Thresholding in Set-Membership Approaches to Fault Isolation Using the Minkowski Functional

Gustave Bainier, Benoît Marx, Jean-Christophe Ponsart

Thresholding in Set-Membership Approaches to Fault Isolation using the Minkowski functional

Gustave Bainier[a,*], Benoît Marx[a], Jean-Christophe Ponsart[a]

[a] *Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France,*

Abstract

In the context of model-based diagnosis of dynamical systems, this paper introduces the Minkowski functional as a new analytical thresholding tool for a generic set-membership approach to FDI. This tool formalizes the threshold computation problem as a function upper- and lower-bounding problem, and provides non-specific inequalities that can be applied to residual signals for real-time system analysis. In particular, this paper provides two new characterizations of the linear transformation of a smooth convex set based on its Minkowski functional. These results are then illustrated by computing several thresholds to residual signals.

Keywords: Minkowski functional, set-membership, fault detection, fault isolation, convex analysis

1. Introduction

1.1. Context of the study

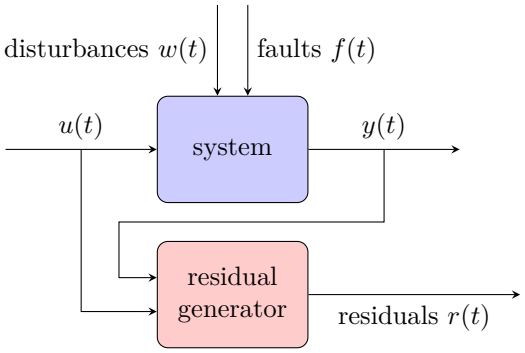

Figure 1: Residuals generation architecture.

Given a monitored dynamical system for which a mathematical model is known, fault diagnosis consists in identifying the faults occurring in the system, its sensors, and its actuators, by analyzing - for example - the discrepancies between the measured inputs and outputs of the system and their expected or estimated behaviour. This task has been approached both by the control theory community using Fault Detection and Isolation (FDI) techniques, and the diagnosis (DX) community, by relying on artificial intelligence techniques. Although this work focuses on techniques from the FDI community, it is worth mentioning the pioneer works of the DX community [48, 41, 19], as well as more recent works bridging the gap between the two communities [26, 12, 61]. In order to achieve fault diagnosis, residual signals are generally generated by

*Corresponding author

Email addresses: gustave.bainier@univ-lorraine.fr (Gustave Bainier), benoit.marx@univ-lorraine.fr (Benoît Marx), jean-christophe.ponsart@univ-lorraine.fr (Jean-Christophe Ponsart)

numerically checking the consistency between the inputs and the outputs of the system onto a single vector, measurable in real-time, nominally centered at the origin, and which diverges from it in case of discrepancies (Figure 1) [25, 39, 27, 13, 21]. From these residuals, the fault diagnosis process is generally handled in three steps: fault detection (detect if at least one fault is active), fault isolation (pinpoint which set of faults is active) and fault identification (evaluate the magnitude of each occurring fault). One obvious way of performing fault diagnosis consists in leveraging physical redundancy in the sensors and actuators monitoring the system (e.g. using triplex sensors), but these solutions usually require costly extra equipment, hence analytical redundancy methods are often preferred. After some pioneer works in aerospace engineering [20] and chemical engineering [34, 56], several analytical residual generating methods have been developed in the literature for both linear and nonlinear systems. The most well-known schemes include the parity space approach [14, 39, 28] and the use of diagnostic observers [54, 6, 24, 39]. Of course, the residual signals obtained through these methods are affected by the system's parametric uncertainties (such as modeling errors) and exogenous disturbances (such as measurement noise), and one of the challenges in performing fault diagnosis is to distinguish the unavoidable noise in the residual signals from the signature of an actual fault affecting the system, while taking into account its parametric uncertainties.

Usually, the presence of noise and parametric uncertainties are handled in two ways [25, 27, 21]:

- firstly, by minimizing or even canceling their influence on the residual generation through some robustness techniques (the *active* approach);
- secondly, by processing the residual signals with consideration for the statistical influence of the noise and of the uncertainties (the *passive* approach).

However, identifying a probabilistic model of the noise and using it with consideration for the parametric uncertainties of the model is generally a difficult task which limits the applicability of the statistical methods of the passive approach. This obstacle has led to the development of so-called *set-membership* (or *bounding*) approaches, where the passive step is achieved through the construction of an adaptive threshold on the residuals which only relies on the knowledge of the noise and uncertainty bounds [36, 17].

The difficulty common to all set-membership approaches - which also includes reachability analysis, robust Model Predictive Control (MPC) and state estimation - concerns the computation of the bounding sets, which are generally hard to compute exactly, especially in real-time. In the literature, geometric results are generally leveraged to characterize their inner- and outer-approximations. The usual method consists in approximating all the sets of the problem by convex sets of a convenient class: orthotopes in the case of interval analysis [1], including interval observer techniques [35, 46, 22]; ellipsoids [31, 49]; parallelotopes [44]; zonotopes [40, 10, 43, 11]; polytopes [9], sometimes considered as constrained zonotopes [55, 45]; or more generic convex shapes, represented using hybrid zonotopes [16, 7] or constrained convex generators [57, 58]. Dedicated algorithms are then used to compute the exact representation of the sets of interest, or at least some inner- or outer-approximations when the exact sets cannot be easily computed. The classes of convex sets and their representations are chosen depending on the system and the task at hand, as a trade-off between their geometric accuracy and the efficiency of the algorithms associated with their representation. These set-memberships techniques have sometimes been merged with results from the probabilistic paradigm mentioned earlier, to enhance their efficiency [38, 5, 15].

Although less common in the literature, there already exist generalized set-membership approaches, which usually rely on the support functions associated with the convex sets of interest [32, 47]. Under these generalized approaches, very little structural properties are assumed on the sets of interests, leading to interesting unified results applicable to all other techniques. This present work falls under the category of these more generalized frameworks.

1.2. Contributions

This paper is an attempt at providing a unified set-theoretic thresholding perspective to FDI, and at establishing results that can be applied to a large class of set-membership approaches, with a possibly higher

degree of accuracy than, for example, by only using ellipsoids. The idea is to establish a flexible framework to the manipulation of convex (or even star-convex) sets, allowing for easier combinations of the already existing set-membership approaches to FDI, while unifying them to a new degree of generality. This paper does not aim at competing with these already existing methodologies, but rather complements them by offering an alternative and unified analytical point of view to represent the sets of interest, similarly to support functions. All algorithm dedicated to special classes of convex sets remain applicable under the proposed methodology, and can be enhanced by leveraging the generic results found in this paper.

To this end, this paper introduces a tool from order theory: a (upper semi-)lattice structure; as well as a tool from convex analysis: the Minkowski functional (or Minkowski function, or gauge function). The Minkowski functional has already gained popularity for the control of dynamical systems due to its practicality, in particular to find so-called Minkowski-Lyapunov functions demonstrating the stability of ordinary differential equations [29, 8, 47]. The authors also believe that these functions are convenient in the context of set-membership approaches, and in particular for FDI, since the Minkowski functional of a set often provides a way to describe this set using a single inequality, which can very easily serve as a threshold that can be scaled up or down. More generally, the Minkowski functional of a set $\mathcal{S}$ possesses algebraic properties linked with the geometric properties of $\mathcal{S}$ which are interesting to know in the context of a set-membership approach. However, as far as the authors are aware, its use has not been adopted in a FDI context yet, and could be very well-suited to complement already existing set-membership approaches, in particular the generic approaches relying on support functions.

