Underapproximative Methods for the Order Reduction of Zonotopes

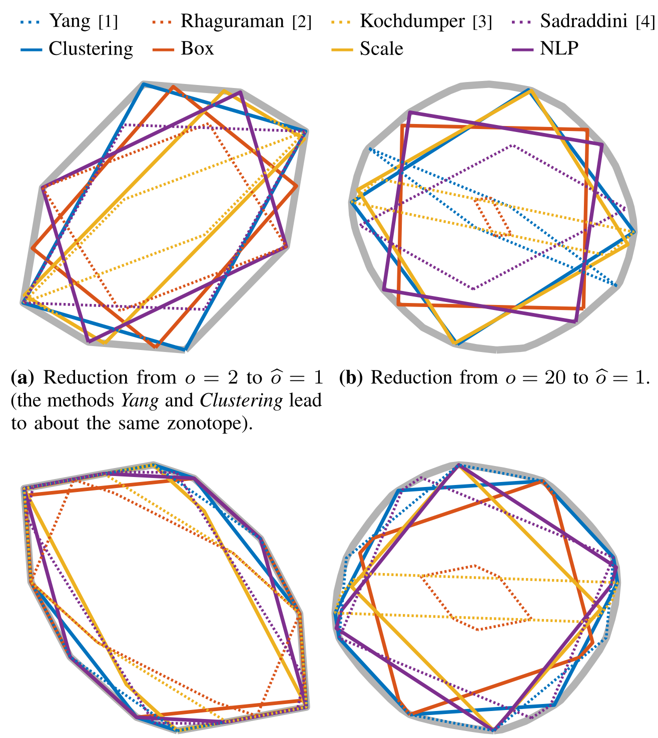
Laura Lützow, *Graduate Student Member, IEEE*, Niklas Kochdumper, and Matthias Althoff, *Member, IEEE*

***Abstract*—Zonotopes are a widely used set representation in set-based computations due to their compact representation size and their closure under many relevant set operations. However, certain set operations, such as the Minkowski sum, increase the zonotope order, which in turn increases the computational cost of further computations. To address this issue, various order reduction techniques have been proposed, most of which focus on overapproximating the original zonotope. While overapproximations are crucial for safety verification, some applications – such as reachset-conformant identification and backward reachability analysis – require underapproximations (also referred to as inner-approximations). Besides providing a comprehensive survey of existing underapproximative order reduction methods, we propose four novel reduction methods in this letter. We analyze the computational cost of all methods and evaluate the tightness of the resulting underapproximations through numerical experiments on more than 2000 randomly generated zonotopes. The results demonstrate that our proposed methods achieve a favorable balance between computational efficiency and approximation accuracy, making them well-suited for applications in control, estimation, and system identification.**

***Index Terms*—Order reduction, set-based computations, underapproximation, zonotopes.**

I. Introduction

SET-BASED computations enable the efficient handling of complex problems by processing entire sets simultaneously. However, the computational efficiency and accuracy of these computations depends heavily on the choice of set representation. A good trade-off between expressiveness and computational efficiency is offered by zonotopes, which are a special type of centrally symmetric polytopes. They can be compactly represented by a center and a number of generators and are closed under several common set operations. Due to these properties, zonotopes are widely used for set-based computations in applications such as reachability analysis [5], [6], safe-by-construction controller synthesis [7], [8], guaranteed state and parameter estimation [9], [10], [11], [12], abstract interpretation [13], and conformant system identification [14], [15].



(a) Reduction from $o = 2$ to $\widehat{o} = 1$ (the methods *Yang* and *Clustering* lead to about the same zonotope). (b) Reduction from $o = 20$ to $\widehat{o} = 1$. (c) Reduction from $o = 3$ to $\widehat{o} = 2$. (d) Reduction from $o = 20$ to $\widehat{o} = 2$.

Fig. 1. Example two-dimensional zonotopes of order o (gray) and underapproximative zonotopes of order $\widehat{o} < o$ computed with different methods.

Certain set operations that are frequently used in set-based computations increase the zonotope order, which is defined as the number of generators divided by the zonotope dimension. Since the computational cost of many algorithms involving zonotopes scales with the zonotope order, reducing the order is crucial for maintaining computational efficiency. Most existing methods address this challenge by overapproximating the original zonotope with a reduced-order zonotope [16], [17]. While

Received 17 March 2025; revised 5 May 2025; accepted 26 May 2025. Date of publication 9 June 2025; date of current version 21 July 2025. This work was supported in part by the German Research Foundation (DFG) – SFB 1608 – under Grant 501798263; in part by the ANR-JST research project CyPhAI under Award ANR-20-JSTM-0001 and Award JPMJCR2012; and in part by a fellowship of the German Academic Exchange Service (DAAD). Recommended by Senior Editor T. Oomen. *(Corresponding author: Laura Lützow.)*

Laura Lützow and Matthias Althoff are with the School of Computation, Information and Technology, Technical University of Munich, 85748 Garching, Germany (e-mail: laura.luetzow@tum.de; althoff@tum.de).

Niklas Kochdumper is with the IRIF Institute, Université Paris Cité, 75013 Paris, France (e-mail: kochdumper@irif.fr).

Digital Object Identifier 10.1109/LCSYS.2025.3577642

overapproximations are essential for problems like safety verification [6] or safe-by-construction controller synthesis [8], other applications, such as safety falsification [18], [19] and the estimation of the guaranteed reachable set [20], require underapproximations. Underapproximations are also critical in reachset-conformant system identification [3], [14], where the model is guaranteed to be conformant if all measurements from the real system are contained in the exact or an underapproximative reachable set of the model. Additionally, they play a key role in backward reachability analysis [1], where the set of states that are guaranteed to reach a given target set is computed.

