begin{smallmatrix} -1 & 0 \\ 0 & -4 \end{smallmatrix}\right]$. If there is zero coordinate shift, i.e., $m = 0$, then $A_s(m) = L_s$ is negative definite. Hence, the system is bounded by Theorem 1, and a conservative estimate on the trapping region radius was given as $R_0 = \frac{\|c\|_2}{|\lambda_1|} = 1$. The optimal cost for the SDP (17) satisfies $a^* < 0$ because $A_s(m) \prec 0$ at $m = 0$. Moreover, we can use Theorem 2 to compute the tightest trapping region $B(0, R_0^*)$. The optimal solution of the SDP (20) is $R_0^* = 0.289$ with $\lambda^* = 1$. Lastly, the critical points computed by (23) are $x^* = y^* = [\pm 0.236,\ 1.67]^\top$. Note that $I + \lambda^* L_s = \left[\begin{smallmatrix} 0 & 0 \\ 0 & -3 \end{smallmatrix}\right]$ is rank one. Therefore, the set of critical points forms a one-dimensional hypersphere, i.e., two endpoints of an interval in $\mathbb{R}^2$.

Figure 2 illustrates these results. The left plot shows state trajectories (light gray) in the state space starting from random initial conditions. All trajectories eventually converge into both valid trapping regions $B(0, R_0)$ and $B(0, R_0^*)$. Note that the estimated radius R_0 is an over-estimation, and the radius R_0^* is the tightest radius that contains the ellipsoid E with nondecreasing energy (18) in green. Two critical points x^* are marked by purple diamonds. The critical points have norm R_0^* and correspond to the points in E that are furthest from the origin. The right plot shows the energy versus time evaluated along the state trajectories shown in the left plot. All energy trajectories eventually decrease below R_0^* and well below R_0, as shown by the analyses. Overall, the proposed SDP-based analysis provides a more accurate characterization of the trapping region than Theorem 1.

Note that this system has a single equilibrium point at $x_{eq} := (0, 0.25)$, and this is marked by a magenta rectangle on the left plot of Figure 2. The energy $\frac{1}{2}\|x - x_0\|_2^2$ is a Lyapunov function that proves this equilibrium point is globally asymptotically stable. This further implies that the system (14) is bounded. However, this is not true in general for lossless quadratic systems, i.e., they can have unstable equilibrium points and still have bounded trajectories (see next example). Moreover, the proposed SDP-based method can be used to verify the boundedness of lossless quadratic systems without directly computing any equilibrium point (as is done in Lyapunov analysis).

Author Manuscript For Peer Review

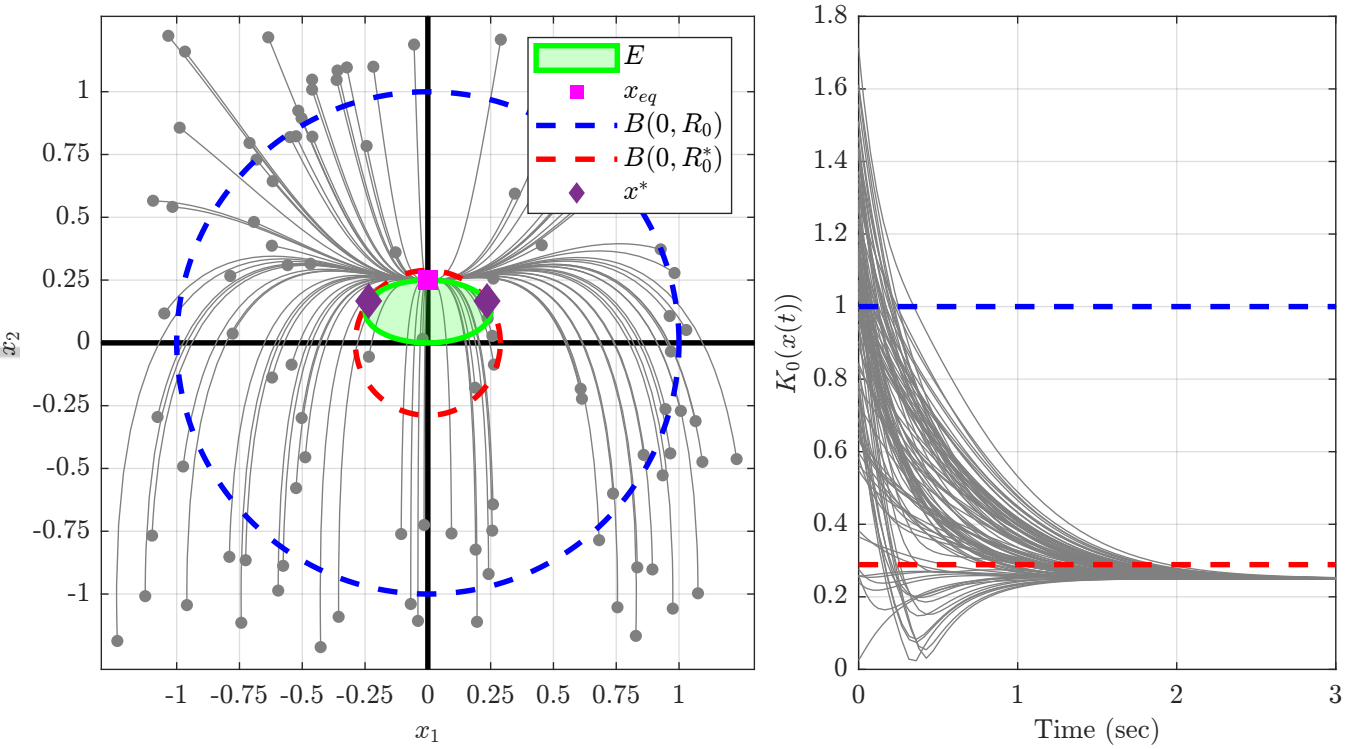


FIGURE 2 Illustrations of trapping regions of the bounded two-state system 24. The left plot shows the ellipsoid with non-decreasing energy E, valid trapping regions $B(0, R_0)$ and $B(0, R_0^*)$, and critical points x^*. Note that the radius R_0 (in blue) is an over-estimation discussed in Section 2.3, and R_0^* (in red) is the tightest radius computed by the SDP-based method proposed in Section 3.2. The critical points x^* (purple diamonds) are computed by the process proposed in Section 3.3. The right plot shows the energy versus time evaluated along the state trajectories shown in the left plot. All energy trajectories decrease below R_0^* in this system and well below R_0.

4.2 | Lorenz Chaotic Attractor

Consider the Lorenz system[1]:

$$\frac{dx_1}{dt} = -\sigma x_1 + \sigma x_2 \tag{25a}$$

$$\frac{dx_2}{dt} = \rho x_1 - x_2 - x_1 x_3 \tag{25b}$$

$$\frac{dx_3}{dt} = -\alpha x_3 + x_1 x_2, \tag{25c}$$

where $\sigma = 10, \rho = 28$ and $\alpha = \frac{8}{3}$. The system has no stable equilibrium points and exhibits chaotic behavior. Nonetheless, the system is shown to be bounded in Schlegel and Noack[5]. Specifically, it can be verified analytically that $A_s(m)$ is negative definite with the translation $m = [0,\ 0,\ \rho + \sigma]^\top$. Moreover, a bound on the trapping region, given by Theorem 1, is $B(m, R_m)$ with $R_m = 101.33$. Note that the estimated radius R_m is different from the radius reported in Schlegel and Noack[5], as there is a minor typo in the estimation formula as pointed out in Kaptanoglu et al[6].

Next, we demonstrate the SDP-based results introduced in Section 3. First, the SDP (17) was solved yielding $a^* = -1$ and $m = [0,\ 0,\ 38]^\top$. It follows from Theorem 2 and $a^* < 0$ that a trapping region exists. Note that the solution m is the same as the translation m used in Schlegel and Noack[5]. This is not a typical case, as the set of m such that $A_s(m)$ is negative definite is not unique, and the solution depends on the exact formulation of both approaches. Next, we solve the SDP (20) and compute the smallest trapping region $B(m, R_m^*)$ with $R_m^* = 39.25 < R_m = 101.33$. Lastly the critical points are characterized by (23) as $x^* = y^* + m$ with $y^* = [0,\ \pm 24.82,\ -30.4]^\top$.

Author Manuscript
For Review

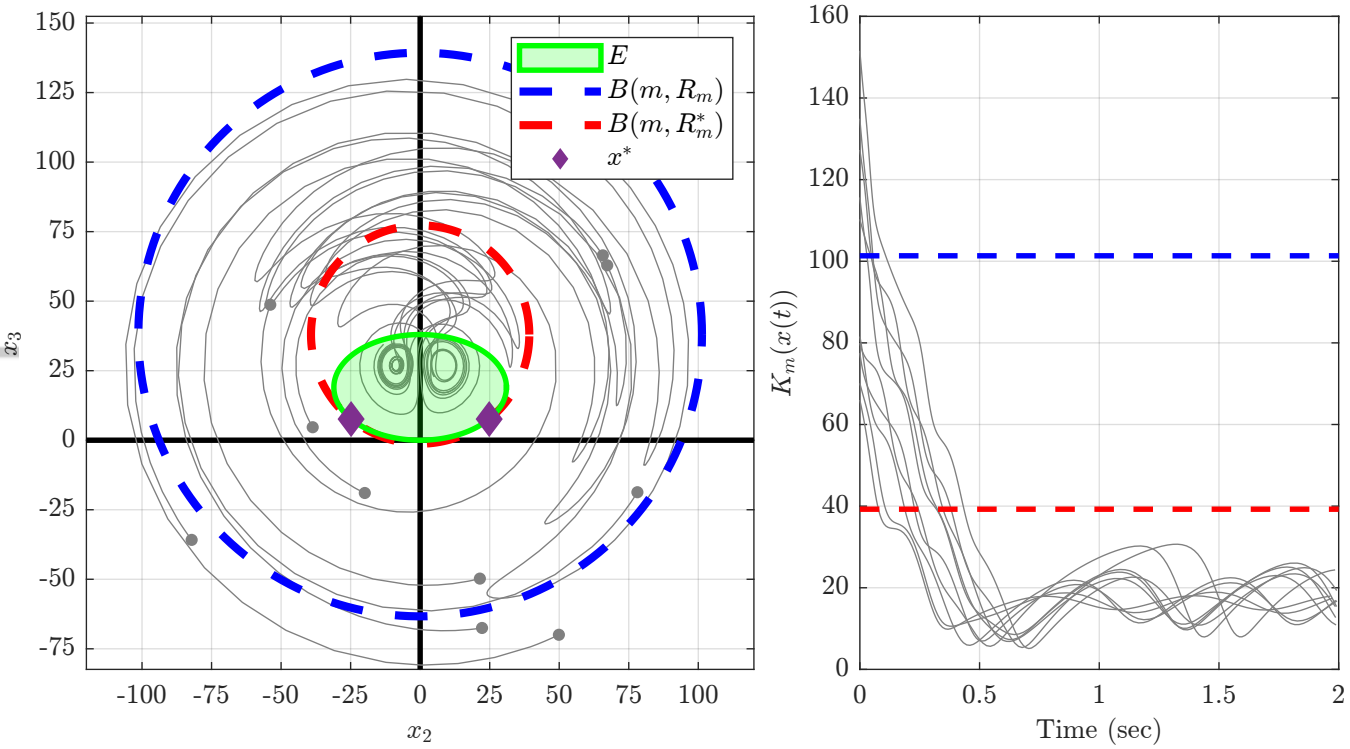


FIGURE 3 Trapping regions of the Lorenz attractor computed by Theorem 1 (blue) and the proposed SDP-based analysis (red). The left plot visualizes trajectories in the $x_2 - x_3$ plane with $x_1 = 0$. The green ellipsoid is the region where energy is instantaneously non-decreasing. The right plot shows the energy versus time evaluated along the state trajectories shown in the left plot. All trajectories are ultimately bounded in the region computed by the proposed SDP-based analysis (red), which captures the boundedness behavior more accurately than Theorem 1 (blue).

