Tight Remainder-Form Decomposition Functions With Applications to Constrained Reachability and Guaranteed State Estimation

Mohammad Khajenejad, *Member, IEEE*, and Sze Zheng Yong, *Member, IEEE*

***Abstract*—In this article, we propose a tractable family of remainder-form mixed-monotone decomposition functions that are useful for overapproximating the image set of nonlinear mappings in reachability and estimation problems. Our approach applies to a new class of nonsmooth, discontinuous nonlinear systems that we call either-sided locally Lipschitz semicontinuous systems, which we show to be a strict superset of locally Lipschitz continuous systems, thus expanding the set of systems that are formally known to be mixed-monotone. In addition, we derive lower and upper bounds for the overapproximation error and show that the lower bound is achieved with our proposed approach, i.e., our approach constructs the tightest, tractable remainder-form mixed-monotone decomposition function. Moreover, we introduce a set inversion algorithm that along with the proposed decomposition functions can be used for constrained reachability analysis and guaranteed state estimation for continuous- and discrete-time systems with bounded noise.**

***Index Terms*—Nonlinear dynamical systems, reachability analysis, mixed-monotonicity, one-sided decomposition functions, ELLS systems.**

I. Introduction

MONOTONICITY properties of systems have proven to be very powerful and useful for analyzing and controlling complex systems [1], [2]. Building upon this idea, it was shown that certain nonmonotone systems can be lifted to higher dimensional monotone systems that can be potentially used to deduce critical information about the original system (see, e.g., [3], [4], and [5]) by decomposing the system dynamics into increasing and decreasing components. Systems that are decomposable in this manner are called *mixed-monotone* and are significantly more general than the class of monotone systems.

Furthermore, mixed-monotonicity has also proven to be very beneficial for system analysis and control. For instance, if mixed-monotonicity holds, it can be concluded that the original system has global asymptotic stability by proving the nonexistence of equilibria of the lifted system, except in a certain lower dimensional subspace [6], [7]. Moreover, forward invariant and attractive sets of the original system can be identified [8] and reachable sets of the original system can be efficiently overapproximated and used for state estimation and abstraction-based control synthesis [8], [9], [10], [11]. However, the usefulness of these system lifting techniques for the analysis and control is highly dependent on the *tightness* of the decomposition approaches, their *computational tractability* [12], [13], and their *applicability to a broad class of systems*; therefore, the capability to compute or construct *tractable* and *tight* mixed-monotone decomposition functions for a broad range of nonlinear, uncertain, and constrained systems is of great interest and will have a significant impact.

Literature review: Mixed-monotone decomposition functions are generally not unique, hence several seminal studies have addressed the issue of identifying and/or constructing appropriate decomposition functions with somewhat different yet highly related definitions and corresponding sufficient conditions for mixed-monotonicity [7], [10], [12], [13], [14], [15]. In particular, recent studies in [12] and [13] provided *tight* decomposition functions for unconstrained discrete-time (DT) and continuous-time (CT) dynamical systems, respectively, whose *computability* rely on the global solvability of nonlinear optimization programs, which is only guaranteed in some specific cases, such as when the vector field is *Jacobian sign-stable* (JSS), or when all its *stationary/critical points* can be computed analytically. On the other hand, the authors in [8], [14], [15], and [16] proposed computable and constructive (but not necessarily tight) decomposition functions for differentiable vector fields with known bounds for the derivatives. Building on these frameworks, we aim to obtain computable/tractable and tighter decomposition functions for a broad class of nonsmooth, discontinuous systems.

Another relevant body of literature pertains to *interval arithmetic* [17], [18], [19], [20], which has been successfully applied to problems in numerical analysis, set estimation, motion planning, etc. Specifically, *inclusion functions* and variations

Manuscript received 20 December 2021; revised 22 September 2022; accepted 11 February 2023. Date of publication 1 March 2023; date of current version 5 December 2023. This work was supported by the National Science Foundation under Grant CNS-1943545 and Grant CNS-1932066. Recommended by Associate Editor M. Guay. *(Corresponding author: Sze Zheng Yong.)*

Mohammad Khajenejad is with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92122 USA (e-mail: mkhajenejad@ucsd.edu).

Sze Zheng Yong is with the Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: s.yong@northeastern.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2023.3250515.

Digital Object Identifier 10.1109/TAC.2023.3250515

thereof (e.g., natural, centered-form, and mixed-form inclusions) that are based on interval arithmetic can be directly related to decomposition functions and can similarly be used for overapproximation of the ranges/image sets of functions and for state observer designs [21], [22], [23], [24], [25]. In addition, various properties of inclusion functions, such as their convergence rates and the *subdivision principle*, have been studied in [18]. Subsequent studies further introduced refinements of interval overapproximations and set inversion algorithms that can incorporate new sources of information about the system, such as state constraints, measurements/observations, manufactured redundant variables, and second-order derivatives, e.g., in [17], [21], [22], [23], and [24], which can be also beneficial for decomposition function-based constrained reachability and set-valued estimation problems that we consider in this work.

Contribution: In this article, we introduce a class of mixed-monotone decomposition functions whose construction is both computationally tractable and tight for a broad range of DT and CT nonsmooth and discontinuous nonlinear systems. The proposed mixed-monotone decomposition functions are in the *remainder-form* (based on the terminology in [18]), and we also show that they result in *difference of monotone functions*, which bear some resemblance with difference of convex functions in DC programming that is also widely used in optimization and state estimation problems, e.g., in [23] and [26].

Our work contributes to the literature on mixed-monotone decomposition functions and more generally, inclusion functions in multiple ways.

1) We introduce a class of remainder-form decomposition functions that are tractable (i.e., computable in closed form without an iterative algorithm), which to the best of our knowledge is novel and unique to our work. Our construction approach is proven to be the tightest for this family of decomposition functions, which we also show to include the methods in [8], [14], and [16], thus generalizing and improving on them.
2) The proposed remainder-form decomposition functions apply to a new broader class of discontinuous and nonsmooth nonlinear systems, which we call *either-sided locally Lipschitz semicontinuous* (ELLS) systems, that is proven in this article to be a strict superset of locally Lipschitz continuous (LLC) systems. This new system class relaxes the (almost everywhere) smoothness, bounded gradient, and continuity requirements and allows nonsmooth vector fields with only one-sided bounded *Clarke Jacobians*, including vector fields with countable and finite-valued discontinuities (jumps). Since the class of nonsmooth, discontinuous ELLS systems include differentiable and LLC systems, our results are also novel contributions to LLC and smooth systems.
3) We further show that the overapproximation of the image sets of ELLS systems converges to the true tightest enclosing interval at least linearly, when the interval domain width goes to zero and moreover, a *subdivision principle* applies for improving the enclosure of the range/image set.
4) We introduce a novel set inversion algorithm based on mixed-monotone decomposition functions as an alternative to SIVIA [17] and the refinement algorithm in [21], which enables the design of algorithms for constrained reachability and interval observers for systems with known constraints, modeling redundancy, and/or sensor measurements.

Moreover, it is noteworthy that the inclusion functions that are computed based on mixed-monotone decomposition functions can be used alongside any existing inclusion functions, where the "best of them" (by virtue of an intersection property) is chosen, since we observed that in general, no single inclusion function consistently outperforms all others. Furthermore, our proposed inclusion function can be directly and simply integrated into existing set inversion, constrained reachability, and interval observer algorithms, e.g., [17], [21], [27], without much modifications. Finally, to demonstrate the effectiveness of the proposed algorithms for decomposition function construction and set inversion, we compare them with existing inclusion/decomposition functions in the literature.

II. Background and Problem Formulation

A. Notation

$\mathbb{N}, \mathbb{N}_a, \mathbb{R}^{n_z}, 0_n$ and $\mathbb{R}^{n\times m}$ denote the set of positive integers, the first a positive integers, the n_z-dimensional Euclidean space, the zero vector in $\mathbb{R}^n$, and the space of n by m real matrices, respectively. Moreover, $\forall z, \underline{z}, \overline{z} \in \mathbb{R}^{n_z}$, $\underline{z} \le \overline{z} \Leftrightarrow \underline{z}_i \le \overline{z}_i$, $\forall i \in \mathbb{N}_{n_z}$, where z_i denotes the ith element of z. Further, $\mathbb{I}\mathcal{Z} \triangleq [\underline{z}, \overline{z}] \triangleq \{z \in \mathbb{R}^{n_z} | \underline{z} \le z \le \overline{z}\}$ and $d(\mathbb{I}\mathcal{Z}) \triangleq \|\overline{z} - \underline{z}\|_\infty$ are called a closed interval/hyperrectangle in $\mathbb{R}^{n_z}$ and the diameter of $\mathbb{I}\mathcal{Z}$, respectively, where $\|z\|_\infty \triangleq \max_i |z_i|$ denotes the ℓ_∞-norm of $z \in \mathbb{R}^{n_z}$. The set of all intervals in $\mathbb{R}^{n_z}$ is denoted by $\mathbb{I}\mathbb{R}^{n_z}$.

B. Definitions and Preliminaries

First, we briefly introduce some of the main concepts that we use throughout the paper, as well as some important existing results that will be used for comparisons and for deriving our main results. We start by introducing decomposition and inclusion functions and some of their typical instances.

1) Mixed-Monotonicity and Inclusion Functions:

Definition 1 (Mixed-monotonicity and decomposition functions): [13, Def. 1], [14, Def. 4] Consider the dynamical system with initial state $x_0 \in \mathbb{I}\mathcal{X}_0 \triangleq [\underline{x}_0, \overline{x}_0]$

$$x_t^+ = \tilde{f}(x_t, u_t, w_t) \triangleq f(z_t), \tag{1}$$

where $x_t^+ \triangleq x_{t+1}$ if (1) is a DT system, and $x_t^+ \triangleq \dot{x}_t$ if (1) is a CT system, $\tilde{f} : \mathcal{X} \times \mathcal{U} \times \mathcal{W} \to \mathbb{R}^{n_x}$ is the vector field with state $x_t \in \mathcal{X} \subset \mathbb{R}^{n_x}$, known input $u_t \in \mathcal{U} \subset \mathbb{R}^{n_u}$, and disturbance input $w_t \in \mathcal{W} \subseteq \mathbb{I}\mathcal{W} \triangleq [\underline{w}, \overline{w}] \in \mathbb{I}\mathbb{R}^{n_w}$. For ease of exposition, we also define $f : \mathcal{Z} \triangleq \mathcal{X} \times \mathcal{W} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ as in (1) that is implicitly dependent on u_t with the augmented state $z_t \triangleq [x_t^\top \; w_t^\top]^\top \in \mathcal{Z}$.

Suppose (1) is a DT system. Then, a mapping $f_d : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^{n_x}$ is a DT mixed-monotone decomposition function with respect to f, if it satisfies the following conditions:

i) f is embedded on the diagonal of f_d, i.e., $f_d(z,z) = f(z)$,

ii) f_d is monotone increasing in its first argument, i.e., $\hat{z} \geq z \Rightarrow f_d(\hat{z}, z') \geq f_d(z, z')$, and

iii) f_d is monotone decreasing in its second argument, i.e., $\hat{z} \geq z \Rightarrow f_d(z', \hat{z}) \leq f_d(z', z)$.

Further, if (1) is a CT system, a mapping $f_d : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^{n_x}$ is a CT mixed-monotone decomposition function with respect to f, if it satisfies the following conditions:

i) f is embedded on the diagonal of f_d, i.e., $f_d(z,z) = f(z)$,

ii) f_d is monotone increasing in its first argument with respect to "off-diagonal" arguments, i.e., $\forall i \in \mathbb{N}_{n_x}, \hat{z}_j \geq z_j, \forall j \in \mathbb{N}_{n_z}, \hat{z}_i = z_i = x_i \Rightarrow f_{d,i}(\hat{z}, z') \geq f_{d,i}(z, z')$, and

iii) f_d is monotone decreasing in its second argument, i.e., $\hat{z} \geq z \Rightarrow f_d(z', \hat{z}) \leq f_d(z', z)$.

In addition, systems that admit mixed-monotone decomposition functions are called mixed-monotone systems.

Moreover, we extend the concept of decomposition functions to *one-sided decomposition functions* via the following definition.

Definition 2 (One-sided decomposition functions): Consider $f : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ and suppose there exist two mixed-monotone mappings $\overline{f}_d, \underline{f}_d : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^{n_x}$ such that for any $\underline{z}, z, \overline{z} \in \mathcal{Z}$, the following statement holds:

$$\underline{z} \leq z \leq \overline{z} \Rightarrow \underline{f}_d(\underline{z}, \overline{z}) \leq f(z) \leq \overline{f}_d(\overline{z}, \underline{z}), \tag{2}$$

where with a slight abuse of notation, we overload the notation of $\underline{f}_d(\underline{z}, \overline{z})$ and $\overline{f}_d(\underline{z}, \overline{z})$ when (1) is a CT system to represent the case with a fixed z_i for the ith function f_i. Then, $\overline{f}_d$ and $\underline{f}_d$ are called upper and lower decomposition functions for f over $\mathbb{I}\mathcal{Z}$, respectively.

Based on the abovementioned definitions, the mixed-monotone decomposition function f_d is the special case when the upper and lower decomposition functions coincide ($\underline{z} \leq z \leq \overline{z} \Rightarrow f_d(\underline{z}, \overline{z}) \leq f(z) \leq f_d(\overline{z}, \underline{z})$). Moreover, one-sided decomposition functions can be obtained from a family of (generally nonunique) upper and lower decomposition functions via the following result that slightly generalizes [28, Prop. 2] to make it applicable to both DT and CT systems.

Corollary 1 (Intersection property): Suppose $\overline{f}^1_d, \overline{f}^2_d$, and $\underline{f}^1_d, \underline{f}^2_d$ are pairs of upper and lower decomposition functions for f, respectively. Then, $\min\{\overline{f}^1_d, \overline{f}^2_d\}$ and $\max\{\underline{f}^1_d, \underline{f}^2_d\}$ (i.e., their *intersection*) are also upper and lower decomposition function for $f(\cdot)$, respectively.

Proof: The proof is similar to the proof of [28, Prop. 2]. ■

Further, we can slightly generalize the notion of *embedding system with respect to* f_d in [13, (7)], to the *embedding system with respect to* $\underline{f}_d, \overline{f}_d$, through the following definition.

Definition 3 (Generalized embedding systems): For an n-dimensional system (1) with any one-sided decomposition functions $\underline{f}_d$ and $\overline{f}_d$, its *generalized embedding system is the* $2n$-dimensional system with initial condition $\begin{bmatrix} \underline{x}_0^\top & \overline{x}_0^\top \end{bmatrix}^\top$

$$\begin{bmatrix} \underline{x}_t^+ \\ \overline{x}_t^+ \end{bmatrix} = \begin{bmatrix} \underline{f}_d\left(\begin{bmatrix} (\underline{x}_t)^\top \ \underline{w}^\top \end{bmatrix}^\top, \begin{bmatrix} (\overline{x}_t)^\top \ \overline{w}^\top \end{bmatrix}^\top\right) \\ \overline{f}_d\left(\begin{bmatrix} (\overline{x}_t)^\top \ \overline{w}^\top \end{bmatrix}^\top, \begin{bmatrix} (\underline{x}_t)^\top \ \underline{w}^\top \end{bmatrix}^\top\right) \end{bmatrix}. \tag{3}$$

Proposition 1 (State framer property): Suppose system (1) with initial state $x_0 \in \mathbb{I}\mathcal{X}_0 \triangleq [\underline{x}_0, \overline{x}_0]$ has a unique solution and is mixed-monotone with a generalized embedding system (3) with respect to $\underline{f}_d$ and $\overline{f}_d$. Then, for all $t \geq 0$, $R^f(t, \mathbb{I}\mathcal{X}_0) \subset \mathbb{I}\mathcal{X}_t \triangleq [\underline{x}_t, \overline{x}_t]$, where $R^f(t, \mathbb{I}\mathcal{X}_0) \triangleq \{\phi(t, x_0, w_{0:t}) \mid x_0 \in \mathbb{I}\mathcal{X}_0 \text{ and } w_t \in \mathbb{I}\mathcal{W}, \forall t \geq 0\}$ is the reachable set at time t of (1) when initialized within $\mathbb{I}\mathcal{X}_0$ and $\mathbb{I}\mathcal{X}_t \triangleq [\underline{x}_t, \overline{x}_t]$ is the solution to the generalized embedding system (3). Consequently, $\underline{x}_t \leq x_t \leq \overline{x}_t, \ \forall t \geq 0$, i.e., system (1) trajectory is *framed* by $\mathbb{I}\mathcal{X}_t$.

Proof: The proof for CT systems is similar to the proof in [29, Prop. 3], whereas the DT result follows from repeatedly applying its definition in (2). ■

Definition 4 (Inclusion functions): [17, Ch. 2.4] Consider a function $f : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$. The interval function $T^f : \mathbb{I}\mathbb{R}^{n_z} \to \mathbb{I}\mathbb{R}^{n_x}$ is an inclusion function for f, if

$$\forall \mathbb{I}\mathcal{Z} \in \mathbb{I}\mathbb{R}^{n_z}, f(\mathbb{I}\mathcal{Z}) \subset T^f(\mathbb{I}\mathcal{Z}),$$

where $f(\mathbb{I}\mathcal{Z})$ is the true image set (or range) of f for the domain $\mathbb{I}\mathcal{Z} \in \mathbb{I}\mathbb{R}^{n_z}$. The tightest enclosing interval of $f(\mathbb{I}\mathcal{Z})$ is denoted by $T^f_O(\mathbb{I}\mathcal{Z}) \triangleq [\underline{f}^{\text{true}}, \overline{f}^{\text{true}}] \triangleq [\min_{z \in \mathbb{I}\mathcal{Z}} f(z), \max_{z \in \mathbb{I}\mathcal{Z}} f(z)] \supset f(\mathbb{I}\mathcal{Z})$; hence, it is the tightest inclusion function, i.e., $T^f_O(\mathbb{I}\mathcal{Z}) \subseteq T^f(\mathbb{I}\mathcal{Z})$. Further, with a slight abuse of notation, we overload the notation of $f(\mathbb{I}\mathcal{Z})$ and $T^f(\mathbb{I}\mathcal{Z})$ when (1) is a CT system to represent $f_c(\mathbb{I}\mathcal{Z}) \subset \mathbb{R}^{n_x}$ and $T^{f_c}(\mathbb{I}\mathcal{Z}) \in \mathbb{I}\mathbb{R}^{n_x}$, respectively, with $f_{c,i}(\mathbb{I}\mathcal{Z}) \triangleq f_i(\mathbb{I}\mathcal{Z}_{c,i})$, $T^{f_c}_i(\mathbb{I}\mathcal{Z}) \triangleq T^f_i(\mathbb{I}\mathcal{Z}_{c,i})$, and $\mathbb{I}\mathcal{Z}_{c,i} \triangleq \{[\underline{z}, \overline{z}] \in \mathbb{I}\mathcal{Z} \mid [\underline{z}_j, \overline{z}_j] = \mathbb{I}\mathcal{Z}_j, \forall j \neq i, \underline{z}_i = \overline{z}_i = z_i\}, \forall i \in \mathbb{N}_{n_x}$.

Next, inspired by the work in [18, Sec. 3], we introduce the notion of remainder-form (additive) inclusion functions.