The main contributions of the paper are listed hereafter:

- The (upper semi-)lattice is introduced as a tool from order theory to isolate the fault occurrences of a system.
- The Minkowski functional of a set is introduced as a thresholding tool, which can represent a set implicitly through a simple inequality. The evaluation of the Minkowksi functional gives an intuitive measurement of the extend to which a set-membership relation is verified. This evaluation is introduced later in this document as the Minkowski signal associated with the set-membership relation.
- The usual operations of union, intersection, Minkowski sum and linear transformation of convex sets are investigated under the lens of the Minkowksi functional.
- The expression of the Minkowski functional is leveraged to obtain analytical criteria on the minimal fault magnitude guaranteeing its isolation.

Although the union and intersection operation are easily dealt with, the Minkowksi sum and the linear transformation of sets represented by their Minkowski functional remain challenging to obtain without relying on pre-existing algorithms dedicated to special classes of convex sets. Some inner- and outer-approximations are nevertheless obtained to approximate the Minkowski sum of convex sets, and a tight result is obtained for ellipsoids. Similarly, two methodologies to analytically characterize the linear transformation of a convex set are derived. The opposite properties are found for support functions associated with convex sets, where the union and intersection set operations are difficult to handle, contrary to the Minkowski sum and the linear transformation which are easily dealt with. These properties probably explain the prevalence of the support functions in the set-membership literature. The support function of a set is however not well-adapted for thresholding purposes. Both tools are complementary in this regard. As a matter of fact, the Minkowski functional of a compact and convex set is the support function associated with its polar set [53, 47].

The paper is organized as follows. Section 3 presents a highly generic set-membership approach to FDI for residuals following an uncertain linear internal structure (a notion introduced in Section 3.1). Section 4 introduces the Minkowski functional as a practical tool to compute inner- and outer-approximations of the sets manipulated in the set-membership approach described earlier, and to generate *Minkowski signals* from the residuals, which are easily thresholded. In particular, Section 4.3 provides two analytical characteriza-

tions of the linear transformation of smooth convex sets. Section 5 illustrates the described methodology with an academic example. Finally, some conclusions and perspectives are discussed in Section 6.

2. Definitions, Notations

$\mathbb{R}$, $\tilde{\mathbb{R}}$, $\mathbb{R}_{>s}$ and $\mathbb{R}_{\geq s}$ denote resp. the field of real numbers, $\mathbb{R}\cup\{-\infty,+\infty\}$, and the intervals $(s,+\infty)$ and $[s,+\infty)$. $\{x\}$ denotes the singleton containing x. $\cup$, $\cap$, $\setminus$, $\oplus$ stand resp. for the union, the intersection, the set difference, and the Minkowski sum between sets. $\subset$ and $\subseteq$ stand for the strict and non-strict inclusion. $\emptyset$ denotes the empty set. Given two sets $\mathcal{S}_1 \subseteq \mathbb{R}^n$, $\mathcal{S}_2 \subseteq \mathbb{R}^m$, and $M \in \mathbb{R}^{m\times n}$ a matrix with m rows and n columns, $-\mathcal{S}_1 \triangleq \{x : -x \in \mathcal{S}_1\}$, hull$(\mathcal{S}_1)$ stands for the convex hull of $\mathcal{S}_1$, $M\mathcal{S}_1$ stands for the linear transformation of $\mathcal{S}_1$ by M, and for all $\star \in \{\cup,\cap,\setminus,\oplus\}$, $\mathcal{S}_1 \star_M \mathcal{S}_2 \triangleq \mathcal{S}_1 \star \{x \in \mathbb{R}^n : Mx \in \mathcal{S}_2\}$. intr$(\mathcal{S}_1)$, cls$(\mathcal{S}_1)$ and $\partial\mathcal{S}_1$ denote respectively the interior, the closure and the boundary of $\mathcal{S}_1$ in $\mathbb{R}^n$. $t\mathcal{S}_1$ denotes the scaled set $\{x \in \mathbb{R}^n : x = ty, y \in \mathcal{S}_1\}$. The sets of symmetric definite positive real matrix is denoted $\mathbb{S}_n^{++}(\mathbb{R})$.

Given two vectors $x, y \in \mathbb{R}^n$, $\langle x|y\rangle$ stands for the usual inner product of the Euclidean space $\mathbb{R}^n$. Let $\mathcal{V}$ be a linear subspaces of $\mathbb{R}^n$. $\mathcal{V}^\perp$ denotes the orthogonal complement of $\mathcal{V}$ in $\mathbb{R}^n$. $\dim(\mathcal{V})$ denotes the dimension of $\mathcal{V}$.

$[\![\cdot,\cdot]\!]$ stands for the integer interval. $\#$ stands for the cardinal. $\mathbf{P}_{n_f}$ stands for the power set of $[\![1,n_f]\!]$, and for all $\mathcal{I} \subseteq [\![1,n_f]\!]$ (i.e. $\mathcal{I} \in \mathbf{P}_{n_f}$), $\overline{\mathcal{I}} \triangleq [\![1,n_f]\!] \setminus \mathcal{I}$. Given two positive integers m, n, $\binom{n}{m}$ stands for the binomial coefficient. $\vee$ and $\wedge$ stand resp. for the logical "OR" and "AND".

Given f, g two maps, $f \circ g$ denotes their composition, f^{-1} denotes the inverse of f, and ∇f denotes the gradient of f. Let M represent a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Im(M) and Ker(M) denote respectively the image and the kernel of M. This document identifies M with its matrix representation in the standard basis. $M^\top$, M^{-1} and $M^\dagger$ denote resp. the transpose, the inverse, and the Moore–Penrose inverse of M.

3. Set-membership diagnosis

After introducing in Section 3.1 the internal structure of the residual that will be studied throughout this paper (Definition 3.1), it is recalled in Section 3.2 how the set-membership approach to diagnosis deals with the question of fault detection using the Direct Image Test (Theorem 3.1), and then a generic extension of this approach to fault isolation is suggested in Section 3.3 by leveraging some order theory results (Theorem 3.2). For now, these results are stated using an all-encompassing set-theoretic framework.

3.1. Residuals internal structure

The residuals synthesized for system diagnosis can be written under two forms [27]:

- their computational form, giving their relationship to the measured signals used in their calculation;
- their internal form, giving their relationship to the noises and faults affecting the system.

It is assumed that the computational form of the residuals provides a vector of signals $r(\theta,t)$, with θ denoting the time varying parametric uncertainties of the system, and where $r(\theta,t) = 0$ is verified in a noise- and fault-free context.

This paper focuses on the set-membership approach to diagnosis schemes where the residuals follow an *uncertain linear internal structure*. This structure is introduced in Definition 3.1 below.