Figure 3 illustrates these results. The left plot shows ten trajectories with randomly sampled initial conditions in the $x_2 - x_3$ plane with $x_1 = 0$. Trajectories do not converge to any steady state or even a limit cycle since the system is chaotic. However, all trajectories are bounded as they are trapped in the boundedness region shown above. The red and blue circles are the computed boundedness region by our method and Theorem 1 respectively. The green ellipsoid is the area of instantaneously non-decreasing energy as characterized by the quadratic inequality in (18). The purple diamonds mark the critical points x^*, which are the furthest points with non-decreasing energy from m. All trajectories converge to and stay inside of both blue and red circles as dictated by Theorem 1 and 2 respectively. The right plot presents the energy $K_m(x(t))$ evolution of the ten trajectories. As predicted, all energies eventually decay below and stay below levels of R_m (blue) and R_m^* (red). Furthermore, our analysis (red) provides a more accurate description of the boundedness behavior of the Lorenz system.

4.3 | High-Dimensional Systems

The example illustrates the proposed method on a higher dimensional system. The example is constructed by stacking multiple Lorenz systems and coupling them via random coordinate rotation. Specifically, consider a collection of K systems with Lorenz dynamics (25):

$$\dot{z}^{(i)}(t) = L^{(i)} z^{(i)}(t) + f^{(i)}\left(z^{(i)}(t)\right) \quad i = 1, \ldots K,$$

Author Manuscript
For Review

where

$$L^{(i)} = \begin{bmatrix} -\sigma & \sigma & 0 \\ \rho & -1 & 0 \\ 0 & 0 & -\alpha \end{bmatrix} \quad \text{and} \quad f^{(i)}(z^{(i)}) = \begin{bmatrix} 0 \\ -z_1^{(i)} z_3^{(i)} \\ z_1^{(i)} z_2^{(i)} \end{bmatrix}. \tag{26}$$

We use the same parameters for each subsystem: $\sigma = 10$, $\rho = 28$, and $\alpha = \frac{8}{3}$. Next, create a larger system by stacking these Lorenz subsystems:

$$\dot{z}(t) = L_z z(t) + f_z(z(t)) \text{ where } z(t) := \begin{bmatrix} z^{(1)}(t) \\ \vdots \\ z^{(K)}(t) \end{bmatrix} \in \mathbb{R}^{3K}.$$

The matrix $L_z \in \mathbb{R}^{3K \times 3K}$ and function $f_z : \mathbb{R}^{3K} \to \mathbb{R}^{3K}$ are obtained by appropriate concatenation of the Lorenz subsystems. Here f_z is a lossless quadratic function as each Lorenz subsystem has a lossless quadratic term. Finally, define the state for our high-dimensional example as $x = Wz$ where $W \in \mathbb{R}^{3K \times 3K}$ is an orthonormal matrix. The dynamics of this stacked and rotated system are:

$$\dot{x}(t) = Lx(t) + f(x(t)) \text{ where } L = WL_zW^\top \text{ and } f(x) = Wf_z(W^\top x). \tag{27}$$

The quadratic function in (27) remains lossless after the rotation because:

$$x^\top f(x) = z^\top f(z) = 0$$

The resulting high-dimensional system (27) has a trapping region and is bounded by construction since each subsystem is bounded. However, the system (27) is not obviously bounded as the dynamics of each Lorenz subsystem are coupled via the orthogonal transformation $x = Wz$.

The boundedness of (27) is studied with 18 values of K chosen on a logarithmic grid from $K = 1$ to $K = 223$. The corresponding state dimension of (27) is $n = 3K$ and this goes from 3 to 669. Ten systems with random orthogonal matrices W are generated for each value of K. The SDP-based trapping region analysis in Theorem 2 is applied to each of these ten $3K$-dimensional systems (27). The analyses verified that the system (27) has a trapping region $B(m, R_m^*)$ for each K and each random W. The solution of the SDP (17) yields $a^* = -1$ for each trial. This is the same result as for a single Lorenz system (Section 4.2) because the combined system (27) is simply multiple copies of Lorenz dynamics in rotated coordinates. However, the coordinate shift m is not unique since the first state in Lorenz subsystems has no nonlinearity. Any translation in this direction does not affect the shifted linear symmetric matrix $A_s(m)$. Hence, optimal solutions m form a set and only one of the solutions is returned by the SDP solver. Furthermore, the radius R_m^* of the combined system (27) is not the same for each trial as the radius R_m^* depends on the coordinate shift m. Nonetheless, all systems are bounded as expected.

Figure 4 illustrates the solve time with respect to the number of states $n = 3K$. The computation time is the time to solve the two SDPs (17) and (20). The results show that the computation time increases as n increases (blue line). For a large number of states, the computation time scales empirically with $O(n^{3.98})$. This empirical estimate (red dashed line) was computed by linear regression on the data with $n > 150$. This matches the theoretical complexity of the primal-dual interior-point SDP solver in MOSEK (Section 13.3 of MOSEK[11]). Specifically, the complexity of a primal-dual interior-point SDP solver is $\max\{c_v c_d^3, c_v^2 c_d^2, c_v^3\}$ (Chapter 11.8.3 of Boyd and Vandenberghe[8]), where c_v is the number of decision variables and c_d is the dimension of LMI. The first SDP (17) has $(c_v, c_d) = (n+1, n)$ and the complexity is $O(n^4)$. The second SDP (20), which solves the QCQP (19), has $(c_v, c_d) = (2, n+1)$ and the complexity is $O(n^3)$. Hence, the overall theoretical complexity is $O(n^4)$ and is close to the observed complexity $O(n^{3.98})$. Note that the computation is dominated by the first SDP (17) to verify the existence of a trapping region. Hence, the radius of the trapping region is relatively cheap to compute by the second SDP (20) once the coordinate shift m is found. Note that the implementation is not optimized and the choice of $K = 223$ is based on the memory size of our desktop platform. The analysis can scale to even larger systems by more efficient implementation, customized SDP solver, and improved computational hardware.

Author Manuscript

For Peer Review

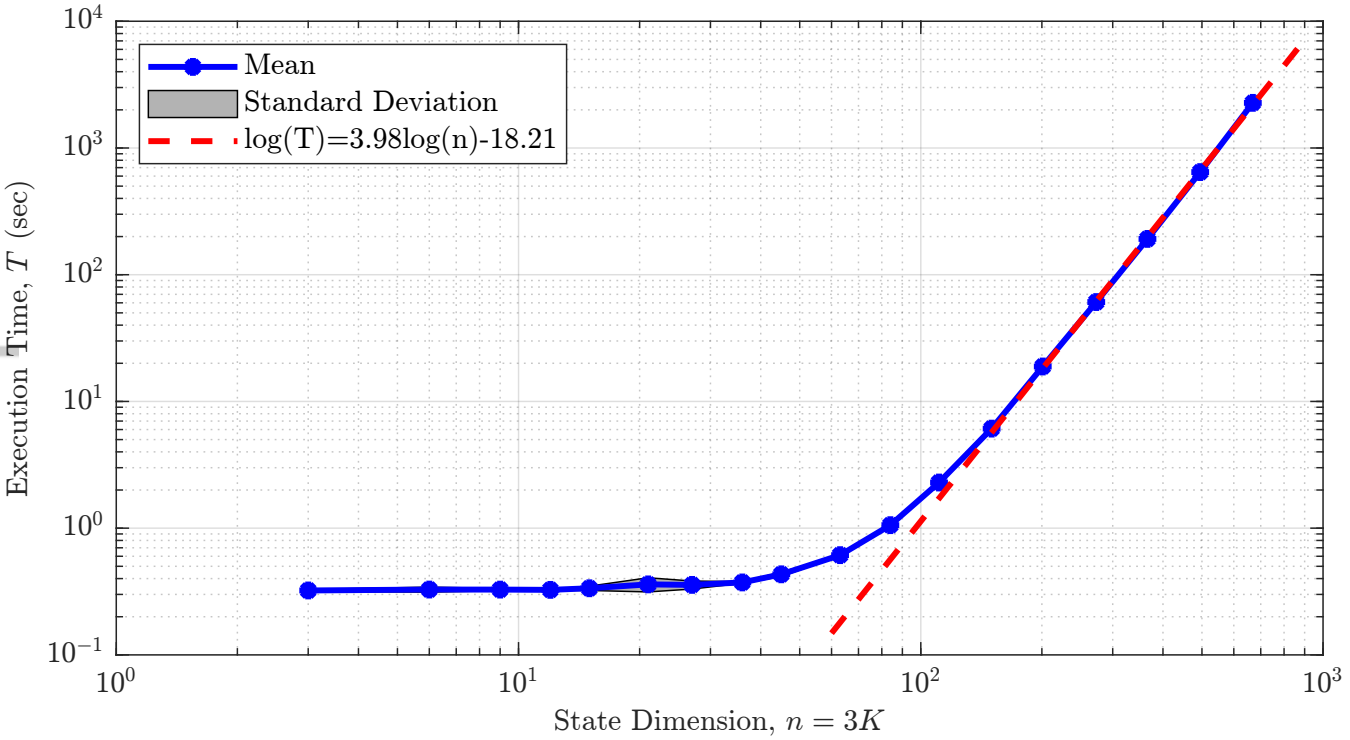


FIGURE 4 Computation times T for the proposed SDP-based analysis versus the number of states $n = 3K$ in the high-dimensional system (27). The computation includes solving the SDP (17) and SDP (20). The mean (blue) increases as number of states grows, and the standard deviation (gray) is negligible for large state dimensions. For large state dimensions, the execution time scales with $O(n^{3.98})$ and is close to the theoretical complexity $O(n^4)$.

5 | CONCLUSION

This paper considers systems with lossless quadratic nonlinearities and proposes a convex optimization approach to analyze boundedness using trapping regions. The contribution is threefold. First, the necessary and sufficient condition for a trapping region to exist is formulated as a convex SDP. This SDP provides theoretical and numerical advantages to verify the condition. Readily available SDP solvers can verify or falsify the existence of a trapping region with guarantees. Second, the least conservative radius of the trapping region at a given coordinate, if one exists, is computed by solving a QCQP exactly using duality theory. The solution of QCQP also characterizes the critical states on the boundary of the trapping region with instantaneously non-decreasing energy. This approach reduces the conservative estimation compared to the literature and provides important testing scenarios for further attention. Lastly, numerical examples are presented to illustrate the analysis process and its computational performance. The proposed method can easily verify that a trapping region exists and provides the precise estimate of the region for systems up to $\sim O(100)$ states.

The proposed convex-optimization-based method provides multiple additional opportunities for future work. The first direction is to incorporate this convex condition into modeling procedures. Several attempts[6,7] use the Schlegel and Noack trapping region condition (Theorem 1) to propose system identification algorithms with a physics-informed prior. Our convex SDP (Theorem 2) could potentially serve as a subroutine and improve the computational efficiency and modeling performance of such algorithms. The second direction is to use the convex condition in controller syntheses. SDPs are widely used to compute stabilizing controllers in both linear systems and nonlinear systems[9]. Notably, the convexity of static output feedback synthesis for lossless quadratic systems has been studied[3]. The trapping region condition being convex implies similar convexity structures potentially exist and can benefit controller synthesis tasks. The last direction here is the robustness analysis of the trapping region condition through convex optimization. For example, one could generalize the global analysis for exactly lossless nonlinearities, as considered in this paper, to local boundedness for weakly lossless nonlinearities[14]. Other classes of nonlinear systems, e.g., systems with higher-order nonlinearities, may be amenable to a similar analysis but this requires further investigation. Moreover,

Author Manuscript
For Review

we could extend our convex condition to analyze robustness by incorporating existing robust stability results[15]. In summary, this convex formulation of the trapping region condition establishes potential connections to a large amount of literature in control theory utilizing convexity.

AUTHOR CONTRIBUTIONS

The authors confirm their contribution to the paper as follows: study conception and design: Shih-Chi Liao, A. Leonid Heide, Maziar S. Hemati, Peter J. Seiler; data collection: Shih-Chi Liao; analysis and interpretation of results: Shih-Chi Liao, Maziar S. Hemati, Peter J. Seiler; visualization: Shih-Chi Liao, A. Leonid Heide; draft manuscript preparation: Shih-Chi Liao, ; manuscript revision: Maziar S. Hemati, Peter J. Seiler. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENTS

This material is based upon work supported by the Army Research Office under grant number W911NF-20-1-0156 and the Air Force Office of Scientific Research under grant number FA9550-21-1-0434. The authors acknowledge valuable discussions with Diganta Bhattacharjee.

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

SUPPORTING INFORMATION

Complete source code for the implementation and examples can be found on GitHub at https://github.com/SCLiao47/Boundedness_LosslessQuadSys.