This letter provides an overview of existing methods for computing reduced-order underapproximations of zonotopes and introduces four novel approaches. We evaluate all methods in terms of computational cost and the tightness of the resulting underapproximations through numerical experiments.

In Section II, we introduce the notation and present fundamentals on zonotopes and order reduction. In Section III, we summarize state-of-the-art underapproximation techniques, while our novel methods are detailed in Section IV. All methods are compared based on their computational cost and the tightness of the underapproximations in Section V.

II. Preliminaries

In this section, we introduce the notation used in this letter and describe the encoding of infinity norm constraints. Furthermore, we present fundamentals on zonotopes and underapproximative order reduction.

A. Notation

We denote sets by calligraphic letters, matrices by uppercase letters, and vectors and scalars by lower-case letters. Furthermore, we use $\mathbf{0}$ for the zero vector. The horizontal concatenation of two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times o}$ is denoted by $[A\ B]$. The i-th element in the j-th column of the matrix A is denoted by $A_{(i,j)}$, and the j-th element of the vector $a \in \mathbb{R}^m$ is denoted by $a_{(j)}$. Removing the i-th and j-th column from the matrix $A \in \mathbb{R}^{m\times n}$ leads to the matrix $A_{\setminus i \setminus j} \in \mathbb{R}^{m\times(n-2)}$. The operation $\operatorname{diag}(a)$ returns a diagonal matrix with the elements of the vector $a \in \mathbb{R}^m$ on its main diagonal, $\det(C)$ computes the determinant of $C \in \mathbb{R}^{m\times m}$, and $\operatorname{sign}(x)$ returns -1 if $x \in \mathbb{R}$ is smaller than 0, and 1 otherwise.

Important set operations are linear transformation and Minkowski sum, where the linear transformation of the set $\mathcal{S} \subset \mathbb{R}^n$ using the matrix $A \in \mathbb{R}^{m\times n}$ is defined as $A\mathcal{S} = \{As | s \in \mathcal{S}\}$. Given two sets $\mathcal{S}_\mathrm{a}, \mathcal{S}_\mathrm{b} \subset \mathbb{R}^n$, their Minkowski sum is defined as $\mathcal{S}_\mathrm{a} \oplus \mathcal{S}_\mathrm{b} = \{s_\mathrm{a} + s_\mathrm{b} | s_\mathrm{a} \in \mathcal{S}_\mathrm{a},\ s_\mathrm{b} \in \mathcal{S}_\mathrm{b}\}$.

B. Infinity Norm Constraints

In the remainder of this letter, we will often use constraints of the form $\|T\|_\infty \le \alpha$ with $T \in \mathbb{R}^{\nu\times\widehat{\nu}}$ and $\alpha \ge 0$, where the infinity norm $\|T\|_\infty$ is defined as $\|T\|_\infty = \max_{1\le i\le \nu} \sum_{j=1}^{\widehat{\nu}} |T_{(i,j)}|$. To encode this constraint, we use the linear formulation from [21, Eq. (6)], which is based on the vertex representation of the L_1-norm ball. In particular, the constraint $\|T\|_\infty \le \alpha$ requires the containment of the vectors formed by each matrix row in the L_1-norm unit-ball scaled by α. The L_1-norm unit-ball in $\mathbb{R}^{\widehat{\nu}}$ is a polytope and can be represented as the convex hull of its $2\widehat{\nu}$ vertices [4, Eq. (3)], which is more efficient than representing the L_1-norm ball via the $2^{\widehat{\nu}}$ inequality constraints for its faces. Since the vertices are given by the positive and negative unit vectors $\pm e_j,\ j = 1, \dots, \widehat{\nu}$, we obtain the linear encoding $\forall i = 1, \dots, \nu,\ \exists \mu_i^-, \mu_i^+ \in \mathbb{R}^{\widehat{\nu}}_{\ge 0}$: $t_i = \sum_{j=1}^{\widehat{\nu}} (\mu^+_{i(j)} - \mu^-_{i(j)})\, e_j \ \wedge\ \sum_{j=1}^{\widehat{\nu}} (\mu^+_{i(j)} + \mu^-_{i(j)}) = \alpha$, where $T = [t_1\ \dots\ t_\nu]^\top$.

C. Zonotopes

Zonotopes can be defined as follows [22, Ch. 7.3]:

Definition 1 (Zonotopes): A zonotope $\mathcal{Z} \subset \mathbb{R}^n$ is defined by a center vector $c \in \mathbb{R}^n$ and a generator matrix $G = [g_1\ \dots\ g_\nu] \in \mathbb{R}^{n\times\nu}$:

$$\mathcal{Z} = \left\{ c + \sum_{i=1}^{\nu} \lambda_{(i)} g_i \,\middle|\, \lambda_{(i)} \in [-1, 1] \right\} = \langle c, G \rangle.$$

The order o of the zonotope is defined as $o = \frac{\nu}{n}$ and represents a measure for the complexity of a zonotope that is independent of the dimension.

The linear transformation of a zonotope $\mathcal{Z} = \langle c, G\rangle \subset \mathbb{R}^n$ using the matrix $A \in \mathbb{R}^{m\times n}$ can be computed as $\langle Ac, AG\rangle \subset \mathbb{R}^m$. Given two zonotopes $\mathcal{Z}_\mathrm{a} = \langle c_\mathrm{a}, G_\mathrm{a}\rangle \subset \mathbb{R}^n$ and $\mathcal{Z}_\mathrm{b} = \langle c_\mathrm{b}, G_\mathrm{b}\rangle \subset \mathbb{R}^n$, their Minkowski sum can be computed as $\langle c_\mathrm{a} + c_\mathrm{b}, [G_\mathrm{a}\ G_\mathrm{b}]\rangle$. Thus, every time two zonotopes are added using the Minkowski sum, the zonotope order increases, which motivates the need for efficient order reduction methods.