Definition 5 (Remainder-form (additive) inclusion functions): Consider a function $f : \mathcal{Z} \in \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$. The interval function $T^f_R : \mathbb{I}\mathbb{R}^{n_z} \to \mathbb{I}\mathbb{R}^{n_x}$ is an additive (remainder-form) inclusion function for f, if there exist two constituent mappings $g, h : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$, such that for any $\mathbb{I}\mathcal{Z} \in \mathbb{I}\mathbb{R}^{n_z}$,

$$f(\mathbb{I}\mathcal{Z}) \subseteq T^f_R(\mathbb{I}\mathcal{Z}) \triangleq g(\mathbb{I}\mathcal{Z}) + h(\mathbb{I}\mathcal{Z}),$$

where addition is based on interval arithmetic (cf. [17]).

The following observation relates the concept of decomposition functions to inclusion functions.

Proposition 2 (Decomposition-based inclusion functions): Given any upper and lower decomposition functions $\overline{f}_d$ and $\underline{f}_d$ (or any decomposition function $f_d = \overline{f}_d = \underline{f}_d$) for f,

$$T^{f_d}(\mathbb{I}\mathcal{Z}) \triangleq [\underline{f}_d(\underline{z}, \overline{z}), \overline{f}_d(\overline{z}, \underline{z})]$$

satisfies $f(\mathbb{I}\mathcal{Z}) \subset T^{f_d}(\mathbb{I}\mathcal{Z})$ with $\mathbb{I}\mathcal{Z} \triangleq [\underline{z}, \overline{z}]$ (including overloading; cf. Definition 4). Consequently, $T^{f_d}(\mathbb{I}\mathcal{Z})$ is an inclusion function (that is based on decomposition functions).

Proof: The results follow directly from Definitions 1–4. ■

As noted earlier, (mixed-monotone) decomposition functions defined in Definition 1 are not unique. Hence, a measure of their tightness is beneficial for comparing these functions.

Definition 6 (Tightness of decomposition functions): [13, Def. 2] A decomposition function f_d^1 for system (1) is tighter than decomposition function f_d^2, if for all $z \leq \hat{z}$,

$$f_d^2(z,\hat{z}) \leq f_d^1(z,\hat{z}) \text{ and } f_d^1(\hat{z},z) \leq f_d^2(\hat{z},z). \tag{4}$$

Then, f_d^O is tight, i.e., it is the tightest possible decomposition function for f, if (4) holds with $f_d^1 = f_d^O$ and any other decomposition function f_d^2. Furthermore, we define the *metric/measure* of tightness as the maximum dimensionwise Hausdorff distance given by

$$q(f(\mathbb{I}\mathcal{Z}), T^{f_d}(\mathbb{I}\mathcal{Z})) \triangleq \max_{i \in \mathbb{N}_{n_x}} \tilde{q}(f_i(\mathbb{I}\mathcal{Z}), T^{f_{d_i}}(\mathbb{I}\mathcal{Z})), \tag{5}$$

where $\tilde{q}(\mathbb{I}\mathcal{X}_1, \mathbb{I}\mathcal{X}_2) \triangleq \max\{|\underline{x}_1 - \underline{x}_2|_\infty, |\overline{x}_1 - \overline{x}_2|_\infty\}$ is the Hausdorff distance between two real intervals $\mathbb{I}\mathcal{X}_1 = [\underline{x}_1, \overline{x}_1]$ and $\mathbb{I}\mathcal{X}_2 = [\underline{x}_2, \overline{x}_2]$, both in $\mathbb{IR}$ [18]. Moreover, the abovementioned tightest decomposition function $T_O^{f_d}(\mathbb{I}\mathcal{Z}) \triangleq [f_d^O(\underline{z}, \overline{z}), f_d^O(\overline{z}, \underline{z})]$ satisfies $q(f(\mathbb{I}\mathcal{Z}), T_O^{f_d}(\mathbb{I}\mathcal{Z})) = 0$.

Further, by Proposition 3, we can obtain inclusion functions from existing decomposition functions for differentiable functions in [14, Th. 2] and [8, Special Case 1], which can also be shown to be equivalent to the one in [16, Prop. 2].

Proposition 3 ($T_L^{f_d}$ inclusion functions): [14, Th. 2], [8, Special Case 1], [16, Prop. 2] For any system in the form of (1), suppose that $f : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ is differentiable and $\frac{\partial f_i}{\partial z_j}(z) \in [a_{ij}, b_{ij}]$, $\forall z \in \mathbb{I}\mathcal{Z} \triangleq [\underline{z}, \overline{z}] \subseteq \mathbb{IR}^{n_z}$. Then, a DT or CT mixed-monotone decomposition function $f_d^L = [f_{d,1}^L \dots f_{d,n_x}^L]$ with respect to f and its corresponding inclusion function $T_L^{f_d}(\mathbb{I}\mathcal{Z}) = [f_d^L(\underline{z}, \overline{z}), f_d^L(\overline{z}, \underline{z})]$ can be constructed as follows. For all $i \in \mathbb{N}_{n_x}$ and $j \in \mathbb{N}_{n_z}$,

$$f_{d,i}^L(z,\hat{z}) = f_i(\zeta) + (\alpha_i - \beta_i)(z - \hat{z}), \tag{6}$$

with $\alpha_i = [\alpha_{i1}, \dots, \alpha_{in_z}]$, $\beta_i = [\beta_{i1}, \dots, \beta_{in_z}]$ and $\zeta = [\zeta_1, \dots, \zeta_{n_z}]^\top$, where $\alpha_{ij} = \begin{cases} 0, & \text{Cases } 1,3,4,5, \\ |a_{ij}|, & \text{Case } 2, \end{cases}$ $\beta_{ij} = \begin{cases} 0, & \text{Cases } 1,2,4,5, \\ -|b_{ij}|, & \text{Case } 3, \end{cases}$ and $\zeta_j = \begin{cases} z_j, & \text{Cases } 1,2,5, \\ \hat{z}_j, & \text{Cases } 3,4, \end{cases}$ with Cases 1 through 4 for DT, and Cases 1 through 5 for CT systems. Moreover, the cases are defined as: Case $1: a_{ij} \geq 0$, Case $2: a_{ij} \leq 0, b_{ji} \geq 0, |a_{ij}| \leq |b_{ij}|$, Case $3: a_{ij} \leq 0, b_{ij} \geq 0, |a_{ij}| \geq |b_{ij}|$, Case $4: b_{ij} \leq 0$, and Case 5: $j = i$.

Proposition 4 (Tight decomposition functions for mixed-monotone systems): [12, Th. 2], [13, Th. 1] For any system in the form of (1) and $\mathbb{I}\mathcal{Z} \triangleq [\underline{z}, \overline{z}]$, a tight (optimal) DT or CT mixed-monotone decomposition function $f_d^O = [f_{d,1}^O \dots f_{d,n_x}^O]$ and its corresponding tight inclusion function $T_O^{f_d}(\mathbb{I}\mathcal{Z}) \triangleq [f_d^O(\underline{z}, \overline{z}), f_d^O(\overline{z}, \underline{z})]$ (i.e., $T_O^f(\mathbb{I}\mathcal{Z}) = T_O^{f_d}(\mathbb{I}\mathcal{Z})$) can be constructed as follows. If (1) is DT, then $\forall i \in \mathbb{N}_{n_x}$,

$$f_{d,i}^O(z,\hat{z}) = \begin{cases} \min_{\zeta \in [z,\hat{z}]} f_i(\zeta) & \text{if } z \leq \hat{z} \\ \max_{\zeta \in [\hat{z},z]} f_i(\zeta) & \text{if } \hat{z} \leq z. \end{cases} \tag{7}$$

Moreover, if (1) is CT, then $\forall i \in \mathbb{N}_{n_x}$,

$$f_{d,i}^O(z,\hat{z}) = \begin{cases} \min_{\zeta \in [z,\hat{z}], \zeta_i = z_i} f_i(\zeta) & \text{if } z \leq \hat{z} \\ \max_{\zeta \in [\hat{z},z], \zeta_i = z_i} f_i(\zeta) & \text{if } \hat{z} \leq z. \end{cases} \tag{8}$$

Remark 1: On the flip side, the DT mixed-monotone decomposition functions can also be directly used as inclusion functions; hence, the proposed decomposition functions can also be relatively easily incorporated into existing analysis, estimation and planning algorithms that are based on *interval arithmetic*, e.g., [17], [18], [19], [20].

Although Proposition 4 provides theoretically tight decomposition functions, it has some limitations in practice (see [29, Sec. III] for a detailed discussion). For instance, exact closed-form solutions to the nonlinear programs in (7) and (8) may not always be available. With this in mind, we define the notion of tractability of decomposition functions as follows.

Definition 7 (Tractable decomposition functions): $\overline{f}_d, \underline{f}_d$ are computationally tractable/computable one-sided decomposition functions for mapping f, if they can be constructed in a closed form, i.e., with a finite number of elementary operations and without differentiation nor an iterative procedure.

Corollary 2 (Tight and tractable decomposition functions for JSS vector fields): Suppose $f(\cdot)$ is continuously differentiable and JSS [14], i.e., $\forall i \in \mathbb{N}_{n_x}$, $\forall j \in \mathbb{N}_{n_z}$, $J_{ij}^f(z) \triangleq \frac{\partial f_i}{\partial z_j}(z) \geq 0$, $\forall z \in \mathbb{I}\mathcal{Z}$, or $J_{ij}^f(z) \triangleq \frac{\partial f_i}{\partial z_j}(z) \leq 0$, $\forall z \in \mathbb{I}\mathcal{Z}$. Then, the following statements hold.

i) $\forall i \in \mathbb{N}_{n_x}$ and $\forall j \in \mathbb{N}_{n_z}$, $f_i(\cdot)$ is either monotonically nondecreasing or monotonically nonincreasing in its jth argument z_j, over the entire domain $\mathbb{I}\mathcal{Z}$.

ii) The optimization programs in (7) and (8) can be tractably and exactly solved by enumerating $f_i(\cdot)$ at the vertices of $\mathbb{I}\mathcal{Z}$ (with fixed $\zeta_i = z_i = x_i$ for CT systems) and choosing the corresponding optima.

Proof: The results are obtained in a straightforward manner by applying Proposition 4 and basic calculus. ■

We conclude this section by briefly introducing *interval arithmetic*-based inclusion functions via Propositions 5 and 6.

Proposition 5 (Natural (T_N^f) inclusion functions): [17, Th. 2.2] Consider $\mathbb{I}\mathcal{Z} \triangleq [\underline{z}, \overline{z}] \in \mathbb{IR}^{n_z}$ and $f \triangleq [f_1, \dots, f_{n_x}]^\top : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$, where each f_j, $j \in \mathbb{N}_{n_x}$, is expressed as a finite composition of the operators $+, -, \times$, and /, and elementary functions (sine, cosine, exponential, square root, etc.). A natural inclusion function $T_N^f : \mathbb{IR}^{n_z} \to \mathbb{IR}^{n_x}$ for f is obtained by replacing each real variable $z_i, i \in \mathbb{N}_{n_z}$, by its corresponding *interval variable* $[z_i] \triangleq \mathbb{I}\mathcal{Z}_i = [\underline{z}_i, \overline{z}_i]$, and each operator or function by its interval counterpart by applying *interval arithmetic* (cf. [17, Ch. 2] for details).

Proposition 6 (Centered (T_C^f) and mixed centered (T_M^f) inclusion functions): [17, Secs. 2.4.3–2.4.4] Let $f \triangleq [f_1, \dots, f_{n_x}]^\top : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ be differentiable over the interval $\mathbb{I}\mathcal{Z} \triangleq [\underline{z}, \overline{z}] \in \mathbb{IR}^{n_z}$. Then, the interval function

$$T_C^f(\mathbb{I}\mathcal{Z}) \triangleq f(m) + \mathbb{I}\mathbf{J}_{\mathbb{I}\mathcal{Z}}^f(\mathbb{I}\mathcal{Z} - m)$$

is an inclusion function for f, called the *centered inclusion function* for f in $\mathbb{I}\mathcal{Z}$, where $m \triangleq \frac{\underline{z}+\overline{z}}{2}$, $\mathbb{I}\mathbf{J}_{\mathbb{I}\mathcal{Z}}^f$ is an *interval Jacobian* *matrix* with domain $\mathbb{I}\mathcal{Z}$ such that $J_f(z) \in \mathbb{I}\mathbf{J}^f_{\mathbb{I}\mathcal{Z}}, \forall z \in \mathbb{I}\mathcal{Z}$, and $J_f(z)$ is the Jacobian matrix of f at point $z \in \mathbb{I}\mathcal{Z}$. Moreover,

$$T^f_M(\mathbb{I}\mathcal{Z}) \triangleq [T^f_{M,i}(\mathbb{I}\mathcal{Z}) \ldots T^f_{M,n_x}(\mathbb{I}\mathcal{Z})]^\top$$

with $T^f_{M,i}(\mathbb{I}\mathcal{Z}) \triangleq f_i(m) + \sum_{j=1}^{n_z} (\mathbb{I}\mathbf{J}^f_{\mathbb{I}\mathcal{Z}_{1\to j}})_{i,j}(\mathbb{I}\mathcal{Z}_j - m_j)$ for all $i \in \mathbb{N}_{n_x}$, is also an inclusion function for f, called the *mixed-centered inclusion function*. Furthermore, $\mathbb{I}\mathcal{Z}_j \triangleq [\underline{z}_j\ \overline{z}_j], j \in \mathbb{N}_{n_z}$, $\mathbb{I}\mathcal{Z}_{1\to j} \triangleq [\mathbb{I}\mathcal{Z}_1 \ldots \mathbb{I}\mathcal{Z}_j\ m_{j+1} \ldots m_{n_z}]^\top$, $(\mathbb{I}\mathbf{J}^f_{\mathbb{I}\mathcal{Z}_{1\to j}})_{i,j}$ is the (i,j)th element of the interval Jacobian matrix $\mathbb{I}\mathbf{J}^f_{\mathbb{I}\mathcal{Z}_{1\to j}}$ with domain $\mathbb{I}\mathcal{Z}_{1\to j}$, and $\underline{z}_j, \overline{z}_j$ and m_j are the j-th elements of the vectors $\underline{z}, \overline{z}$ and m, respectively.

2) Novel Class of Mixed-Monotone Systems: To describe our modeling framework, we formally define a novel class of nonsmooth systems, which we show to include a wide range of nonlinearities and prove in Section III to be mixed-monotone.

Definition 8 (ELLS Systems): System (1) is ELLS at $z \in \mathcal{Z}$, if there exists an open neighborhood $\mathcal{N}_z \subset \mathcal{Z}$ of z, and for all $i \in \mathbb{N}_{n_x}$, there exist vectors $\kappa_i \triangleq [\kappa_{i1}, \ldots, \kappa_{in_z}]^\top \in \mathbb{R}^{n_z}$ with nonzero elements (i.e., $\kappa_{ij} \neq 0, \ \forall j \in \mathbb{N}_{n_z}$) and constants $\rho_i \in \mathbb{R}$, such that $\forall z', z'' \in \mathcal{N}_z$,

$$\langle \kappa_i(f_i(z') - f_i(z'')), z' - z'' \rangle \leq \rho_i \|z' - z''\|_2^2, \quad \forall i \in \mathbb{N}_{n_x}, \quad (9)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product operator. Further, if (9) holds $\forall i \in \mathbb{N}_{n_x}$, we also call f an ELLS function on $\mathcal{Z}$.

Note that (9) holds for a very broad range of nonlinear systems. Particularly, ELLS systems reduce to one-sided locally Lipschitz systems,[1] when $n_z = n_x, \kappa_{ii} > 0, \ \forall i$, and $(f_i(z) - f_i(z')(z_{ij} - z'_{ij}))$ has the same sign as κ_{ij}, uniformly, $\forall i \neq j, z \neq z'$, with Lipschitz constant $\sum_{i=1}^{n_x} \frac{\rho_i}{\kappa_{ii}}$. Moreover, the class of ELLS systems is a strict superset of the class of LLC systems[2] and can even have finite-valued and countable discontinuities (jumps), as shown in the following Proposition 7 and Corollary 3.

Proposition 7: An LLC system is also an ELLS system. Moreover, the converse is not true. Consequently, the class of LLC systems is a strict subset of the class of ELLS systems, i.e., LLC $\subsetneq$ ELLS.

Proof: Suppose a system in the form of (1) is LLC. Then, clearly each f_i is also LLC with some $\tilde{\rho}_i$. Next, by applying the Cauchy–Schwartz inequality and with each f_i being LLC and any κ_i with nonzero elements, we have $\langle \kappa_i(f_i(z') - f_i(z)), z' - z \rangle \leq \|\kappa_i\|_2 |f_i(z') - f_i(z)| \|z' - z\|_2 \leq \rho_i \|z' - z\|_2^2$, where $\rho_i \triangleq \|\kappa_i\|_2 \tilde{\rho}_i$.

To show that the converse is not true, we provide two counterexamples: 1) Consider the *semicontinuous*[3] "sign" function $s(x) = 0$ if $x < 0$ and $s(x) = 1$ if $x \geq 0$, which has a discontinuity at $x = 0$. First, note that s is not LLC at 0, since given any open interval $\mathbb{I}\mathbb{V} = (-r, r)$, $r > 0$, and any $L > 0$, one can pick $m > \max(\frac{1}{Lr}, 2), x_1 = \frac{1}{mL} \in \mathbb{I}\mathbb{V}, x_2 = -\frac{1}{mL} \in \mathbb{I}\mathbb{V}$, and hence, $|s(x_1) - s(x_2)| = 1 > L|x_1 - x_2| = L\frac{2}{mL} = \frac{2}{m}$. On the other hand, we can show that s is ELLS at any x with $\rho = 0$ and any $\kappa < 0$, as follows. Let z', z'' be arbitrarily picked from an open neighborhood of x, and consider the following three possible cases:

i) $z', z'' < 0$ or $z', z'' > 0$. In this case, the left-hand side of (9) is zero and the right-hand side is nonnegative, and hence, the inequality in (9) holds.

ii) $z' < 0$ and $z'' > 0$. In this case, $s(z') - s(z'') = -1$, $z' - z'' \leq 0$ and $\kappa < 0$. So, the left-hand side of (9) is nonpositive, and the right-hand side is zero, and thus, the inequality in (9) holds.

iii) $z' > 0$ and $z'' < 0$. In this case, $s(z') - s(z'') = 1$, $z' - z'' \geq 0$, and $\kappa < 0$. So, the left-hand side of (9) is nonpositive, and the right-hand side is zero; thus, (9) holds.

2) Similarly, it can be shown that the *continuous* function $q(x) = \sqrt{x}$ if $x \geq 0$ and $q(x) = 0$ if $x < 0$ is not LLC at $x = 0$, since $\lim_{x\to 0^+} \frac{dq}{dx}(x) = \infty$, but is ELLS with $\rho = 0$ and any $\kappa < 0$. ■

Corollary 3: Upper or lower semicontinuous[3] functions that are LLC almost everywhere, except at a nonempty, countable set of finite-valued discontinuities are ELLS but not LLC.

Proof: Since any finite-valued discontinuity (jump) can be characterized using an appropriately scaled sign function, any semicontinuous function that is LLC except at the discontinuities can be rewritten as a sum of an LLC function and scaled sign functions. Then, by the proof of Proposition 7 that sign functions are ELLS, combined with the fact that LLC functions are ELLS functions (with Clarke Jacobians that are bounded above and below; cf. Corollary 4), by Proposition 7, any semicontinuous function that is LLC almost everywhere except at the discontinuities is ELLS, but is clearly not LLC. ■

Next, we review a notion of generalized gradients used in *nonsmooth* analysis in systems and control theory [30] when vector fields of the systems are not necessarily differentiable.

