Automatica 69 (2016) 126–136

Contents lists available at ScienceDirect

Automatica

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

Constrained zonotopes: A new tool for set-based estimation and fault detection☆

Joseph K. Scott [a], Davide M. Raimondo [b], Giuseppe Roberto Marseglia [b], Richard D. Braatz [c]

[a] Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC, USA

[b] Identification and Control of Dynamic Systems Laboratory, University of Pavia, Italy

[c] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form
21 December 2015
Accepted 15 February 2016

Keywords:
State estimation
Fault detection
Set-based computing
Zonotopes
Reachability analysis

ABSTRACT

This article introduces a new class of sets, called *constrained zonotopes*, that can be used to enclose sets of interest for estimation and control. The numerical representation of these sets is sufficient to describe arbitrary convex polytopes when the complexity of the representation is not limited. At the same time, this representation permits the computation of exact projections, intersections, and Minkowski sums using very simple identities. Efficient and accurate methods for computing an enclosure of one constrained zonotope by another of lower complexity are provided. The advantages and disadvantages of these sets are discussed in comparison to ellipsoids, parallelotopes, zonotopes, and convex polytopes in halfspace and vertex representations. Moreover, extensive numerical comparisons demonstrate significant advantages over other classes of sets in the context of set-based state estimation and fault detection.

1. Introduction

Many modern control algorithms make use of sets (e.g., intervals, ellipsoids, zonotopes, polytopes) as basic computational objects, with the aim of characterizing some sets of interest, such as reachable or invariant sets of dynamical systems, or sets of states or parameters consistent with a bounded-error model (Althoff, Stursberg, & Buss, 2010; Ingimundarson, Bravo, Puig, Alamo, & Guerra, 2009; Le, Stoica, Alamo, Camacho, & Dumur, 2013; Mayne, Rakovic, Findeisen, & Allgower, 2006; Scott & Barton, 2013). The true set of interest is often difficult or impossible to represent exactly with finite data, so its enclosure by an element of a class of simpler sets is sought instead. The choice of class for a given application is based on a tradeoff between (i) the accuracy with which a member of the class can represent the set of interest, and (ii) the complexity of the required computations. For linear estimation and control problems, the required computations typically involve standard set operations such as Minkowski sums, linear mappings, intersections, and Pontryagin differences. Another important consideration in (ii) is the convenience of the approximating set for its end use, which may involve checking for the inclusion of given points, as in model invalidation and fault diagnosis (Rosa, Silvestre, Shamma, & Athans, 2010), checking for intersection with another set, as in system verification and safety analysis (Althoff et al., 2010), or using the set as a constraint in an optimization problem, as in robust optimal control and active fault diagnosis (Mayne et al., 2006; Raimondo, Marseglia, Braatz, & Scott, 2013).

This article introduces a new class of sets, *constrained zonotopes*, and demonstrates that this class provides a better tradeoff between accuracy and efficiency than existing classes for some representative problems of interest. Although these new sets potentially have broad applicability, their performance is demonstrated here by considering the classical set-based state estimation problem for discrete-time linear systems with bounded noise (Schweppe, 1968), and its application to set-based fault diagnosis (Scott, Findeisen, Braatz, & Raimondo, 2014). Some notable advantages of the constrained zonotope representation are:

- (*Accuracy*) When the complexity of the representation is not limited, it can describe arbitrary convex polytopes;
- (*Efficiency*) Standard set operations, including intersections, can be computed exactly through simple identities;
- (*Tunability*) Effective techniques are provided to conservatively reduce the complexity of a given set, enabling a highly tunable tradeoff between efficiency and accuracy.

☆ BP is acknowledged for financial support of this project. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Mario Sznaier under the direction of Editor Richard Middleton.

E-mail addresses: jks9@clemson.edu (J.K. Scott), davide.raimondo@unipv.it (D.M. Raimondo), groberto.marseglia@gmail.com (G.R. Marseglia), braatz@mit.edu (R.D. Braatz).

http://dx.doi.org/10.1016/j.automatica.2016.02.036

To motivate this new class of sets, common set representations are reviewed in Section 2 and their advantages and disadvantages are discussed with respect to common set operations. Constrained zonotopes are introduced in Section 3, and associated computations are described in Sections 3.1–4. Numerical results are presented in Sections 5–6, and Section 7 concludes the paper.