D. Underapproximative Order Reduction

We consider order reduction methods that underapproximate the zonotope $\mathcal{Z} = \langle c, G\rangle$, $G \in \mathbb{R}^{n\times\nu}$ of order $o = \frac{\nu}{n}$ by a zonotope $\widehat{\mathcal{Z}} = \langle c, \widehat{G}\rangle \subseteq \mathcal{Z}$, $\widehat{G} \in \mathbb{R}^{n\times\widehat{\nu}}$ of order $\widehat{o} = \frac{\widehat{\nu}}{n} < o$. Note that order reduction results in a new generator matrix $\widehat{G}$ while the center vector c remains unchanged.

Two general approaches exist for computing reduced-order underapproximations: *Optimization-based methods* aim to maximize a measure of the volume of $\widehat{\mathcal{Z}}$, while ensuring the containment $\widehat{\mathcal{Z}} \subseteq \mathcal{Z}$ via the sufficient condition [4, Th. 3]

$$\exists T \in \mathbb{R}^{\nu\times\widehat{\nu}}\colon \widehat{G} = GT,\ \ \|T\|_\infty \le 1. \tag{1}$$

The constraints in (1) are linear in T and $\widehat{G}$ since $\|T\|_\infty \le 1$ can be encoded as described in Section II-B.

Heuristic methods satisfy the containment condition in (1) by replacing generators in G by their sum or difference, e.g., the combination of the two generators g_i and g_j would lead to the reduced-order generator matrix

$$\widehat{G} = \left[G_{\setminus i \setminus j}\ g_i + a g_j \right], \tag{2}$$

with $a \in \{1, -1\}$. The generators that are combined in (2) are selected with the aim of maximizing the volume of $\widehat{\mathcal{Z}}$.

When deriving the computational complexity of the heuristic methods with respect to the dimension n, we make the assumption that the zonotope order o is a constant that is independent of n, while the number of generators is $\nu = o\,n$. We do not provide the computational complexity for the optimization-based methods, as their complexity depends on the specific solver and algorithm used.

III. Existing Order Reduction Methods

We identified the following four order reduction methods in the literature, where the first three methods are based on heuristics and the last method uses optimization.

A. Yang

Yang and Ozay [1, Sec. IV.B] propose a heuristic underapproximation method, where generators that are closely aligned or small in 2-norm are greedily combined. This leads to the following procedure for computing the reduced-order generator matrix $\widehat{G}$ from G:

1) Find

$$(i, j) = \underset{1 \leq i < j \leq \nu}{\arg\min} \|g_i\|_2 \left\| g_j - \frac{g_i}{\|g_i\|_2^2} g_j^\top g_i \right\|_2 .$$

2) Set $G \leftarrow$ (2), where

$$a = \begin{cases} 1 & \text{if } \frac{\|H(g_i+g_j)\|_2}{\|H(g_i-g_j)\|_2} \geq 1, \\ -1 & \text{otherwise,} \end{cases}$$

and H is the right inverse of $G_{\setminus i \setminus j}$, which we compute as $H = G_{\setminus i \setminus j}^\top (G_{\setminus i \setminus j} G_{\setminus i \setminus j}^\top)^{-1}$.

3) The number of generators in G decreases to $\nu \leftarrow \nu - 1$. If $\nu > \widehat{\nu}$, continue with Step 1). Otherwise, set $\widehat{G} \leftarrow G$.

Assuming $\nu = o\,n$, the computational complexity of this method is $\mathcal{O}(n^4)$, since the computation of the inverse matrix in each iteration has complexity $\mathcal{O}(n^3)$, and the number of iterations is proportional to n.

B. Raghuraman

Raghuraman and Koeln [2, Sec. 5.1] split the generators into large generators $g_{s,j}$, $j = 1, \dots, \widehat{\nu}$, and small generators $g_{s,i}$, $i = \widehat{\nu} + 1, \dots, \nu$ according to the 2-norm. The reduced-order generator matrix $\widehat{G}$ is created from the large generators, where each small generator is added to or subtracted from the large generator that it is most aligned to in terms of the 2-norm. This idea can be implemented as follows:

1) Create the sorted generator matrix $G_s = [g_{s,1} \ \dots \ g_{s,\nu}]$ by sorting the generators in G according to their 2-norm, i.e., $\|g_{s,i}\|_2 \geq \|g_{s,i+1}\|_2, \ \forall i = 1, \dots, \nu - 1$.
2) Compute $P_{(j,i)} = g_{s,i}^\top g_{s,j}, \ \forall i = 1, \dots, \widehat{\nu}, \ \forall j = \widehat{\nu} + 1, \dots, \nu$.
3) Compute the reduced-order generator matrix $\widehat{G} = G_s T$, with $T \in \mathbb{R}^{\nu \times \widehat{\nu}}$ and

$$T_{(j,i)} = \begin{cases} 1 & \text{if } j = i \text{ and } j \leq \widehat{\nu}, \\ \frac{P_{(j,i)}}{|P_{(j,i)}|} & \text{if } j > \widehat{\nu} \text{ and } \forall k \neq i : |P_{(j,i)}| > |P_{(j,k)}|, \\ 0 & \text{otherwise.} \end{cases}$$

Using $\nu = o\,n$, the computational complexity is $\mathcal{O}(n^3)$ due to the $\widehat{\nu}(\nu - \widehat{\nu})$ dot products $g_{s,i}^\top g_{s,j}$ in Step 2), each of which has complexity $\mathcal{O}(n)$.

C. Kochdumper

Kochdumper and Bak [3, Proposition 1] propose to sum the generators with the smallest 2-norm, where the first element $g_{i(1)}$ of each generator g_i determines the sign in the sum. This results in the following procedure:

1) Create the sorted generator matrix G_s as done in Step 1) of the *Rhaguraman* method in Section III-B.
2) Compute the reduced-order generator matrix $\widehat{G}$ as $\widehat{G} = [g_{s,1} \ \dots \ g_{s,\widehat{\nu}-1} \ \sum_{i=\widehat{\nu}}^{\nu} \operatorname{sign}(g_{s,i(1)})\, g_{s,i}]$.