Definition 9 (Clarke generalized directional derivatives): [31, Ch. II] Given a function $f : \mathcal{Z} \subseteq \mathbb{R}^{n_z} \to \mathbb{R}$,

$$f_C^{\uparrow}(z, v) \triangleq \limsup_{t\to z, \lambda\downarrow 0} \frac{f(t + \lambda v) - f(t)}{\lambda} = \sup_{\xi \in \partial_C f(z)} \xi^\top v,$$

$$f_C^{\downarrow}(z, v) \triangleq \liminf_{t\to z, \lambda\downarrow 0} \frac{f(t + \lambda v) - f(t)}{\lambda} = \inf_{\xi \in \partial_C f(z)} \xi^\top v \quad (10)$$

are the (generalized) Clarke upper and lower directional derivatives/gradients of f at $z \in \mathcal{Z}$ in the direction $v \in \mathbb{R}^{n_z}$, respectively, where the set

$$\partial_C f(z) \triangleq \{\xi \in \mathbb{R}^{n_z} | f_C^{\downarrow}(z, v) \leq \xi^\top v \leq f_C^{\uparrow}(z, v) \forall v \in \mathbb{R}^{n_z}\}$$

is the Clarke subdifferential (set) of f at $z \in \mathcal{Z}$.

Note that, by definition, $f_C^{\uparrow}(z, v)$ and $f_C^{\downarrow}(z, v)$ are the *upper* and *lower support functions* of the set $\partial_C f(z)$. Further, as shown in [31, Appendix I], $\partial_C f(z)$ for an LLC system is nonempty, convex, and compact for all $z \in \mathcal{Z}$, and consequently, at each $z \in \mathcal{Z}$, the Clarke directional derivatives are bounded in each direction v. However, this does not hold in general for ELLS systems. Nonetheless, by the following proposition, the Clarke

[1]System (1) with $n_x = n_z$ is one-sided LLC if for all $z \in \mathcal{Z}$, there exist an open neighborhood $\mathcal{N}_z \subset \mathcal{Z}$ of z and a $\rho \in \mathbb{R}$ such that $(f(z') - f(z''))^\top(z' - z'') \leq \rho\|z' - z''\|_2^2, \ \forall z', z'' \in \mathcal{N}_z$.

[2]System (1) is LLC if for all $z \in \mathcal{Z}$, there exist an open neighborhood $\mathcal{N}_z$ of z and a $\rho \geq 0$ such that $\|f(z') - f(z'')\|_2 \leq \rho\|z' - z''\|_2, \ \forall z', z'' \in \mathcal{N}_z$.

[3]System (1) is upper or lower semicontinuous if for all $z \in \mathcal{Z}$, $\limsup_{z'\to z} f(z') \leq f(z)$ or $\liminf_{z'\to z} f(z') \geq f(z)$, respectively.

directional derivatives in some specific directions are bounded from above *or* from below.

Proposition 8: Suppose $f : \mathcal{Z} \subseteq \mathbb{R}^{n_z} \to \mathbb{R}$ is ELLS on $\mathcal{Z}$, with $\kappa = [\kappa_1, \ldots, \kappa_{n_z}]^\top \in \mathbb{R}^{n_z}$ and $\rho \in \mathbb{R}$. Let e_i, $\forall i \in \mathbb{N}_{n_z}$ denote the standard unit vector/basis in the ith coordinate direction. Then, $f_C^{\uparrow}(z, e_i)$ is bounded from above, or $f_C^{\downarrow}(z, e_i)$ is bounded from below. In particular, for $i \in \mathbb{N}_{n_z}$, if $\kappa_i > 0$, then $f_C^{\uparrow}(z, e_i) \leq \frac{\rho}{\kappa_i}$, and if $\kappa_i < 0$, then $f_C^{\downarrow}(z, e_i) \geq \frac{\rho}{\kappa_i}$.

Proof: Setting $z' = z + \lambda e_i$ in (9), where $\lambda > 0$ is sufficiently small, we obtain $\kappa_i(f(z + \lambda e_i) - f(z))\lambda \leq \rho\lambda^2$. Then, dividing both sides by $\kappa_i \neq 0$ and taking the $\limsup$ if $\kappa_i > 0$, or the $\liminf$ if $\kappa_i < 0$, from both sides when $\lambda \to 0^+$, imply that $f_C^{\uparrow}(z, e_i)$ is bounded from above or $f_C^{\downarrow}(z, e_i)$ is bounded from below, by $\frac{\rho}{\kappa_i}$, respectively [cf. (10)]. ∎

Corollary 4: Consider the mapping $f = [f_1, \ldots, f_{n_x}]^\top : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$, where $\forall i \in \mathbb{N}_{n_x}$, f_i is ELLS on $\mathcal{Z}$, with $\kappa_i = [\kappa_{i1}, \ldots, \kappa_{in_z}]^\top \in \mathbb{R}^{n_z}$, and $\rho_i \in \mathbb{R}$. We consider upper and lower Clarke Jacobian matrices of f at $z \in \mathcal{Z}$, $\overline{J}_C^f(z) \triangleq [(\overline{J}_C^f(z))_{ij}]$, and $\underline{J}_C^f(z) \triangleq [(\underline{J}_C^f(z))_{ij}]$, with the upper and lower partial Clarke derivatives at point $z \in \mathcal{Z}$ for each $i \in \mathbb{N}_{n_x}$ and $j \in \mathbb{N}_{n_z}$, respectively, defined as

$$(\overline{J}_C^f(z))_{ij} \triangleq f_{i,C}^{\uparrow}(z, e_j),\ (\underline{J}_C^f(z))_{ij} \triangleq f_{i,C}^{\downarrow}(z, e_j). \quad (11)$$

Then, $\forall (i,j) \in \mathbb{N}_{n_x} \times \mathbb{N}_{n_z}$, either $\overline{J}_C^f(z)_{ij} \leq \frac{\rho_i}{\kappa_{ij}}$ (bounded above) or $\underline{J}_C^f(z)_{ij} \geq \frac{\rho_i}{\kappa_{ij}}$ (bounded below), i.e., we cannot simultaneously have $(\overline{J}_C^f(z))_{ij} = \infty$ and $(\underline{J}_C^f(z))_{ij} = -\infty$.

Proof: The results follow from applying Proposition 8 in a dimensionwise manner. ∎

Next, we provide a slight modification of the results in [31, Prop. 1.12], which plays an important role in our main results later and its proof goes precisely along the lines of the proof of [31, Prop. 1.12].

Proposition 9: Let $f : \mathcal{Z} \subseteq \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ be decomposable into $f = f^1 + f^2$, where $f^1, f^2 : \mathcal{Z} \subseteq \mathbb{R}^{n_z} \to \mathbb{R}$, and f_i, f_i^1, f_i^2, $\forall i \in \mathbb{N}_{n_x}$, are ELLS. Then, $\forall z \in \mathcal{Z}$, $\forall i \in \mathbb{N}_{n_x}$, and $\forall j \in \mathbb{N}_{n_z}$,

$$(\overline{J}_C^f(z))_{ij} \leq (\overline{J}_C^{f^1}(z))_{ij} + (\overline{J}_C^{f^2}(z))_{ij}$$
$$(\underline{J}_C^f(z))_{i,j} \geq (\underline{J}_C^{f^1}(z))_{ij} + (\underline{J}_C^{f^2}(z))_{ij}.$$

Proof: The results follow from (11) and the facts that $f_C^{\uparrow}(z, v) \leq f_C^{1\uparrow}(z, v) + f_C^{2\uparrow}(z, v)$ [31, Ch. II, Prop. 1.12] and $f_C^{\downarrow}(z, v) = -f_C^{\uparrow}(z, -v)$ [31, Ch. II, Prop. 1.7], $\forall z \in \mathcal{Z}$ and $\forall v \in \mathbb{R}^{n_z}$. ∎

It is worth noticing that when f is differentiable, then $\nabla f(x) \in \partial_C f(x)$, and if f is continuously differentiable or strictly differentiable, then $\partial_C f(x) = \{\nabla f(x)\}$.

Now, we are ready to explain the notion of *Clarke Jacobian sign-stability* through the following definition, which is a generalization of Jacobian sign-stability in [14] (cf. Corollary 2).

Definition 10 (Clarke Jacobian sign-stability): A mapping $f : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ is called Clarke Jacobian sign-stable (CJSS) over $\mathcal{Z}$, if $\forall z \in \mathcal{Z}$, $\forall i \in \mathbb{N}_{n_x}$, and $\forall j \in \mathbb{N}_{n_z}$,

$$(\overline{J}_C^f(z))_{ij} \leq 0 \ \vee \text{(or)}\ (\underline{J}_C^f(z))_{ij} \geq 0. \quad (12)$$

Finally, we present an extension of Corollary 2, which we will apply later in our derivations.

Proposition 10: Suppose f is ELLS and CJSS over $\mathbb{I}\mathcal{Z}$. Then, $\forall i \in \mathbb{N}_{n_x}$ and $\forall j \in \mathbb{N}_{n_z}$, f_i is either monotonically nondecreasing or monotonically nonincreasing in its jth argument z_j, over the entire domain $\mathbb{I}\mathcal{Z}$, and consequently, the optima of (7) and (8) are attained at some vertices of $\mathbb{I}\mathcal{Z}$.

Proof: For $i \in \mathbb{N}_{n_x}$, consider any arbitrary z^1 from the interior of $\mathbb{I}\mathcal{Z}$ and construct $z^2 = z^1 + \lambda e_j \in \mathbb{I}\mathcal{Z}$, for some small enough $\lambda > 0$. Then, by an identical proof to [31, Ch. II, Th. 1.3], there exists a $z_\theta \in \mathbb{I}\mathcal{Z}$ on the connecting line between z^1 and z^2, such that $\lambda f_{i,C}^{\downarrow}(z_\theta, e_j) = f_{i,C}^{\downarrow}(z_\theta, \lambda e_j) \leq f_i(z^2) - f_i(z^1) \leq f_{i,C}^{\uparrow}(z_\theta, \lambda e_j) = \lambda f_{i,C}^{\uparrow}(z_\theta, e_j)$, where the equalities hold by [31, Ch. II, Prop. 1.5]. Using this and the CJSS assumption [cf. (12)], $f_i(z^2) \leq f_i(z_2)$ if $f_{i,C}^{\uparrow}(z, \lambda e_j) = (\overline{J}_C^f(z))_{ij} \leq 0$ holds for all $z \in \mathbb{I}\mathcal{Z}$ (including $z = z_\theta$), and similarly, $f_i(z^2) \geq f_i(z_2)$ if $f_{i,C}^{\downarrow}(z, \lambda e_j) = (\underline{J}_C^f(z))_{ij} \geq 0$ holds for all $z \in \mathbb{I}\mathcal{Z}$ (including $z = z_\theta$). Hence, by moving along the coordinate directions, one can always increase or decrease each of the f_i's. Using a similar argument, this result also holds for $\lambda < 0$. So, the optimum for each f_i is attained at some vertices of the interval $\mathbb{I}\mathcal{Z}$. ∎

C. Modeling Framework And Problem Statement

We consider constrained dynamical systems of the form

$$x_t^+ = \tilde{f}(x_t, u_t, w_t) \triangleq f(z_t),\quad \mu(x_t, u_t) \triangleq \nu(x_t) \in \mathcal{Y}_t, \quad (13)$$

where $x_t^+ = \dot{x}_t$ if (13) is a CT system and $x_t^+ = x_{t+1}$ if (13) is a DT system. $x_t \in \mathcal{X} \subset \mathbb{R}^{n_x}$ is the state, $u_t \in \mathcal{U} \subset \mathbb{R}^{n_x}$ is the known input, and $w_t \in \mathcal{W} \subseteq \mathbb{I}\mathcal{W} \triangleq [\underline{w}, \overline{w}] \in \mathbb{I}\mathbb{R}^{n_w}$ is the augmentation of all exogenous inputs, e.g., bounded disturbance/noise and internal uncertainties, such as uncertain parameters, with known bounds $\underline{w}, \overline{w}$, while $f : \mathcal{X} \times \mathcal{U} \times \mathcal{W} \to \mathbb{R}^{n_x}$ and $\mu : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^{n_y}$ are the nonlinear vector field and the observation/constraint mapping, respectively, while $\mathcal{Y}_t$ is the known or measured time-varying constraint/observation set. The mapping μ and the set $\mathcal{Y}_t \subseteq [\underline{y}_t, \overline{y}_t]$ describe system constraints that can represent prior or additional knowledge about the system states, e.g., sensor observations or measurements with bounded noise, known state constraints, manufactured constraints from modeling redundancy [25] (cf. Section VI-B for an example), etc. For ease of exposition, we further define $f : \mathcal{Z} \triangleq \mathcal{X} \times \mathcal{W} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ and $\nu : \mathcal{X} \subset \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ as in (13) that is implicitly dependent on u_t with the augmented state $z_t \triangleq [x_t^\top\ w_t^\top]^\top \in \mathcal{Z}$. Further, we assume the following.

Assumption 1: The mappings f and ν are ELLS (cf. Definition 8) and have a countable number of (finite-valued) discontinuities.

Note that it is straightforward to show that Assumption 1 implies that the vector field f is locally essentially bounded (LEB), i.e., it is bounded on a bounded neighborhood of every point, excluding a set of measure zero. This implies that there exists a solution for (13) by [32, Prop. 3]. Moreover, if the ELLS condition holds, $\kappa_{ii} > 0$, $\forall i$, and $(f_i(z) - f_i(z')(z_{ij} - z'_{ij}))$ has the same sign as κ_{ij}, uniformly, $\forall i \neq j, z \neq z'$, then the vector field is one-sided locally Lipschitz, and hence, (13) exhibits a unique solution by [32, Corollary 1]. So Proposition 1, i.e., the state framer property, will be applicable with any decomposition functions, including the remainder-form decomposition functions that will be introduced in Section III-A. Further, in the even more general case that some of the κ_i's are negative where the uniqueness of solutions is not guaranteed, we can still apply a weaker version of Proposition 1 that will be shown later in Lemma 4 to hold for remainder-form decomposition functions.

Assumption 2: For the mappings f and ν, there exist known bounds on their Clarke Jacobian matrices, $\overline{J}^f_C, \underline{J}^f_C \in \{\mathbb{R} \cup \pm\infty\}^{n_x \times n_z}$ and $\overline{J}^\nu_C, \underline{J}^\nu_C \in \{\mathbb{R} \cup \pm\infty\}^{n_y \times n_x}$, that satisfy

$$((J^f_C(z))_{ij} \leq (\overline{J}^f_C)_{ij} < \infty) \vee ((J^f_C(z))_{ij} \geq (\underline{J}^f_C)_{ij} > -\infty)$$

$$((J^\nu_C(x))_{pq} \leq (\overline{J}^\nu_C)_{pq} < \infty) \vee ((J^\nu_C(x))_{pq} \geq (\underline{J}^\nu_C)_{pq} > -\infty)$$

$\forall z \in \mathbb{I}\mathcal{Z}, \forall x \in \mathbb{I}\mathcal{X}, \forall i, q \in \mathbb{N}_{n_x}, \forall j \in \mathbb{N}_{n_z}$, and $\forall p \in \mathbb{N}_{n_y}$.

Under the abovementioned modeling framework and assumptions, this article seeks to find tight and tractable (i.e., computable) remainder-form upper and lower decomposition functions and their induced inclusion functions (cf. Definitions 2, 4, and 5 and Proposition 2) as well as to develop set inversion algorithms based on (mixed-monotone) decomposition functions.

Problem 1 (Decomposition functions): Suppose Assumptions 1 and 2 hold. Construct and quantify the tightness [via the metric (5)] of remainder-form decomposition functions, by solving the following subproblems.

1.1 Given an ELLS vector field $f : \mathcal{Z} \subset \mathcal{R}^{n_z} \to \mathbb{R}^{n_x}$, construct a tractable family of mixed-monotone remainder-form (i.e., additive) decomposition functions for f.

1.2 Derive lower and upper bounds for the *tightness* (quantified via (5)) of the family of remainder-form decomposition functions obtained in 1.1.

1.3 Find the tightest decomposition function(s) among the family of remainder-form decomposition functions obtained in 1.1 and compare them with the decomposition function in [14] (cf. Proposition 3), and natural, centered and mixed-centered natural inclusions (cf. Proposition 6).

Problem 2 (Set inversion algorithm): Suppose Assumptions 1 and 2 hold. Given a prior/propagated interval $\mathbb{I}\mathcal{X}^p_t \in \mathbb{I}\mathbb{R}^{n_x}$, a constraint/observation function and set, $\mu(x_t, u_t)$ and $\mathcal{Y}_t \subseteq [\underline{y}_t, \overline{y}_t]$ with known u_t, develop an algorithm to find an interval superset of all x_t that are compatible with μ, $[\underline{y}_t, \overline{y}_t]$, and $\mathbb{I}\mathcal{X}^p_t$, i.e., to find the updated/refined interval $\mathbb{I}\mathcal{X}^u_t$ such that

$$\{x \in \mathbb{I}\mathcal{X}^p_t \mid \mu(x, u_t) \in [\underline{y}_t, \overline{y}_t]\} \subseteq \mathbb{I}\mathcal{X}^u_t \subseteq \mathbb{I}\mathcal{X}^p_t. \quad (14)$$

In the context of constrained reachability analysis and guaranteed state estimation (cf. Section V), the solution of the generalized embedding system (cf. Definition 3) based on the decomposition functions obtained from solving Problem 1 provides the unconstrained reachable set (or propagated set), $\mathbb{I}\mathcal{X}^p_t$, whereas the set inversion algorithm in Problem 2 finds the constrained reachable set (or updated set) $\mathbb{I}\mathcal{X}^u_t$.

III. Main Results

To address the aforementioned problems, we first describe our proposed construction approach to find a tractable family of mixed-monotone remainder-form decomposition functions. Then, by characterizing their tightness, we can determine the tightest decomposition function among the proposed family. Further, we present a novel set inversion algorithm that serves as an alternative and improves on existing approaches, e.g., SIVIA in [17, Ch. 3]) and $\mathcal{I}_G$ in [21, Algorithm 1]).

A. Remainder-Form Decomposition Functions

To solve Problem 1.1, we provide a constructive procedure for computing the family of remainder-form decomposition functions in a tractable manner (i.e., in closed-form). Intuitively, our approach is based on the idea of decomposing each ELLS function f into the remainder/additive form, i.e., $f = g + h$, such that g is a CJSS function (cf. Definition 10; so that Proposition 10 applies) by "shifting" the Clarke directional gradients of f and accounting for the "error" using h. Since there are several "shift" directions, we obtain a family of decomposition functions. Note that the construction below is to be independently performed for each dimension of the ELLS function f; hence, without loss of generality, we only consider a scalar ELLS function f_i throughout this subsection.