ORCID

Shih-Chi Liao https://orcid.org/0000-0001-8311-3753
A. Leonid Heide https://orcid.org/0009-0006-1522-245X
Maziar S. Hemati https://orcid.org/0000-0003-1831-051X
Peter J. Seiler https://orcid.org/0000-0003-3423-1109

References

1. Lorenz EN. Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences.* 1963;20(2):130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2
2. Schmid PJ, Henningson DS. *Stability and Transition in Shear Flows*. 142. Springer Science & Business Media, 2012.
3. Mushtaq T, Seiler P, Hemati MS. On the convexity of static output feedback control synthesis for systems with lossless nonlinearities. *Automatica.* 2024;159:111380. doi: 10.1016/j.automatica.2023.111380
4. Khalil HK. *Nonlinear Systems. 3rd ed.* Prentice Hall, 2002.
5. Schlegel M, Noack BR. On long-term boundedness of Galerkin models. *Journal of Fluid Mechanics.* 2015;765:325-352. doi: 10.1017/jfm.2014.736
6. Kaptanoglu AA, Callaham JL, Aravkin A, Hansen CJ, Brunton SL. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *Physical Review Fluids.* 2021;6. doi: 10.1103/PhysRevFluids.6.094401
7. Goyal P, Duff IP, Benner P. Guaranteed Stable Quadratic Models and their applications in SINDy and Operator Inference. *arXiv preprint arXiv:2308.13819.* 2023.
8. Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press, 2004
9. Boyd S, El Ghaoui L, Feron E, Balakrishnan V. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994
10. Sturm JF. Using SeDuMi 1.02, A MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software.* 1999;11(1-4):625-653. doi: 10.1080/10556789908805766

Author Manuscript
For Peer Review

11. MOSEK ApS . *The MOSEK optimization toolbox for MATLAB manual. Version 9.2.* 2019.

12. Tütüncü RH, Toh KC, Todd MJ. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical programming.* 2003;95:189–217. doi: 10.1007/s10107-002-0347-5

13. Grant M, Boyd S. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. http://cvxr.com/cvx; 2014.

14. Peng M. Local Stability Guarantees for Data-Driven Quadratically Nonlinear Models. Master's thesis. University of Washington. 2023.

15. Dullerud GE, Paganini F. *A Course in Robust Control Theory: a Convex Approach*. 36. Springer Science & Business Media, 2013

Author Manuscript

For Peer Review

Novelty File

Title

A Convex Optimization Approach to Compute Trapping Regions for Lossless Quadratic Systems

Authors:

Shih-Chi Liao (shihchil@umich.edu),
A. Leonid Heide (heide116@umn.edu),
Maziar S. Hemati (mhemati@umn.edu),
Peter J. Seiler (pseiler@umich.edu)

Summary

Quadratic systems with lossless quadratic terms arise in many applications, including models of atmosphere and incompressible fluid flows. Such systems have a trapping region if all trajectories eventually converge to and stay within a bounded set. Conditions for the existence and characterization of trapping regions have been established in prior works for boundedness analysis. However, prior solutions have used non-convex optimization methods, resulting in conservative estimates. In this paper, we build on this prior work and provide a convex semidefinite programming condition for the existence of a trapping region. The condition allows precise verification or falsification of the existence of a trapping region. If a trapping region exists, then we provide a second semidefinite program to compute the least conservative trapping region in the form of a ball. Two low-dimensional systems are provided as examples to illustrate the results. A third high-dimensional example is also included to demonstrate that the computation required for the analysis can be scaled to systems of up to ≈O(100) states. The proposed method provides a precise and computationally efficient numerical approach for computing trapping regions. We anticipate this work will benefit future studies on modeling and control of lossless quadratic dynamical systems.

Why the paper is relevant to the journal as described in the aims and scope?

The journal webpage states the following: “The International Journal of Robust and Nonlinear Control aims to encourage the development of analysis and design techniques for uncertain linear and nonlinear systems.” Our paper focuses on a method to analyze the boundedness of a class of nonlinear systems with theoretical guarantees. We believe this fall within the state aims and scope of the journal.

Why the theoretical contributions in the paper are novel?

The key contribution of our paper is to provide a convex optimization approach to analyze trapping regions of lossless quadratic systems. Specifically, the paper provides a convex semidefinite programming condition for the existence of a trapping region. If a trapping region exists, the paper further provides a second semidefinite program to compute the least conservative trapping region in the form of a ball. These contributions provide precise and computationally efficient numerical approaches for computing trapping regions.

List of keywords

Boundedness, Trapping Region, Lyapunov Method, Semidefinite Programming

Author Manuscript
For Peer Review

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

A Convex Optimization Approach to Compute Trapping Regions for Lossless Quadratic Systems

Shih-Chi Liao[1] | **A. Leonid Heide**[2] | **Maziar S. Hemati**[2] | **Peter J. Seiler**[1]

[1]Electircal and Computer Engineering, University of Michigan, Ann Arbor, Michigan, USA

[2]Aerospace Engineering and Mechanics, University of Minnesota, Minnesota, USA

Correspondence
Shih-Chi Liao, Electrical and Computer Engineering, University of Michigan, Ann Arbor.
Email: shihchil@umich.edu

Funding Information
This research was supported by the This material is based upon work supported by the Army Research Office under grant number W911NF-20-1-0156 and the Air Force Office of Scientific Research under grant number FA9550-21-1-0434.

Abstract

Quadratic systems with lossless quadratic terms arise in many applications, including models of atmosphere and incompressible fluid flows. Such systems have a trapping region if all trajectories eventually converge to and stay within a bounded set. Conditions for the existence and characterization of trapping regions have been established in prior work for boundedness analysis. However, prior solutions have used non-convex optimization methods, resulting in conservative estimates. In this paper, we build on this prior work and provide a convex semidefinite programming condition for the existence of a trapping region. The condition allows for precise verification or falsification of the existence of a trapping region. If a trapping region exists, then we provide a second semidefinite program to compute the least conservative radius of the spherical trapping region. Two low-dimensional systems are provided as examples to illustrate the results. A third high-dimensional example is also included to demonstrate that the computation required for the analysis can be scaled to systems of up to $\sim O(100)$ states. The proposed method provides a precise and computationally efficient numerical approach for computing trapping regions. We anticipate this work will benefit future studies on modeling and control of lossless quadratic dynamical systems.

KEYWORDS
Boundedness, Trapping Region, Lyapunov Method, Semidefinite Programming

1 | INTRODUCTION

Many physical phenomena, including those of the atmosphere[1] and incompressible fluid flows[2], can be modeled by nonlinear dynamics with lossless quadratic nonlinearities. Here "lossless" means the quadratic term neither introduces nor dissipates energy. Such models should exhibit long-term boundedness as this is a feature of the corresponding physical phenomena. Similarly, closed-loop control of lossless quadratic systems has been studied in the literature[3]. In such cases, the boundedness can be used to maintain the closed-loop system in a certain operation condition[4]. Hence, methods to verify boundedness are required for this class of models. One straightforward approach is establishing an equilibrium's global stability, entailing all trajectories to converge and be bounded. However, a system can have an unstable equilibrium yet still be bounded, e.g., the Lorenz system[1]. Thus, the typical Lyapunov stability analysis is often too strict for such systems, and a dedicated boundedness analysis is required.

A trapping region is a bounded set that all trajectories eventually converge into and stay within. This notion has been proposed in the literature as an approach to analyze boundedness[1,5]. Lorenz[1] proposed an energy function as a certificate for the existence of a trapping region to study the boundedness of the chaotic three-state system. Schlegel and Noack[5] formalized this idea with a necessary and sufficient condition for the existence of a spherical trapping region. They applied the condition to demonstrate the boundedness of several reduced-order models of incompressible fluid flows with $\sim O(10)$ state variables. Furthermore, these trapping region conditions have been exploited in physics-informed system identification algorithms[6,7] to obtain long-term bounded models of fluid flows with $\sim O(10)$ state variables from data.

While trapping regions provide a certificate of boundedness, verifying the existence of a trapping region is non-trivial. Several approaches are discussed in Section 7 of Schlegel and Noack[5]. However, they are either limited by the state dimension of systems (as they involve an algebraic condition on the eigenvalues of the system matrices) or involve non-convex optimization

Author Manuscript
For Review

relying on a good initial seed for the algorithm. The approaches offer no guarantees despite using physical understanding and simulations to aid the analysis. Thus the approaches may fail to find a trapping region even if one exists. Furthermore, even if the existence of a trapping region is verified, Schlegel and Noack provide a conservative radius estimation of the spherical trapping region. These limitations restrict the application of the boundedness analysis using trapping regions.

This paper provides three contributions to the trapping region analysis. First, a convex semidefinite programming (SDP) condition is proposed as the necessary and sufficient condition for the system to have a trapping region. This convex condition provides theoretical guarantees and corresponding efficient numerical solutions. Hence, readily available SDP solvers can be used to verify or falsify the existence of a trapping region. Second, a quadratically constrained quadratic program (QCQP) is derived to compute the least conservative radius of the spherical trapping region if one exists. This QCQP can be solved exactly by its SDP dual problem. Also, the solution of the dual SDP characterizes the set of points with instantaneously non-decreasing energy on the boundary of the trapping region. This set of points can be useful for further study via simulations. Lastly, the proposed SDP-based trapping region analysis is demonstrated by three examples. The first two examples are low-dimensional systems to illustrate the concepts and the analysis process. The last example highlights that the proposed method can scale to $\sim O(100)$ states. The SDP-based method can easily verify the existence of the trapping region and provide the least conservative estimate for the trapping region. This convex formulation can potentially benefit the relevant modeling[6,7] and control[3] approaches for lossless quadratic systems.

The rest of the paper is structured in four sections. Section 2 formulates the problem of this study and presents relevant prior work. Section 3 presents the proposed SDP-based trapping region analysis, including verification, computing the least conservative radius, and characterizing critical points. Section 4 provides numerical examples to illustrate the process and computational load of the analysis. Lastly, Section 5 summarizes the paper and discusses possible future directions.

2 | PRELIMINARIES

2.1 | Lossless Quadratic Systems

Let $c \in \mathbb{R}^n$, $L \in \mathbb{R}^{n\times n}$, and symmetric matrices $\{Q^{(i)}\}_{i=1}^n \subset \mathbb{R}^{n\times n}$ be given. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a quadratic function with i^{th} entry defined by $f_i(x) = x^\top Q^{(i)} x$ for $i = 1 \dots n$. Given this notation, define the following n-state nonlinear system:

$$\dot{x}(t) = c + Lx(t) + f(x(t)), \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector at time t. The time dependency of the state vector $x(t)$ is occasionally omitted to simplify the notation. Furthermore, we assume the quadratic nonlinearity is lossless: $x^\top f(x) = 0$ for all $x \in \mathbb{R}^n$. The lossless condition is:

$$0 = x^\top f(x) = \sum_{i=1}^{n} x_i \left(x^\top Q^{(i)} x\right) = \sum_{i,j,k=1}^{n} Q^{(i)}_{jk} x_i x_j x_k \qquad \forall x \in \mathbb{R}^n. \tag{2}$$

Note that every term in the product $x^\top f(x)$ is cubic in the entries of x. The lossless condition holds if and only if the coefficient of each term $x_i x_j x_k$ is zero. Accounting for the fact that $x_i x_j x_k = x_j x_k x_i = x_k x_i x_j$, the nonlinearity $f(x)$ is lossless if and only if the symmetric matrices $\{Q^{(i)}\}_{i=1}^n$ satisfy the following condition:

$$Q^{(i)}_{jk} + Q^{(j)}_{ik} + Q^{(k)}_{ij} = 0, \qquad \forall i, j, k \in \{1, \dots, n\}. \tag{3}$$

The nonlinear term is called "lossless" as it does not change the energy over time. Specifically, assume the energy in state x is $K_0(x) = \frac{1}{2}\|x\|_2^2$. Then the time derivative of the energy along trajectories of (1) is

$$\frac{d}{dt} K_0(x(t)) = x(t)^\top \left(c + Lx(t) + f(x(t))\right) \tag{4}$$

$$= x(t)^\top \left(c + Lx(t)\right). \tag{5}$$

The nonlinear term does not affect the time derivative of energy due to the lossless property. This property is used in subsequent sections to develop sufficient conditions for the boundedness of (1).