2. Set representations and operations

Definition 1. Let $P, Z, E \subset \mathbb{R}^n$. P is a *convex polytope* if it is bounded and (1) holds; Z is a *zonotope* if (2) holds, and E is an *ellipsoid* if (3) holds:

$$\exists(\mathbf{H},\mathbf{k}) \in \mathbb{R}^{n_h\times n}\times\mathbb{R}^{n_h}: \quad P=\{\mathbf{z}\in\mathbb{R}^n:\mathbf{Hz}\le\mathbf{k}\}, \tag{1}$$

$$\exists(\mathbf{G},\mathbf{c}) \in \mathbb{R}^{n\times n_g}\times\mathbb{R}^{n} \quad : \quad Z=\{\mathbf{G}\boldsymbol{\xi}+\mathbf{c}:\|\boldsymbol{\xi}\|_\infty\le 1\}, \tag{2}$$

$$\exists(\mathbf{Q},\mathbf{c}) \in \mathbb{R}^{n\times n}\times\mathbb{R}^{n} \quad : \quad E=\{\mathbf{Q}\boldsymbol{\xi}+\mathbf{c}:\|\boldsymbol{\xi}\|_2\le 1\}. \tag{3}$$

Z is a *parallelotope* if (2) holds with $n_g = n$ and an *interval* if (2) holds with $\mathbf{G} = \mathbf{I}_{n\times n}$.

Eq. (1) is called the *halfspace-representation* (H-rep) of P. P can also be represented as the convex hull of its vertices (V-rep). Zonotopes are convex polytopes that are *centrally symmetric*; every chord through $\mathbf{c}$ is bisected by $\mathbf{c}$. Moreover, a convex polytope is a zonotope if and only if every 2-face is centrally symmetric (McMullen, 1971). This symmetry makes the representation (2) possible. The vector $\mathbf{c}$ is called the *center*, the columns of $\mathbf{G}$ are called the *generators*, and (2) is called the *generator-representation* (G-rep). The G-rep of a zonotope is often much more compact than the equivalent H- or V-rep. Both zonotopes and ellipsoids are affine images of a unit ball. However, zonotopes use the ∞-norm and n_g need not equal n. The representation (3) captures degenerate ellipsoids when $\mathbf{Q}$ is singular and is equivalent to the familiar form $E = \{\mathbf{z} : (\mathbf{z}-\mathbf{c})^{\mathrm{T}}(\mathbf{QQ}^{\mathrm{T}})^{-1}(\mathbf{z}-\mathbf{c}) \le 1\}$ whenever $\mathbf{Q}$ is invertible.

Note that intervals, parallelotopes, and ellipsoids all have fixed complexity for fixed n. In contrast, convex polytopes and zonotopes can be made arbitrarily complex by increasing the number of halfspaces and generators, respectively, which makes these sets more flexible, but also more cumbersome. The complexity of a zonotope is described by its *order*, n_g/n.

For the estimation and fault diagnosis problems considered in Sections 5–6, as well as many other problems in linear control theory, the accuracy and efficiency of the below set operations are of primary concern:

Definition 2. Let $Z, W \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^k$, $\mathbf{R} \in \mathbb{R}^{k\times n}$, and define

$$\mathbf{R}Z \equiv \{\mathbf{Rz} : \mathbf{z} \in Z\}, \tag{4}$$

$$Z + W \equiv \{\mathbf{z}+\mathbf{w} : \mathbf{z}\in Z,\ \mathbf{w}\in W\}, \tag{5}$$

$$Z \cap_{\mathbf{R}} Y \equiv \{\mathbf{z}\in Z : \mathbf{Rz}\in Y\}. \tag{6}$$

Eq. (4) is a linear mapping of Z, (5) is the *Minkowski sum*, and (6) is a generalized intersection that arises in state estimation (e.g., with Z containing the current state and Y a bounded-error measurement; see Section 5). Note that $\cap_{\mathbf{R}}$ is the standard intersection when $k = n$ and $\mathbf{R} = \mathbf{I}$.