Theorem 1 (Family of remainder-form decomposition functions): Consider an ELLS vector field $f_i : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}$ and suppose that Assumptions 1 and 2 hold. Then, f_i admits a family of mixed-monotone remainder-form decomposition functions denoted as $\{f_{d,i}(z, \hat{z}; \mathbf{m}, h(\cdot))\}_{\mathbf{m} \in \mathbf{M}_i, h(\cdot) \in \mathcal{H}_{\mathbf{M}_i}}$, that is parameterized by a supporting vector $\mathbf{m} \in \mathbf{M}_i$ and an ELLS remainder function $h \in \mathcal{H}_{\mathbf{M}_i}$, where

$$f_{d,i}(z, \hat{z}; \mathbf{m}, h) = h(\zeta_{\mathbf{m}}(\hat{z}, z)) + f_i(\zeta_{\mathbf{m}}(z, \hat{z})) - h(\zeta_{\mathbf{m}}(z, \hat{z})) \quad (15)$$

with $\zeta_{\mathbf{m}}(z, \hat{z}) = [\zeta_{\mathbf{m},1}(z, \hat{z}), \ldots, \zeta_{\mathbf{m},n_z}(z, \hat{z})]^\top, \forall j \in \mathbb{N}_{n_z}$,

$$\zeta_{\mathbf{m},j}(z, \hat{z}) = \begin{cases} \hat{z}_j, & \text{if } \mathbf{m}_j \geq \max((\overline{J}^f_C)_{ij}, 0) \\ z_j, & \text{if } \mathbf{m}_j \leq \min((\underline{J}^f_C)_{ij}, 0) \end{cases} \quad (16)$$

and the set of supporting vectors $\mathbf{M}_i$ is defined as

$$\mathbf{M}_i \triangleq \{\mathbf{m} \in \mathbb{R}^{n_z} \mid \mathbf{m}_j \geq \max((\overline{J}^f_C)_{ij}, 0) \vee \mathbf{m}_j \leq \min((\underline{J}^f_C)_{ij}, 0) \quad \forall j \in \mathbb{N}_{n_z}\} \quad (17)$$

if (13) is a DT system, and with

$$\zeta_{\mathbf{m},j}(z, \hat{z}) = \begin{cases} \hat{z}_j, & \text{if } \mathbf{m}_j \geq \max((\overline{J}^f_C)_{ij}, 0) \vee j = i, \\ z_j, & \text{if } \mathbf{m}_j \leq \min((\underline{J}^f_C)_{ij}, 0) \wedge j \neq i, \end{cases} \quad (18)$$

$$\mathbf{M}_i \triangleq \{\mathbf{m} \in \mathbb{R}^{n_z} \mid \mathbf{m}_j \geq \max((\overline{J}^f_C)_{ij}, 0) \vee \mathbf{m}_j \leq \min((\underline{J}^f_C)_{ij}, 0), \quad \forall j \in \mathbb{N}_{n_z}, j \neq i, \mathbf{m}_i = 0\}, \quad (19)$$

if (13) is a CT system, whereas the set of remainder functions $\mathcal{H}_{\mathbf{M}_i}$ is the family of all ELLS remainder functions whose Clarke subdifferential set over $\mathcal{Z}$ (cf. Definition 9) is a subset of $\mathbf{M}_i$ and is given by

$$\mathcal{H}_{\mathbf{M}_i} \triangleq \{h : \mathcal{Z} \to \mathbb{R} \mid [\underline{J}^h_C(z), \overline{J}^h_C(z)] \subseteq \mathbf{M}_i \quad \forall z \in \mathcal{Z}\}. \quad (20)$$

Consequently, the resulting family of decomposition-based inclusion functions is given by

$$T^{f_{d,i}}_{\underline{\mathbf{m}},\overline{\mathbf{m}}}(\mathbb{I}\mathcal{Z}) \triangleq [f_{d,i}(\underline{z},\overline{z};\underline{\mathbf{m}},\underline{h}), f_{d,i}(\overline{z},\underline{z};\overline{\mathbf{m}},\overline{h})] \tag{21}$$

for all $\underline{\mathbf{m}},\overline{\mathbf{m}} \in \mathbf{M}_i$, and the corresponding $\underline{h},\overline{h} \in \mathcal{H}_{\mathbf{M}_i}$.

Note that the small difference in the definitions of the set of supporting vectors, $\mathbf{M}$, for DT and CT systems in Theorem 1 originates from the subtle difference between the definitions of decomposition functions for DT and CT systems (cf. Definition 1). To prove the abovementioned theorem, we first prove the following lemma.

Lemma 1: Any remainder function $h \in \mathcal{H}_{\mathbf{M}_i}$ [cf. (20)] is CJSS and the function $g_i \triangleq f_i - h$ is also CJSS. Moreover, the pair $(g_i, -h)$ is aligned, i.e., $\forall j \in \mathbb{N}_{n_z}$:

1) $(\overline{J}^{g_i}_C(z))_j \leq 0$ if and only if $(\underline{J}^{h}_C(z))_j \geq 0$, or
2) $(\underline{J}^{g_i}_C(z))_j \geq 0$ if and only if $(\overline{J}^{h}_C(z))_j \leq 0$.

Proof: Since $\mathbf{m} \in \mathbf{M}_i$, by construction of $\mathbf{M}_i$ [cf. (17), (19)], $h \in \mathcal{H}_{\mathbf{M}_i}$ is a CJSS function. Next, by applying Proposition 9 to $g_i \triangleq f_i - h$, we know that $\forall z \in \mathcal{Z}$, $\forall j \in \mathbb{N}_{n_z}$,

$$(\underline{J}^{g_i}_C(z))_j \leq (\underline{J}^{f}_C(z))_{ij} - (\underline{J}^{h}_C(z))_j \tag{22}$$

$$(\overline{J}^{g_i}_C(z))_j \geq (\overline{J}^{f}_C(z))_{ij} - (\overline{J}^{h}_C(z))_j. \tag{23}$$

Then, since $h \in \mathcal{H}_{\mathbf{M}_i}$, according to (16)–(20), we consider the following two cases (the case with $\mathbf{m}_i = 0$ for CT systems is trivial):

1) $\forall z \in \mathcal{Z}, (\underline{J}^{h}_C(z))_j \geq \max((\overline{J}^{f}_C)_{ij}, 0) \geq 0$. Then, from (22), we obtain $\forall z \in \mathcal{Z}, (\underline{J}^{g_i}_C(z))_j \leq 0 \Rightarrow (\underline{J}^{g_i}_C)_j \leq 0$.
2) $\forall z \in \mathcal{Z}, (\overline{J}^{h}_C(z))_j \leq \min((\underline{J}^{f}_C)_{ij}, 0) \leq 0$. Then, from (23), we have $\forall z \in \mathcal{Z}, (\overline{J}^{g_i}_C(z))_j \geq 0 \Rightarrow (\overline{J}^{g_i}_C)_j \geq 0$.

The reverse can be similarly deduced. Finally, since $(\overline{J}^{g_i}_C)_j \geq 0$ or $(\underline{J}^{g_i}_C)_j \leq 0$ holds, g_i is CJSS by Definition 10. ■

Remark 2: Since the pair $(g_i, -h)$ is aligned and $f_i = g_i + h$, the proposed remainder-form decomposition function can also be viewed as the decomposition of f_i into a *difference of monotone functions*, which is similar in spirit to difference of convex functions in DC programming, e.g., [23] and [26].

Proof of Theorem 1: Armed by Lemma 1, we now prove that (15) is mixed-monotone (cf. Definition 1). Having defined $g_i \triangleq f_i - h$, (15) can be rewritten as

$$f_{d,i}(z,\hat{z};\mathbf{m},h) = h(\zeta_{\mathbf{m}}(\hat{z},z)) + g_i(\zeta_{\mathbf{m}}(z,\hat{z})). \tag{24}$$

First, it directly follows from (16) that $\zeta_{\mathbf{m}}(z,z) = z$ and $f_{d,i}(z,z;\mathbf{m},h) = f_i(z)$. Hence, it remains to show that $f_{d,i}(z,\hat{z};\mathbf{m},h)$ is nondecreasing in z and nonincreasing in $\hat{z}$. To do this, consider $z, \tilde{z}, \hat{z} \in \mathcal{Z}$, where $\tilde{z} \geq z$. Let $j_0 \in \mathbb{N}_{n_z}$, and suppose case (i) in the proof of Lemma 1 holds for dimension j_0. Then, by the first case in (16),

$$\zeta_{\mathbf{m},j_0}(\hat{z},\tilde{z}) = \tilde{z}_{j_0} \geq z_{j_0} = \zeta_{\mathbf{m},j_0}(\hat{z},z)$$

$$\zeta_{\mathbf{m},j_0}(\tilde{z},\hat{z}) = \hat{z}_{j_0} = \zeta_{\mathbf{m},j_0}(z,\hat{z}). \tag{25}$$

Next, we define $z^1 \in \mathcal{Z}$ as follows: $z^1_{j_0} = \tilde{z}_{j_0}$ and $z^1_j = z_j, \forall j \neq j_0$. Thus, $z^1 \geq z$, and by (25), $\zeta_{\mathbf{m}}(z^1,\hat{z}) = \zeta_{\mathbf{m}}(z,\hat{z})$, $\zeta_{\mathbf{m},j}(\hat{z},z^1) = \zeta_{\mathbf{m},j}(\hat{z},z)$, $\forall j \neq j_0$, and $\zeta_{\mathbf{m},j_0}(\hat{z},z^1) \geq \zeta_{\mathbf{m},j_0}(\hat{z},z)$. Moreover, by case (i) in the proof of Lemma 1 and Proposition 10, h is nondecreasing in the dimension j_0, and thus, $h(\zeta_{\mathbf{m}}(\hat{z},z^1)) \geq h(\zeta_{\mathbf{m}}(\hat{z},z))$ and $g_i(\zeta_{\mathbf{m}}(z^1,\hat{z})) = g_i(\zeta_{\mathbf{m}}(z,\hat{z}))$. Then, it follows from (24) that $f_{d,i}(z^1,\hat{z};\mathbf{m},h) \geq f_{d,i}(z,\hat{z};\mathbf{m},h)$. Repeating this procedure sequentially for all dimensions j for which case (i) in Lemma 1 holds (where τ is the size of this set), we obtain

$$f_{d,i}(z^\tau,\hat{z};\mathbf{m},h) \geq f_{d,i}(z^{\tau-1},\hat{z};\mathbf{m},h) \geq \cdots$$
$$\geq f_{d,i}(z^1,\hat{z};\mathbf{m},h) \geq f_{d,i}(z,\hat{z};\mathbf{m},h). \tag{26}$$

Next, we consider the rest of the dimensions j' that satisfy case (ii) in Lemma 1. It follows from the second case in (16) that for such a dimension $j'_0 \in \mathbb{N}_{n_z}$,

$$\zeta_{\mathbf{m},j'_0}(\hat{z},\tilde{z}) = \hat{z}_{j'_0} = \zeta_{\mathbf{m},j'_0}(\hat{z},z^\tau)$$

$$\zeta_{\mathbf{m},j'_0}(\tilde{z},\hat{z}) = \tilde{z}_{j'_0} \geq z^\tau_{j'_0} = \zeta_{\mathbf{m},j'_0}(z^\tau,\hat{z}). \tag{27}$$

Repeating a similar procedure as for case (i), we define $z^{\tau+1} \in \mathcal{Z}$ as follows: $z^{\tau+1}_{j'_0} = \tilde{z}_{j'_0}$ and $z^{\tau+1}_j = z^\tau_j$, $\forall j \neq j'_0$. Thus, $z^{\tau+1} \geq z^\tau$, and by (25), $\zeta_{\mathbf{m}}(\hat{z},z^{\tau+1}) = \zeta_{\mathbf{m}}(\hat{z},z^\tau)$, $\zeta_{\mathbf{m},j}(z^{\tau+1},\hat{z}) = \zeta_{\mathbf{m},j}(z^\tau,\hat{z})$, $\forall j \neq j'_0$, and $\zeta_{\mathbf{m},j'_0}(z^{\tau+1},\hat{z}) \geq \zeta_{\mathbf{m},j'_0}(z^\tau,\hat{z})$. Moreover, by case (ii) in the proof of Lemma 1 and Proposition 10, g_i is nondecreasing in the dimension j'_0, and thus, $g_i(\zeta_{\mathbf{m}}(z^{\tau+1},\hat{z})) \geq g_i(\zeta_{\mathbf{m}}(z^\tau,\hat{z}))$, and $h(\zeta_{\mathbf{m}}(\hat{z},z^{\tau+1})) = h(\zeta_{\mathbf{m}}(\hat{z},z^\tau))$. Then, it follows from (24) that $f_{d,i}(z^{\tau+1},\hat{z};\mathbf{m},h) \geq f_{d,i}(z^\tau,\hat{z};\mathbf{m},h)$.

Repeating this procedure sequentially for all dimensions j for which case (ii) in Lemma 1 holds, we obtain

$$f_{d,i}(z^\tau,\hat{z};\mathbf{m},h) \leq f_{d,i}(z^{\tau+1},\hat{z};\mathbf{m},h) \leq \ldots \leq f_{d,i}(\tilde{z},\hat{z};\mathbf{m},h),$$

where the last term is $f_{d,i}(\tilde{z},\hat{z};\mathbf{m},h)$ since there exist only two possible cases (i) or (ii) for each dimension. Combining this and (26) yields $f_{d,i}(\tilde{z},\hat{z};\mathbf{m},h) \geq f_{d,i}(z,\hat{z};\mathbf{m},h)$, which means that $f_{d,i}$ is nondecreasing in its first argument. An almost identical argument shows that $f_{d,i}$ is nonincreasing in its second argument. Thus, $f_{d,i}$ is mixed-monotone. ■

Theorem 1 mathematically introduces a family of decomposition functions (cf. Definition 1), but the results are not yet tractable (cf. Definition 7), since to build such a family, we have to search over $\mathbf{M}_i$ [cf. (17) and (19)], which is an unbounded and infinite set, as well as over $\mathcal{H}_{\mathbf{M}_i}$ [cf. (20)], which is an infinite-dimensional space of functions. To overcome this problem, we propose tractable upper and lower decomposition functions that only require a search over a *finite* set of supporting vectors $\mathbf{M}^c_i \subset \mathbf{M}_i$ with the choice of linear remainder functions $h(\zeta) = \langle \mathbf{m}, \zeta \rangle = \mathbf{m}^\top \zeta$, and prove that these tractable decomposition functions are the tightest among the family of decomposition functions in Theorem 1.

Theorem 2 (Tight and tractable remainder-form upper and lower decomposition functions): Consider an ELLS vector field $f_i : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}$ and let Assumptions 1 and 2 hold. Then, the tightest tractable (mixed-monotone) remainder-form upper and lower decomposition functions with $z \geq \hat{z}$ are

$$\overline{f}_{d,i}(z,\hat{z}) = \min_{\mathbf{m}\in\mathbf{M}^c} f_i(\zeta^+_{\mathbf{m}}) + \mathbf{m}^\top(\zeta^-_{\mathbf{m}} - \zeta^+_{\mathbf{m}}),$$

$$\underline{f}_{d,i}(\hat{z},z) = \max_{\mathbf{m}\in\mathbf{M}^c} f_i(\zeta^-_{\mathbf{m}}) + \mathbf{m}^\top(\zeta^+_{\mathbf{m}} - \zeta^-_{\mathbf{m}}), \tag{28}$$

where $\zeta^+_{\mathbf{m}} \triangleq \zeta_{\mathbf{m}}(z,\hat{z})$ and $\zeta^-_{\mathbf{m}} \triangleq \zeta_{\mathbf{m}}(\hat{z},z)$ with $\zeta_{\mathbf{m}}(\cdot,\cdot)$ defined in (16) and the *finite* set of supporting vectors $\mathbf{M}^c_i$ defined as

$$\mathbf{M}^c_i \triangleq \{\mathbf{m} \in \mathbb{R}^{n_z} \mid \mathbf{m}_j = \max((\overline{J}^f_C)_{ij},0) \vee \mathbf{m}_j = \min((\underline{J}^f_C)_{ij},0) \quad \forall j \in \mathbb{N}_{n_z}\}, \quad (29)$$

if (13) is a DT system, and with

$$\mathbf{M}^c_i \triangleq \{\mathbf{m} \in \mathbb{R}^{n_z} \mid \mathbf{m}_j = \max((\overline{J}^f_C)_{ij},0) \vee \mathbf{m}_j = \min((\underline{J}^f_C)_{ij},0) \quad \forall j \in \mathbb{N}_{n_z}, j \neq i, \mathbf{m}_i = 0\} \quad (30)$$

and $\zeta_{\mathbf{m}}(\cdot,\cdot)$ defined in (18) if (13) is a CT system.

Moreover, we call the resulting inclusion function for an interval domain $\mathbb{I}\mathcal{Z} = [\underline{z},\overline{z}]$ the remainder-form inclusion function $T^{f_{d,i}}_R \triangleq [\underline{f}_{d,i}(\underline{z},\overline{z}), \overline{f}_{d,i}(\overline{z},\underline{z})]$ (cf. Definition 5).

Remark 3: Note that for the special case of LLC functions, Theorems 1 and 2 hold trivially by replacing generalized Clarke Jacobians and their bounds with regular Jacobians. Thus, all results in this article for constructing tight and tractable remainder-form decomposition functions also contribute to the literature on LLC functions, where such decompositions have not been considered before.

We will prove the above theorem in two steps with the help of the following lemmas, where the two steps show that we can restrict our search for the tightest upper and lower decomposition functions to a *finite* set of supporting vectors and the set of *linear* remainder functions, respectively, without introducing any conservatism.

Lemma 2 (Finite set of supporting vectors): Suppose the assumptions in Theorem 2 hold. Then, $\forall z,\hat{z} \in \mathcal{Z}$, $\forall h \in \mathcal{H}_{\mathbf{M}_i}$ and for both opt $\in \{\min,\max\}$,

$$\operatorname*{opt}_{\mathbf{m}\in\mathbf{M}_i, h\in\mathcal{H}_{\mathbf{M}_i}} f_{d,i}(z,\hat{z};\mathbf{m},h) = \operatorname*{opt}_{\mathbf{m}\in\mathbf{M}^c_i, h\in\mathcal{H}_{\mathbf{M}_i}} f_{d,i}(z,\hat{z};\mathbf{m},h),$$

where $f_{d,i}(z,\hat{z};\mathbf{m},h)$ is defined in (15), $\mathbf{M}_i$ in (17) or (19), $\mathbf{M}^c_i$ in (29) or (30), and $\mathcal{H}_{\mathbf{M}_i}$ in (20).