Definition 3.1 (Uncertain linear internal structure)**.** *The residual signal $r(\theta,t) \in \mathbb{R}^{n_r}$ is said to follow an uncertain linear internal structure if its internal form can be written as:*

$$r(\theta,t) = \sum_{i=1}^{n_w} W_i(\theta,t)w_i(t) + \sum_{i=1}^{n_f} F_i(\theta,t)f_i(t) \tag{1}$$

with $w_i(t) \in \mathcal{W}_i \subset \mathbb{R}^{n_{w_i}}$ *the bounded noises affecting the residual signals,* $f_i(t) \in \mathcal{F}_i \subset \mathbb{R}^{n_{f_i}}$ *the potential faults to be detected,* $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$ *the time-varying uncertainties of the system, and* $W_i \in \mathbb{R}^{n_r \times n_{w_i}}$, $F_i \in \mathbb{R}^{n_r \times n_{f_i}}$ *uncertain time-varying matrices. Let* $\Omega(t) \subseteq \Theta$ *denote the bounding set for the uncertainties* θ *of the system at time* t*. The union over all the uncertainties* $\theta \in \Omega(t)$ *of the residual signals is referred to as the residual set and is denoted* $R(t)$*:*

$$R(t) \triangleq \bigcup_{\theta \in \Omega(t)} \{r(\theta, t)\} \tag{2}$$

This structure is common among residuals obtained for Linear Time Invariant, Time Varying, and Parameter Varying (LTI, LTV, LPV) discrete-time systems. In particular, residual signals following this internal structure can be obtained with parity space approaches or some simple diagnostic observer designs [27, 39, 40, 11]. Note that this structure can also be leveraged for residuals with a more complex internal structure by means of linearization, or by considering the residuals' internal nonlinearities as new parametric uncertainties.

Remark 3.1. *Despite (1) being an instantaneous expression, delays on the disturbances, on the faults and on the time-varying components of the matrices are not difficult to handle by considering* $\tilde{w}_i(t) = w_i(t-\tau_i)$, $\tilde{f}_i(t) = f_i(t-\tau_i)$, $\tilde{W}_i(\theta, t) = W_i(\theta, t-\tau_i)$ *and* $\tilde{F}_i(\theta, t) = F_i(\theta, t-\tau_i)$*, with* $\tau_i \geq 0$*. However, if the parametric uncertainties* $\theta \in \Omega(t)$ *are also time-varying and subject to delays, the residuals analysis can become more challenging, and some assumptions on the variation rate of* θ *may become extremely useful to approximate the bounding set* $\Omega(t)$*.*

Remark 3.2. *The choice of writing the residuals in the form (1) might be surprising, as they can be simply rewritten as the uncertain linear transformation of a single fault and noise vector. The sums are introduced in order to distinguish between several types of additive faults and noises (actuator, sensor, or parametric faults and noises), at potentially different instant of the past. Moreover, the faults and noises are assumed to be vectors instead of scalars in order to limit the number of terms in the sums, hence reducing the complexity and conservatism of the FDI methods described in this paper. The model described above contains in fact* m_w *scalar noises and* m_f *scalar faults, with:*

$$m_w \triangleq \sum_{i=1}^{n_w} n_{w_i}, \qquad m_f \triangleq \sum_{i=1}^{n_f} n_{f_i} \tag{3}$$

Example 3.1 (Parity Relations [40])**.** *Considering the following uncertain discrete-time linear system:*

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t + G_1(\theta)f_t + V_1(\theta)w_t \tag{4a}$$
$$y_t = C(\theta)x_t + D(\theta)u_t + G_2(\theta)f_t + V_2(\theta)w_t \tag{4b}$$

where $x_t \in \mathbb{R}^{n_x}$ *is the state of the system,* $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$ *is a vector of parametric uncertainties,* $u_t \in \mathbb{R}^{n_u}$ *is the known input vector,* $y_t \in \mathbb{R}^{n_y}$ *is the measured output vector,* $f_t \in \mathcal{F} \subset \mathbb{R}^{n_x}$ *are potential faults to detect, and* $w_t \in \mathcal{W} \subset \mathbb{R}^{n_x}$ *are the bounded noises affecting the system. In the following parity space approach,* θ *is assumed to be a constant and unknown uncertainty at a horizon of* h *time steps. It is moreover assumed that an estimation of* θ*, denoted* $\hat{\theta}$*, is known. The residual* $r([\theta, \hat{\theta}], t)$ *is a vector signal with the following computational and internal form resp.:*

$$r([\theta, \hat{\theta}], t) \triangleq N(\hat{\theta}) \left(Y_t - \Gamma_u(\hat{\theta}) U_t \right) \tag{5a}$$
$$r([\theta, \hat{\theta}], t) = N(\hat{\theta}) \left(\mathcal{O}(\theta)x_t + \Gamma_w(\theta)W_t + \Gamma_f(\theta)F_t + \left(\Gamma_u(\theta) - \Gamma_u(\hat{\theta}) \right) U_t \right) \tag{5b}$$

with for all $(Z, z) \in \{(U, u), (Y, y), (W, w), (F, f)\}$, $Z_t = \begin{pmatrix} z_{t-h}^\top & \dots & z_t^\top \end{pmatrix}^\top$*, and where the matrix* N *is*

synthesized such that $N(\hat{\theta})\mathcal{O}(\theta)x_t \approx 0$, *with*

$$\Gamma_u = \begin{bmatrix} D & & & 0 \\ CB & \ddots & & \\ \vdots & \ddots & & \\ CA^{h-1}B & \dots & CB & D \end{bmatrix}, \quad \Gamma_w = \begin{bmatrix} V_2 & & & 0 \\ CV_1 & \ddots & & \\ \vdots & \ddots & & \\ CA^{h-1}V_1 & \dots & CV_1 & V_2 \end{bmatrix},$$
$$\Gamma_f = \begin{bmatrix} G_2 & & & 0 \\ CG_1 & \ddots & & \\ \vdots & \ddots & & \\ CA^{h-1}G_1 & \dots & CG_1 & G_2 \end{bmatrix}, \quad \mathcal{O} = \begin{bmatrix} C^\top & [CA]^\top & \dots & [CA^h]^\top \end{bmatrix}^\top \tag{6}$$

where the dependence on θ *or* $\hat{\theta}$ *has been omitted for concision. The residual* r *follows the uncertain linear internal structure introduced in Definition 3.1 if* $N(\hat{\theta})\mathcal{O}(\theta)x_t$ *and* $(\Gamma_u(\theta) - \Gamma_u(\hat{\theta}))U_t$ *are assumed to be bounded noise signals.*

Example 3.2 (Luenberger observer [24])**.** *The system (4) is considered once again, but* θ *is now assumed to be an uncertainty measured in real-time. The Luenberger observer (7) is designed with a matrix* $L \in \mathbb{R}^{n_x \times n_y}$ *chosen such that the system generating the state estimation error* $e_{t+1} = (A(\theta_t) + LC(\theta_t))e_t$ *is globally asymptotically stable.*

$$\hat{x}_{t+1} = A(\theta_t)\hat{x}_t + B(\theta_t)u_t + L(y_t - C(\theta_t)\hat{x}_t - D(\theta_t)u_t) \tag{7}$$

A residual $r([\theta_t, \theta_{t-1}], t)$ *can be obtained with the following computational and internal form resp.:*

$$r([\theta_t, \theta_{t-1}], t) \triangleq y_t - C(\theta_t)\hat{x}_t - D(\theta_t)u_t \tag{8a}$$
$$r([\theta_t, \theta_{t-1}], t) = C(\theta_t)[A(\theta_{t-1})(x_{t-1} - \hat{x}_{t-1}) - Lr([\theta_{t-1}, \theta_{t-2}], t-1) + G_1(\theta_{t-1})f_{t-1} + V_1(\theta_{t-1})w_{t-1}]$$
$$\dots + G_2(\theta_t)f_t + V_2(\theta_t)w_t \tag{8b}$$

The residual r *follows the uncertain linear internal structure introduced in Definition 3.1 if the state estimation error* $(x_{t-1} - \hat{x}_{t-1})$ *and the residual at the previous time-step* $r([\theta_{t-1}, \theta_{t-2}], t-1)$ *are assumed to be noise signals, with bounds that can be evaluated recursively, e.g. using set-membership state estimation [54, 6].*

Remark 3.3. *From now on, the time-dependence in* $r(\theta, t)$, $R(t)$, $W(\theta, t)$, $w_i(t)$, $F(\theta, t)$, *and* $f_i(t)$ *is omitted, resp.* $r(\theta)$, R, $W(\theta)$, w_i, $F(\theta)$ *and* f_i.