Manuscript For Peer Review

One important feature of a quadratic system is that the nonlinearities are preserved under coordinate translation. Specifically, define $y := x - m$, where $m \in \mathbb{R}^n$ is a constant translation. The dynamics after translation are:

$$\dot{y}(t) = d(m) + A(m)\, y(t) + f(y(t)), \tag{6}$$

where $d(m) := c + Lm + f(m)$ and

$$A(m) := L + 2 \begin{bmatrix} m^\top Q^{(1)} \\ \vdots \\ m^\top Q^{(n)} \end{bmatrix}. \tag{7}$$

The quadratic nonlinearity in (6) is exactly the same as in (1). Hence the nonlinearity in (6) is also lossless. The symmetric part of $A(m)$ is given by $A_s(m) := \frac{1}{2}(A(m) + A(m)^\top)$. This symmetric part plays an important role in the boundedness analysis introduced later. It can be expressed simply as:

$$A_s(m) = L_s - \sum_{k=1}^{n} m_k Q^{(k)}, \tag{8}$$

where $L_s = \frac{1}{2}(L + L^\top)$ is the symmetric part of L and m_k is the kth element of the coordinate shift m. Equation 8 is due to the lossless property of the nonlinearity (3):

$$\begin{bmatrix} m^\top Q^{(1)} \\ \vdots \\ m^\top Q^{(n)} \end{bmatrix}_{i,j} + \left[Q^{(1)} m \ \cdots \ Q^{(n)} m \right]_{i,j} = \sum_{k=1}^{n} m_k Q^{(i)}_{jk} + \sum_{k=1}^{n} m_k Q^{(j)}_{ik} = -\sum_{k=1}^{n} m_k Q^{(k)}_{ij}. \tag{9}$$

2.2 | Boundedness of a System

Let $\phi(t; x_0, t_0)$ denote the solution of (1) at time $t \geq t_0$ starting from the initial condition $x_0 \in \mathbb{R}^n$ at time t_0. As the stability of a nonlinear system is hard to analyze in general, the boundedness of solutions $\phi(t; x_0, t_0)$ serves as an alternative notion to characterize the behavior of a system. The boundedness of the system can be defined as in Chapter 4.8 of Khalil[4].

Definition 1 (Boundedness)**.** The solutions of system (1) are globally uniformly ultimately bounded if there exists a scalar $\beta > 0$, independent of $t_0 \geq 0$, and a function $T : \mathbb{R} \to \mathbb{R}$ such that:

$$\|\phi(t; x_0, t_0)\|_2 \leq \beta \qquad \forall x_0 \in \mathbb{R}^n \text{ and } \forall t \geq t_0 + T(\|x_0\|_2). \tag{10}$$

In words, a system is globally uniformly ultimately bounded with bound $\beta > 0$ if the trajectory from any initial condition eventually converges to within a spherical ball centered at the origin and with radius β. Moreover, $T(\|x_0\|_2)$ is the time required to converge within the bound β. This time is independent of t_0 but depends on the norm of the initial condition. In this paper, only globally uniformly ultimate boundedness is investigated and we will simply refer to this as the "boundedness" in the remainder of the paper.

Note that the boundedness of the shifted system (6) is equivalent to the boundedness of the original system (1) as the coordinates only differ by translation. In addition, the triangle inequality implies $\|x(t)\|_2 \leq \|y(t)\|_2 + \|m\|_2$. Hence, if a trajectory in shifted coordinates is bounded by β, i.e. $\|y(t)\|_2 \leq \beta$, then the corresponding trajectory in the original coordinates is bounded by $\beta + \|m\|_2$. Hence, one can show the boundedness of the original system (1) by showing the boundedness of the shifted system (6).

2.3 | A Sufficient Condition for Boundedness

This section reviews a sufficient condition to certify the boundedness of quadratic systems (1). This section is mainly based on early results discussed by Lorenz[1], which were further formalized by Schlegel and Noack[5]. First, we introduce the notion of a monotonically attracting trapping region as defined in Schlegel and Noack[5].

Manuscript
For Peer Review

Author Manuscript

Definition 2 (Trapping region)**.** A trapping region $D \subseteq \mathbb{R}^n$ is a compact set that is forward invariant, i.e., if $x(t_0) \in D$ then $x(t) \in D$ for all $t \geq t_0$. In other words, once a trajectory enters a trapping region, it remains in the trapping region for all future time. A trapping region is termed globally monotonically attracting if an energy function is strictly monotonically decreasing along all trajectories starting from an arbitrary state outside of D.

The existence of a monotonically attracting trapping region for a system implies that the system is bounded, as all trajectories will enter the compact set D. We will focus on globally monotonically attracting trapping regions and will refer to this simply as a "trapping region." Furthermore, we will only consider spherical regions as candidates of a trapping region in this paper.

Section 2 of Lorenz's seminal 1963 paper[1] provides a condition for the existence of a trapping region for the dynamical system (1). The condition uses the energy and states that a trapping region exists if $L_s < 0$. A more precise characterization for a trapping region is presented in Schlegel and Noack[5]. These results are summarized in the following lemma with the definition of a closed-ball with center ν and radius R given by $B(\nu, R) := \{x \in \mathbb{R}^n \,:\, \|x - \nu\|_2 \leq R\}$.

Lemma 1. *If $L_s \prec 0$, a trapping region exists, and the system* (1) *is bounded. Moreover, if λ_1 is the largest (least negative) eigenvalue of L_s and $R_0 := \frac{\|c\|_2}{|\lambda_1|}$, then $B(0, R_0)$ is a trapping region.*

Proof. The energy was previously defined in Section 2.1 as $K_0 := \frac{1}{2}\|x\|_2^2$. The derivative of this energy function along trajectories of the system (1) is:

$$\frac{d}{dt}K_0\left(x(t)\right) := \frac{1}{2}\left(x(t)^\top \dot{x}(t) + \dot{x}(t)^\top x(t)\right) = c^\top x(t) + x(t)^\top L_s x(t). \tag{11}$$

As noted previously, the derivative of $K_0(x(t))$ does not depend on the quadratic dynamics f as the nonlinearity is lossless.

Assume $L_s \prec 0$. The energy evolution (11) is dominated by the quadratic term and is negative if $\|x\|_2$ is sufficiently large. The energy can increase only when it is close to the origin. Therefore, a trapping region around the origin exists as all trajectories converge toward the origin or stay close to the origin. In summary, if Ls is negative definite, a trapping region exists, and the system is bounded.

Next, the time derivative of K_0 is bounded using the Cauchy-Schwartz inequality and the largest (least negative) eigenvalue λ_1:

$$\frac{d}{dt}K_0(x(t)) \leq \|c\|_2 \|x(t)\|_2 + \lambda_1 \|x(t)\|_2^2 = \left(\|c\|_2 + \lambda_1 \|x(t)\|_2\right) \cdot \|x(t)\|_2. \tag{12}$$

Thus, the derivative of K_0 is negative if $\|x(t)\|_2 > \frac{\|c\|_2}{|\lambda_1|}$. Hence $B(0, R_0)$ is a trapping region for the system (1) in original coordinates with the radius $R_0 := \frac{\|c\|_2}{|\lambda_1|}$. □

Lemma 1 provides a sufficient condition to check the boundedness of the system (1). The condition can be easily verified by eigenvalue computation. Note that the condition is independent of the equilibrium point and depends only on c and L_s. This allows analyzing the boundedness of system (1) even when it has multiple equilibrium points or chaotic behavior. Furthermore, if a trapping region exists, all equilibrium points would be inside the region as the energy evolution at an equilibrium is always zero.

Nonetheless, the radius bound in Lemma (1) is conservative in general and is attained only when the vector c aligns exactly with the eigenvector corresponding to λ_1. This is due to the derivation using Cauchy-Schwartz inequality in (12). To demonstrate the conservatism, consider the set of states where the energy is (instantaneously) non-decreasing, $E := \left\{x(t) \in \mathbb{R}^n \,:\, \frac{d}{dt}K_0(x(t)) \geq 0\right\}$. We can characterize E using completion of squares to rewrite (11) as follows:

$$\frac{d}{dt}K_0(x(t)) = \left(x(t) + \frac{1}{2}L_s^{-1}c\right)^\top L_s \left(x(t) + \frac{1}{2}L_s^{-1}c\right) - \frac{1}{4}c^\top L_s^{-1} c. \tag{13}$$

The boundary of E corresponds to the states where $\frac{d}{dt}K_0(x(t)) = 0$. It follows from (13) that this boundary is an ellipsoid.† The point $x(t) = -\frac{1}{2}L_s^{-1}c \in E$ is the ellipsoid center and is the point of maximum energy growth: $\frac{d}{dt}K_0(x(t)) = -\frac{1}{4}c^\top L_s^{-1}c > 0$ (assuming $c \neq 0$). By Lemma 1, $B(0, R_0)$ is a trapping region, where the energy is strictly decreasing outside this ball. Hence $E \subseteq B(0, R_0)$.

† The matrix L_s is symmetric and negative definite. Thus, it has an eigenvalue decomposition $L_s = U\Lambda U^\top$ where U is orthogonal and $\Lambda < 0$ is a diagonal matrix with the eigenvalues $\lambda_n \leq \cdots \leq \lambda_1 < 0$ along the diagonal. Define the rotated coordinates $z(t) := U^\top x(t)$. The condition $\frac{d}{dt}K_0(x(t)) = 0$ can be written in these rotated coordinates as $(z(t) - v)^\top \Lambda (z(t) - v) = \frac{1}{4}c^\top L_s^{-1} c$ where $v := -\frac{1}{2}U^\top L_s^{-1} c \in \mathbb{R}^n$ is the ellipsoid center in the rotated coordinates. This can be further simplified to $\sum_{i=1}^n \frac{1}{\alpha_i^2}(z_i - v_i)^2 = 1$ where $\alpha_i := \frac{1}{2}\sqrt{\frac{c^\top L_s^{-1} c}{\lambda_i}}$ are the semi-axis lengths.

For Peer Review

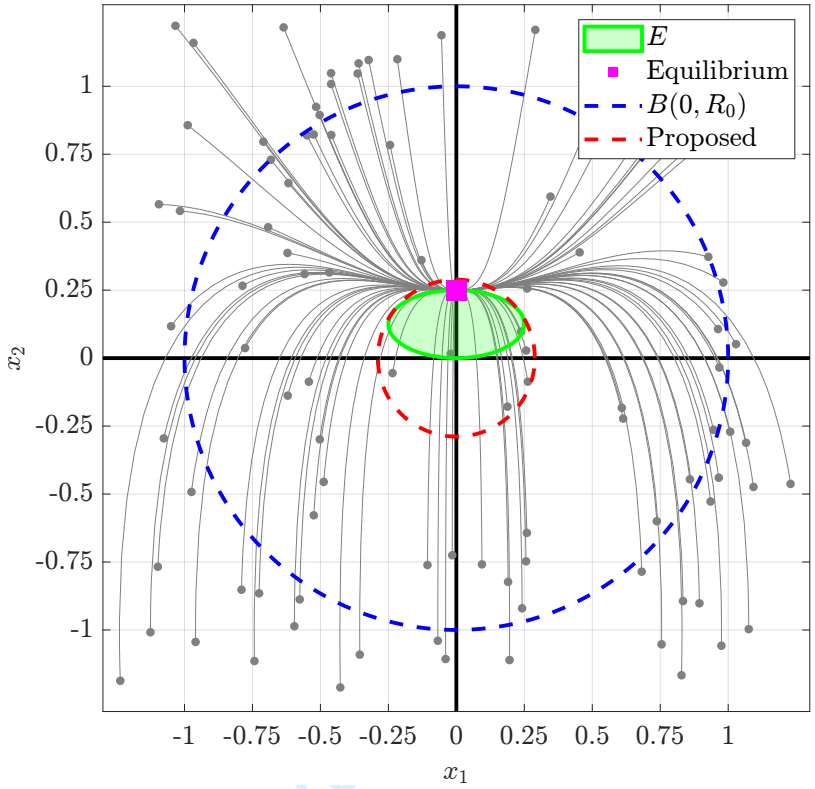


FIGURE 1 An illustration of trapping region for the two-state system (14) with kinetic energy $K_0(x) = \frac{1}{2}x^\top x$. All trajectories with random initial conditions (gray) converge to the equilibrium point at $(0, 0.25)$ (magenta square). The set of states with non-decreasing energy E (green solid ellipsoid) and a conservative trapping region $B(0, R_0)$ (blue-dashed circle) are shown. The tightest trapping region centered at the origin can be computed by our proposed method (Section 3.2) and is shown in the red-dashed circle.