Proof: We consider $\mathbf{m} \in \mathbf{M}_i$ and construct $\tilde{\mathbf{m}} \in \mathbf{M}^c_i$ as follows: for all $j \in \mathbb{N}_{n_z}$: $\tilde{\mathbf{m}}_j = \max((\overline{J}^f_C)_{ij},0)$ if $\mathbf{m}_j \geq \max((\overline{J}^f_C)_{ij},0)$ and $\tilde{\mathbf{m}}_j = \min((\underline{J}^f_C)_{ij},0)$ if $\mathbf{m}_j \leq \min((\underline{J}^f_C)_{ij},0)$. Then, it can be easily verified from (16) that $\forall z^1, z^2 \in \mathcal{Z}, \zeta_{\mathbf{m}}(z^1,z^2) = \zeta_{\tilde{\mathbf{m}}}(z^1,z^2)$, and hence, by (15), $f_{d,i}(z^1,z^2;\mathbf{m},h) = f_{d,i}(z^1,z^2;\tilde{\mathbf{m}},h)$, $\forall h \in \mathcal{H}_{\mathbf{M}_i}$. Hence, for any $\mathbf{m} \in \mathbf{M}_i$, there exists $\tilde{\mathbf{m}} \in \mathbf{M}^c_i$ that admits an equivalent decomposition function and correspondingly, the optimization over $\mathbf{M}_i$ and $\mathbf{M}^c_i$ are equivalent. ■

Lemma 3 (Linear remainder functions): Suppose the assumptions in Theorem 2 hold. Then, $\forall z,\hat{z} \in \mathcal{Z}, z \geq \hat{z}$,

$$\begin{aligned}\min_{\mathbf{m}\in\mathbf{M}^c_i, h\in\mathcal{H}_{\mathbf{M}_i}} f_{d,i}(z,\hat{z};\mathbf{m},h) &= \min_{\mathbf{m}\in\mathbf{M}^c} f_i(\zeta^+_{\mathbf{m}}) + \mathbf{m}^\top(\zeta^-_{\mathbf{m}} - \zeta^+_{\mathbf{m}}),\\ \max_{\mathbf{m}\in\mathbf{M}^c_i, h\in\mathcal{H}_{\mathbf{M}_i}} f_{d,i}(\hat{z},z;\mathbf{m},h) &= \max_{\mathbf{m}\in\mathbf{M}^c} f_i(\zeta^-_{\mathbf{m}}) + \mathbf{m}^\top(\zeta^+_{\mathbf{m}} - \zeta^-_{\mathbf{m}}),\end{aligned} \quad (31)$$

where $\zeta^+_{\mathbf{m}} \triangleq \zeta_{\mathbf{m}}(z,\hat{z})$ and $\zeta^-_{\mathbf{m}} \triangleq \zeta_{\mathbf{m}}(\hat{z},z)$ with $f_{d,i}(z,\hat{z};\mathbf{m},h)$ being defined in (15), $\zeta_{\mathbf{m}}(\cdot,\cdot)$ in (16), $\mathbf{M}^c_i$ in (29) or (30), and $\mathcal{H}_{\mathbf{M}_i}$ in (20).

Proof: Consider any $h \in \mathcal{H}_{\mathbf{M}_i}$, $\mathbf{m} \in \mathbf{M}_i$ and from (15),

$$f^+_{d,i} \triangleq f_{d,i}(z,\hat{z};\mathbf{m},h) = f_i(\zeta^+_{\mathbf{m}}) + \Delta h_{\mathbf{m}}, \quad (32)$$

$$f^-_{d,i} \triangleq f_{d,i}(\hat{z},z;\mathbf{m},h) = f_i(\zeta^-_{\mathbf{m}}) - \Delta h_{\mathbf{m}}, \quad (33)$$

where $\Delta h_{\mathbf{m}} \triangleq h(\zeta^-_{\mathbf{m}}) - h(\zeta^+_{\mathbf{m}})$. Then, by applying the Clarke mean value theorem [31, Ch. II, Th. 1.3] to $\Delta h_{\mathbf{m}}$, there exists $\xi \in [\underline{J}^h_C, \overline{J}^h_C] \subset \mathbf{M}_i$ such that $\Delta h_{\mathbf{m}} = \langle \xi, (\zeta^+_{\mathbf{m}} - \zeta^-_{\mathbf{m}})\rangle$. Since $\xi \in \mathbf{M}_i$, by (17) and (19), we know that $\xi_j \leq \min((\underline{J}^f_C)_{ij},0)$ or $\xi_j \geq \max((\overline{J}^f_C)_{ij},0)$, $\forall j \in \mathbb{N}_{n_z}$ (excluding $j = i$ for CT systems where $\xi_j = 0$).

Then, similar to the proof of Lemma 1, $\forall j \in \mathbb{N}_{n_x}$, we can consider two cases corresponding to the two cases in (16).

i) $\xi_j \geq \max((\overline{J}^f_C)_{ij},0) \geq 0$: From (16), $\zeta^-_{\mathbf{m},j} = \zeta_{\mathbf{m},j}(\hat{z},z) = z_j$, $\zeta^+_{\mathbf{m},j} = \zeta_{\mathbf{m},j}(z,\hat{z}) = \hat{z}_j$, and $z_j \geq \hat{z}_j$; thus, we have $\zeta^-_{\mathbf{m},j} - \zeta^+_{\mathbf{m},j} \geq 0$ and $\xi_j(\zeta^-_{\mathbf{m},j} - \zeta^+_{\mathbf{m},j}) \geq \max((\overline{J}^f_C)_{ij},0)(\zeta^-_{\mathbf{m},j} - \zeta^+_{\mathbf{m},j})$. Then, the minimum of $f^+_{d,i}$ in (32) and the maximum of $f^-_{d,i}$ in (33) are attained in (31) when $\xi_j = \max((\overline{J}^f_C)_{ij},0) \in (\mathbf{M}^c_i)_j$.

ii) $\xi_j \leq \min((\underline{J}^f_C)_{ij},0) \leq 0$: From (16), $\zeta^-_{\mathbf{m},j} = \zeta_{\mathbf{m},j}(\hat{z},z) = \hat{z}_j$, $\zeta^+_{\mathbf{m},j} = \zeta_{\mathbf{m},j}(z,\hat{z}) = z_j$, and $z_j \geq \hat{z}_j$; thus, we have $\zeta^-_{\mathbf{m},j} - \zeta^+_{\mathbf{m},j} \leq 0$ and $\xi_j(\zeta^-_{\mathbf{m},j} - \zeta^+_{\mathbf{m},j}) \geq \min((\underline{J}^f_C)_{ij},0)(\zeta^-_{\mathbf{m},j} - \zeta^+_{\mathbf{m},j})$. Then, the minimum of $f^+_{d,i}$ in (32) and the maximum of $f^-_{d,i}$ in (33) are attained in (31) when $\xi_j = \min((\underline{J}^f_C)_{ij},0) \in (\mathbf{M}^c_i)_j$.

Finally, we can restrict our search to the class of linear remainder functions $h(\zeta) = \langle \mathbf{m},\zeta\rangle = \mathbf{m}^\top\zeta$ with $\mathbf{m} \in \mathbf{M}^c_i$, since it can achieve the optima in (31). ■

Proof of Theorem 2: First, by repeatedly applying Corollary 1 on all the decomposition functions in the family (15) and the fact that the upper and lower decomposition functions can be optimized independently, it can be seen that the tightest upper and lower decomposition functions with $z,\hat{z} \in \mathcal{Z}, z \geq \hat{z}$ are

$$\overline{f}_{d,i}(z,\hat{z}) = \min_{\mathbf{m}\in\mathbf{M}_i, h(\cdot)\in\mathcal{H}_{\mathbf{M}_i}} f_{d,i}(z,\hat{z};\mathbf{m},h),$$
$$\underline{f}_{d,i}(\hat{z},z) = \max_{\mathbf{m}\in\mathbf{M}_i, h(\cdot)\in\mathcal{H}_{\mathbf{M}_i}} f_{d,i}(\hat{z},z;\mathbf{m},h).$$

Then, by Lemmas 2 and 3, we obtain the tractable and tight upper and lower decomposition functions in (28). ■

Theorem 2 guarantees that in order to obtain the tightest possible decomposition function in the form of (15), it is sufficient to only search over a *finite* set of supporting vectors $\mathbf{M}^c_i$ and the class of *linear* remainder functions with (Clarke) gradients from $\mathbf{M}^c_i$, i.e., $h(\zeta) = \langle \mathbf{m},\zeta\rangle = \mathbf{m}^\top\zeta$, $\forall \mathbf{m} \in \mathbf{M}^c_i$, where the search space is the finite and countable set $\mathbf{M}^c_i$. Hence, the optimal search for the tightest decomposition functions is computable/tractable according to Definition 7.

Moreover, the result in Theorem 2 can be applied to each f_i, $i \in \mathbb{N}_{n_x}$, of the function f to obtain the tightest remainder-form decomposition functions from the family of *remainder-form CJSS decomposition functions* in (15). This is summarized in Algorithm 1, which takes an interval domain $\mathbb{I}\mathcal{Z} = [\underline{z},\overline{z}]$,

Algorithm 1: Remainder-Form Decomposition Functions.

```
function T_R^{f_d}(f, J_C^f (lower), J_C^f (upper), z (lower), z (upper))
  Initialize: f_d (lower) ← −∞, f_d (upper) ← ∞;
  for i = 1 to n_x do
    Compute M_i^c via (29) or (30);
    for m ∈ M_i^c do
      for j = 1 to n_z do
        if m_j = min((J_C^f (lower))_ij, 0) then
          ζ^+_{m,j} ← z_j (upper); ζ^-_{m,j} ← z_j (lower);
        else ζ^+_{m,j} ← z_j (lower); ζ^-_{m,j} ← z_j (upper);
        end if
        ζ^{c+}_{m,j} ← ζ^+_{m,j}; ζ^{c-}_{m,j} ← ζ^-_{m,j};
      end for
      if (13) is a continuous-time system then
        ζ^{c+}_{m,i} ← z_i (lower); ζ^{c-}_{m,i} ← z_i (upper); m_i ← 0;
      end if
      f_{d,i} (lower) ← min(f_{d,i} (lower), f_i(ζ^-_m) + m^T(ζ^+_m − ζ^-_m));
      f_{d,i} (upper) ← min(f_{d,i} (upper), f_i(ζ^+_m) + m^T(ζ^-_m − ζ^+_m));
    end for
  end for
  return f_d (lower), f_d (upper);
end function
```

the function f and its Clarke Jacobians, $\overline{J}^f_C$ and $\underline{J}^f_C$, as inputs, and outputs the remainder-form inclusion function $T^{f_d}_R \triangleq [\underline{f}_d(\underline{z},\overline{z}), \overline{f}_d(\overline{z},\underline{z})]$ (cf. Definition 5). It is worth mentioning that the computation of the tightest reminder-form mixed-monotone decomposition function via Algorithm 1 requires n_z^2 function evaluations, which may not scale well to high-dimensional functions. To reduce computational burden, one may replace the set $\mathbf{M}^c_i$ in (28) with a potentially fixed-size subset $\tilde{\mathbf{M}}^c_i \subset \mathbf{M}^c_i$ that can be selected randomly or by empirically including "good" cases, e.g., the case corresponding to [14, Th. 2], [8, Special Case 1] (cf. Theorem 5) without losing the inclusion/framer property (cf. Definition 4).

Further, note that a DT embedding system (3) has unique solutions, whereas a CT embedding system (3) that is constructed based on the decomposition functions (28) exhibits the existence of solutions by [32, Prop. 3] and the fact that the vector fields in (28) are LEB since they are pointwise summations of an LEB vector field f and a linear remainder function h. So, the CT remainder-form embedding system has a *nonempty* solution set. Although this set (and the solution set of (1)) are not guaranteed to be singletons for the most general CT ELLS systems, we can still show through the following lemma that thanks to the remainder-form decomposition functions, a weaker version of the framer property in Proposition 1 holds.

Lemma 4: Suppose the assumptions in Theorem 2 hold. Then, for any reachable set at time $t \geq 0$ of the CT system (1) initialized at $x_0 \in \mathbb{I}\mathcal{X}_0 \triangleq [\underline{x}_0, \overline{x}_0]$, $R^f(t, \mathbb{I}\mathcal{X}_0)$, there exists $\mathbb{I}\mathcal{X}_t \triangleq [\underline{x}_t, \overline{x}_t]$, where $[\underline{x}_t^\top \; \overline{x}_t^\top]^\top$ is in the solution set of the generalized embedding system (3) constructed based on the remainder-form decomposition functions (28) and $R^f(t, \mathbb{I}\mathcal{X}_0) \subset \mathbb{I}\mathcal{X}_t$.

Proof: With a slight abuse of notation, let $x_t \in \Phi_x(x_0, t)$ denote a known (given) solution to (1), picked from its solution set $\Phi_x(x_0, t)$. Our strategy is to construct a $[\underline{x}_t^\top \; \overline{x}_t^\top]^\top$ that is guaranteed to be in the solution set of (3) and satisfies $\underline{x}_t \leq x_t \leq \overline{x}_t$, $\forall t \geq 0$. To do so, given x_t, consider the ordinary differential equation

$$[\dot{\underline{e}}_t^\top \; \dot{\overline{e}}_t]^\top = \delta(e_t) \tag{34}$$

with the initial values $e_0 \triangleq [\underline{e}_0^\top \; \overline{e}_0^\top]^\top \triangleq [(x_0 - \underline{x}_0)^\top \; (\overline{x}_0 - x_0)^\top]^\top \geq 0_{2n}$, where $e_t \triangleq [\underline{e}_t^\top \; \overline{e}_t^\top]^\top \in E \subset \mathbb{R}^{2n}$, $\delta_i(e_t) \triangleq f_i(x_t) - \underline{f}_{d,i}(x_t - \underline{e}_t, x_t + \overline{e}_t)$ for $i = 1, \ldots, n$, and $\delta_i(e_t) \triangleq \overline{f}_{d,i}(x_t + \overline{e}_t, x_t - \underline{e}_t) - f_i(x_t)$ for $i = n+1, \ldots, 2n$ denote the augmented state and dimensionwise vector fields of (34), respectively. Note that $\delta \triangleq [\delta_1^\top \ldots \delta_{2n}^\top]^\top$ is an ELLS mapping, since by construction, $\underline{f}_{d,i}, \overline{f}_{d,i}$ are pointwise summations of ELLS and linear functions. Hence, by (9), for all $i \in \mathbb{N}_{2n}$ and all $e', e'' \in E$, there exist $\kappa_i \in \mathbb{R}^{2n}, \kappa_{ij} \neq 0, \rho_i \in \mathbb{R}$, such that $\langle \kappa_i(\delta_i(e') - \delta_i(e'')), e' - e'' \rangle \leq \rho_i \|e' - e''\|_2^2$. Expanding this, in addition to setting $e'' = 0_{2n}$ and given that $\delta_i(0_{2n}) = f_i(x_t) - \underline{f}_{d,i}(x_t, x_t) = \overline{f}_{d,i}(x_t, x_t) - f_i(x_t) = f_i(x_t) - f_i(x_t) = 0$, we obtain $\delta_i(e')e'_i + (\kappa_{ii} - 1)\delta_i(e')e'_i + \sum_{j=1, j\neq i}^{2n} \kappa_{ij}\delta_j(e')e'_j \leq \rho_i \|e'\|_2^2$ for all $e' \in E$ and all $i \in \mathbb{N}_{2n}$. Consequently, $\delta_i(e')e'_i \leq \rho'_i \|e'\|_2^2 + c_i$, where $c_i \triangleq \sup_{\tilde{e} \in E}((1 - \kappa_{ii})\delta_i(\tilde{e})\tilde{e}_i - \sum_{j=1, j\neq i}^{2n} \kappa_{ij}\delta_j(\tilde{e})\tilde{e})$. Summing up both sides of the $2n$ inequalities returns $\langle \delta(e'), e' \rangle \leq c_1 \|e'\|_2^2 + c_2$ for all $e' \in E$, where $c_1 \triangleq \sum_{i=1}^{2n} \rho_i$, $c_2 \triangleq \max(\epsilon, \sum_{i=1}^{2n} c_1) > 0$, and ϵ is a very small positive number. With this, the dynamical system (34) satisfies the conditions in [33, Eq. (4)]. Moreover, since decomposition functions are nondecreasing in their first argument and nonincreasing in their second, for any ordered pair of $e^1 \triangleq [\underline{e}^{1\top} \; \overline{e}^{1\top}]^\top \leq e^2 \triangleq [\underline{e}^{2\top} \; \overline{e}^{2\top}]^\top, e^1_i = e^2_i, i \in \mathbb{N}_{2n}$, we have $\underline{f}_{d,i}(x - \underline{e}^1, x + \overline{e}^1) \leq \underline{f}_{d,i}(x - \underline{e}^2, x + \overline{e}^2) \Rightarrow \delta_i(e^1) = f(x) - \underline{f}_{d,i}(x - \underline{e}^1, x + \overline{e}^1) \leq \delta_i(e^2) = f(x) - \underline{f}_{d,i}(x - \underline{e}^2, x + \overline{e}^2), i = 1, \ldots, n$, and also $\overline{f}_{d,i}(x + \overline{e}^1, x - \underline{e}^1) \leq \overline{f}_{d,i}(x + \overline{e}^2, x - \underline{e}^2) \Rightarrow \delta_i(e^1) = f_i(x) + \overline{f}_{d,i}(x + \overline{e}^1, x - \underline{e}^1) \leq \delta_i(e^2) = f_i(x) + \overline{f}_{d,i}(x + \overline{e}^2, \; x - \underline{e}^2), \; i = n+1, \ldots, 2n$. In other words, the conditions in [33, Eq. (5)] are also satisfied. Further, by plugging "zero" initial values into (34), we obtain $[\dot{\underline{e}}_t^\top \; \dot{\overline{e}}_t^\top]^\top = 0_{2n}$, which means that the only possible solution is zero, i.e., $\Phi_e(0, t) = \{0\}$, where $\Phi_e(., t)$ denotes the solution set of (34). All of these in addition to [33, Lemma 2] imply that there exists $e^*_t \triangleq [\underline{e}^{*\top}_t \; \overline{e}^{*\top}_t]^\top \in \Phi_e(e_0, t)$ such that $0_n \leq \underline{e}^*_t, 0_n \leq \overline{e}^*_t$, for all $t \geq 0$. Consequently, $\underline{x}_t \triangleq x_t - \underline{e}^*_t \leq x_t \leq \overline{x}_t \triangleq x_t + \overline{e}^*_t$. By taking derivatives of all sides and (34), we obtain $\dot{\underline{x}}_t = \dot{x}_t - \dot{\underline{e}}^*_t = f(x_t) - f(x_t) + \underline{f}_d(x_t - \underline{e}^*_t, x_t + \overline{e}^*_t) = \underline{f}_d(\underline{x}_t, \overline{x}_t)$ and $\dot{\overline{x}}_t = \dot{x}_t + \dot{\overline{e}}^*_t = f(x_t) + \overline{f}_d(x_t + \overline{e}^*_t, x_t - \underline{e}^*_t) - f(x_t) = \overline{f}_d(\overline{x}_t, \underline{x}_t)$. Hence, $[\underline{x}_t^\top \; \overline{x}_t^\top]^\top$ belongs to the solution set of (3). ■

B. Error Bounds

Next, we formally characterize the tightness of our proposed family of remainder-form decomposition functions in (15), where we use the Hausdorff distance-based tightness metric/measure in (5). In particular, we derive lower and upper bounds on the *overapproximation error* of the image set/range of function f, where the lower bound is achievable by our tight and tractable decomposition functions in Theorem 2.