The computational complexity is $\mathcal{O}(n^2)$, because computing the 2-norm of one generator is $\mathcal{O}(n)$, and the number of generators is proportional to n.

D. Sadraddini

Sadraddini and Tedrake [4, Corollary 6] aim to compute the reduced-order generator matrix $\widehat{G}$ that minimizes the Hausdorff distance δ [4, Eq. (36)] between the original zonotope $\mathcal{Z}$ and the reduced-order zonotope $\widehat{\mathcal{Z}}$ via the following optimization problem:

$$\min_{\substack{\delta \in \mathbb{R}_{\geq 0},\, \widehat{G} \in \mathbb{R}^{n \times \widehat{\nu}},\, T \in \mathbb{R}^{\nu \times \widehat{\nu}}, \\ T_\Delta \in \mathbb{R}^{\widehat{\nu} \times \nu},\, \Delta \in \mathbb{R}^{n \times \nu}}} \delta \quad \text{s.t. } (1),\ G = \widehat{G} T_\Delta + \Delta, \quad \|T_\Delta\|_\infty \leq 1,\ \|\Delta\|_\infty \leq \delta. \tag{3}$$

Due to the bilinear expression $\widehat{G} T_\Delta$, this optimization problem is nonlinear. To solve it efficiently, the authors propose the following procedure [4, Sec. V.B]:

1) Randomly initialize $\widehat{G} \in \mathbb{R}^{n \times \widehat{\nu}}$.
2) Find diagonal matrix $\operatorname{diag}(\lambda)$, with $\lambda \in \mathbb{R}^{\widehat{\nu}}$, such that

$$\exists T \in \mathbb{R}^{\nu \times \widehat{\nu}} \colon \widehat{G} \operatorname{diag}(\lambda) = G T \ \wedge \ \|T\|_\infty \leq 1.$$

3) Set $\widehat{G} \leftarrow \widehat{G} \operatorname{diag}(\lambda)$.
4) Alternate between solving (3) for T_Δ and $\widehat{G}$ until convergence. With fixed $\widehat{G}$ or fixed T_Δ, the optimization problem in (3) becomes a linear program that can be solved efficiently.

Although this method explicitly searches for the underapproximative zonotope that has the minimum Hausdorff distance to the original zonotope, the resulting reduced-order zonotope might not be optimal due to the following reasons: First, there is no guarantee that the iterative procedure above converges to the optimal solution of problem (3). Second, the zonotope containment constraints in (1) are sufficient but not necessary.

IV. Novel Underapproximation Methods

This section presents four novel methods for underapproximative order reduction: one heuristic method based on (2) and three optimization-based approaches that incorporate (1) within their constraints.

A. Clustering

Most heuristic underapproximation methods, such as the methods from Yang and Ozay [1] and Raghuraman and Koeln [2], try to find generators that are closely aligned as their combination via (2) promises a less conservative underapproximation. To find globally optimal groupings of aligned generators rather than relying on a purely greedy strategy, we propose to use kmeans++ clustering [23], which extends the kmeans algorithm by a sophisticated method for cluster center initialization.

1) We partition the generators in G into $\widehat{\nu}$ clusters $\mathcal{C}_k$, $k = 1, \dots, \widehat{\nu}$, using the k-means++ algorithm [23], where

the distance between a generator g_i and the cluster center g_j is computed as $1 - d_{\cos}(g_i, g_j)$ with

$$d_{\cos}\big(g_i, g_j\big) = \operatorname{sign}\big(g_{i(1)} g_{j(1)}\big) \frac{g_i^\top g_j}{\sqrt{\big(g_i^\top g_i\big)\big(g_j^\top g_j\big)}}. \quad (4)$$

Eq. (4) computes the cosine of the angle between $\operatorname{sign}(g_{i(1)})g_i$ and $\operatorname{sign}(g_{j(1)})g_j$, where we include the signum functions to ensure that both vectors are pointing in the same direction in the first dimension (multiplication of generators with -1 does not change the zonotope). Using the first element is a heuristic choice inspired by the *Kochdumper* method.

2) We combine generators assigned to the same cluster $\mathcal{C}_k$ with $\widehat{g}_k = \sum_{g \in \mathcal{C}_k} \operatorname{sign}(g_{(1)})g$.
3) The reduced generator matrix is given as $\widehat{G} = [\widehat{g}_1 \ \dots \ \widehat{g}_{\widehat{\nu}}]$.

The complexity of evaluating (4) for $\nu = o\,n$ generators g_i and $\widehat{\nu} = \widehat{o}\,n$ cluster centers g_j with dimension n is $\mathcal{O}(n^3)$.

B. Box

Another idea is to reduce the zonotope by computing a box underapproximation, which can be represented by the zonotope $\mathcal{B} = \langle c, \operatorname{diag}(\gamma)\rangle$ with $\gamma \in \mathbb{R}^n$.