3.2. Fault detection

As discussed in the introduction, fault detection consists in performing a consistency test to a monitored dynamical system in order to detect if the system is subject to at least one occurring fault. In a fault-free context, a residual $r(\theta)$ following the structure (1) is only affected by noises, and belongs to $\mathcal{R}_\theta$, a finite Minkowski sum of uncertain linear transformations of the sets $(\mathcal{W}_i)_{1 \le i \le n_w}$. Knowing precisely these bounding sets, the set-membership approach provides $r(\theta) \in \mathcal{R}_\theta$ as an elementary consistency test across all uncertainties $\theta \in \Theta$, where $\mathcal{R}_\theta$ is the set of all the residual values possibly due to the noises, defined by:

$$\mathcal{R}_\theta \triangleq \bigoplus_{i=1}^{n_w} W_i(\theta)\mathcal{W}_i, \text{ and } \mathcal{R}_\Theta \triangleq \bigcup_{\theta \in \Theta} \mathcal{R}_\theta \tag{9}$$

The boundary of $\mathcal{R}_\theta$ can be considered as an *exact* or *clear-cut* threshold for fault detection, since it provides the sharpest criterion to detect a fault which also avoids any false detection. Checking if the residuals verify $\cup_{\theta \in \Theta}(r(\theta) \setminus \mathcal{R}_\theta) \neq \emptyset$ in order to detect a fault is usually called the Direct Image Test (DIT) [43, 62], and it relies on the following property:

Theorem 3.1 (Direct Image Test)**.** *Considering that the following statement holds:*

$$\forall i \in [\![1, n_f]\!], f_i = 0 \Rightarrow \forall \theta \in \Theta, r(\theta) \in \mathcal{R}_\theta \tag{10}$$

the contrapositive statement provides:

$$\exists \theta \in \Theta, r(\theta) \notin \mathcal{R}_\theta \Rightarrow \exists i \in [\![1, n_f]\!], f_i \neq 0 \tag{11}$$

which is to say: if there exists an uncertainty $\theta \in \Theta$ for which the residual $r(\theta)$ does not belong to the set $\mathcal{R}_\theta$ (or more succinctly, if $\cup_{\theta\in\Theta}(r(\theta) \setminus \mathcal{R}_\theta) \neq \emptyset$), then at least one fault has occurred in the system.

Remark 3.4. *Note that no internal model of the faults influence on the residuals is leveraged at this point, hence this fault detection scheme can be applied to any residuals signals for which the influence of the noises on the residuals reduces to the internal structure (1) in a fault-free context.*

Remark 3.5 (Robust Direct Image Test)**.** *This test is not equivalent to checking if $(\cup_{\theta\in\Theta} r(\theta))\setminus(\cup_{\theta\in\Theta}\mathcal{R}_\theta)) \triangleq R \setminus \mathcal{R}_\Theta \neq \emptyset$ (with R defined in (2)). This later test, although not as sharp as (11), is still interesting as it does not require to know which uncertainty θ is responsible for each residual in R, which is often the case in fault diagnosis schemes. This paper refers to this latter test as the Robust DIT (RDIT).*

The DIT still works using inner-approximations of Θ or outer-approximations of $\mathcal{R}_\Theta$. In particular, the literature sometimes evaluates $\hat{\Omega}(t)$, an estimated set of values for θ at time t (such that $\Omega(t) \subseteq \hat{\Omega}(t) \subseteq \Theta$ in (2)) using some additional assumptions on θ (e.g. known dynamic, bounded rate of variation, partial estimation, etc). The DIT described above is then performed with this set estimate in order to detect a fault. This approach is sometimes called the *inverse image test*, where the term *inverse* refers to the dynamical construction of $\hat{\Omega}(t)$, which can involve the computation of the pre-image set of some output signals [43, 62]. This paper will mainly focus on DIT, but its extension to inverse image tests is possible as long as residuals of the form (1) can be constructed. In that case, instead of Θ, the set $\hat{\Omega}(t)$ should be considered.

3.3. Fault isolation

Following the ideas previously introduced, checking what are the occurring faults in the system can again be achieved by means of consistency tests. Moving forward, $\mathbf{P}_{n_f}$ stands for the power set of $[\![1, n_f]\!]$, i.e. the set of all the subsets of $[\![1, n_f]\!]$, and for all $\mathcal{I} \subseteq [\![1, n_f]\!]$ (i.e. $\mathcal{I} \in \mathbf{P}_{n_f}$), $\overline{\mathcal{I}} \triangleq [\![1, n_f]\!] \setminus \mathcal{I}$. Assuming $\mathcal{I} \in \mathbf{P}_{n_f}$ a list of potentially active faults indices, the obtainable residuals are given by the set $\mathcal{R}_{\theta,\mathcal{I}}$ defined hereafter.

Definition 3.2 (Feasible set)**.** *The feasible set $\mathcal{R}_{\theta,\mathcal{I}}$ is the set of residuals obtainable with $\mathcal{I} \in \mathbf{P}_{n_f}$ a list of potentially active faults indices. The residuals structure (1) provides:*

$$\mathcal{R}_{\theta,\mathcal{I}} \triangleq \mathcal{R}_\theta \oplus \left(\bigoplus_{i\in\mathcal{I}} F_i(\theta)\mathcal{F}_i \right), \text{ and } \mathcal{R}_{\Theta,\mathcal{I}} \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_{\theta,\mathcal{I}} \tag{12}$$

where it is considered that $\mathcal{R}_{\theta,\emptyset} = \mathcal{R}_\theta$ in a fault-free context. The collections of all the feasible sets are denoted FS_θ and FS_Θ, with:

$$FS_\theta \triangleq \{\mathcal{R}_{\theta,\mathcal{I}} : \mathcal{I} \in \mathbf{P}_{n_f}\}, \qquad FS_\Theta \triangleq \{\mathcal{R}_{\Theta,\mathcal{I}} : \mathcal{I} \in \mathbf{P}_{n_f}\} \tag{13}$$

It is easily verified that there are $\#\mathrm{FS}_\theta = 2^{n_f}$ feasible sets, hence their calculation can rapidly become computationally heavy, especially over all uncertainties θ. Moreover, in a faulty situation, the fault isolation process has to discriminate between all the feasible sets containing $r(\theta)$ in order to estimate the list of possible active faults. This task is difficult, and to the authors' knowledge, the literature has only tackled this problem for specific fault isolation schemes. In particular, fault isolation has motivated the design of Dedicated and Generalized Observer Schemes (the DOS and GOS architectures) [24] and the use of structured residuals and fault matrices [28]. Leveraging some tools from order theory, this section provides a