Theorem 3 (Error bounds): Suppose that all the assumptions in Theorem 1 are satisfied for each $f_i : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}, i \in \mathbb{N}_{n_x}$. Let $T^f_O(\mathbb{I}\mathcal{Z}) \triangleq [\underline{f}^{\text{true}}, \overline{f}^{\text{true}}] \triangleq [\min_{z \in \mathbb{I}\mathcal{Z}} f(z),$

$\max_{z\in\mathbb{I}\mathcal{Z}} f(z)] = T_O^{f_d}(\mathbb{I}\mathcal{Z})$ be the tightest enclosing interval of the true image set/range of f over $\mathbb{I}\mathcal{Z} = [\underline{z}, \overline{z}] \in \mathbb{IR}^{n_z}$ (cf. Definition 4), $T^{f_{d,i}}_{\underline{\mathbf{m}},\overline{\mathbf{m}}}(\mathbb{I}\mathcal{Z}) \supseteq T_O^{f_i}(\mathbb{I}\mathcal{Z})$, $\forall i \in \mathbb{N}_{n_x}$, given in (21), be inclusion functions using the family of decomposition functions in (15), and $T^{f_d}_{\mathbb{M}}(\mathbb{I}\mathcal{Z}) \supseteq T_O^f(\mathbb{I}\mathcal{Z})$ be any inclusion function such that $T^{f_{d,i}}_{\mathbb{M}}(\mathbb{I}\mathcal{Z}) = T^{f_{d,i}}_{\underline{\mathbf{m}},\overline{\mathbf{m}}}(\mathbb{I}\mathcal{Z})$, $\forall i \in \mathbb{N}_{n_x}$, for some $\underline{\mathbf{m}}, \overline{\mathbf{m}} \in \mathbf{M}_i$. Then, the following inequalities hold:

$$\underline{q}_{f_d}(\mathbb{I}\mathcal{Z}) \le q(T^{f_d}_{\mathbb{M}}(\mathbb{I}\mathcal{Z}), T_O^f(\mathbb{I}\mathcal{Z})) \le \overline{q}_{f_d}(\mathbb{I}\mathcal{Z}) \le \hat{\overline{q}}_{f_d}(\mathbb{I}\mathcal{Z}), \tag{35}$$

with the tightness metric $q([\underline{v}, \overline{v}], [\underline{w}, \overline{w}]) = \max_{i\in\mathbb{N}_{n_x}} \max\{|\overline{v}_i - \overline{w}_i|, |\underline{v}_i - \underline{w}_i|\}$ defined in (5), and with the bounds $\underline{q}_{f_d}(\mathbb{I}\mathcal{Z}) \triangleq \max_{i\in\mathbb{N}_{n_x}} \underline{q}_{f_{d,i}}(\mathbb{I}\mathcal{Z})$, $\overline{q}_{f_d}(\mathbb{I}\mathcal{Z}) \triangleq \max_{i\in\mathbb{N}_{n_x}} \overline{q}_{f_{d,i}}(\mathbb{I}\mathcal{Z})$ and $\hat{\overline{q}}_{f_d}(\mathbb{I}\mathcal{Z}) \triangleq \max_{i\in\mathbb{N}_{n_x}} \hat{\overline{q}}_{f_{d,i}}(\mathbb{I}\mathcal{Z})$, where for each $i \in \mathbb{N}_{n_x}$,

$$\begin{aligned}
\underline{q}_{f_{d,i}}(\mathbb{I}\mathcal{Z}) &\triangleq \max\left\{\min_{\mathbf{m}\in\mathbf{M}_i^c} \Delta^1_{i,\mathbf{m}} - \overline{f}_i^{\text{true}}, \underline{f}_i^{\text{true}} - \max_{\mathbf{m}\in\mathbf{M}_i^c} \Delta^2_{i,\mathbf{m}}\right\}\\
\hat{\overline{q}}_{f_{d,i}}(\mathbb{I}\mathcal{Z}) &\triangleq \min_{\mathbf{m}\in\mathbf{M}_i^c} \Delta^3_{i,\mathbf{m}}\\
\overline{q}_{f_{d,i}}(\mathbb{I}\mathcal{Z}) &\triangleq \min\left\{\min_{\mathbf{m}\in\mathbf{M}_i^c} \Delta^3_{i,\mathbf{m}}, \min_{\mathbf{m}\in\mathbf{M}_i^c} \Delta^3_{i,\mathbf{m}} + \Delta^4_{i,\mathbf{m}}\right\}
\end{aligned}$$

with $\mathbf{M}_i$ and $\mathbf{M}_i^c$ in (17) and (29) or (19) and (30), respectively,

$$\begin{aligned}
&\Delta^1_{i,\mathbf{m}} \triangleq f_i(\zeta^+_{i,\mathbf{m}}) + \Delta^3_{i,\mathbf{m}}, \Delta^2_{i,\mathbf{m}} \triangleq f_i(\zeta^-_{i,\mathbf{m}}) - \Delta^3_{i,\mathbf{m}}\\
&\Delta^3_{i,\mathbf{m}} \triangleq h_i(\zeta^-_{i,\mathbf{m}}) - h_i(\zeta^+_{i,\mathbf{m}}), \Delta^4_{i,\mathbf{m}} \triangleq f_i(\zeta^+_{i,\mathbf{m}}) - f_i(\zeta^-_{i,\mathbf{m}})
\end{aligned} \tag{36}$$

as well as $\zeta^+_{i,\mathbf{m}} \triangleq \zeta_{i,\mathbf{m}}(\overline{z}, \underline{z}), \zeta^-_{i,\mathbf{m}} \triangleq \zeta_{i,\mathbf{m}}(\underline{z}, \overline{z})$ and $\zeta_{i,\mathbf{m}}(.,.)$ defined in Theorem 1. Further, without loss of tightness (cf. Lemma 3), we can replace $\Delta^3_{i,\mathbf{m}}$ in (36) with $\Delta^3_{i,\mathbf{m}} \triangleq \mathbf{m}^\top(\zeta^-_{i,\mathbf{m}} - \zeta^+_{i,\mathbf{m}})$, whereas $\underline{q}_{f_d}(\mathbb{I}\mathcal{Z})$ is attained by the upper and lower decomposition functions (i.e., $T_R^{f_d}$) in Theorem 2.

Proof: First, from (15), (21), and (36), we find

$$\begin{aligned}
&\tilde{q}(T^{f_{d,i}}_{\underline{\mathbf{m}},\overline{\mathbf{m}}}(\mathbb{I}\mathcal{Z}), T_O^{f_i}(\mathbb{I}\mathcal{Z})) = \max\left\{\Delta^1_{i,\mathbf{m}} - \overline{f}_i^{\text{true}}, \underline{f}_i^{\text{true}} - \Delta^2_{i,\mathbf{m}}\right\}\\
&\ge \max\left\{\min_{\mathbf{m}\in\mathbf{M}_i^c} \Delta^1_{i,\mathbf{m}} - \overline{f}_i^{\text{true}}, \underline{f}_i^{\text{true}} - \max_{\mathbf{m}\in\mathbf{M}_i^c} \Delta^2_{i,\mathbf{m}}\right\}
\end{aligned} \tag{37}$$

where the first inequality in (35) follows from independently searching over $\mathbf{M}_i^c \subset \mathbf{M}_i$ to minimize each argument of the maximization, as well as the fact that by Lemmas 2 and 3, we can apply $\Delta^3_{i,\mathbf{m}} \triangleq \mathbf{m}^\top(\zeta^-_{i,\mathbf{m}} - \zeta^+_{i,\mathbf{m}})$ and search only over $\mathbf{M}_i^c \subset \mathbf{M}_i$ without any conservatism, and it can be verified that by construction, the lower bound $\underline{q}_f(\mathbb{I}\mathcal{Z})$ is attained by the decomposition functions in Theorem 2. To obtain the second and third inequalities, we apply [18, Th. 4-(b)], which proved that for any remainder-form inclusion functions with remainder function $r_i(\cdot)$ satisfying $r_i(\mathbb{I}\mathcal{Z}) \subset [\underline{r}_i, \overline{r}_i]$, $\tilde{q}(W^R_{f_{d,i}}(\mathbb{I}\mathcal{Z}), V_{f_i}(\mathbb{I}\mathcal{Z})) \le \overline{r}_i - \underline{r}_i$ holds. For the third inequality in (35), only h_i is considered as the remainder function, whereas in the second inequality, both h_i and $g_i \triangleq f_i - h_i$ are considered as remainder functions, separately, with the minimum chosen as the bound. Moreover, since h_i is CJSS, by (16), $\overline{h}_i$ and $\underline{h}_i$ are attained at the corner points given by $\zeta^-_{i,\mathbf{m}}$ and $\zeta^+_{i,\mathbf{m}}$, respectively, with $\Delta^3_{i,\mathbf{m}} \triangleq \overline{h}_i - \underline{h}_i$. Further, since g_i is aligned with $-h_i$ by Lemma 1, $\overline{g}_i$ and $\underline{g}_i$ are attained at the corner points given by $\zeta^+_{i,\mathbf{m}}$ and $\zeta^-_{i,\mathbf{m}}$, respectively, with $\Delta^3_{i,\mathbf{m}} + \Delta^4_{i,\mathbf{m}} \triangleq \overline{g}_i - \underline{g}_i$. ■

The abovementioned result holds for both DT and CT systems (with overloading described in Definition 4). Further, note that lower bound $\underline{q}_{f,d}(\mathbb{I}\mathcal{Z})$ is attainable by $T_R^{f_d}$ but since it is a function of the unknown $\overline{f}^{\text{true}}$ and $\underline{f}^{\text{true}}$, it cannot be computed. Thus, its upper bounds $\overline{q}_{f_d}(\mathbb{I}\mathcal{Z})$ and $\hat{\overline{q}}_{f_d}(\mathbb{I}\mathcal{Z})$ in (ii) that are independent of $\overline{f}^{\text{true}}$ and $\underline{f}^{\text{true}}$ are more useful, e.g., as worst-case function overapproximation error bounds in reachability and robust control problems.

C. Convergence Rate and Subdivision Principle

In this section, we study the convergence rate of our proposed $T_R^{f_d}$, i.e., the rate at which its approximation error goes to zero, when the domain interval diameter $d(\mathbb{I}\mathcal{Z})$ shrinks. We show that when using $T_R^{f_d}$, the error converges at least linearly, which is also the convergence rate of natural inclusions T_N^f [19, Ch. 6]. Further, we show that the *subdivision principle* introduced in [20] can be applied to improve the convergence. We first introduce the notion of convergence rate, inspired by [20], and then, we present the convergence rate and the subdivision principle for our proposed $T_R^{f_d}$.

Definition 11 (Convergence rate): An inclusion function $T^f : \mathbb{IR}^{n_z} \to \mathbb{IR}^{n_x}$ for an ELLS vector field $f : \mathcal{Z} \subset \mathbb{R}_{n_z} \to \mathbb{R}^{n_x}$ has a convergence rate $\alpha > 0$, if

$$q(T^f(\mathbb{I}\mathcal{Z}), T_O^f(\mathbb{I}\mathcal{Z})) \le \beta\, d(\mathbb{I}\mathcal{Z})^\alpha \tag{38}$$

for some $\beta > 0$, where $T^f(\mathbb{I}\mathcal{Z})$ is the interval overapproximation of the range of f over $\mathbb{I}\mathcal{Z}$ (i.e., an inclusion function), $T_O^f(\mathbb{I}\mathcal{Z})$ is the tightest inclusion function (cf. Definition 4), $q(\cdot,\cdot)$ is defined in (5) and $d(\mathbb{I}\mathcal{Z}) \triangleq \|\overline{z} - \underline{z}\|_\infty$.

Theorem 4 (Convergence rate and subdivision principle for $T_R^{f_d}$): The $T_R^{f_d}$ inclusion function for any ELLS vector field $f : \mathcal{Z} \subset \mathbb{R}_{n_z} \to \mathbb{R}^{n_x}$ satisfies

$$\underline{q}_{f_d}(\mathbb{I}\mathcal{Z}) = q(T_R^{f_d}(\mathbb{I}\mathcal{Z}), T_O^f(\mathbb{I}\mathcal{Z})) \le \beta_R^f d(\mathbb{I}\mathcal{Z}) \tag{39}$$

for some $\beta_R^f > 0$, with a convergence rate $\alpha = 1$. Moreover, applying the subdivision principle, we have

$$q(T_R^{f_d}(\mathbb{I}\mathcal{Z}; k), T_O^f(\mathbb{I}\mathcal{Z}; k)) \le \frac{\gamma_R^f}{k} \tag{40}$$

where $\mathbb{I}\mathcal{Z}$ is subdivided into k^{n_z} interval vectors $\mathbb{I}\mathcal{Z}^l, l \in \mathbb{N}_{k^{n_z}}$ (i.e., with k divisions in each dimension such that $d(\mathbb{I}\mathcal{Z}^l_j) = \frac{d(\mathbb{I}\mathcal{Z}_j)}{k}$ for $j \in \mathbb{N}_{n_z}$, $l \in \mathbb{N}_{k^{n_z}}$), $T_R^{f_d}(\mathbb{I}\mathcal{Z}; k) \triangleq \bigcup_{l=1}^{k^{n_z}} T_R^{f_d}(\mathbb{I}\mathcal{Z}^l)$ and $T_O^f(\mathbb{I}\mathcal{Z}; k) \triangleq \bigcup_{l=1}^{k^{n_z}} T_O^f(\mathbb{I}\mathcal{Z}^l)$.

Proof: For a linear remainder function $h(z) = \mathbf{m}^\top z$ with $\mathbf{m} \triangleq \begin{bmatrix}\mathbf{m}_1 & \dots & \mathbf{m}_{n_x}\end{bmatrix}$ (cf. Lemma 3), $\hat{\overline{q}}_{f_d}(\mathbb{I}\mathcal{Z})$ in Theorem 3 can be upper bounded by triangle inequality by (39) with $\beta_R^f = \max_{i\in\mathbb{N}_{n_x}} \|\mathbf{m}_i\|_\infty$. Further, the proof of (40) follows the same lines as the proof of [34, Th. 4.1]. ■

D. Set Inversion Algorithm

The remainder-form decomposition functions returned by Algorithm 1 can be used with the generalized embedding system in Definition 3 to overapproximate (unconstrained) reachable sets of a dynamic system governed by the vector field f, which corresponds to the propagated/predicted sets in state observers/estimators. However, when additional state constraint information is available (e.g., sensor observations/measurements in state estimation problems, known safety constraints from system design and manufactured constraints from modeling redundancy [21], [22], [23], [24], [25]), an additional set inversion (also known as update or refinement) step will allow us to take the advantage of the constraints to shrink the propagated sets, i.e., to obtain a tighter subset of the propagated set that is compatible/consistent with the given constraints. Further details about the application of the reachable/propagated set and set inversion algorithms will be described in Section V.

Formally, given the constraint/observation function $\mu(x,u)$ in (13) with known u, a constraint/observation set $\mathcal{Y}$ with maximal and minimal values $\overline{y}, \underline{y}$ (satisfying $\mu(x,u) \subset \mathcal{Y} \subseteq [\underline{y}, \overline{y}]$) and a prior (or propagated/predicted) reachable interval $\mathbb{I}\mathcal{X}^p = [\underline{x}^p, \overline{x}^p]$, we wish to find an updated/refined interval $\mathbb{I}\mathcal{X}^u \subseteq \mathbb{I}\mathcal{X}^p$, such that (14) holds, i.e., to solve Problem 2. Finding $\mathbb{I}\mathcal{X}^u$ in (14) is called the *set inversion* problem [17]. To the best of our knowledge, existing set inversion algorithms/operators either compute subpavings (i.e., unions of intervals) instead of an interval using (conservative) natural inclusions (SIVIA [17, Ch. 3]) or are only applicable if relatively restrictive monotonicity assumptions hold ($\mathcal{I}_G$ [21, Algorithm 1]).

In this section, leveraging our proposed *DT* decomposition-based inclusion functions for an ELLS function $\nu(x) \triangleq \mu(x,u)$ with known u, i.e., $T_R^{\nu_d}$, we develop a novel set inversion algorithm that solves Problem 2, which is summarized in Algorithm 2. The main idea behind Algorithm 2 is based on the observation that a candidate interval $\mathbb{I}\Xi \triangleq [\underline{\xi}, \overline{\xi}] \subseteq \mathbb{I}\mathcal{X}^p$ that satisfies $T_R^{\nu_d}(\mathbb{I}\Xi) \cap [\underline{y}, \overline{y}] = \emptyset$ (i.e., if $\underline{\nu}_d(\underline{\xi}, \overline{\xi}) > \overline{y}$ or if $\overline{\nu}_d(\overline{\xi}, \underline{\xi}) < \underline{y}$) is *incompatible/inconsistent* with the set $\{x \in \mathbb{I}\mathcal{X}^p \,|\, \underline{y} \le \mu(x,u) \le \overline{y}\}$ and can be eliminated/ruled out from $\mathbb{I}\mathcal{X}^p$, and thus shrinking $\mathbb{I}\mathcal{X}^u$.

Using this idea, starting from the prior/propagated interval and using bisection for each dimension, Algorithm 2 shrinks the compatible interval from below and/or above if $\underline{\nu}_d$ is strictly greater than $\overline{y}$ or if $\overline{\nu}_d$ is strictly smaller than $\underline{y}$ (cf. Lines 8 and 18). Repeating this procedure along with bisections with a threshold ϵ, the candidate intervals that are determined to be *inconsistent* with the constraint/observation set are ruled out. Note that the ordering of the dimensions in the "for" loop on Line 3 may have an impact on the tightness of the returned interval $\mathbb{I}\mathcal{X}^u = [\underline{x}^u, \overline{x}_u]$ and so, it may be desirable to tailor the order to the problem at hand, to randomize the order or to repeat the algorithm with the previous $\mathbb{I}\mathcal{X}^u$ as $\mathbb{I}\mathcal{X}^p$ multiple times.

The following result shows that Algorithm 2 returns $\mathbb{I}\mathcal{X}^u = [\underline{x}^u, \overline{x}_u]$ that satisfies (14), i.e., solves Problem 2.

Proposition 11: Suppose Assumptions 1 and 2 hold and consider an ELLS constraint/observation function $\nu : \mathcal{X} \subset \mathbb{R}^{n_x} \to \mathbb{R}^{n_\mu}$, where $\nu(x) \triangleq \mu(x,u)$ with known u, a constraint/observation set $\mathcal{Y}_t \subseteq [\underline{y}_t, \overline{y}_t]$ and a prior/propagated interval $\mathbb{I}\mathcal{X}^p \triangleq [\underline{x}_p, \overline{x}_p] \in \mathbb{I}\mathbb{R}^{n_x}$. Then, the updated/refined interval $\mathbb{I}\mathcal{X}^u \triangleq [\underline{x}^u, \overline{x}^u]$ returned by Algorithm 2 satisfies (14).

Algorithm 2: Set Inversion Based on DT $T_R^{\nu_d}$.