1) The zonotope $\mathcal{Z} = \langle c, G\rangle$ is split into the zonotope $\mathcal{Z}_\mathrm{a} = \langle c, G_\mathrm{a}\rangle$ that is not changed and the zonotope $\mathcal{Z}_\mathrm{b} = \langle \mathbf{0}, G_\mathrm{b}\rangle$ that will be underapproximated, i.e., $\mathcal{Z} = \mathcal{Z}_\mathrm{a} \oplus \mathcal{Z}_\mathrm{b}$, where $G_\mathrm{a} \in \mathbb{R}^{n\times(\widehat{\nu}-n)}$ and $G_\mathrm{b} \in \mathbb{R}^{n\times\nu_\mathrm{b}}$, $\nu_\mathrm{b} = \nu - \widehat{\nu} + n$, are created from partitioning the generators in G, such that $\forall i = 1, \dots, \widehat{\nu} - n$ and $\forall j = 1, \dots, \nu_\mathrm{b}$

$$\|g_{\mathrm{a},i}\|_1 - \|g_{\mathrm{a},i}\|_\infty \geq \|g_{\mathrm{b},j}\|_1 - \|g_{\mathrm{b},j}\|_\infty. \quad (5)$$

2) Even though the metric used to sort the generators in (5) is expected to select generators for reduction that are close to a box [5, Sec. 3.4], an even better underapproximating box might still be obtained if the zonotope $\mathcal{Z}_\mathrm{b}$ is further rotated into a more favorable orientation. According to [16, Algorithm 2], a suitable rotation matrix $U \in \mathbb{R}^{n\times n}$ can be computed via singular value decomposition of $C = [G_\mathrm{b} \ \ -G_\mathrm{b}][G_\mathrm{b} \ \ -G_\mathrm{b}]^\top$, i.e., $C = U\Sigma V^\top$, which leads to the rotated zonotope $\mathcal{Z}_\mathrm{br} = U^\top \mathcal{Z}_\mathrm{b}$.
3) The rotated zonotope $\mathcal{Z}_\mathrm{br} = \langle \mathbf{0}, G_\mathrm{br}\rangle$ is underapproximated by a box $\mathcal{B} = \langle \mathbf{0}, \operatorname{diag}(\gamma)\rangle \subseteq \mathcal{Z}_\mathrm{br}$. A tight box underapproximation can be computed via optimization, where we maximize the volume of the box and use the sufficient linear condition in (1) to guarantee $\mathcal{B} \subseteq \mathcal{Z}_\mathrm{br}$:

$$\max_{\substack{\gamma\in\mathbb{R}^n_{\geq 0},\\ T\in\mathbb{R}^{\nu_\mathrm{b}\times n}}} 2^n \prod_{i=1}^n \gamma_{(i)} \quad \text{s.t.} \quad \operatorname{diag}(\gamma) = G_\mathrm{br}\, T, \ \ \|T\|_\infty \leq 1.$$

According to [7, Sec. V.A], this optimization problem can be formulated as a second-order cone program using the approach in [24, Sec. 2.3 e)], because maximizing the concave function $\sqrt[n]{\prod_i \gamma_{(i)}}$ yields the same optimal scaling factors as maximizing the volume $2^n \prod_i \gamma_{(i)}$ since the function $\sqrt[n]{x}$ is strictly monotonically increasing in $x \in \mathbb{R}_{\geq 0}$.

4) We rotate the underapproximating box $\mathcal{B}$ back using U. Since $\mathcal{B} \subseteq \mathcal{Z}_\mathrm{br} = U^\top \mathcal{Z}_\mathrm{b}$ and U is orthogonal, i.e., $U^{-1} = U^\top$, we obtain $U\mathcal{B} \subseteq \mathcal{Z}_\mathrm{b}$ after multiplication with U from the left. The resulting underapproximative zonotope of order $\widehat{o}$ is computed as $\widehat{\mathcal{Z}} = \mathcal{Z}_\mathrm{a} \oplus U\mathcal{B} = \langle c, [G_\mathrm{a} \ U \operatorname{diag}(\gamma)]\rangle$.

C. Scale

One can also scale the generators of a suitable zonotope template, where good templates can be generated using one of the numerous methods for overapproximative order reduction of zonotopes [16], [17]:

1) Given the zonotope $\mathcal{Z} = \langle c, G\rangle$, we first compute a reduced-order zonotope $\mathcal{Z}_\mathrm{o} = \langle c, G_\mathrm{o}\rangle$ with $G_\mathrm{o} \in \mathbb{R}^{n\times\widehat{\nu}}$ that overapproximates the zonotope $\mathcal{Z} \subseteq \mathcal{Z}_\mathrm{o}$ using an overapproximative order reduction method [16], [17].
2) The underapproximative zonotope $\widehat{\mathcal{Z}} \subseteq \mathcal{Z}$ is obtained by scaling each generator $g_{\mathrm{o},i}$ of $\mathcal{Z}_\mathrm{o}$ with a scaling factor $\lambda_{(i)}$, i.e., $\widehat{\mathcal{Z}} = \langle c, G_\mathrm{o} \operatorname{diag}(\lambda)\rangle$, $\lambda \in \mathbb{R}^{\widehat{\nu}}_{\geq 0}$. We can compute the scaling factors with the following linear program, where we use the sufficient linear condition in (1) to enforce $\widehat{\mathcal{Z}} \subseteq \mathcal{Z}$ while maximizing the 2-norm of each generator:

$$\max_{\substack{\lambda\in\mathbb{R}^{\widehat{\nu}}_{\geq 0},\\ T\in\mathbb{R}^{\nu\times\widehat{\nu}}}} \sum_{i=1}^{\widehat{\nu}} \lambda_{(i)} \|g_{\mathrm{o},i}\|_2 \quad \text{s.t.} \quad \begin{aligned} &G_\mathrm{o} \operatorname{diag}(\lambda) = G\,T, \\ &\|T\|_\infty \leq 1. \end{aligned}$$