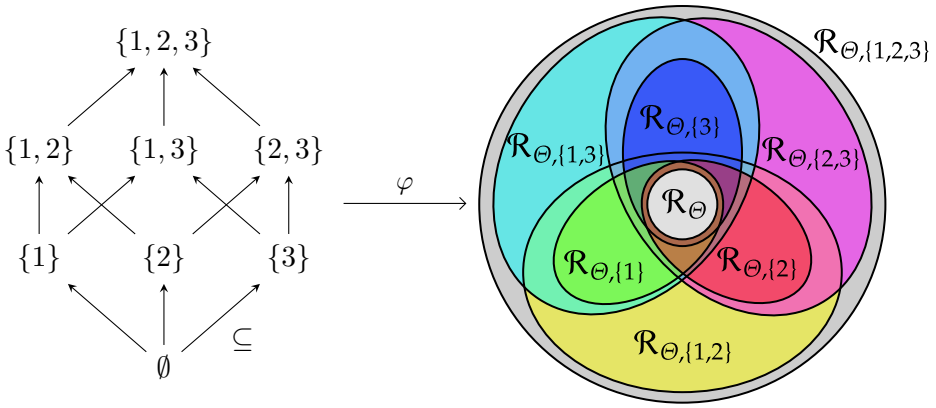

Figure 2: Hasse diagram of the lattice $(\mathbf{P}_3, \subseteq)$ in relation to the inclusion hierarchy of the feasible sets FS_Θ.

straightforward and all-encompassing procedure to fault isolation, which generalizes the DIT discussed previously.

As a first observation, one can notice that the inclusion hierarchy of the feasible sets (13) contains *at least* the same lattice structure as the power set $\mathbf{P}_{n_f}$ taken with the inclusion $\subseteq$ as a lattice ordering (Figures 2 and 3). This is formalized in the following property.

Property 3.1 (Order structure of the feasible sets)**.** *Let FS_θ be considered with the partial order $\preceq$ defined by $\mathcal{R}_{\theta,\mathcal{I}_1} \preceq \mathcal{R}_{\theta,\mathcal{I}_2}$ if and only if $\mathcal{I}_1 \subseteq \mathcal{I}_2$. $(FS_\theta, \preceq)$ and $(\mathbf{P}_{n_f}, \subseteq)$ have an isomorphic lattice structure. Moreover if $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$ (i.e., each fault can be null), then:*

$$\forall \mathcal{I}_1, \mathcal{I}_2 \in \mathbf{P}_{n_f},\ \mathcal{R}_{\theta,\mathcal{I}_1} \preceq \mathcal{R}_{\theta,\mathcal{I}_2} \Rightarrow \mathcal{R}_{\theta,\mathcal{I}_1} \subseteq \mathcal{R}_{\theta,\mathcal{I}_2} \tag{14}$$

meaning $\preceq$ is only a restriction of the partial order $\subseteq$ on FS_θ.

Proof. The map $\varphi : \mathbf{P}_{n_f} \to \mathrm{FS}_\theta$ defined such that for all $\mathcal{I} \in \mathbf{P}_{n_f}$, $\varphi(\mathcal{I}) = \mathcal{R}_{\theta,\mathcal{I}}$, is by definition an order isomorphism, and order isomorphisms preserve lattice structures, hence, since $(\mathbf{P}_{n_f}, \subseteq)$ is a lattice, then $(\mathrm{FS}_\theta, \preceq)$ has an isomorphic lattice structure. Moreover if $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$, from equation (2) and by definition of the Minkowski sum, (14) holds, and φ is an order homomorphism between $(\mathbf{P}_{n_f}, \subseteq)$ and $(\mathrm{FS}_\theta, \subseteq)$. □

Remark 3.6. *This property is easily extended to FS_Θ (Figure 2).*

Remark 3.7. *Notice that the reciprocal to (14) is not verified in general, since non-trivial inclusions such as $\mathcal{R}_{\theta,\{2\}} \subseteq \mathcal{R}_{\theta,\{1,3\}}$ can hold true in practice. Hence, even if $(FS_\theta, \subseteq)$ contains at least the lattice structure of $(\mathbf{P}_{n_f}, \subseteq)$ by restriction to $(FS_\theta, \preceq)$, it is not necessarily a lattice itself (Figure 3).*

As discussed previously, the DIT described in Theorem 3.1 can be generalized to the feasible sets (13). In that case, in order to prove that at least one fault of index $i \in \mathcal{I}$ has occurred in the system, one must show that a residual $r(\theta)$ does not belong to the feasible set which has the complement of $\mathcal{I}$ (denoted $\overline{\mathcal{I}}$) as a list of potentially active faults. Succinctly, there must exists a $\theta \in \Theta$ such that $r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}}$ to demonstrate that one of the fault of index $i \in \mathcal{I}$ is active. Contrary to the DIT to fault detection, all sets of potentially active faults $\mathcal{I} \in \mathbf{P}_{n_f}$ have to be considered, and the question of which elements of $\mathbf{P}_{n_f}$ verifying the DIT provide the most meaningful results must be raised. It is shown below that the sets of FS_θ satisfying the DIT follow an upper semi-lattice structure [18], where the strongest statements about the system's faults are found at the bottom of the semi-lattice (Example 3.3 illustrated by Figure 4).

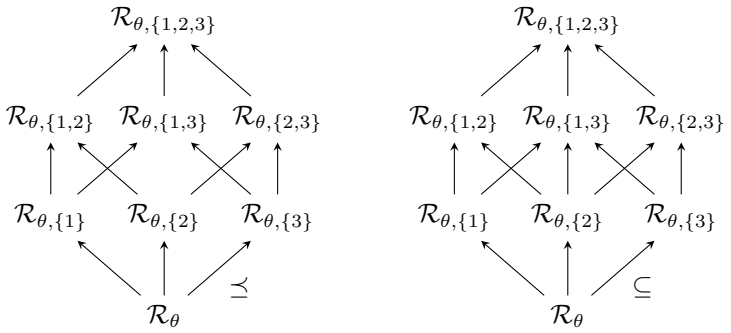

Figure 3: Hasse diagram of the lattice $(\mathrm{FS}_\theta, \preceq)$ (on the left) and of the partially ordered set $(\mathrm{FS}_\theta, \subseteq)$ (on the right) with $n_f = 3$, where the non-trivial inclusion $\mathcal{R}_{\theta,\{2\}} \subseteq \mathcal{R}_{\theta,\{1,3\}}$ is verified.

Theorem 3.2 (Generalized Direct Image Test)**.** *If $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$ (i.e., each fault can be null), then, the DIT set defined below is an upper semi-lattice with inclusion taken as partial order.*

$$DIT = \left\{ \mathcal{I} \in \mathbf{P}_{n_f} \ : \ \exists \theta \in \Theta \ s.t. \ r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}} \right\} \tag{15}$$

Denoting P_i with $i \in, [\![1, n_f]\!]$ and $Q_\mathcal{I}$ with $\mathcal{I} \triangleq \{i_1, \ldots, i_p\} \in \mathbf{P}_{n_f}$ the following statements:

$$P_i \triangleq [f_i \neq 0] : \text{“ The } i\text{-th vector fault is active”} \tag{16a}$$

$$Q_\mathcal{I} \triangleq P_{i_1} \vee \cdots \vee P_{i_p} : \text{“ At least one of the } i_1\text{-th, } \ldots, \ i_p\text{-th vector fault is active”} \tag{16b}$$

the DIT set demonstrates that the statement $\bigwedge_{\mathcal{I} \in DIT} Q_\mathcal{I}$ holds true. Moreover, the following equivalence holds:

$$\left(\bigwedge_{\mathcal{I} \in DIT} Q_\mathcal{I} \right) \Leftrightarrow \left(\bigwedge_{\mathcal{I} \in m_e(DIT)} Q_\mathcal{I} \right) \tag{17}$$

where $m_e(DIT)$ stands for the sets of minimal elements of $(DIT, \subseteq)$. These elements belong to $\mathbf{P}_{n_f}$ and summarize all the information about the faults' activity that can be deduced from DIT.