1: **function** Set-Inv($\nu, \underline{J}_C^\nu, \overline{J}_C^\nu, \underline{x}^p, \overline{x}^p, \underline{y}, \overline{y}, \epsilon$)
2: Initialize: $\underline{x}^u \leftarrow \underline{x}^p, \overline{x}^u \leftarrow \overline{x}^p$;
3: **for** $i = 1$ to n_x **do**
4: $\underline{\zeta} \leftarrow \underline{x}_i^u; \overline{\zeta} \leftarrow \overline{x}_i^u$;
5: **while** $\overline{\zeta} - \underline{\zeta} > \epsilon$ **do**
6: $\zeta_m \leftarrow \frac{1}{2}(\overline{\zeta} + \underline{\zeta}); \overline{\xi} \leftarrow \overline{x}^u; \underline{\xi} \leftarrow \underline{x}^u; \underline{\xi}_i \leftarrow \zeta_m$;
7: $(\underline{\nu}_d, \overline{\nu}_d) \leftarrow T_R^{\nu_d}(\nu, \underline{J}_C^\nu, \overline{J}_C^\nu, \underline{\xi}, \overline{\xi})$; (Algorithm 1)
8: **if** $(\overline{\nu}_d < \underline{y}) \vee (\underline{\nu}_d > \overline{y})$ $\overline{\zeta} \leftarrow \zeta_m; \overline{x}_i^u \leftarrow \overline{\zeta}$; **then**
9: **else** $\underline{\zeta} \leftarrow \zeta_m$;
10: **end if**
11: **end while**
12: $\underline{\zeta} \leftarrow \underline{x}_i^u; \overline{\zeta} \leftarrow \overline{x}_i^u$;
13: **while** $\overline{\zeta} - \underline{\zeta} > \epsilon$ **do**
14: $\zeta_m \leftarrow \frac{1}{2}(\overline{\zeta} + \underline{\zeta}); \overline{\xi} \leftarrow \overline{x}^u; \underline{\xi} \leftarrow \underline{x}^u; \overline{\xi}_i \leftarrow \zeta_m$;
15: $(\underline{\nu}_d, \overline{\nu}_d) \leftarrow T_R^{\nu_d}(\nu, \underline{J}_C^\nu, \overline{J}_C^\nu, \underline{\xi}, \overline{\xi})$; (Algorithm 1)
16: **if** $(\overline{\nu}_d < \underline{y}) \vee (\underline{\nu}_d > \overline{y})$ $\underline{\zeta} \leftarrow \zeta_m; \underline{x}_i^u \leftarrow \underline{\zeta}$; **then**
17: **else** $\overline{\zeta} \leftarrow \zeta_m$;
18: **end if**
19: **end while**
20: **end for**
21: **return** $\underline{x}^u, \overline{x}^u$;
22: **end function**

Proof: Obviously, $\mathbb{I}\mathcal{X}^u \subseteq \mathbb{I}\mathcal{X}^p$ (i.e., $\underline{x}^p \le \underline{x}^u$ and $\overline{x}^p \ge \overline{x}^u$) by initialization and construction (cf. Lines 2, 9, and 19). Further, we show that $\mathbb{I}\mathcal{X}^u \supseteq \mathbb{I}\mathcal{X}^* \triangleq \{x \in \mathbb{I}\mathcal{X}^p \,|\, \underline{y} \le \nu(x) \le \overline{y}\}$. To use contradiction, suppose that it does not hold. Then, $\exists \zeta \in \mathbb{I}\mathcal{X}^*$ such that $\zeta \notin \mathbb{I}\mathcal{X}^u$, i.e., $\exists i \in \mathbb{N}_{n_x}$ such that $\zeta_i > \overline{x}_i^u$ or $\zeta_i < \underline{x}_i^u$. Without loss of generality, suppose the first case holds, i.e., $\zeta_i > \overline{x}_i^u$ (the proof for $\zeta_i < \underline{x}_i^u$ is similar). Then, $\zeta \in [\underline{x}^m, \overline{x}^p]$, where $\underline{x}_i^m > \overline{x}_i^u$ and $\underline{x}_{i'}^m = \underline{x}_{i'}^p$, $\forall i' \neq i$. Hence,

$$\underline{\nu}_d^R(\underline{x}^m, \overline{x}^p) \le \nu(\zeta) \le \overline{\nu}_d^R(\overline{x}^p, \underline{x}^m), \tag{41}$$

where $\overline{\nu}_d^R(\cdot,\cdot)$ and $\underline{\nu}_d^R(\cdot,\cdot)$ are the proposed upper and lower remainder-form decomposition functions in Algorithm 1. On the other hand, note that $\mathcal{X}^u \cap [\underline{x}^m, \overline{x}^p] = \emptyset$; hence, the interval $[\underline{x}^m, \overline{x}^p]$ has been "ruled out" by Algorithm 2. In other words, one of the "or" conditions in line 8 of Algorithm 2 must hold for this interval, i.e., $\overline{\nu}_d^R(\overline{x}^p, \underline{x}^m) < \underline{y} \ \vee \ \underline{\nu}_d^R(\underline{x}^m, \overline{x}^p) < \overline{y}$. Combining this and (41), we obtain $\nu(\zeta) < \underline{y} \vee \mu(\zeta) > \overline{y}$, which contradicts with $\zeta \in \mathbb{I}\mathcal{X}^*$ (i.e., $\underline{y} \le \nu(\zeta) \le \overline{y}$). ■

It is noteworthy that our set inversion algorithm can also be used with any applicable inclusion functions (such as $T_N^f, T_C^f, T_M^f, T_L^{f_d}, T_O^{f_d}$) or the best of them (i.e., by independently computing the reachable sets of all inclusion functions and intersecting them; cf. Corollary 1) in place of $T_R^{f_d}$ in Lines 7 and 17. On the other hand, the proposed $T_R^{f_d}$ (as well as $T_L^{f_d}, T_O^{f_d}$) can also be directly used in place of or in combination with natural inclusions within SIVIA [17, Ch. 3] to obtain subpavings (i.e., a union of intervals).

IV. Comparison With Existing Inclusion Functions

A. Comparison With the $T_L^{f_d}$ Inclusion Function

In this section, we compare the performance of the proposed $T_R^{f_d}$ with $T_L^{f_d}$ (cf. Proposition 3) through the following Theorem 5. We show that the decomposition function f_d^L, introduced in [8], [14], and [16] and recapped in Proposition 3, belongs to the family of the remainder-form decomposition functions in (15), and hence, $T_L^{f_d}$ cannot be tighter than $T_R^{f_d}$ that is the tightest decomposition function that belongs to (15).

Theorem 5 ($T_L^{f_d}$ versus $T_R^{f_d}$): Suppose all the assumptions in Theorem 1 hold. Then, the following statements are true.

i) f_d^L belongs to the family of decomposition functions in (15), i.e., for each $i \in \mathbb{N}_{n_x}$, a specific pair of $\mathbf{m}_i^L \in \mathbf{M}_i$ and $h_i^L \in \mathcal{H}_{\mathbf{M}_i}$ corresponds to the decomposition function f_d^L in [14, Th. 2], [16, Proposition 2], and [8, Special Case 1] (cf. Proposition 3).

ii) The optimal remainder-form decomposition function T_R^f is always tighter than (at least as good as) the inclusion function T_L^f, induced by the decomposition function f_d^L.

Proof: To prove (i), consider a specific decomposition function from the family of remainder functions in (15) that is constructed, for each $i \in \mathbb{N}_{n_x}$, with a supporting vector $\mathbf{m}_i^L$ and a linear remainder function $h_i^L(\cdot) = \langle \mathbf{m}_i^L, \cdot \rangle$ as follows:

$$(\mathbf{m}_i^L)_j = \begin{cases} \min((\underline{J}_C^f)_{ij}, 0), \text{ if } |\min((\underline{J}_C^f)_{ij}, 0)| \le |\max((\overline{J}_C^f)_{ij}, 0)| \\ \max((\overline{J}_C^f)_{ij}, 0), \text{if } |\max((\overline{J}_C^f)_{ij}, 0)| < |\min((\underline{J}_C^f)_{ij}, 0)| \end{cases} \quad (42)$$

for all $j \in \mathbb{N}_{n_z}$. Clearly, $\mathbf{m}_i^L \in \mathbf{M}_i^c \subset \mathbf{M}_i$ by its definition. Furthermore, it is easy to observe that $\mathbf{m}_i^L$ can be rewritten as

$$(\mathbf{m}_i^L)_j = \begin{cases} 0, & \text{if } (a_{ij} \ge 0) \vee (b_{ij} \le 0) \vee (j = i) \\ a_{ij}, & \text{if } (a_{ij} < 0) \wedge (b_{ij} > 0) \wedge (|a_{ij}| \le |b_{ij}|) \\ b_{ij} & \text{if } (a_{ij} < 0) \wedge (b_{ij} > 0) \wedge (|b_{ij}| \le |a_{ij}|) \end{cases} \quad (43)$$

where $a_{ij} \triangleq (\underline{J}_{Cij}^f)$ and $b_{ij} \triangleq (\overline{J}_C^f)_{ij}$. Recall that $(a_{ij} \ge 0)$, $(a_{ij} < 0) \wedge (b_{ij} > 0) \wedge (|a_{ij}| \le |b_{ij}|)$, $(a_{ij} < 0) \wedge (b_{ij} > 0) \wedge (|b_{ij}| \le |a_{ij}|)$, $(b_{ij} \le 0)$, and $j = i$ (CT systems only) correspond to Cases 1–5 in Proposition 3, respectively. Then, by (42) and (43), we find that $\zeta_{\mathbf{m}_i^L, j}(z, \hat{z})$ in (16) coincides with ζ_j in Proposition 3. Moreover, by (43),

$$h_i^L(\zeta_{\mathbf{m}_i^L}(\hat{z}, z)) = \langle \mathbf{m}_i^L, \zeta_{\mathbf{m}_i^L}(\hat{z}, z) \rangle = \sum_{j=1}^{\mathbb{N}_{n_z}} \phi_j^i$$
$$h_i^L(\zeta_{\mathbf{m}_i^L}(z, \hat{z})) = \langle \mathbf{m}_i^L, \zeta_{\mathbf{m}_i^L}(z, \hat{z}) \rangle = \sum_{j=1}^{\mathbb{N}_{n_z}} \psi_j^i \quad (44)$$

where $\phi_j^i \triangleq \begin{cases} 0, & \text{Cases } 1,4,5, \\ a_{ij}\hat{z}_j, & \text{Case } 2, \\ b_{ij} z_j, & \text{Case } 3, \end{cases}$ and $\psi_j^i \triangleq \begin{cases} 0, & \text{Cases } 1,4,5, \\ a_{ij} z_j, & \text{Case } 2, \\ b_{ij}\hat{z}_j, & \text{Case } 3. \end{cases}$ Consequently,

$$h_i^L(\zeta_{\mathbf{m}_i^L}(\hat{z}, z)) - h_i^L(\zeta_{\mathbf{m}_i^L}(z, \hat{z})) = \sum_{j=1}^{\mathbb{N}_{n_z}} \phi_j^i - \psi_j^i = \sum_{j=1}^{\mathbb{N}_{n_z}} \theta_j^i$$

where $\theta_j^i = \begin{cases} 0, & \text{Cases } 1, 4, 5, \\ a_{ij}(\hat{z}_j - z_j), & \text{Case } 2, \\ b_{ij}(z_j - \hat{z}_j), & \text{Case } 3. \end{cases}$ Then, defining two indicator functions $\alpha^i, \beta^i \in \mathbb{R}^{n_z}$, where for all $j \in \mathbb{N}_{n_z}$, $\alpha_j^i \triangleq \begin{cases} 0, & \text{Cases } 1, 3, 4, 5, \\ |a_{ij}|, & \text{Case } 2, \end{cases}$ and $\beta_j^i \triangleq \begin{cases} 0, & \text{Cases } 1, 2, 4, 5, \\ -|b_{ij}|, & \text{Case } 3, \end{cases}$ θ_j^i can be rewritten as $\theta_j^i = (\alpha_j^i - \beta_j^i)(z_j - \hat{z}_j)$, and hence,

$$h_i^L(\zeta_{\mathbf{m}_i^L}(\hat{z}, z)) - h_i^L(\zeta_{\mathbf{m}_i^L}(z, \hat{z})) = \sum_{j=1}^{\mathbb{N}_{n_z}} \theta_j^i = \langle \alpha^i - \beta^i, z - \hat{z} \rangle.$$

Finally, since $\zeta_{\mathbf{m}_i^L}(z, \hat{z})$ coincides with ζ in Proposition 3, by (15), $f_d^i(z, \hat{z}; \mathbf{m}_i^L, h_i^L) = h_i^L(\zeta_{\mathbf{m}_i^L}(\hat{z}, z)) + f_i(\zeta_{\mathbf{m}_i^L}(z, \hat{z})) - h_i(\zeta_{\mathbf{m}_i^L}(z, \hat{z}))) = f_i(\zeta) + \langle \alpha^i - \beta^i, z - \hat{z} \rangle = f_{d,i}^L(z, \hat{z})$, where $f_{d,i}^L(z, \hat{z})$ is the decomposition function introduced in Proposition 3 and is defined in (6).

ii) The result directly follows from (i) and Theorem 2. ∎

From the abovementioned, we know that $T_L^{f_d}$, which only considers a specific $\mathbf{m}_i \in \mathbf{M}_i$ for all $i \in \mathbb{N}_{n_x}$, cannot be tighter than $T_L^{f_d}$, which considers all $\mathbf{m}_i \in \mathbf{M}_i$. Nonetheless, since $T_L^{f_d}$ requires less computation, it can still be useful for systems with large dimensions, and can also be tighter than T_N^f, T_C^f, and T_M^f (see Examples 1 and 2 in the following). Further, this suggests that when computational resources are limited, it is also possible to consider a strict subset of $\mathbf{M}_i$ on top of the one in $T_L^{f_d}$ to obtain a tighter decomposition function than $T_L^{f_d}$. In addition, the abovementioned theorem indirectly proves that $T_L^{f_d}$ in Proposition 3 also applies to ELLS systems.

B. Comparison With T_N^f, T_C^f, and T_M^f Inclusion Functions

In this section, we compare the performance of interval arithmetic-based natural inclusions and some of their modifications, i.e., T_N^f, T_C^f, and T_M^f with the (DT) $T_R^{f_d}$, via computing the overapproximation for the range of some example functions. It is worth mentioning that we were not able to derive any theoretical results that show the superiority of one over the others. In fact, our simulation results showed that depending on the considered function and its corresponding domain, one of them can be tighter than the others in some cases and the opposite holds for other cases. However, in some cases, reflected in the following examples, $T_R^{f_d}$ typically returns tighter intervals.

1) Composition of Nonelementary Functions: In cases where the considered vector field is not a composition of "elementary functions" (e.g., simple monomials, sin, cos, monotone functions, etc.), T_N^f, T_C^f, and T_M^f are known to be hard to compute and result in conservative overapproximations for bounding the constituent functions that are often needed, which lead to poor inclusion functions, i.e., large errors. In these cases, it is most likely that $T_R^{f_d}$ returns better bounds. The following example describes one such function.

Example 1: Consider $f(x) = x \arctan(x^2 - 2x + 5)$, which is composed of nonelementary functions, and an interval domain $\mathbb{I}\mathcal{X} = [1, 3]$. In this case, T_N^f, T_C^f, T_M^f, $T_L^{f_d}$, and $T_R^{f_d}$ return $[-4.7124, 4.7124]$, $[1.3258,\ 4.3393]$, $[1.3187,\ 4.2475]$, $[1.2835, 2.9461]$, and $[1.1760, 2, 7468]$, respectively, where the final interval (corresponding to $T_R^{f_d}$) is a subset of all others.

2) "Almost" Sign-Stable Functions: In cases where f can be decomposed into a CJSS constituent and a relatively small additive perturbation, T_R^f will most likely return tighter bounds than the bounds returned by T_N^f, T_C^f, and T_M^f. For instance, consider the following example.

Example 2: Consider $f(x) = x^3 - 0.1x$, which is a monotone increasing (and hence CJSS) function on its interval domain $\mathbb{I}\mathcal{X} = [-1, 3]$, except on the short interval $[-\sqrt{\frac{0.1}{3}}, \sqrt{\frac{0.1}{3}}]$. For this example, T_N^f, T_C^f, T_M^f, $T_L^{f_d}$, and $T_R^{f_d}$ return $[-8.9000, 27.0100]$, $[-49.9000, 54.7000]$, $[-49.9000, 54.7000]$, $[-1.0300, 26.0100]$, and $[-1.0300, 26.0100]$, respectively, where $T_L^{f_d}$ and $T_R^{f_d}$ are much tighter than T_N^f, T_C^f and T_M^f.

3) Vector Fields With Several Additive Terms: It is also well known in the literature that natural, centered, and mixed-centered inclusions perform worse for the functions with many additive terms, compared with the ones with fewer additive terms [17], [19]. This is not necessarily true for the performance of $T_R^{f_d}$. The following example illustrates this fact, where a function with several additive terms is considered.

Example 3: Consider $f(x) = x_1x_2x_3 + x_1^2x_2 + x_2^2x_3 + x_3^2x_1 + x_1^2x_3 + x_3^2x_2 + x_2^2x_1 + x_1^3 + x_2^3 + x_3^3$ with an interval domain $\mathbb{I}\mathcal{X} = [-2, 2] \times [-2, 2] \times [-2, 2]$. Then, T_N^f, T_C^f, T_M^f, $T_L^{f_d}$, and $T_R^{f_d}$ return $[-80, 80]$, $[-76.45, 76.45]$, $[-73.62, 73.62]$, $[-176, 176]$, and $[-54.4, 54.4]$, respectively, where the final interval from $T_R^{f_d}$ is the tightest among all.

4) Existence of Closed-Form Decomposition Functions: Finally, it is notable that our proposed $T_R^{f_d}$ approach (and $T_L^{f_d}$) enables us to find closed-form inclusion functions for a wide class of vector fields. This can be analytically beneficial, e.g., in convergence analysis for reachable sets or stability analysis in interval observer designs [35], [36]. This is in contrast to natural, centered, and mixed-centered inclusions (and also $T_O^{f_d}$ in general), where a closed-form inclusion function for general classes of functions, is often not available.

V. Applications to Constrained Reachability Analysis and Guaranteed State Estimation

Consider the following constrained bounded-error system:

$$x_t^+ = f(x_t, u_t, w_t), \quad \mu(x_t, u_t) \in \mathcal{Y}_t \tag{45}$$

where $x_t^+ \triangleq x_{t+1}$ if (45) is a DT system (with sampling time δt) and $x_t^+ \triangleq \dot{x}_t$ if (45) is a CT system, $x_t \in \mathbb{R}^{n_x}$ with $x_0 \in [\underline{x}_0, \overline{x}_0]$ and $u_t \in \mathbb{R}^{n_u}$ are state and known input signals, $w_t \in [\underline{w}, \overline{w}] \in \mathbb{IR}^{n_w}$ is a bounded process disturbance signals, $\mathcal{Y}_t \subseteq [\underline{y}_t, \overline{y}_t] \in \mathbb{IR}^{n_y}$ is the time-varying, and *uncertain* state interval constraint and $f : \mathbb{R}^{n_x+n_u+n_w} \to \mathbb{R}^{n_x}$, $\mu : \mathbb{R}^{n_x+n_u} \to \mathbb{R}^{n_\mu}$ are known vector fields. The following proposition shows how to apply Algorithms 1 and 2, i.e., the mixed-monotone remainder-form decomposition function construction and the set inversion algorithms, to compute overapproximations of the reachable sets/framers of the states for the system in (45).