However, $B(0, R_0)$ is not necessarily the tightest one that contains E as pointed out. In fact, any forward-invariant set containing E renders a trapping region as the energy is decreasing for any states outside of E. In Section 3.2, we will derive a convex optimization problem to reduce conservatism and find the tightest trapping region containing the non-decreasing energy set E.

The following two-state system with lossless quadratic nonlinearities serves as an example to illustrate the boundedness analysis using Lemma 1 and its conservatism:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix} x + \begin{bmatrix} -x_1 x_2 \\ x_1^2 \end{bmatrix}. \tag{14}$$

The trajectories (light gray lines) in Figure 1 are generated with random initial conditions and converge to the single stable equilibrium $(0, 0.25)$. The constant vector in the dynamics is $c = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ and the largest (least negative) eigenvalue of L_s is $\lambda_1 = -1$. Hence, $B(0, R_0)$ is a trapping region with $R_0 = \frac{\|c\|_2}{|\lambda_1|} = 1$ as derived above. This is shown as the blue-dashed circle in Figure 1. All trajectories eventually converge inside this ball as expected. The green ellipsoid in this figure corresponds to the non-decreasing energy states E. Based on (13), the boundary of this ellipse is defined by:

$$\frac{d}{dt}K_0(x(t)) = -x_1^2 - 4\left(x_2 - \frac{1}{8}\right)^2 + \frac{1}{16} = 0.$$

As shown in Figure 1, the circle $B(0, R_0)$ is a conservative estimate as any ball centered at the origin containing the ellipsoid E is a trapping region. Note that this conservatism is rooted in the derivation using Cauchy-Schwartz inequality in (12). We will derive in Section 3.2 a convex optimization to compute the tightest ball containing the ellipsoid E. This tightest (least conservative) ball is shown as the red-dashed circle in Figure 1.

Author Manuscript
For Review

2.4 | Boundedness Using Coordinate Translations

The previous subsection presented a sufficient condition for boundedness using the energy centered at the state origin. Schlegel and Noack[5] further generalized the condition using coordinate translations. Consider the translated system (6) with $y(t) := x(t) - m$. The energy for the translated system is $K_m(y) := \frac{1}{2}\|y\|_2^2$ with corresponding time derivative:

$$\frac{d}{dt}K_m(y(t)) := \frac{1}{2}\left(y(t)^\top \dot{y}(t) + \dot{y}(t)^\top y(t)\right) = d(m)^\top y(t) + y(t)^\top A_s(m) y(t). \tag{15}$$

Using the same argument as in Section 2.3, if the matrix $A_s(m)$ is negative definite, then the shifted system (6) has a trapping region $B(0, R_m)$ in the shifted coordinate for some R_m. This implies the original system (1) is bounded by $\|m\|_2 + R_m$. Furthermore, a valid radius is $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$ where λ_1 is the largest (least negative) eigenvalue of $A_s(m)$. The following theorem in Schlegel and Noack[5] formalizes these discussions for monotonically attracting trapping regions.

Theorem 1 (Theorem 1 of Schlegel and Noack[5])**.** *The system* (1) *has a monotonically attracting trapping region* $B(m, R_m)$ *if and only if there exists an* $m \in \mathbb{R}^n$ *such that the real, symmetric matrix* $A_s(m)$ *is negative definite. If* $A_s(m)$ *has eigenvalues* $\lambda_n \leq \cdots \leq \lambda_1 < 0$*, then a trapping region is given by* $B(m, R_m)$ *with the radius* $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$.

A formal proof of this necessary and sufficient trapping region condition is given in Appendix B of Schlegel and Noack[5]. Note that the radius R_m given in Theorem 1 is different from the statement in Schlegel and Noack[5] as there is a minor typo that is also pointed out in Kaptanoglu et al[6]. Also, the proof in Schlegel and Noack[5] regarding the estimated radius R_m used a geometrical argument instead of the Cauchy-Schwartz inequality approach presented in (12). The two estimates are the same, and the Cauchy-Schwartz inequality is adopted in this paper for simplicity.

The key point of Theorem 1 is that the boundedness of the system (1) can be determined by an algebraic condition. In other words, if m exists such that $A_s(m)$ is negative definite, then the system has a trapping region $B(m, R_m)$. This corresponds to a bound R_m in the shifted coordinates and $\|m\|_2 + R_m$ in the original coordinates. This approach does not require solving the nonlinear differential equation for all possible initial conditions. Note that the converse is not true, i.e., a bounded system does not imply the existence of a trapping region. Consider a trivial system $\dot{x} = 0$. All solutions are stationary and thus bounded. However, this system does not have a monotonically attracting trapping region, i.e., one cannot find an m such that $A_s(m)$ is negative definite. If no trapping region exists, then Schlegel and Noack[5] point out that additional analysis of the linear and nonlinear terms is needed to determine whether the system is bounded.

Even though Theorem 1 provides an approach to verify the boundedness of the system, there are two remaining caveats. First, verifying this algebraic condition is non-trivial as it involves a search over $m \in \mathbb{R}^n$. An inf-sup problem is formulated in Schlegel and Noack[5] to minimize the largest eigenvalue of $A_s(m)$ and simulated annealing is proposed as a solution method. However, this formulation is a non-convex optimization, and hence there are no guarantees of convergence to the global optimum. Thus the test is inconclusive if the simulated annealing fails to find a feasible m with the largest eigenvalue of $A_s(m)$ being negative. In other words, a trapping region may exist even if the simulated annealing algorithm fails. The second caveat of Theorem 1 is that the estimated bound R_m is an approximate upper bound and is usually conservative. As mentioned in Section 2.3, the approximation is made by the analysis of eigenvalues of symmetric matrix $A_s(m)$ with the worst-case alignment of direction $d(m)$ in the Cauchy-Schwartz inequality (12). Hence, the upper bound approximation R_m is often conservative. We address both of these caveats in the next section by providing: (i) a convex optimization to compute the trapping region, and (ii) a tight bound on the trapping region radius.

3 | COMPUTATION OF TRAPPING REGIONS VIA CONVEX OPTIMIZATION

This section revisits the boundedness condition in Theorem 1. First, the condition is reformulated as a constraint in the form of a linear matrix inequality (LMI), whose feasible set is a convex set. The feasibility of this LMI condition can be tested numerically, and a feasible solution, if one exists, verifies the existence of a trapping region. This avoids the non-convex inf-sup formulation in Schlegel and Noack[5]. Second, a quadratically constrained quadratic program (QCQP) is derived to compute the exact radius of the bounded region without conservative approximation. Lastly, the set of furthest points with non-decreasing energy can be identified by the dual solution of the QCQP and shown to be a hypersphere.

Author Manuscript

For Peer Review

Author Manuscript

3.1 | Trapping Condition as a Convex Constraint

By Theorem 1, the system (1) has a trapping region if and only if there exists $m \in \mathbb{R}^n$ such that $A_s(m)$ is negative definite, i.e.,

$$A_s(m) = L_s - \sum_{i=1}^{N} m_i Q^{(i)} \prec 0. \tag{16}$$

Observe that $A_s(m)$ is an affine function of m. Hence, the negative definite constraint (16) is a linear matrix inequality (LMI) in the variable m[8,9]. LMIs have been widely studied in the optimization community and are known to have a convex feasible set, i.e., a convex combination of two solutions is also a solution (Chapter 2.3.2 of Boyd and Vandenberghe[8]). Hence, the set of $m \in \mathbb{R}^n$ that satisfies (16) is a convex set. LMI feasibility problems can be efficiently solved with many numerical optimization solvers, e.g., SeDuMi[10], MOSEK[11], and SDPT3[12]. There is also software that enables easy implementation and interface with these solvers, e.g., CVX[13]. Finally, LMI feasibility problems have useful duality results such as the theorem of alternatives (Chapter 5.8 of Boyd and Vandenberghe[8]). Thus, the numerical algorithm will return a coordinate shift $m \in \mathbb{R}^n$ satisfying (16) (and hence verifying the existence of a trapping region) or it will return a dual variable that certifies that no such feasible m exists.

The following lemma states the application of this LMI condition to verify the existence of a trapping region in system (1):

Lemma 2. *Consider system* (1) *and a semidefinite programming (SDP) to minimize the largest eigenvalue of* $A_s(m)$*:*

$$\begin{aligned} a^* = \min_{m \in \mathbb{R}^n, a \in \mathbb{R}} \quad & a \\ \text{s.t.} \quad & A_s(m) \preceq a I_n. \end{aligned} \tag{17}$$

A trapping region exists and the system (1) *is bounded if the optimal cost of the SDP* (17) *satisfies* $a^* < 0$*. If* $a^* \geq 0$*, there is no trapping region in the system* (1)*.*

Proof. The optimization (17) has a single LMI constraint with the linear objective given by the slack variable a. It falls into the category of SDPs (Chapter 4.6.2 of Boyd and Vandenberghe[8]). In addition, the SDP (17) is always feasible by choosing any $m \in \mathbb{R}^n$ and a sufficiently large value of a. Finally, there exists an m such that $A_s(m)$ is negative definite if the optimal cost $a^* < 0$. By Theorem 1, the system (1) has a trapping region and is bounded. If $a^* \geq 0$, there is no feasible m such that LMI (16) is satisfied. By Theorem (1), no trapping region exists in the system (1). □

SDPs are convex optimizations that can be efficiently solved to the global optimum (within numerical tolerances) using numerical solvers. This particular SDP (17) can be solved on a standard desktop for problems with dimensions up to $n \sim O(100)$. Specialized implementations can be made for larger problems. In contrast to the inf-sup formulation proposed in Schlegel and Noack[5], this SDP formulation is guaranteed to converge to the global optimum and provide a conclusive trapping region analysis.

3.2 | Computing the Tightest Trapping Region Radius with SDP

Assume the SDP (17) has an optimal val $a^* < 0$ so that there exists an m such that $A_s(m) \prec 0$. By Theorem 1, the system (1) has a monotonically attracting trapping region. Let λ_1 denote the largest (least negative) eigenvalue of $A_s(m)$. A conservative bound on the trapping region is given by $B(m, R_m)$ with the radius $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$. In this subsection, we utilize duality in optimization theory to compute the tightest (least conservative) radius R_m^* such that $B(m, R_m^*)$ is a trapping region for a certain choice of m. The discussion of choosing the best m to obtain the smallest R_m^* is briefly addressed at the end of the section.

If $\frac{d}{dt} K_m(y(t)) \geq 0$, then the energy is instantaneously non-decreasing. It follows from (15) that the set of all states where the energy is instantaneously non-decreasing (in shifted coordinates) is:

$$E := \left\{ y \in \mathbb{R}^n \,:\, d(m)^\top y + y^\top A_s(m) y \geq 0 \right\}. \tag{18}$$

For Peer Review

The tightest (least conservative) radius R_m^* corresponds to the smallest ball (in shifted coordinates) that contains all states in E. Thus the smallest radius is given by:

$$\begin{aligned} R_m^* := \min_{R_m} \quad & R_m \\ \text{s.t.} \quad & E \subseteq B(0, R_m). \end{aligned}$$

Equivalently, we can find the state in E with the largest squared norm, i.e., maximize $\|y\|_2^2$ subject to $y \in E$. This equivalent formulation leads to the following optimization:

$$(R_m^*)^2 = \max_y \quad f_0(y) := y^\top y \tag{19a}$$

$$\text{s.t.} \quad f_1(y) := d(m)^\top y + y^\top A_s(m) y \geq 0. \tag{19b}$$

Note that (19) is a non-convex problem as we are maximizing a convex objective. Non-convex problems can, in general, have multiple local optima, and solving for global optima can be difficult.