D. Nonlinear Programming (NLP)

Finding the reduced-order underapproximation $\widehat{\mathcal{Z}} = \langle c, \widehat{G}\rangle$ with the maximum volume can be written as the following nonlinear optimization problem:

$$\max_{\widehat{G}\in\mathbb{R}^{n\times\widehat{\nu}},\ T\in\mathbb{R}^{\nu\times\widehat{\nu}}} \operatorname{volume}\big(\langle c, \widehat{G}\rangle\big) \quad \text{s.t.} \quad (1), \quad (6)$$

where (1) enforces the containment $\widehat{\mathcal{Z}} \subseteq \mathcal{Z}$. Assuming a full-rank generator matrix, the volume of a zonotope can be computed as [25, Corollary 3.4]

$$\operatorname{volume}\big(\langle c, \widehat{G}\rangle\big) = 2^n \sum_{1\leq j_1<\dots<j_n\leq\widehat{\nu}} \big|\det\big(\widehat{G}_{j_1,\dots,j_n}\big)\big|, \quad (7)$$

where the summation considers all possible choices of $j_1, \dots, j_n$ that satisfy $1 \leq j_1 < \dots < j_n \leq \widehat{\nu}$, and $\widehat{G}_{j_1,\dots,j_n}$ is constructed from columns $j_1, \dots, j_n$ of $\widehat{G}$. Note that we can remove the factor 2^n when using (7) as a cost function since constant factors do not change the location of the optimum.

This approach searches for the optimal underapproximation. However, since the cost function is not concave, optimizers using gradient ascend are not guaranteed to find the global maximum. Furthermore, the computational cost of evaluating cost function (7) is high for a large number of generators $\widehat{\nu}$ due to the large number of choices of $j_1, \dots, j_n$. To obtain a faster runtime, we can use the following procedure:

1) We split the zonotope $\mathcal{Z} = \langle c, G\rangle$ into the zonotope $\langle c, G_\mathrm{a}\rangle$, whose generator matrix $G_\mathrm{a} \in \mathbb{R}^{n\times\widehat{\nu}_\mathrm{a}}$ consists of the $(\widehat{\nu} - n)$ generators with the largest 2-norm, and the zonotope $\langle \mathbf{0}, G_\mathrm{b}\rangle$, whose generator matrix $G_\mathrm{b} \in \mathbb{R}^{n\times(\nu-\widehat{\nu}+n)}$ consists of the remaining generators, i.e., $\mathcal{Z} = \langle c, G_\mathrm{a}\rangle \oplus \langle \mathbf{0}, G_\mathrm{b}\rangle$.

2) We underapproximate $\langle \mathbf{0}, G_b \rangle$ with $\langle \mathbf{0}, \widehat{G}_b \rangle$, where $\widehat{G}_b \in \mathbb{R}^{n \times n}$, using the nonlinear program in (6). Since $\widehat{G}_b$ is a square matrix, the cost function in (7) simplifies to

$$\mathrm{volume}\big(\langle c, \widehat{G}_b \rangle\big) = 2^n |\det\big(\widehat{G}_b\big)|.$$

3) The resulting underapproximation of order $\widehat{o}$ is computed as $\widehat{\mathcal{Z}} = \langle c, G_a \rangle \oplus \langle \mathbf{0}, \widehat{G}_b \rangle = \langle c, [G_a \; \widehat{G}_b] \rangle$.

V. Comparison

We compare the performance of the presented methods for different dimensions and zonotope orders. For each dimension-order configuration, we create 50 random zonotopes, where the direction of the generators is determined by randomly generating points on a sphere, and their length is sampled from a certain distribution [16, Sec. IV.A]. Here, we use the uniform distribution over the interval [0, 1], the exponential distribution with a mean of one, and the gamma distribution with a shape parameter of one and a scale parameter of two for one third of the zonotopes each. We measure the computation time and the tightness of the resulting underapproximations, where the tightness of an underapproximation is assessed using the volume ratio [16, Sec. IV.B]

$$R = \left(\frac{\mathrm{volumeApprox}\big(\widehat{\mathcal{Z}}\big)}{\mathrm{volumeApprox}(\mathcal{Z})} \right)^{\frac{1}{n}}. \tag{8}$$

For zonotopes of dimension less than 10 with at most 150 generators, the function volumeApprox$(\cdot)$ computes the exact volume according to (7). In all other cases, volumeApprox$(\cdot)$ approximates the zonotope volume with the volume of the tightest box overapproximation according to [5, Sec. 3.4] because evaluating (7) becomes extremely expensive for high dimensions and a large number of generators.

All methods are implemented in the MATLAB toolbox CORA [26], and the experiments are conducted on an i9-12900HK processor (2.5GHz) with 64GB memory. We solve the linear, nonlinear, and second-order cone programs with the built-in solvers `linprog` using the dual-simplex algorithm, `fmincon` using the sequential quadratic programming algorithm, and `coneprog`. Note that in our implementation of the *Sadraddini* method [4], we perform a fixed number of 10 alternations in Step 4) since this yields a good trade-off between accuracy and computation time. Furthermore, we apply the overapproximative order reduction method *CoOptdir* [16, Sec. III.D] in Step 1) of the *Scale* method because it provides a good trade-off between accuracy and computational cost [16, Table 1].

Example zonotopes and their underapproximations are displayed in Fig. 1, while the volume ratios as well as the computation times are visualized in Fig. 2(a) for reduction to order $\widehat{o} = 1$ and in Fig. 2(b) for reduction to order $\widehat{o} = 2$. We do not display the volume ratios and computation times for methods that take longer than $50s$.

The *NLP* method achieves the tightest underapproximations in low-dimensional settings for reductions to order $\widehat{o} = 1$. However, all four optimization-based methods (*Sadraddini*, *Scale*, *Box*, and *NLP*) incur high computational cost due to the need to solve the corresponding optimization problems, exceeding the feasible computation time for zonotopes with $n \geq 10$ or $o \geq 50$. The fastest methods are *Rhaguraman* and *Kochdumper*, from which *Kochdumper* yields the least conservative underapproximations. The *Yang* and *Clustering* methods provide a good balance between approximation accuracy and efficiency by computing tight underapproximations in reasonable computation time for reductions to orders 1 and 2. While the *Yang* method is more efficient for low-dimensional zonotopes, *Clustering* scales better and is able to successfully handle zonotopes with $n \geq 20$ and $o \geq 200$, where the *Yang* method fails.