A class of sets is *closed* under a set operation if performing the operation on members of the class results in another member of the class. The convex polytopes are closed under (4)–(6) and, using H-rep, both (4) and (6) can be computed efficiently if $\mathbf{R}$ is invertible. However, the complexity of (5) is exponential in n, as is the worst-case number of halfspaces describing $Z + W$ (Hagemann, 2015; Tiwary, 2008). The same is true of (4) and (6) when $\mathbf{R}$ is not invertible (e.g., polytope projection) (Jones, Kerrigan, & Maciejowski, 2008). In V-rep, (4)–(5) are much simpler, but (6) is NP-hard (Tiwary, 2008), and existing algorithms for interconversion between H- and V-rep have worst-case exponential run-time. Consequently, working with convex polytopes is very costly and numerically unstable when n exceeds about 5 or the number of halfspaces or vertices is large.

In contrast, intervals, parallelotopes, and ellipsoids all provide low-complexity set representations and relatively low-cost set operations. However, the intervals are not closed under (4) unless $\mathbf{R}$ is diagonal, the parallelotopes and ellipsoids are not closed under (5), and none of these classes are closed under (6) except intervals when $\mathbf{R}$ is diagonal. Thus, the results of these operations must be conservatively enclosed, which can ultimately lead to very inaccurate enclosures of the set of interest. The optimal interval enclosures of these operations are easily computed (Neumaier, 1990), but are often very weak enclosures of the true sets. For ellipsoids, cheap heuristic enclosure methods are given in Schweppe (1968). Optimal enclosures are given in Chernousko (1980), Durieu, Walter, and Polyak (2001) and Fogel and Huang (1982), but (6) requires the solution of a convex optimization when $k > 1$. Cheap heuristic enclosures for parallelotopes are given in Chisci, Garulli, and Zappa (1996), and numerical results there show that these are tighter than even the optimal ellipsoidal enclosures in the context of state estimation.

Over the past decade, zonotopes have gained popularity within the control community, particularly because (4)–(5) can be computed exactly and efficiently in G-rep (Kuhn, 1998). Define the shorthand $Z = \{\mathbf{G},\mathbf{c}\} \subset \mathbb{R}^n$ for Z defined by (2). Then, with $Z = \{\mathbf{G}_z, \mathbf{c}_z\}$ and $W = \{\mathbf{G}_w, \mathbf{c}_w\}$,

$$\mathbf{R}Z = \{\mathbf{RG}_z, \mathbf{Rc}_z\}, \tag{7}$$

$$Z + W = \{[\mathbf{G}_z\ \mathbf{G}_w], \mathbf{c}_z + \mathbf{c}_w\}. \tag{8}$$

Clearly, these computations can be done efficiently and reliably, even in high dimensions. Like general convex polytopes, these operations are nonconservative, but lead to an increase in the complexity of the set representation. However, in contrast to the worst-case exponential increase in the size of the H-rep under (4)–(5), the increase in the complexity of the G-rep is modest; $\mathbf{R}Z$ has the same n_g as Z, while the n_g of $Z + W$ is simply the sum of the n_g's of Z and W. Moreover, conservative *order reduction* techniques are available that enclose a given zonotope within a zonotope of lower order (Althoff et al., 2010; Combastel, 2003). Similar techniques have also been proposed for convex polytopes, but the required computations are much more complex (Hagemann, 2015). For zonotopes, these techniques provide a tunable mechanism for balancing accuracy and complexity that has proven to be effective in reachability analysis (Althoff et al., 2010; Kuhn, 1998), identification (Bravo, Alamo, & Camacho, 2006), state estimation (Alamo, Bravo, & Camacho, 2005), and fault detection (Ingimundarson et al., 2009; Scott et al., 2014).

However, zonotopes are not closed under intersection, and tight enclosures are difficult to compute, which leads to serious complications in many applications, such as state estimation and hybrid systems verification (Althoff & Krogh, 2011; Bravo et al., 2006). Indeed, the symmetry of zonotopes, as well as intervals, parallelotopes, and ellipsoids, implies that they cannot accurately represent sets that are strongly centrally asymmetric, which are readily generated by (6). This has led some researchers to use a combination of G- and H-rep, although the conversion from G- to H-rep can be costly; it scales as $n_g\binom{n_g}{n-1}$ (Althoff & Krogh, 2011).

Set representations based on collections of sets have also been proposed, such as unions of intervals (Neumaier, 1990) and intersections of ellipsoids (Kurzhanski, 2011) or zonotopes (Althoff & Krogh, 2011). These can be very accurate, but the associated cost increases with the number of sets required, which can be large. We restrict the scope of the comparisons herein to 'single-set' representations.