Proposition 12: Consider the system (45) with initial state $x_0 \in \mathbb{I}\mathcal{X}_0 \triangleq [\underline{x}_0, \overline{x}_0]$ and let $f(z_t) \triangleq \tilde{f}(x_t, u_t, w_t)$ and $\nu(x_t) \triangleq \mu(x_t, u_t)$ with $z_t \triangleq [x_t^\top \ \ w_t^\top]^\top$ and known u_t. Suppose that Assumptions 1 and 2 hold and ϵ is a chosen small positive threshold. Then, for all $t \geq 0, CR^f(t, \mathbb{I}\mathcal{X}_0) \subset \mathbb{I}\mathcal{X}_t^u \triangleq [\underline{x}_t^u, \overline{x}_t^u]$, where $CR^f(t, \mathbb{I}\mathcal{X}_0) \triangleq \{\phi(t, x_0, w_t) \mid x_0 \in \mathbb{I}\mathcal{X}_0, \mu(x_t, u_t) \in \mathcal{Y}_t \subseteq [\underline{y}_t, \overline{y}_t]$ and $w_t \in \mathbb{I}\mathcal{W}, \ \ \forall t \geq 0\}$ is the constrained reachable set at time t of (45) when initialized within $\mathbb{I}\mathcal{X}_0$ and $\mathbb{I}\mathcal{X}_t^u \triangleq [\underline{x}_t^u, \overline{x}_t^u]$ is the solution to the following constrained embedding system:

$$\begin{bmatrix} \underline{x}_t^{p+} \\ \overline{x}_t^{p+} \end{bmatrix} = \begin{bmatrix} \underline{f}_d\left(\left[(\underline{x}_t^u)^\top \ \underline{w}^\top\right]^\top, \left[(\overline{x}_t^u)^\top \ \overline{w}^\top\right]^\top\right) \\ \overline{f}_d\left(\left[(\overline{x}_t^u)^\top \ \overline{w}^\top\right]^\top, \left[(\underline{x}_t^u)^\top \ \underline{w}^\top\right]^\top\right) \end{bmatrix}, \ \begin{bmatrix} \underline{x}_0^p \\ \overline{x}_0^p \end{bmatrix} = \begin{bmatrix} \underline{x}_0 \\ \overline{x}_0 \end{bmatrix}$$

$$(\underline{x}_t^u, \overline{x}_t^u) = \text{SET-INV}(\nu, \underline{J}_C^\nu, \overline{J}_C^\nu, \underline{x}_t^p, \overline{x}_t^p, \underline{y}_t, \overline{y}_t, \epsilon)$$

where $(\underline{f}_d(\cdot, \cdot), \overline{f}_d(\cdot, \cdot)) = T_R^{f_d}(f, \underline{J}_C^f, \overline{J}_C^f, \cdot, \cdot)$ is the DT or CT decomposition-based inclusion function in Algorithm 1, and the SET-INV function in Algorithm 2 is based on the DT $T_R^{\nu_d}(\nu, \underline{J}_C^\nu, \overline{J}_C^\nu, \cdot, \cdot)$ from Algorithm 1. Consequently, the constrained system state trajectory x_t satisfies $\underline{x}_t^u \leq x_t \leq \overline{x}_t^u$ at all times t.

Proof: The results directly follow from applying Propositions 1 and 2, Theorems 1 and 2, and Lemmas 3–11. ■

Remark 4: It can be easily shown that the abovementioned results also apply to guaranteed state estimation (with the standard propagation and update steps) by noticing that uncertain observation functions $y_t = \mu(x_t, u_t) + Vv_t$ can be written in the form of $\mu(x_t, u_t) \in \mathcal{Y}_t \triangleq [y_t - \overline{s}, y_t - \underline{s}]$ in (45), where $\overline{s} = V^\oplus \overline{v} - V^\ominus \underline{v}$, $\underline{s} = V^\oplus \underline{v} - V^\ominus \overline{v}$, $V^\oplus \triangleq \max(V, \mathbf{0}_{n_v})$ and $V^\ominus \triangleq V^\oplus - V$, with $\mathbf{0}_{n_v}$ being a zero vector in $\mathbb{R}^{n_v}$ [37]. In this setting, $[\underline{x}_t^p, \overline{x}_t^p]$ is the state interval from the prediction/propagation step and $[\underline{x}_t^u, \overline{x}_t^u]$ is the updated state interval after a measurement update step. Further, if (45) is a sampled-data system, i.e., the system dynamics is continuous and the observations are sampled in DT, our proposed approach still directly applies with minor modifications. The stability analysis of the guaranteed state estimation approach and the synthesis of stable interval/set-valued estimators are subjects of our future work.

VI. Simulations

In this section, we compare the performances of T_N^f (natural inclusions; cf. Proposition 5), T_C^f, T_M^f (centered and mixed-centered inclusions; cf. Proposition 6), $T_L^{f_d}$ (decomposition functions proposed in [14]; cf. Proposition 3), $T_R^{f_d}$ (our proposed remainder-form decomposition function in Algorithm 1), $T^f_{S_1}$ (the first proposed bounding approach in [21, Theorem 1], if applicable), $T^f_{S_2}$ (the second proposed approach in [21, Theorem 2], if applicable), and $T^{f_d}_O$ (the tight decomposition functions proposed in [12, Th. 2] for DT and in [13, Th. 1] for CT systems, when computable; cf. Proposition 4) in computing the reachable sets of several unconstrained and constrained dynamical systems in the form of (13). Further, by the intersection property in Corollary 1, we can also intersect the reachable sets of all applicable inclusion functions (except $T^{f_d}_O$) for comparison. Note that we only compare methods in the literature that use *first-order* (generalized) gradient information. The consideration of higher order information is a subject of our future work.

Due to space limitations, only a subset of the simulation examples are provided here. The readers are referred to the arXiv version of this article [38] for the rest of the simulation examples. In brief, for both CT and DT systems, our proposed $T^{f_d}_R$ often outperforms all applicable methods, except $T^{f_d}_O$ (that is often not computable), and the best of T^f_N–$T^{f_d}_R$ additionally provides significant improvements. Moreover, the tightness of the approaches can be further improved with modeling redundancy and when sensor observations and constraints are available (using the set inversion approach in Section III-D).

A. Generic Transport Longitudinal Model

We consider NASA's Generic Transport Model (GTM) [39], a remote-controlled 5.5% scale commercial aircraft [40], with the following main parameters: wing area $S = 5.902$ ft^2, mean aerodynamic chord $\bar{c} = 0.9153$ ft, mass $m = 1.542$ slugs, pitch axis moment of inertia $I_{yy} = 4.254$ slugs/ft^2, air density $\rho = 0.002375$ slugs/ft^3, and gravitational acceleration $g = 32.17$ ft/s^2. The longitudinal dynamics of the GTM can be described as the following CT dynamical system:

$$\dot{V}_t = \frac{-D_t - mg\sin(\theta_t - \alpha_t) + T_{x,t}\cos\alpha_t + T_{z,t}\sin\alpha_t}{m}$$

$$\dot{\alpha}_t = q_t + \frac{-L_t + mg\cos(\theta_t - \alpha_t) - T_{x,t}\sin\alpha_t + T_{z,t}\cos\alpha_t}{mV_t}$$

$$\dot{q}_t = \frac{M_t + T_{m,t}}{I_{yy}},\ \dot{\theta}_t = q_t \tag{46}$$

where V_t, α_t, q_t, and θ_t are air speed (ft/s), angle of attack (rad), pitch rate (rad/s), and pitch angle (rad), respectively. Moreover, $T_{x,y}$ (lbs), $T_{z,t}$ (lbs), $T_{m,t}$ (lbs–ft), D_t (lbs), L_t (lbs), and M_t (lbs) denote the projection of the total engine thrust along the body's X-axis, the projection of the total engine thrust along the body's Z-axis, the pitching moment due to both engines, the drag force, the lift force, and the aerodynamic pitching moment, respectively, with their nominal values given in [41]. Defining $x_t \triangleq [V_t\ \alpha_t\ q_t\ \theta_t]^\top$ with $x_0 \in [147, 158] \times [0.04, 0.05] \times [0.1, 0.2] \times [0.04, 0.05]$, Fig. 1(a) depicts the reachable set approximations for $x_{1,t}$ and $x_{2,t}$ of the system (46). For this system, $T^{f_d}_O$ is not computable, since the stationary/critical points of the vector fields cannot be obtained analytically, and as shown in Fig. 1(a), $T^{f_d}_R$ obtains a tighter overapproximation than T^f_N, T^f_C, T^f_M, and $T^{f_d}_L$, with the best of T^f_N–$T^{f_d}_R$ showing further improvement.

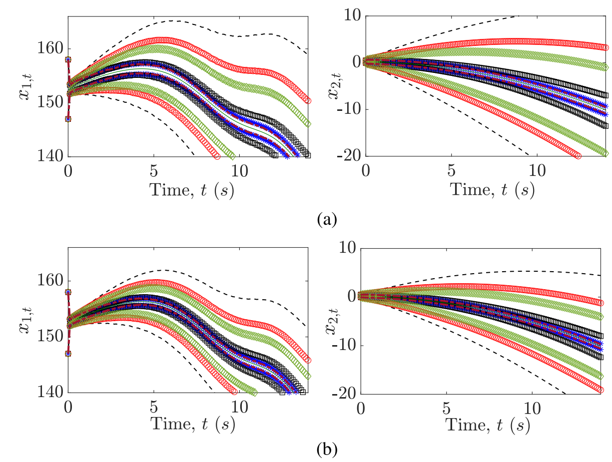

Fig. 1. Upper and lower bounds on x_1 and x_2 in the GTM system (Example F in [38]), when applying $T^f_N(--)$, T^f_C (○), T^f_M (◇), T^f_L (□), $T^{f_d}_R$ (∗), the best of T^f_N–$T^{f_d}_R$ (..), as well as the midpoint trajectory (_). (a) GTM system without observations (46). (b) GTM system with observations (46) and (47).

Next, we consider an additional set of measurements in the form of a linear observation equation

$$y_t = x_{1,t} + x_{2,t} - x_{3,t} + v_t,\ v_t \in [-0.01, 0.01]. \tag{47}$$

Then, applying T^f_N–$T^{f_d}_R$ along with the set inversion approach in Algorithm 2 to the constrained system (46) and (47), we observe considerably tighter intervals for all approaches with observations [Fig. 1(b)] when compared with the approximations of the reachable sets without observations [Fig. 1(a)].

VII. Conclusion

A tractable family of remainder-form mixed-monotone decomposition functions was proposed in this article for a relatively large class of nonsmooth, discontinuous systems called ELLS systems that is proven to include LLC systems. We characterized the lower and upper bounds for the overapproximation errors when using the proposed remainder-form decomposition functions to overapproximate the true range/image set of ELLS nonlinear mappings, where the lower bound is achieved by our proposed tight and tractable decomposition function, which was further proven to be tighter than the ones introduced in [8] and [14]. Further, a novel set inversion algorithm based on decomposition functions was developed to further refine/update the reachable sets when knowledge of state constraints and/or when measurements are available, which can be applied for constrained reachability analysis and guaranteed state estimation for bounded-error CT, DT, or sampled-data systems. Finally, the effectiveness of our proposed mixed-monotone decomposition functions was demonstrated using several benchmark examples. In future work, we will consider higher order information about derivatives of functions and a more principled way to design modeling redundancy to further improve the tightness of decomposition and inclusion functions. We will also extend our proposed tools to perform reachability analysis with *polytopes* and to compute inner and outerapproximations of reach-avoid-stay and (controlled) invariant sets.

References

[1] M. Hirsch and H. Smith, "Monotone dynamical systems," in *Handbook of Differential Equations: Ordinary Differential Equations*. Boston, MA, USA: Acad. Press, 2006, pp. 239–357.

[2] D. Angeli and E. Sontag, "Monotone control systems," *IEEE Trans. Autom. Control*, vol. 48, no. 10, pp. 1684–1698, Oct. 2003.

[3] M. Kulenović and O. Merino, "A global attractivity result for maps with invariant boxes," *Discrete Continuous Dyn. Syst.-B*, vol. 6, no. 1, pp. 97–110, 2006.

[4] G. Enciso, H. Smith, and E. Sontag, "Nonmonotone systems decomposable into monotone systems with negative feedback," *J. Differ. Equ.*, vol. 224, no. 1, pp. 205–227, 2006.

[5] J. Gouzé and K. Hadeler, "Monotone flows and order intervals," *Nonlinear World*, vol. 1, no. 1, pp. 23–34, 1994.

[6] H. Smith, "Global stability for mixed monotone systems," *J. Difference Equ. Appl.*, vol. 14, no. 10–11, pp. 1159–1164, 2008.

[7] T. Chu and L. Huang, "Mixed monotone decomposition of dynamical systems with application," *Chin. Sci. Bull.*, vol. 43, no. 14, pp. 1171–1175, 1998.

[8] M. Abate and S. Coogan, "Computing robustly forward invariant sets for mixed-monotone systems," in *Proc. IEEE Conf. Decis. Control*, 2020, pp. 4553–4559.

[9] S. Coogan and M. Arcak, "Stability of traffic flow networks with a polytree topology," *Automatica*, vol. 66, pp. 246–253, 2016.

[10] S. Coogan and M. Arcak, "Efficient finite abstraction of mixed monotone systems," in *Proc. Intl. Conf. Hybrid Syst., Comput. Control*, 2015, pp. 58–67.

[11] L. Yang, A. Karnik, B. Pence, M. Waez, and N. Ozay, "Fuel cell thermal management: Modeling, specifications, and correct-by-construction control synthesis," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 5, pp. 1638–1651, Sep. 2019.

[12] L. Yang and N. Ozay, "Tight decomposition functions for mixed monotonicity," in *Proc. IEEE Conf. Decis. Control*, 2019, pp. 5318–5322.

[13] M. Abate, M. Dutreix, and S. Coogan, "Tight decomposition functions for continuous-time mixed-monotone systems with disturbances," *IEEE Control Syst. Lett.*, vol. 5, no. 1, pp. 139–144, Jan. 2021.

[14] L. Yang, O. Mickelin, and N. Ozay, "On sufficient conditions for mixed monotonicity," *IEEE Trans. Autom. Control*, vol. 64, no. 12, pp. 5080–5085, Dec. 2019.

[15] P. Meyer, A. Devonport, and M. Arcak, *Interval Reachability Analysis: Bounding Trajectories of Uncertain Systems With Boxes for Control and Verification*. Cham, Switzerland: Springer, 2021.

[16] P. Meyer, "Reachability analysis of neural networks using mixed monotonicity," *IEEE Control Syst. Lett.*, vol. 6, pp. 3068–3073, 2022.

[17] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis*. London, U.K.: Springer, 2001.

[18] H. Cornelius and R. Lohner, "Computing the range of values of real functions with accuracy higher than second order," *Computing*, vol. 33, no. 3/4, pp. 331–347, 1984.

[19] R. Moore, R. Kearfott, and M. Cloud, *Introduction to Interval Analysis*. Philadelphia, PA, USA: SIAM, 2009.

[20] G. Alefeld and G. Mayer, "Interval analysis: Theory and applications," *J. Comput. Appl. Math.*, vol. 121, no. 1/2, pp. 421–464, 2000.

[21] X. Yang and J. Scott, "Accurate uncertainty propagation for discrete-time nonlinear systems using differential inequalities with model redundancy," *IEEE Trans. Autom. Control*, vol. 65, no. 12, pp. 5043–5057, Dec. 2020.

[22] T. Alamo, J. Bravo, and E. Camacho, "Guaranteed state estimation by zonotopes," *Automatica*, vol. 41, no. 6, pp. 1035–1043, 2005.

[23] T. Alamo, J. Bravo, M. Redondo, and E. Camacho, "A set-membership state estimation algorithm based on DC programming," *Automatica*, vol. 44, no. 1, pp. 216–224, 2008.

[24] J. Scott and P. Barton, "Bounds on the reachable sets of nonlinear control systems," *Automatica*, vol. 49, no. 1, pp. 93–100, 2013.

[25] K. Shen and J. Scott, "Rapid and accurate reachability analysis for nonlinear dynamic systems by exploiting model redundancy," *Comput. Chem. Eng.*, vol. 106, pp. 596–608, 2017.

[26] R. Horst and N. V. Thoai, "DC programming: Overview," *J. Optim. Theory Appl.*, vol. 103, no. 1, pp. 1–43, 1999.

[27] M. Khajenejad, Z. Jin, and S. Z. Yong, "Interval observers for simultaneous state and model estimation of partially known nonlinear systems," in *Proc. Amer. Control Conf.*, 2021, pp. 2848–2854.

[28] M. Abate and S. Coogan, "Improving the fidelity of mixed-monotone reachable set approximations via state transformations," in *Proc. Amer. Control Conf.*, 2021, pp. 4674–4679.

[29] S. Coogan, "Mixed monotonicity for reachability and safety in dynamical systems," in *Proc. IEEE Conf. Decis. Control*, 2020, pp. 5074–5085.

[30] F. Clarke, *Nonsmooth Analysis in Systems and Control Theory*. New York, NY, USA: Springer, 2009.

[31] V. Demyanov and A. Rubinov, *Constructive Nonsmooth Analysis*. Frankfurt, Germany: Peter Lang, 1995.

[32] J. Cortes, "Discontinuous dynamical systems," *IEEE Control Syst. Mag.*, vol. 28, no. 3, pp. 36–73, Jun. 2008.

[33] T. Caraballo, J. Langa, and J. Valero, "Asymptotic behaviour of monotone multi-valued dynamical systems," *Dyn. Syst.*, vol. 20, no. 3, pp. 301–321, Jun. 2005.

[34] H. Ratschek and J. Rokne, *Computer Methods for the Range of Functions*. Chichester, New York, USA: Horwood, 1984.

[35] M. Khajenejad and S. Z. Yong, "Simultaneous input and state interval observers for nonlinear systems with full-rank direct feedthrough," in *Proc. IEEE Conf. Decis. Control*, 2020, pp. 5443–5448.

[36] M. Khajenejad and S. Z. Yong, "Simultaneous input and state interval observers for nonlinear systems with rank-deficient direct feedthrough," in *Proc. Eur. Control Conf.*, 2021, pp. 2311–2316.

[37] D. Efimov, T. Raïssi, S. Chebotarev, and A. Zolghadri, "Interval state observer for nonlinear time varying systems," *Automatica*, vol. 49, no. 1, pp. 200–205, 2013.

[38] M. Khajenejad and S. Z. Yong, "Tight remainder-form decomposition functions with applications to constrained reachability and interval observer design," 2021, *arXiv:2103.08638*, extended version.

[39] E. Summers et al., "Quantitative local L2-gain and reachability analysis for nonlinear systems," *Int. J. Robust Nonlinear Control*, vol. 23, no. 10, pp. 1115–1135, 2013.

[40] A. Murch and J. Foster, "Recent NASA research on aerodynamic modeling of post-stall and spin dynamics of large transport airplanes," in *Proc. 45th AIAA Aerosp. Sci. Meeting Exhibit*, 2007, p. 463.

[41] B. Stevens, F. Lewis, and E. Johnson, *Aircraft Control and Simulation: Dynamics, Controls Design, and Autonomous Systems*. Hoboken, NJ, USA: Wiley, 2015.

Mohammad Khajenejad (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from The University of Tehran, Tehran, Iran, in 2006 and 2009, respectively, and the Ph.D. degree in mechanical engineering from Arizona State University, Tempe, AZ, USA, in 2021.

He is currently a Postdoctoral Scholar with Mechanical and Aerospace Engineering Department, University of California, San Diego, CA, USA. He is the author or coauthor of diverse papers published in refereed conference proceedings and journals. His current research interests include set-theoretic control, resiliency and privacy of networked cyber-physical systems, and robust game theory.

Dr. Khajenejad was the recipient of the ASU Dean's Dissertation Award for his Ph.D. dissertation.

Sze Zheng Yong (Member, IEEE) received the Dipl.-Ing.(FH) degree in automotive engineering from the Esslingen University of Applied Sciences, Esslingen, Germany, in 2008, and the S.M. and Ph.D. degrees in mechanical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA in 2010 and 2016, respectively.

He is currently an Associate Professor with the Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, USA. Prior to that, he was an Assistant Professor with the School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, USA. His research interests include the broad areas of control, estimation, planning, identification and analysis of hybrid systems, with applications to autonomous, robotic and cyber-physical dynamic systems and their safety, robustness and resilience.

Dr. Yong was the recipient of the DARPA Young Faculty Award in 2018, the NSF CAREER and NASA Early Career Faculty awards in 2020, and the ONR Young Investigator Program Award in 2022.