In fact, (19) is a quadratically constrained quadratic program (QCQP) with only one quadratic constraint. This particular problem, while nonconvex, can be efficiently solved for the global optimum. The following lemma provides a convex SDP to compute the tightest radius R_m^* by solving QCQP (19):

Lemma 3. *Given system* (1) *and an m such that* $A_s(m) \prec 0$*. The following convex SDP exactly solves the QCQP* (19)*:*

$$(R_m^*)^2 = \min_{\lambda \geq 0, \gamma} \quad \gamma \tag{20a}$$

$$s.t. \quad f_2(\lambda, \gamma) := \begin{bmatrix} I + \lambda A_s(m) & \frac{\lambda}{2} d(m) \\ \frac{\lambda}{2} d(m)^\top & -\gamma \end{bmatrix} \preceq 0. \tag{20b}$$

Then, $B(m, R_m^*)$ *is the tightest (least conservative) trapping region of system* (1) *that centers at m.*

Proof. The SDP (20) is the Lagrange dual of the QCQP (19) (Appendix B of Boyd and Vandenberghe [8]). Two cases are discussed to show the optimal cost for this dual SDP (20) is identical to the optimal cost of the primal QCQP (19). First, if $d(m) \neq 0$, the strong duality holds between QCQP (19) and dual SDP (20). In particular, the point $y_0 = -\frac{1}{2} \frac{d(m)^\top d(m)}{d(m)^\top A_s(m) d(m)} d(m)$ is strictly feasible for the QCQP (19), i.e. $f_1(y_0) < 0$. Hence, the QCQP satisfies Slater's constraint qualification and the strong duality holds (Appendix B of Boyd and Vandenberghe [8]). Second, if $d(m) = 0$, then the constraint $f_1(y) \leq 0$ does not have a constant term and is not strictly feasible. The only feasible point of the QCQP (19) in this case is $y_0 = 0$, and therefore the optimal cost is $R_m^* = 0$. The dual (20) has the same optimal cost of 0 by the solution of $\gamma = 0$ with $\lambda > 0$ sufficiently large such that $I + \lambda A_s(m) \preceq 0$. This is possible since $A_s(m) \prec 0$. Thus, QCQP and dual SDP achieve the same cost.

Finally, $(R_m^*)^2$ is both the optimal cost of the primal problem (19) and the dual (20). The primal form implies that there exists a state y^* with $\|y^*\|_2 = R_m^*$ such that $f_1(y^*) = 0$ and it is contained in the trapping region. Any state y_0 such that $\|y_0\|_2 > R_m^*$ must have $f_1(y_0) < 0$. In other words, $\frac{d}{dt} K_m(y(t)) < 0$ for any state $y(t) = y_0$ with $\|y_0\|_2 > R_m^*$. Hence, the trapping region $B(m, r_m^*)$ of system (1) is the tightest (least conservative) trapping region centered at m.

□

Lemma 3 provides a computationally feasible approach to solve for the global optimum. The dual SDP (20) can be solved by the numerical solvers mentioned above. Note that for the case $d(m) = 0$, the optimal cost $R_m^* = 0$ and the smallest trapping region is the origin in the shifted coordinates. Thus, $A_s(m) \prec 0$ and $d(m) = 0$ implies that the state m is a globally stable equilibrium point for the system (1).

In summary, we can characterize, with no conservatism, the smallest radius R_m^* such that the trapping region $B(m, R_m^*)$ contains all the states where the energy is instantaneously non-decreasing. All trajectories of the original system (1) will eventually converge to and be tightly bounded in the region $B(m, R_m^*)$. The global optimum R_m^* from the dual SDP (20) provides a tighter bound than the value $R_m := \frac{\|d(m)\|_2}{|\lambda_1|}$ given in Theorem 1. The discussion in this section applies to any m such that $A_s(m) \prec 0$. It can also be applied to the optimal m^* obtained from the SDP (17) assuming it has an optimal cost $a^* < 0$.

Note that one might potentially want to find the tightest trapping region among all possible coordinate shifts. This corresponds to jointly optimizing the problem (19) over (m, y) with the additional constraint (16) such that a trapping region exists. Recall that $d(m) = c + Lm + f(m)$. Hence, constraints of (19) have a term $f(m)^\top y$ that is cubic in the variables (m, y). The problem no longer

Author Manuscript For Peer Review

has the same strong duality results, in general. We are not aware of an efficient approach to jointly optimize the problem (19) over (m, y). Nonlinear optimization algorithms can be applied, but they are not guaranteed to converge to the global optima. Further investigation of solving this problem is left to future work.

3.3 | Recovering Furthest Points with Non-Decreasing Energy

In this subsection, we characterize the set of optimal solutions y^* of the QCQP (19). These correspond to states y^* with norm of R_m^* and non-decreasing energy. The case $d(m) = 0$ has the trivial optimal solution $y^* = 0$ as discussed in the previous subsection. This subsection focuses on the case of $d(m) \neq 0$.

Strong duality holds between the QCQP (19) and the dual SDP (20) when $d(m) \neq 0$. Hence any primal and dual optimal solutions must satisfy the first-order optimality conditions. Specifically, any primal and dual optimal variables $(y^*, \gamma^*, \lambda^*)$ must satisfy the following Karush–Kuhn–Tucker (KKT) conditions (Chapter 5.5.3 of Boyd and Vandenberghe[8]):

$$\textit{Stationary condition:} \qquad \nabla_y \left[f_0(y) + \lambda^* f_1(y)\right]\big|_{y=y^*} = 0 \tag{21a}$$

$$\textit{Primal feasibility:} \qquad f_1(y^*) \leq 0 \tag{21b}$$

$$\textit{Dual feasibility:} \qquad \lambda^* \geq 0, \quad f_2(\lambda^*, \gamma^*) \succeq 0 \tag{21c}$$

$$\textit{Complementary slackness:} \qquad \lambda^* f_1(y^*) = 0 \tag{21d}$$

The stationary condition (21a) for the given f_0 and f_1 is:

$$0 = \nabla_y \left[f_0(y) + \lambda^* f_1(y)\right]\big|_{y=y^*} = 2(I + \lambda^* A_s(m))y^* + \lambda^* d(m).$$

Let r denote the rank of $(I + \lambda^* A_s(m))$ and $V \in \mathbb{R}^{n\times(n-r)}$ be a matrix whose columns are orthonormal and span the null space of $(I + \lambda^* A_s(m))$. Then for any optimal point y^* there exists $c \in \mathbb{R}^{n-r}$ such that:

$$y^*(c) = -\frac{\lambda^*}{2}(I + \lambda^* A_s(m))^\dagger d(m) + Vc,$$

where $(\cdot)^\dagger$ is the pseudo-inverse of a matrix.

Next, observe that the dual constraint $f_2(\lambda^*, \gamma^*) \succeq 0$ in (21c) is infeasible when $\lambda^* = 0$. Hence, the dual constraints (21c) imply that $\lambda^* > 0$. The complementary slackness condition (21d) then further implies that $f_1(y^*(c)) = 0$. Note that $f_1(y^*(c)) = 0$ also implies the primal feasibility (21b). Therefore, the primal-dual point $(y^*(c), \lambda^*, \gamma^*)$ satisfies KKT conditions (21) for any $c \in \mathbb{R}^{n-r}$. The primal points $y^*(c)$ are only candidates for the optimal solution of the QCQP (19) as the KKT conditions are only necessary for optimality. The null space vector c^* yields an optimal point if $y^*(c^*)$ achieves the optimal cost:

$$(R_m^*)^2 = f_0(y^*(c^*)) = \frac{(\lambda^*)^2}{4} d(m)^\top \left((I + \lambda^* A_s(m))^\dagger\right)^2 d(m) + (c^*)^\top c^*. \tag{22}$$

Therefore, the set of optimal solutions $y^*(c^*)$ forms a $(n-r)$-dimensional hypersphere with radius of $\|c^*\|_2$:

$$\left\{ y^*(c^*) = -\frac{\lambda^*}{2}(I + \lambda^* A_s(m))^\dagger d(m) + Vc^* \;:\; \|c^*\|_2^2 = (R_m^*)^2 - \frac{(\lambda^*)^2}{4} d(m)^\top \left((I + \lambda^* A_s(m))^\dagger\right)^2 d(m) \right\}, \tag{23}$$

where the first term of $y^*(c^*)$ is the center of the hypersphere, columns of V are $(n-r)$ directions of the hypersphere in n-dimensional space. This $(n-r)$-dimensional hypersphere corresponds to the points y^* furthest from the origin of the shifted system (6) where the energy $K_m(y^*(c^*))$ is instantaneously non-decreasing.

3.4 | Comparison to Existing Results

The following theorem summarizes the proposed convex optimization method for computing a trapping region.

Theorem 2. *The system* (1) *has a monotonically attracting trapping region* $B(m, R_m^*)$ *if and only if the optimal cost of SDP* (17) *satisfies* $a^* < 0$. *Moreover, if* $a^* < 0$ *then there exists at least one* m *such that* $A_s(m) \prec 0$. *The tightest trapping region centered at any such* m *is given by* $B(m, R_m^*)$ *where* $(R_m^*)^2$ *is the optimal cost of SDP* (20).

Author Manuscript For Peer Review

Proof. The first statement follows from Lemma 2. The second statement follows from Lemma 3. □

The proposed method is to first solve the SDP (17) and, if $a* < 0$, then solve for R_m^* using the SDP in (20). As an aside, the SDP (20) also yields the optimal Lagrange multiplier λ^*. This can be used to compute the points on the boundary of $B(m, R_m^*)$ with non-decreasing energy. These correspond to the points $x^* = y^*(c) + m$ with $y*$ in the set (23).

As a comparison, this paper has three key differences to Theorem 1 (result in Schlegel and Noack[5]). First, this paper provides a computational approach to verify the existence of a trapping region with guarantees, while Schlegel and Noack's non-convex approach does not provide such guarantees. Second, the tightest trapping region is identified in this paper instead of an upper bounding radius in Schlegel and Noack. Third, this paper characterizes the set of points on the boundary of the tightest trapping region with non-decreasing energy, which is not discussed in Schlegel and Noack.

4 | NUMERICAL EXAMPLES

Here, we provide numerical examples to illustrate the proposed SDP-based analysis. First, the trapping region of the two-dimensional example system (14) is revisited. Second, the Lorenz attractor, which highlights the complex dynamics that can arise from models with lossless quadratic nonlinearities, is presented as a benchmark to compare our method against that given in Schlegel and Noack[5]. Lastly, a system with an adjustable number of states is used to study the scalability of the proposed method. All analyses are implemented in Matlab with CVX[13] and SDP solver MOSEK[11]. The examples are run on a standard desktop with an Intel(R) i7-9700 CPU @ 3.00GHz and 16 GB memory. The complete source code for the implementation and examples can be found on GitHub at https://github.com/SCLiao47/Boundedness_LosslessQuadSys.

4.1 | Bounded Two-Dimensional System

Recall the two-dimensional system previously introduced in Section 2.3:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix} x + \begin{bmatrix} -x_1 x_2 \\ x_1^2 \end{bmatrix}. \tag{24}$$

The linear symmetric part is $L_s = \left[\begin{smallmatrix} -1 & 0 \\ 0 & -4 \end{smallmatrix}\right]$. If there is zero coordinate shift, i.e., $m = 0$, then $A_s(m) = L_s$ is negative definite. Hence, the system is bounded by Theorem 1, and a conservative estimate on the trapping region radius was given as $R_0 = \frac{\|c\|_2}{|\lambda_1|} = 1$. The optimal cost for the SDP (17) satisfies $a^* < 0$ because $A_s(m) \prec 0$ at $m = 0$. Moreover, we can use Theorem 2 to compute the tightest trapping region $B(0, R_0^*)$. The optimal solution of the SDP (20) is $R_0^* = 0.289$ with $\lambda^* = 1$. Lastly, the critical points computed by (23) are $x^* = y^* = [\pm 0.236,\ 1.67]^\top$. Note that $I + \lambda^* L_s = \left[\begin{smallmatrix} 0 & 0 \\ 0 & -3 \end{smallmatrix}\right]$ is rank one. Therefore, the set of critical points forms a one-dimensional hypersphere, i.e., two endpoints of an interval in $\mathbb{R}^2$.

Figure 2 illustrates these results. The left plot shows state trajectories (light gray) in the state space starting from random initial conditions. All trajectories eventually converge into both valid trapping regions $B(0, R_0)$ and $B(0, R_0^*)$. Note that the estimated radius R_0 is an over-estimation, and the radius R_0^* is the tightest radius that contains the ellipsoid E with nondecreasing energy (18) in green. Two critical points x^* are marked by purple diamonds. The critical points have norm R_0^* and correspond to the points in E that are furthest from the origin. The right plot shows the energy versus time evaluated along the state trajectories shown in the left plot. All energy trajectories eventually decrease below R_0^* and well below R_0, as shown by the analyses. Overall, the proposed SDP-based analysis provides a more accurate characterization of the trapping region than Theorem 1.

Note that this system has a single equilibrium point at $x_{eq} := (0, 0.25)$, and this is marked by a magenta rectangle on the left plot of Figure 2. The energy $\frac{1}{2}\|x - x_0\|_2^2$ is a Lyapunov function that proves this equilibrium point is globally asymptotically stable. This further implies that the system (14) is bounded. However, this is not true in general for lossless quadratic systems, i.e., they can have unstable equilibrium points and still have bounded trajectories (see next example). Moreover, the proposed SDP-based method can be used to verify the boundedness of lossless quadratic systems without directly computing any equilibrium point (as is done in Lyapunov analysis).