Proof. This proof is split in three parts. First, $(\mathrm{DIT}, \subseteq)$ is shown to be an upper semi-lattice. Then $\bigwedge_{\mathcal{I} \in \mathrm{DIT}} Q_\mathcal{I}$ is shown to hold. Finally, the equivalence (17) is proved.

To prove that $(\mathrm{DIT}, \subseteq)$ is an upper semi-lattice, as $\mathrm{DIT} \subseteq \mathbf{P}_{n_f}$, it is sufficient to demonstrate the stability of DIT with respect to inclusion [18]. Let $\mathcal{I}_1 \in \mathrm{DIT}$ and $\mathcal{I}_2 \in \mathbf{P}_{n_f}$ such that $\mathcal{I}_1 \subseteq \mathcal{I}_2$. By definition there exits $\theta \in \Theta$ for which $r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}_1}$. Since $\overline{\mathcal{I}_2} \subseteq \overline{\mathcal{I}}_1$, by an application of the lattice isomorphism φ, $\mathcal{R}_{\theta,\overline{\mathcal{I}_2}} \preceq \mathcal{R}_{\theta,\overline{\mathcal{I}}_1}$, and finally if $0 \in \mathcal{F}_i$ for all $i \in [\![1, n_f]\!]$, (14) provides $\mathcal{R}_{\theta,\overline{\mathcal{I}_2}} \subseteq \mathcal{R}_{\theta,\overline{\mathcal{I}}_1}$, hence $r(\theta) \notin \mathcal{R}_{\theta,\overline{\mathcal{I}_2}}$, meaning $\mathcal{I}_2 \in \mathrm{DIT}$.

The statement $\bigwedge_{\mathcal{I} \in \mathrm{DIT}} Q_\mathcal{I}$ holds by a direct and repeated application of the usual DIT to the sets $\mathcal{R}_{\Theta,\overline{\mathcal{I}}}$ with $\mathcal{I} \in \mathbf{P}_{n_f}$.

Finally the equivalence (17) is shown by double implication.

$\boxed{\Rightarrow}$ This implication is an easy consequence of $m_e(\mathrm{DIT}) \subseteq \mathrm{DIT}$.

$\boxed{\Leftarrow}$ On one hand, for all $\mathcal{I}_1, \mathcal{I}_2 \in \mathrm{DIT}$, $(\mathcal{I}_1 \subseteq \mathcal{I}_2) \Rightarrow (Q_{\mathcal{I}_1} \Rightarrow Q_{\mathcal{I}_2})$ is trivial considering that $Q_{\mathcal{I}_j}$ holds if and only if at least one of the $(P_i)_{i \in \mathcal{I}_j}$ holds, with $j \in \{1, 2\}$. On the other hand, considering

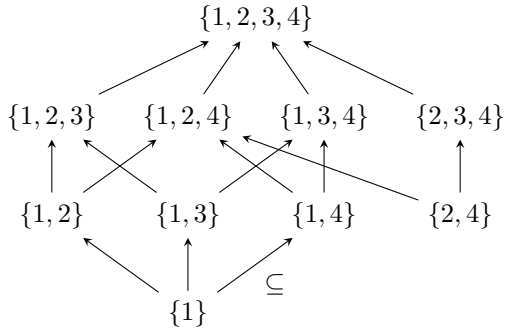

Figure 4: Hasse diagram of upper semi-lattice $(\mathrm{DIT}, \subseteq)$ defined in Example 3.3.

$\{\mathcal{I}_1 \in \mathrm{DIT} : \mathcal{I}_1 \subseteq \mathcal{I}_2\}$ a subset of DIT, since this set is finite, it must contain at least one minimal element, which will also be a minimal element of DIT by definition. Hence for all $\mathcal{I}_2 \in \mathrm{DIT}$, there exists $\mathcal{I}_1 \in m_e(\mathrm{DIT})$ such that $\mathcal{I}_1 \subseteq \mathcal{I}_2$, which implies $(Q_{\mathcal{I}_1} \Rightarrow Q_{\mathcal{I}_2})$. Repeating the process for all $Q_{\mathcal{I}}$ in $\left(\bigwedge_{\mathcal{I}\in\mathrm{DIT}} Q_{\mathcal{I}}\right)$ demonstrates the implication. □

Corollary 3.1 (Generalized Robust Direct Image Test)**.** *The following inclusion is verified (with R defined in (2)):*

$$RDIT \triangleq \left\{\mathcal{I} \in \mathbf{P}_{n_f} \,:\, R \setminus \mathcal{R}_{\Theta,\overline{\mathcal{I}}} \neq \emptyset\right\} \subseteq DIT \tag{18}$$

The set RDIT (Robust Direct Image Test) on the left also follows an upper semi-lattice structure with respect to inclusion, where the minimal elements summarize all the information about the faults' activity that can be deduced from the elements in the set.

Remark 3.8. *As noted in Remark 3.5, the tests performed to obtain the elements of RDIT do not require to know which uncertainty θ is responsible for each residual $r(\theta)$ in R.*

Corollary 3.2 (Fault isolation)**.** *If there exists $\mathcal{I} \in DIT$ (or $\mathcal{I} \in RDIT$) such that $\#\mathcal{I} = 1$, i.e. $\mathcal{I} = \{i\}$ with $i \in [\![1, n_f]\!]$, then the i-th fault has been isolated and $f_i \neq 0$ is guaranteed.*

Remark 3.9. *No matter how many faults are active in the system, $r(\theta) \in \mathcal{R}_{\theta,[\![1,n_f]\!]}$ and $R \subseteq \mathcal{R}_{\Theta,[\![1,n_f]\!]}$ are always verified by definition, and the edge cases $\emptyset \in DIT$ and $\emptyset \in RDIT$ do not need to be considered. If these edge cases happen anyway, it means that the model used for the residual is incorrect or incomplete.*

Remark 3.10. *As for the DIT and the RDIT, these generalized DIT and RDIT still work using inner-approximations of Θ or outer-approximations of $\mathcal{R}_{\theta,\overline{\mathcal{I}}}$.*

Example 3.3. *Consider a system with $n_f = 4$ potentially active faults, and a residual R such that the DIT results in:*

$$DIT = \{\{1\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \{1,2,3,4\}\} \tag{19}$$

The upper semi-lattice structure of DIT is illustrated by Figure 4. The minimal elements of DIT are given below:

$$m_e(DIT) = \{\{1\}, \{2,4\}\} \tag{20}$$

Hence, the generalized DIT provides the truth statement $P_1 \wedge (P_2 \vee P_4)$, i.e. $[f_1 \neq 0] \wedge ([f_2 \neq 0] \vee [f_4 \neq 0])$ which is to say: the first fault of the system is active, and the second or the fourth (or both) faults are active.

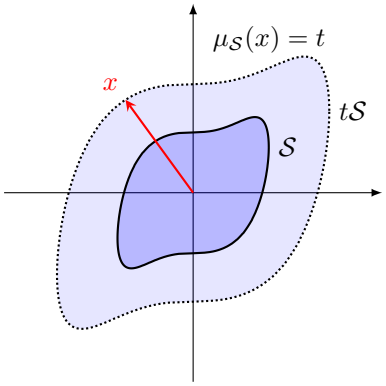

Figure 5: Minkowski functional of a set $\mathcal{S}$ of $\mathbb{R}^2$, star-convex at 0.