VI. Conclusion

This letter provides a comprehensive survey of underapproximative order reduction methods for zonotopes, reviewing four existing approaches from the literature and introducing four novel methods. Through numerical experiments, we compared the approximation accuracy and computational efficiency of all methods and demonstrated that the proposed techniques can achieve significant improvements. Among the presented methods, we find the *Clustering* method the most practical choice for most applications due to its good trade-off between tightness and efficiency. When tight underapproximations are essential and computational cost is not a concern, we recommend utilizing the *NLP* method for reductions to order 1. Conversely, the *Kochdumper* method is well-suited for scenarios with constrained computational resources. Future work may refine these methods through a detailed analysis of various heuristics or by tailoring them to specific applications, such as reachset-conformant system identification and backward reachability analysis.

References

[1] L. Yang and N. Ozay, "Scalable zonotopic under-approximation of backward reachable sets for uncertain linear systems," *IEEE Control Syst. Lett.*, vol. 6, pp. 1555–1560, 2022.

[2] V. Raghuraman and J. P. Koeln, "Set operations and order reductions for constrained zonotopes," *Automatica*, vol. 139, May 2022, Art. no. 110204.

[3] N. Kochdumper and S. Bak, "Conformant synthesis for Koopman operator linearized control systems," in *Proc. 61st IEEE Conf. Decision Control*, 2022, pp. 7327–7332.

[4] S. Sadraddini and R. Tedrake, "Linear encodings for polytope containment problems," in *Proc. 58th IEEE Conf. Decision Control*, 2019, pp. 4367–4372, (References with respect to the extended arXiv version https://doi.org/10.48550/arXiv.1903.05214).

[5] A. Girard, "Reachability of uncertain linear systems using zonotopes," in *Proc. 8th ACM Int. Conf. Hybrid Syst., Comput. Control*, 2005, pp. 291–305.

[6] M. Althoff, O. Stursberg, and M. Buss, "Reachability analysis of linear systems with uncertain parameters and inputs," in *Proc. 46th IEEE Conf. Decision Control*, 2007, pp. 726–732.

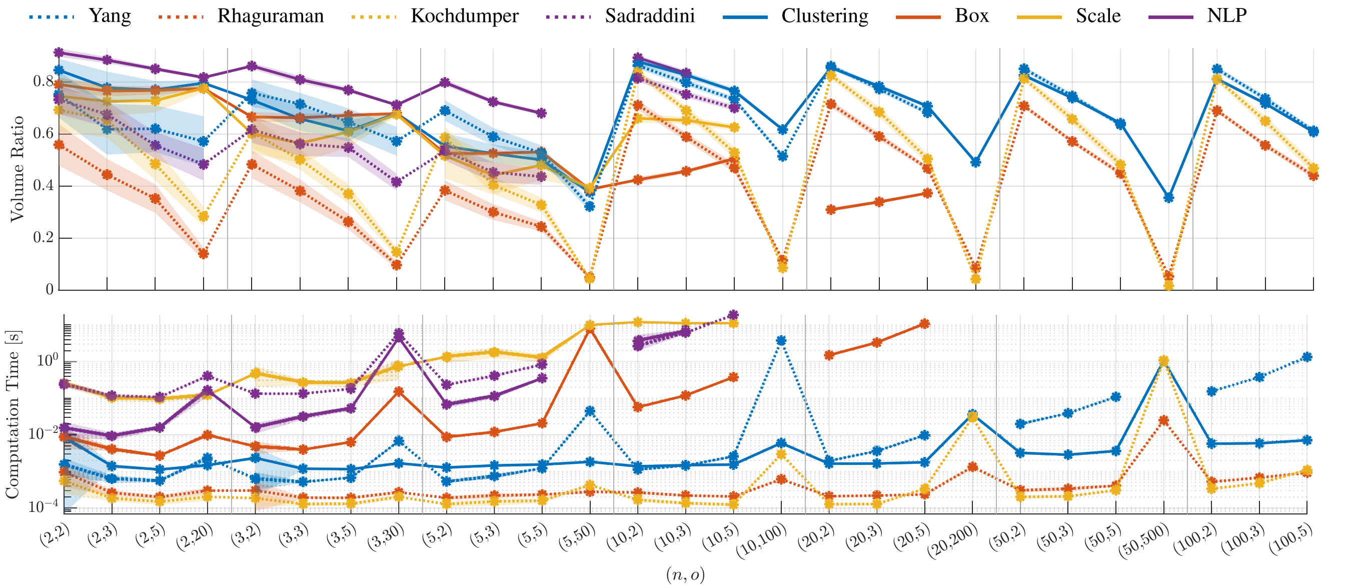
[7] L. Schäfer, F. Gruber, and M. Althoff, "Scalable computation of robust control invariant sets of nonlinear systems," *IEEE Trans. Autom. Control*, vol. 69, no. 2, pp. 755–770, Feb. 2024.

[8] J. M. Bravo, T. Alamo, and E. F. Camacho, "Robust MPC of constrained discrete-time nonlinear systems based on approximated reachable sets," *Automatica*, vol. 42, no. 10, pp. 1745–1751, 2006.