Author Manuscript For Peer Review

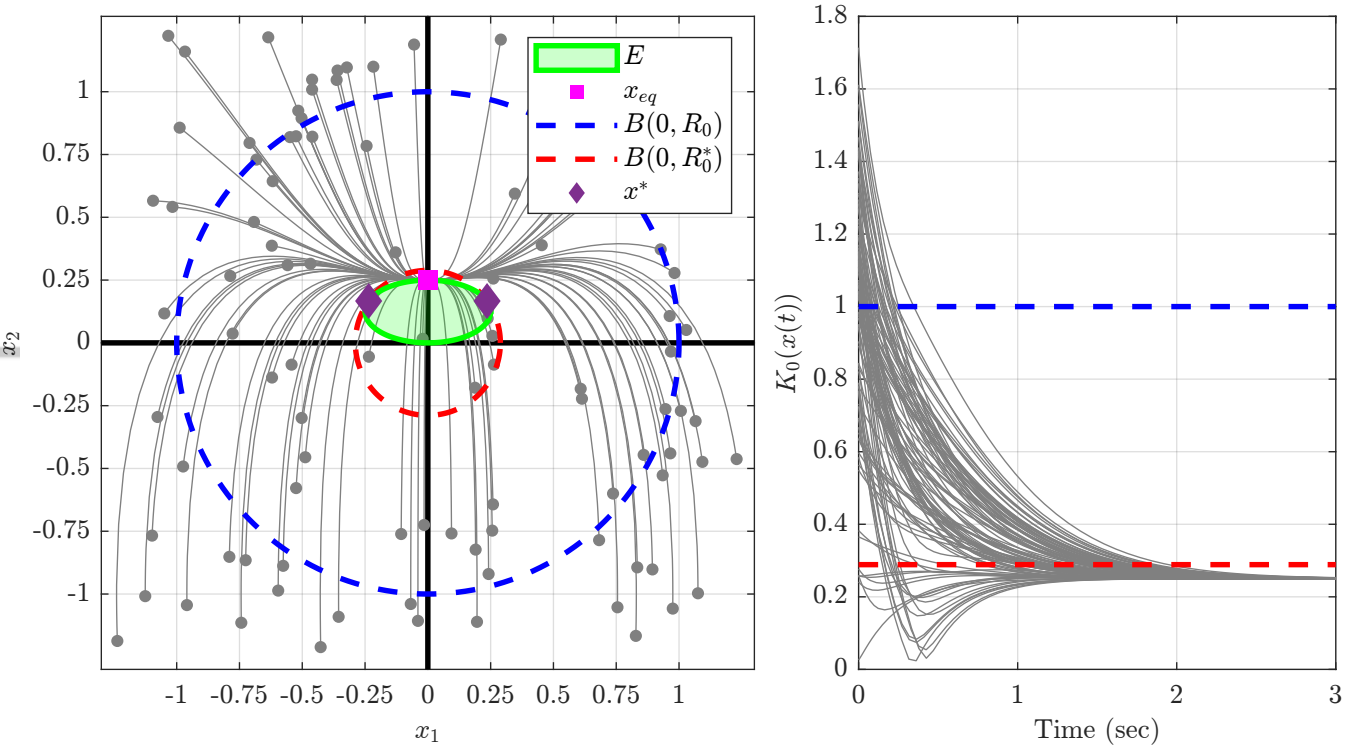


FIGURE 2 Illustrations of trapping regions of the bounded two-state system 24. The left plot shows the ellipsoid with non-decreasing energy E, valid trapping regions $B(0, R_0)$ and $B(0, R_0^*)$, and critical points x^*. Note that the radius R_0 (in blue) is an over-estimation discussed in Section 2.3, and R_0^* (in red) is the tightest radius computed by the SDP-based method proposed in Section 3.2. The critical points x^* (purple diamonds) are computed by the process proposed in Section 3.3. The right plot shows the energy versus time evaluated along the state trajectories shown in the left plot. All energy trajectories decrease below R_0^* in this system and well below R_0.

4.2 | Lorenz Chaotic Attractor

Consider the Lorenz system[1]:

$$\frac{dx_1}{dt} = -\sigma x_1 + \sigma x_2 \tag{25a}$$

$$\frac{dx_2}{dt} = \rho x_1 - x_2 - x_1 x_3 \tag{25b}$$

$$\frac{dx_3}{dt} = -\alpha x_3 + x_1 x_2, \tag{25c}$$

where $\sigma = 10, \rho = 28$ and $\alpha = \frac{8}{3}$. The system has no stable equilibrium points and exhibits chaotic behavior. Nonetheless, the system is shown to be bounded in Schlegel and Noack[5]. Specifically, it can be verified analytically that $A_s(m)$ is negative definite with the translation $m = [0,\ 0,\ \rho + \sigma]^\top$. Moreover, a bound on the trapping region, given by Theorem 1, is $B(m, R_m)$ with $R_m = 101.33$. Note that the estimated radius R_m is different from the radius reported in Schlegel and Noack[5], as there is a minor typo in the estimation formula as pointed out in Kaptanoglu et al[6].

Next, we demonstrate the SDP-based results introduced in Section 3. First, the SDP (17) was solved yielding $a^* = -1$ and $m = [0,\ 0,\ 38]^\top$. It follows from Theorem 2 and $a^* < 0$ that a trapping region exists. Note that the solution m is the same as the translation m used in Schlegel and Noack[5]. This is not a typical case, as the set of m such that $A_s(m)$ is negative definite is not unique, and the solution depends on the exact formulation of both approaches. Next, we solve the SDP (20) and compute the smallest trapping region $B(m, R_m^*)$ with $R_m^* = 39.25 < R_m = 101.33$. Lastly the critical points are characterized by (23) as $x^* = y^* + m$ with $y^* = [0,\ \pm 24.82,\ -30.4]^\top$.

Author Manuscript
For Review

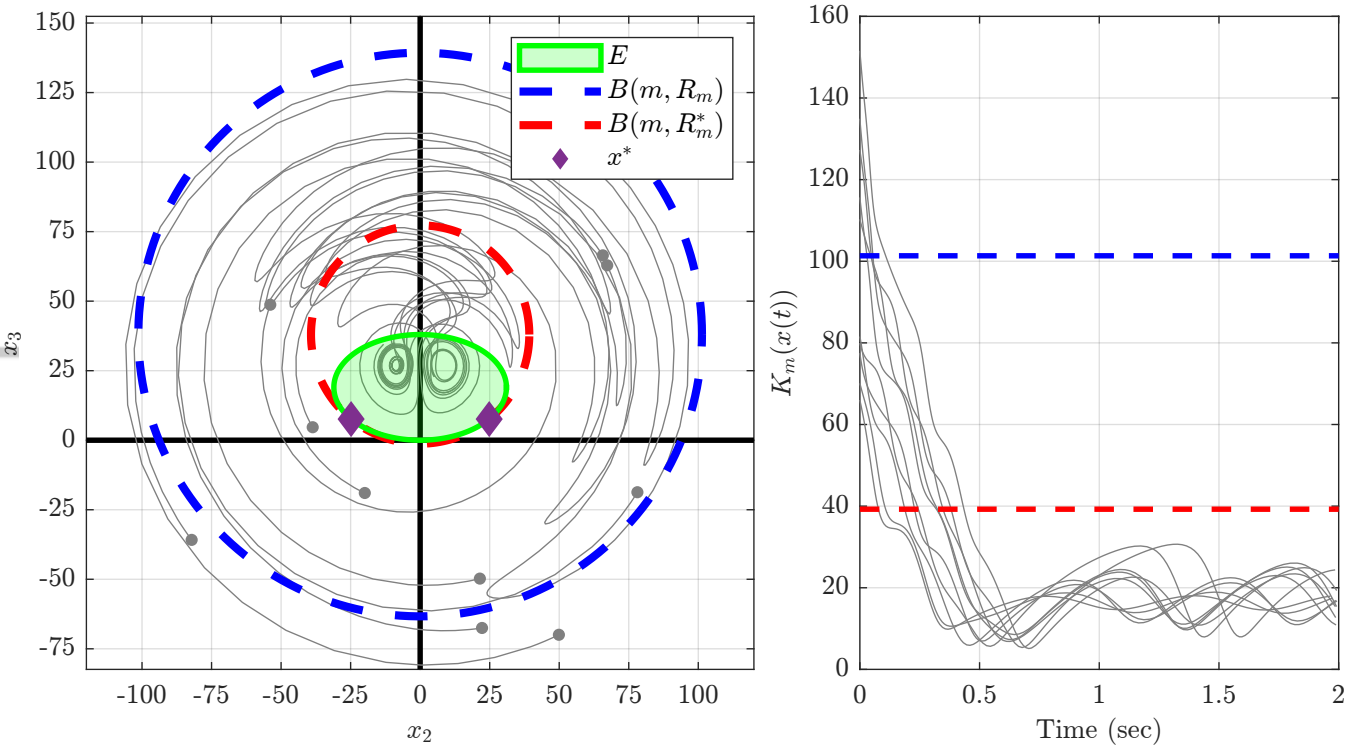


FIGURE 3 Trapping regions of the Lorenz attractor computed by Theorem 1 (blue) and the proposed SDP-based analysis (red). The left plot visualizes trajectories in the $x_2 - x_3$ plane with $x_1 = 0$. The green ellipsoid is the region where energy is instantaneously non-decreasing. The right plot shows the energy versus time evaluated along the state trajectories shown in the left plot. All trajectories are ultimately bounded in the region computed by the proposed SDP-based analysis (red), which captures the boundedness behavior more accurately than Theorem 1 (blue).

Figure 3 illustrates these results. The left plot shows ten trajectories with randomly sampled initial conditions in the $x_2 - x_3$ plane with $x_1 = 0$. Trajectories do not converge to any steady state or even a limit cycle since the system is chaotic. However, all trajectories are bounded as they are trapped in the boundedness region shown above. The red and blue circles are the computed boundedness region by our method and Theorem 1 respectively. The green ellipsoid is the area of instantaneously non-decreasing energy as characterized by the quadratic inequality in (18). The purple diamonds mark the critical points x^*, which are the furthest points with non-decreasing energy from m. All trajectories converge to and stay inside of both blue and red circles as dictated by Theorem 1 and 2 respectively. The right plot presents the energy $K_m(x(t))$ evolution of the ten trajectories. As predicted, all energies eventually decay below and stay below levels of R_m (blue) and R_m^* (red). Furthermore, our analysis (red) provides a more accurate description of the boundedness behavior of the Lorenz system.

4.3 | High-Dimensional Systems

The example illustrates the proposed method on a higher dimensional system. The example is constructed by stacking multiple Lorenz systems and coupling them via random coordinate rotation. Specifically, consider a collection of K systems with Lorenz dynamics (25):

$$\dot{z}^{(i)}(t) = L^{(i)} z^{(i)}(t) + f^{(i)}\left(z^{(i)}(t)\right) \quad i = 1, \ldots K,$$

Author Manuscript
For Review

where

$$L^{(i)} = \begin{bmatrix} -\sigma & \sigma & 0 \\ \rho & -1 & 0 \\ 0 & 0 & -\alpha \end{bmatrix} \quad \text{and} \quad f^{(i)}(z^{(i)}) = \begin{bmatrix} 0 \\ -z_1^{(i)} z_3^{(i)} \\ z_1^{(i)} z_2^{(i)} \end{bmatrix}. \tag{26}$$

We use the same parameters for each subsystem: $\sigma = 10$, $\rho = 28$, and $\alpha = \frac{8}{3}$. Next, create a larger system by stacking these Lorenz subsystems:

$$\dot{z}(t) = L_z z(t) + f_z(z(t)) \text{ where } z(t) := \begin{bmatrix} z^{(1)}(t) \\ \vdots \\ z^{(K)}(t) \end{bmatrix} \in \mathbb{R}^{3K}.$$

The matrix $L_z \in \mathbb{R}^{3K \times 3K}$ and function $f_z : \mathbb{R}^{3K} \to \mathbb{R}^{3K}$ are obtained by appropriate concatenation of the Lorenz subsystems. Here f_z is a lossless quadratic function as each Lorenz subsystem has a lossless quadratic term. Finally, define the state for our high-dimensional example as $x = Wz$ where $W \in \mathbb{R}^{3K \times 3K}$ is an orthonormal matrix. The dynamics of this stacked and rotated system are:

$$\dot{x}(t) = Lx(t) + f(x(t)) \text{ where } L = WL_zW^\top \text{ and } f(x) = Wf_z(W^\top x). \tag{27}$$

The quadratic function in (27) remains lossless after the rotation because:

$$x^\top f(x) = z^\top f(z) = 0$$

The resulting high-dimensional system (27) has a trapping region and is bounded by construction since each subsystem is bounded. However, the system (27) is not obviously bounded as the dynamics of each Lorenz subsystem are coupled via the orthogonal transformation $x = Wz$.