4. The Minkowski functional for set-membership fault isolation

In practice, the feasible sets (13) introduced in Definition 3.2 can be difficult to compute, and when their direct calculation is not possible, inner- and outer-approximation of these sets are often used as fault detection thresholds. In this section, the authors suggest a unifying approach where the geometric problem of finding inner- and outer-approximations to the feasible sets of the previous section is translated into the problem of upper- and lower-bounding their Minkowski functionals.

Section 4.1 begins with the introduction of the general properties of the Minkowski functional, and, in particular, how to use it for the usual set operations (union, intersection, Minkowski sum and left-invertible linear transformations). In Section 4.3, two new characterizations of the (non-invertible) linear transformation of smooth convex sets are provided. Finally, these results are leveraged in Section 4.4 to obtain a generic solution to the inner- and outer-approximation problem of the feasible sets (13), providing upper-bounds on the minimal faults magnitude guaranteeing fault isolation through the generalized image test.

4.1. The Minkowski functional

Intuitively, the Minkowski functional $\mu_{\mathcal{S}}$ associated with the set $\mathcal{S}$ is defined such that for all x, $\mu_{\mathcal{S}}(x)$ provides the smallest scaling t of the set $\mathcal{S}$ with respect to the origin such that the scaled set $t\mathcal{S}$ reaches x (Figure 5). This definition is formalized below, and the expression of the Minkowski functional is specified for the usual classes of convex sets.

Definition 4.1 (Minkowski functional [53])**.** *Given $\mathcal{S}$ a non-empty set of $\mathbb{R}^n$, the Minkowski functional associated to $\mathcal{S}$ is the map $\mu_{\mathcal{S}} : \mathbb{R}^n \to \tilde{\mathbb{R}}_{\geq 0}$ defined by:*

$$\mu_{\mathcal{S}}(x) \triangleq \inf \left\{ t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{S} \right\} \tag{21}$$

Remark 4.1. *The set $\mathcal{S}$ is said to be absorbing if $\mu_{\mathcal{S}} < +\infty$.*

Remark 4.2. *Alternative functionals can be used for off-centered sets $\mathcal{S} \subset \mathbb{R}^n$, where the scaling is performed with respect to a point $x_0 \in \mathbb{R}^n$ instead of the origin, such that $\mu_{\mathcal{S}}^{x_0} \triangleq \mu_{\mathcal{S} \oplus \{-x_0\}}$.*

Example 4.1 (p-Ball [53])**.** *The Minkowski functional of $\mathcal{B}_p \subset \mathbb{R}^n$ the unit ball of norm $p \in \tilde{\mathbb{R}}_{\geq 1}$ centered at the origin is given by:*

$$\mu_{\mathcal{B}_p}(x) = \|x\|_p \tag{22}$$

where $\|x\|_p$ stands for the p-norm of x.

Example 4.2 (Ellipsoid [8]). *The Minkowski functional of an ellipsoid $\mathcal{E} \subset \mathbb{R}^n$ centered at the origin is given by:*

$$\mu_{\mathcal{E}}(x) = \mu_{M\mathcal{B}_2}(x) = \sqrt{x^\top Q x} \tag{23}$$

with $M \in \mathbb{R}^{n\times m}$ a full row rank matrix, and $Q \in \mathbb{S}_n^{++}(\mathbb{R})$, the positive definite matrix given by $Q = (MM^\top)^{-1}$.

Example 4.3 (Zonotope [3]). *The Minkowski functional of a zonotope $\mathcal{Z} \subset \mathbb{R}^n$ centered at the origin is given by:*

$$\mu_{\mathcal{Z}}(x) = \mu_{G\mathcal{B}_\infty}(x) = \|Lx\|_\infty \tag{24}$$

with $G \in \mathbb{R}^{n\times m}$ a full row rank matrix, and $L \in \mathbb{R}^{p\times n}$ the matrix whose $p = \binom{m}{n-1}$ rows are given by $l_{\mathcal{I}}/d_{\mathcal{I}}$ for all $\mathcal{I} \subseteq [\![1, m]\!]$ with $\#\mathcal{I} = m - n + 1$, and such that

$$l_{\mathcal{I}} = \frac{\times^n(G^{\langle\overline{\mathcal{I}}\rangle})^\top}{\|\times^n(G^{\langle\overline{\mathcal{I}}\rangle})\|_2} \tag{25a}$$

$$d_{\mathcal{I}} = \|l_{\mathcal{I}} G\|_1 \tag{25b}$$

with $G^{\langle\overline{\mathcal{I}}\rangle}$ the matrix G taken without its columns of index in $\mathcal{I}$, and $\times^n$ the generalized vector product defined by:

$$\forall H \in \mathbb{R}^{n\times(n-1)},\ \times^n(H) \triangleq \begin{pmatrix} \dots & (-1)^{i+1}\det H^{[i]} & \dots \end{pmatrix}^\top \tag{26}$$

where $H^{[i]}$ stands for the sub-matrix defined by H without its i-th row.

Remark 4.3. *A parallelotope is a zonotope where $m = n$. An orthotope is a zonotope where $G = I_n$.*

Example 4.4 (Polytope [23]). *Given $\mathcal{P} \subset \mathbb{R}^n$ a compact and convex polytope whose interior contains the origin and whose halfspace-representation is given by:*

$$\mathcal{P} = \{x \in \mathbb{R}^n : \forall k \in [\![1, m]\!],\ \langle h_k | x\rangle \leq 1\} \tag{27}$$

with $(h_k)_{1\leq k\leq m}$ as set of m vectors of $\mathbb{R}^n$, then, the Minkowski functional of $\mathcal{P}$ is given by:

$$\mu_{\mathcal{P}}(x) = \max_{k\in[\![1,m]\!]} \langle h_k | x\rangle \tag{28}$$

Remark 4.4. *As discussed in [55], it is possible to extract the polytopic halfspace-representation of constrained zonotopes, and it is therefore also possible to obtain their Minkowski functional.*

Remark 4.5. *At their final set-based evaluation, hybrid zonotopes can be considered as the union of a finite number of constrained zonotopes. As such, it is also possible to obtain an explicit expression of their Minkowksi functional [16, 7]. It is the minimum of all the Minkowski functionals associated with the constrained zonotopes of this union (see Property 4.3 and Remark 4.6).*

Example 4.5 (Radially parameterized set). *Let $\mathcal{O} \subset \mathbb{R}^n$ be a star-convex set at 0 whose boundary can be parameterized radially by the map $\rho : [0, \pi)^{n-2} \times [0, 2\pi) \to \mathbb{R}_{>0}$ such that*

$$\partial\mathcal{O} = \left\{ \rho(\varphi) \begin{pmatrix} \cos(\varphi_1) \\ \sin(\varphi_1)\cos(\varphi_2) \\ \vdots \\ \sin(\varphi_1)\dots\sin(\varphi_{n-2})\cos(\varphi_{n-1}) \\ \sin(\varphi_1)\dots\sin(\varphi_{n-2})\sin(\varphi_{n-1}) \end{pmatrix} : \varphi \in [0, \pi)^{n-2} \times [0, 2\pi) \right\} \tag{29}$$

with $\varphi \triangleq (\varphi_1, \dots, \varphi_{n-1})$. The Minkowski functional of $\mathcal{O}$ is given by:

$$\mu_{\mathcal{O}}(x) = \begin{cases} \|x\|_2/\rho(\varphi(x)) & \text{if } x \neq 0 \\ 0 & \text{else} \end{cases} \tag{30}$$

where $\varphi(x) = (\varphi_1(x), \dots, \varphi_{n-1}(x))$ *stands for the following angular coordinates of* x*:*

$$\varphi_k(x) = \arccos \frac{x_k}{\sqrt{\sum_{i=k}^{n} x_i^2}}, \qquad \forall k \in [\![1, n-2]\!] \tag{31a}$$

$$\varphi_{n-1}(x) = 2 \operatorname{arccot} \frac{x_{n-1} + \sqrt{x_n^2 + x_{n-1}^2}}{x_n} \tag{31b}$$