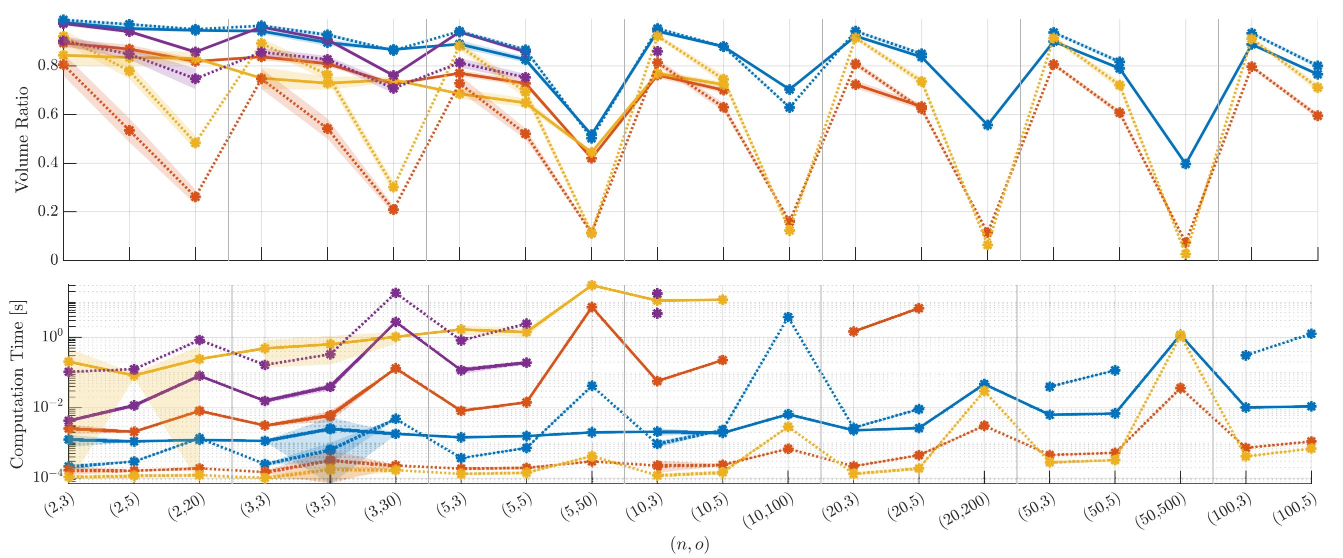
[9] T. Alamo, J. Bravo, and E. Camacho, "Guaranteed state estimation by zonotopes," in *Proc. 42nd IEEE Int. Conf. Decision Control*, 2003, pp. 5831–5836.

[10] C. Combastel, "Zonotopes and Kalman observers: Gain optimality under distinct uncertainty paradigms and robust convergence," *Automatica*, vol. 55, pp. 265–273, May 2015.

[11] S. B. Chabane, C. S. Maniu, T. Alamo, E. Camacho, and D. Dumur, "Improved set-membership estimation approach based on zonotopes and ellipsoids," in *Proc. Eur. Control Conf.*, 2014, pp. 993–998.



(a) Reduction to order $\widehat{o} = 1$.



(b) Reduction to order $\widehat{o} = 2$.

Fig. 2. Resulting volume ratios, computed with (8), and computation times (mean values $\pm \frac{1}{3}$ standard deviation) for different order reduction methods, which were applied to 50 random zonotopes for each dimension n and zonotope order o.

[12] H. Wang, I. V. Kolmanovsky, and J. Sun, "Zonotope-based recursive estimation of the feasible solution set for linear static systems with additive and multiplicative uncertainties," *Automatica*, vol. 95, pp. 236–245, Sep. 2018.

[13] K. Ghorbal, E. Goubault, and S. Putot, "The zonotope abstract domain Taylor1+," in *Proc. Int. Conf. Comput. Aided Verificat.*, 2009, pp. 627–633.

[14] L. Lützow and M. Althoff, "Scalable reachset-conformant identification of linear systems," *IEEE Control Syst. Lett.*, vol. 8, pp. 520–525, 2024.

[15] L. Lützow and M. Althoff, "Recursive identification of reachset-conformant models using constraint underapproximation," in *Proc. Amer. Control Conf.*, 2025, pp. 2530–2537.

[16] A.-K. Kopetzki, B. Schürmann, and M. Althoff, "Methods for order reduction of zonotopes," in *Proc. 56th IEEE Conf. Decision Control*, 2017, pp. 5626–5633.

[17] X. Yang and J. K. Scott, "A comparison of zonotope order reduction techniques," *Automatica*, vol. 95, pp. 378–384, Sep. 2018.

[18] M. Wetzlinger, N. Kochdumper, S. Bak, and M. Althoff, "Fully automated verification of linear systems using inner and outer approximations of reachable sets," *IEEE Trans. Autom. Control*, vol. 68, no. 12, pp. 7771–7786, Dec. 2023.

[19] L. Geretti, P. Collins, P. Nuzzo, and T. Villa, "Constraint-driven nonlinear reachability analysis with automated tuning of tool properties," *Nonlin. Anal. Hybrid Syst.*, vol. 54, Nov. 2024. Art. no. 101532.

[20] D. Ren, Z. Liang, C. Wu, J. Ding, T. Wu, and B. Xue, "Inner-approximate reachability computation via zonotopic boundary analysis," in *Proc. Int. Conf. Comput. Aided Verificat.*, 2024, pp. 307–328.

[21] A. Cristofari, M. De Santis, S. Lucidi, and F. Rinaldi, "Minimization over the ℓ_1-ball using an active-set non-monotone projected gradient," *Comput. Optim. Appl.*, vol. 83, no. 2, pp. 693–721, 2022.

[22] G. M. Ziegler, *Lectures on Polytopes*. New York, NY, USA: Springer, 1995.

[23] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[24] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Math. Program.*, vol. 95, no. 1, pp. 3–51, 2003.

[25] E. Gover and N. Krikorian, "Determinants and the volumes of parallelotopes and zonotopes," *Linear Algebra Appl.*, vol. 433, no. 1, pp. 28–40, 2010.

[26] M. Althoff, "An introduction to CORA 2015," in *Proc. Workshop Appl. Verificat. Continuous Hybrid Syst.*, 2015, pp. 120–151.