The boundedness of (27) is studied with 18 values of K chosen on a logarithmic grid from $K = 1$ to $K = 223$. The corresponding state dimension of (27) is $n = 3K$ and this goes from 3 to 669. Ten systems with random orthogonal matrices W are generated for each value of K. The SDP-based trapping region analysis in Theorem 2 is applied to each of these ten $3K$-dimensional systems (27). The analyses verified that the system (27) has a trapping region $B(m, R_m^*)$ for each K and each random W. The solution of the SDP (17) yields $a^* = -1$ for each trial. This is the same result as for a single Lorenz system (Section 4.2) because the combined system (27) is simply multiple copies of Lorenz dynamics in rotated coordinates. However, the coordinate shift m is not unique since the first state in Lorenz subsystems has no nonlinearity. Any translation in this direction does not affect the shifted linear symmetric matrix $A_s(m)$. Hence, optimal solutions m form a set and only one of the solutions is returned by the SDP solver. Furthermore, the radius R_m^* of the combined system (27) is not the same for each trial as the radius R_m^* depends on the coordinate shift m. Nonetheless, all systems are bounded as expected.

Figure 4 illustrates the solve time with respect to the number of states $n = 3K$. The computation time is the time to solve the two SDPs (17) and (20). The results show that the computation time increases as n increases (blue line). For a large number of states, the computation time scales empirically with $O(n^{3.98})$. This empirical estimate (red dashed line) was computed by linear regression on the data with $n > 150$. This matches the theoretical complexity of the primal-dual interior-point SDP solver in MOSEK (Section 13.3 of MOSEK[11]). Specifically, the complexity of a primal-dual interior-point SDP solver is $\max\{c_v c_d^3, c_v^2 c_d^2, c_v^3\}$ (Chapter 11.8.3 of Boyd and Vandenberghe[8]), where c_v is the number of decision variables and c_d is the dimension of LMI. The first SDP (17) has $(c_v, c_d) = (n+1, n)$ and the complexity is $O(n^4)$. The second SDP (20), which solves the QCQP (19), has $(c_v, c_d) = (2, n+1)$ and the complexity is $O(n^3)$. Hence, the overall theoretical complexity is $O(n^4)$ and is close to the observed complexity $O(n^{3.98})$. Note that the computation is dominated by the first SDP (17) to verify the existence of a trapping region. Hence, the radius of the trapping region is relatively cheap to compute by the second SDP (20) once the coordinate shift m is found. Note that the implementation is not optimized and the choice of $K = 223$ is based on the memory size of our desktop platform. The analysis can scale to even larger systems by more efficient implementation, customized SDP solver, and improved computational hardware.

Author Manuscript

For Peer Review

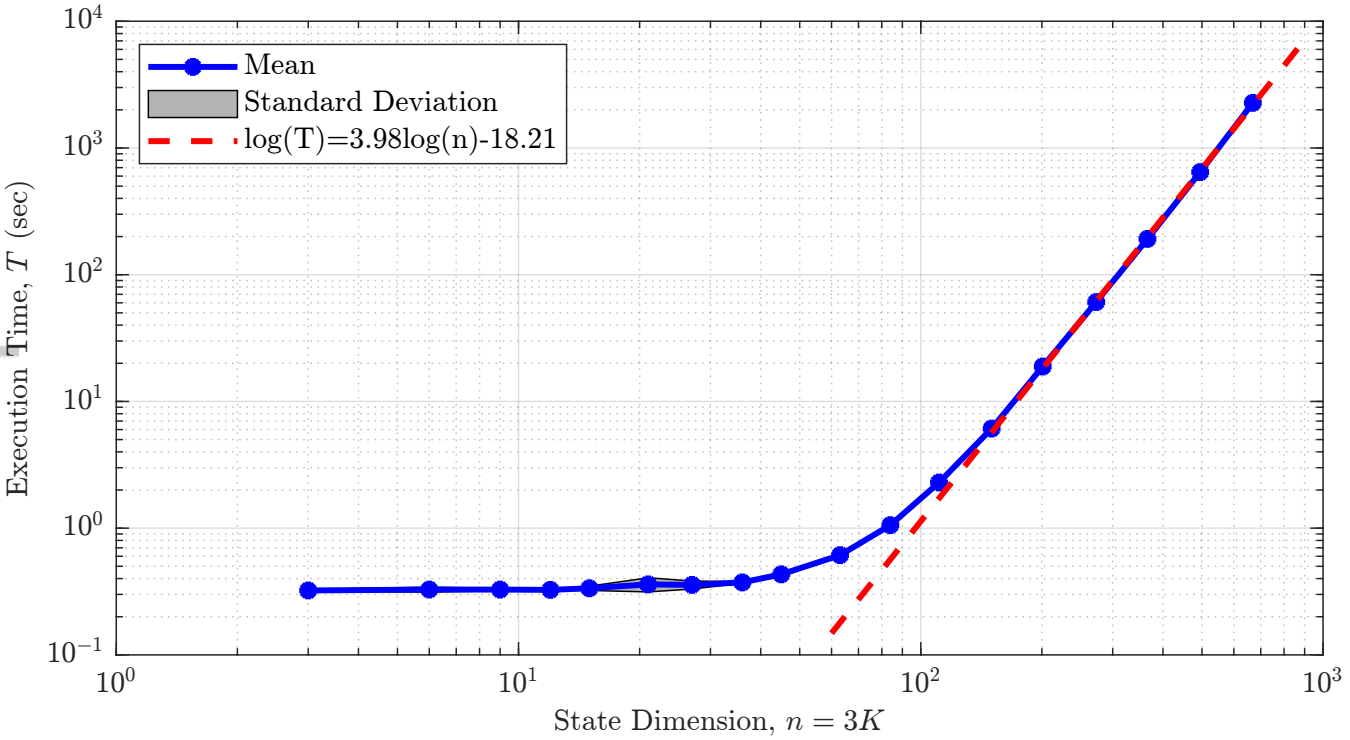


FIGURE 4 Computation times T for the proposed SDP-based analysis versus the number of states $n = 3K$ in the high-dimensional system (27). The computation includes solving the SDP (17) and SDP (20). The mean (blue) increases as number of states grows, and the standard deviation (gray) is negligible for large state dimensions. For large state dimensions, the execution time scales with $O(n^{3.98})$ and is close to the theoretical complexity $O(n^4)$.

5 | CONCLUSION

This paper considers systems with lossless quadratic nonlinearities and proposes a convex optimization approach to analyze boundedness using trapping regions. The contribution is threefold. First, the necessary and sufficient condition for a trapping region to exist is formulated as a convex SDP. This SDP provides theoretical and numerical advantages to verify the condition. Readily available SDP solvers can verify or falsify the existence of a trapping region with guarantees. Second, the least conservative radius of the trapping region at a given coordinate, if one exists, is computed by solving a QCQP exactly using duality theory. The solution of QCQP also characterizes the critical states on the boundary of the trapping region with instantaneously non-decreasing energy. This approach reduces the conservative estimation compared to the literature and provides important testing scenarios for further attention. Lastly, numerical examples are presented to illustrate the analysis process and its computational performance. The proposed method can easily verify that a trapping region exists and provides the precise estimate of the region for systems up to $\sim O(100)$ states.

The proposed convex-optimization-based method provides multiple additional opportunities for future work. The first direction is to incorporate this convex condition into modeling procedures. Several attempts[6,7] use the Schlegel and Noack trapping region condition (Theorem 1) to propose system identification algorithms with a physics-informed prior. Our convex SDP (Theorem 2) could potentially serve as a subroutine and improve the computational efficiency and modeling performance of such algorithms. The second direction is to use the convex condition in controller syntheses. SDPs are widely used to compute stabilizing controllers in both linear systems and nonlinear systems[9]. Notably, the convexity of static output feedback synthesis for lossless quadratic systems has been studied[3]. The trapping region condition being convex implies similar convexity structures potentially exist and can benefit controller synthesis tasks. The last direction here is the robustness analysis of the trapping region condition through convex optimization. For example, one could generalize the global analysis for exactly lossless nonlinearities, as considered in this paper, to local boundedness for weakly lossless nonlinearities[14]. Other classes of nonlinear systems, e.g., systems with higher-order nonlinearities, may be amenable to a similar analysis but this requires further investigation. Moreover,

Author Manuscript
For Review

we could extend our convex condition to analyze robustness by incorporating existing robust stability results [15]. In summary, this convex formulation of the trapping region condition establishes potential connections to a large amount of literature in control theory utilizing convexity.

AUTHOR CONTRIBUTIONS

The authors confirm their contribution to the paper as follows: study conception and design: Shih-Chi Liao, A. Leonid Heide, Maziar S. Hemati, Peter J. Seiler; data collection: Shih-Chi Liao; analysis and interpretation of results: Shih-Chi Liao, Maziar S. Hemati, Peter J. Seiler; visualization: Shih-Chi Liao, A. Leonid Heide; draft manuscript preparation: Shih-Chi Liao, ; manuscript revision: Maziar S. Hemati, Peter J. Seiler. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENTS

This material is based upon work supported by the Army Research Office under grant number W911NF-20-1-0156 and the Air Force Office of Scientific Research under grant number FA9550-21-1-0434. The authors acknowledge valuable discussions with Diganta Bhattacharjee.

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

SUPPORTING INFORMATION

Complete source code for the implementation and examples can be found on GitHub at https://github.com/SCLiao47/Boundedness_LosslessQuadSys.

ORCID

Shih-Chi Liao https://orcid.org/0000-0001-8311-3753
A. Leonid Heide https://orcid.org/0009-0006-1522-245X
Maziar S. Hemati https://orcid.org/0000-0003-1831-051X
Peter J. Seiler https://orcid.org/0000-0003-3423-1109

References

1. Lorenz EN. Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences.* 1963;20(2):130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2
2. Schmid PJ, Henningson DS. *Stability and Transition in Shear Flows*. 142. Springer Science & Business Media, 2012.
3. Mushtaq T, Seiler P, Hemati MS. On the convexity of static output feedback control synthesis for systems with lossless nonlinearities. *Automatica.* 2024;159:111380. doi: 10.1016/j.automatica.2023.111380
4. Khalil HK. *Nonlinear Systems. 3rd ed.* Prentice Hall, 2002.
5. Schlegel M, Noack BR. On long-term boundedness of Galerkin models. *Journal of Fluid Mechanics.* 2015;765:325-352. doi: 10.1017/jfm.2014.736
6. Kaptanoglu AA, Callaham JL, Aravkin A, Hansen CJ, Brunton SL. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *Physical Review Fluids.* 2021;6. doi: 10.1103/PhysRevFluids.6.094401
7. Goyal P, Duff IP, Benner P. Guaranteed Stable Quadratic Models and their applications in SINDy and Operator Inference. *arXiv preprint arXiv:2308.13819.* 2023.
8. Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press, 2004
9. Boyd S, El Ghaoui L, Feron E, Balakrishnan V. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994
10. Sturm JF. Using SeDuMi 1.02, A MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software.* 1999;11(1-4):625-653. doi: 10.1080/10556789908805766

Author Manuscript
For Peer Review

11. MOSEK ApS . *The MOSEK optimization toolbox for MATLAB manual. Version 9.2.* 2019.

12. Tütüncü RH, Toh KC, Todd MJ. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical programming.* 2003;95:189–217. doi: 10.1007/s10107-002-0347-5

13. Grant M, Boyd S. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. http://cvxr.com/cvx; 2014.

14. Peng M. Local Stability Guarantees for Data-Driven Quadratically Nonlinear Models. Master's thesis. University of Washington. 2023.

15. Dullerud GE, Paganini F. *A Course in Robust Control Theory: a Convex Approach*. 36. Springer Science & Business Media, 2013

Author Manuscript

For Peer Review