Proof. It is easily verified that $\mu_{\mathcal{O}}(0) = 0$. Moreover, for all $x \in \mathbb{R}^n \setminus \{0\}$, the following equalities stand:

$$\begin{aligned} \mu_{\mathcal{O}}(x) &= \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{O}\} \\ &= \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\rho(\varphi(x))\mathcal{B}_2\} \\ &= \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : \frac{1}{t\rho(\varphi(x))} x \in \mathcal{B}_2\} \\ &= \mu_{\mathcal{B}_2}(x/\rho(\varphi(x))) = \|x\|_2/\rho(\varphi(x)) \end{aligned} \tag{32}$$

This concludes the proof. □

4.2. Properties

First, the classical properties on the Minkowski functional associated with star-convex and convex sets are recalled. In particular, when $\mathcal{S}$ is a star-convex set at 0 (i.e. $t\mathcal{S} \subseteq \mathcal{S}$ for all $t \in [0, 1]$ [53]) of $\mathbb{R}^n$ whose interior contains 0, checking if $x \in \mathbb{R}^n$ belongs - in a loose sense - to $t\mathcal{S}$, can simply be achieved by verifying the inequality $\mu_{\mathcal{S}}(x) \leq t$. This fact is formalized, among others, in the two properties below.

Property 4.1 (Star-convex properties [52, 51])**.** *Let* $\mathcal{S}$ *be a bounded star-convex set at* 0 *of* $\mathbb{R}^n$ *whose interior contains* 0*. The Minkowski functional* $\mu_{\mathcal{S}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ *associated with* $\mathcal{S}$ *satisfies:*

1. *For all* $x \in \mathbb{R}^n$, $0 \leq \mu_{\mathcal{S}}(x) < +\infty$,
2. *For all* $x \in \mathbb{R}^n$ *and* $t \in \mathbb{R}_{\geq 0}$, $\mu_{\mathcal{S}}(tx) = t\mu_{\mathcal{S}}(x)$, *and* $\mu_{\mathcal{S}}(tx) = \mu_{\frac{1}{t}\mathcal{S}}(x)$ *if* $t \neq 0$,
3. $\mu_{\mathcal{S}}^{-1}([0, 1)) \subseteq \mathcal{S} \subseteq \mu_{\mathcal{S}}^{-1}([0, 1])$.

Property 4.2 (Convex properties [33])**.** *Let* $\mathcal{S}$ *be a convex set whose interior contains* 0*. The Minkowski functional* $\mu_{\mathcal{S}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ *associated with* $\mathcal{S}$ *satisfies:*

1. *For all* $x_1, x_2 \in \mathbb{R}^n$, $\mu_{\mathcal{S}}(x_1 + x_2) \leq \mu_{\mathcal{S}}(x_1) + \mu_{\mathcal{S}}(x_2)$,
2. $\mu_{\mathcal{S}} \in \mathcal{C}^0(\mathbb{R}^n, \mathbb{R}_{\geq 0})$,
3. $\mu_{\mathcal{S}}^{-1}([0, 1)) = \operatorname{intr}(\mathcal{S})$, $\mu_{\mathcal{S}}^{-1}([0, 1]) = \operatorname{cls}(\mathcal{S})$, $\mu_{\mathcal{S}}^{-1}(\{1\}) = \partial\mathcal{S}$.

Item 1 hereabove combined with item 2 of Property 4.1 makes $\mu_{\mathcal{S}}$ *a convex function.*

For all non-empty sets $\mathcal{S}_1$ and $\mathcal{S}_2$, it is easily established from Definition 4.1 that $\mathcal{S}_1 \subseteq \mathcal{S}_2$ provides $\mu_{\mathcal{S}_2} \leq \mu_{\mathcal{S}_1}$, however $\mathcal{S}_1$ and $\mathcal{S}_2$ need to be star-convex at 0 for the reciprocal to hold (this is a particular case of (33b) stated below). Moreover, the usual set operations of union, intersection and Minkowski sum can also be expressed in terms of Minkowski functionals as follows.

Property 4.3 (Usual set operations)**.** *Given* M *a linear map from* $\mathbb{R}^n$ *to* $\mathbb{R}^m$ *in matrix form, and* $\mathcal{S}_1$, $\mathcal{S}_2$ *two bounded star-convex sets at* 0 *of* $\mathbb{R}^n$ *and* $\mathbb{R}^m$ *respectively and whose interior contains* 0*. For all* $x \in \mathbb{R}^n$*, the following identities hold:*

$$\mu_{\mathcal{S}_1 \cup_M \mathcal{S}_2}(x) = \min\{\mu_{\mathcal{S}_1}(x), \mu_{\mathcal{S}_2}(Mx)\} \tag{33a}$$

$$\mu_{\mathcal{S}_1 \cap_M \mathcal{S}_2}(x) = \max\{\mu_{\mathcal{S}_1}(x), \mu_{\mathcal{S}_2}(Mx)\} \tag{33b}$$

$$\mu_{\mathcal{S}_1 \oplus_M \mathcal{S}_2}(x) = \inf_{x_1 + x_2 = x} \{\max\{\mu_{\mathcal{S}_1}(x_1), \mu_{\mathcal{S}_2}(Mx_2)\}\} \tag{33c}$$

Given $\star \in \{\cup, \cap, \oplus\}$*, then the set operation* $\mathcal{S}_1 \star_M \mathcal{S}_2$ *stands for* $\mathcal{S}_1 \star \{x \in \mathbb{R}^n : Mx \in \mathcal{S}_2\}$.

Proof. Given any set operation $\star \in \{\cup, \cap, \oplus\}$, by definition of $\mathcal{S}_1 \star_M \mathcal{S}_2$, and since $\mu_{\{x \in \mathbb{R}^n : Mx \in \mathcal{S}_2\}} = \mu_{\mathcal{S}_2} \circ M$ holds, then, the proof reduces to showing that (33) holds when $m = n$ and $M = I_n$. Note that (33a) holds even if $\mathcal{S}_1$ and $\mathcal{S}_2$ are simply non-empty sets without further assumptions: this follows from Definition 4.1 and from $\inf \cup = \min \inf$ in this context (see Exercise 1.5 from [60], which can be solved by contradiction):

$$\inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t(\mathcal{S}_1 \cup \mathcal{S}_2)\} = \inf \cup_{i \in \{1,2\}} \{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{S}_i\} = \min_{i \in \{1,2\}} \inf\{t \in \tilde{\mathbb{R}}_{\geq 0} : x \in t\mathcal{S}_i\} \tag{34}$$

The proof of (33b) can be found in Lemma 5.49, page 192 of [2]. Finally, the proof of (33c) can be found in Lemma 1, page 9 of [30] in the convex and balanced case, but it is easily noticed that the proof only leverages Property 4.1, hence Lemma 1 at page 9 of [30] also holds in the star-convex case. □

Remark 4.6. *The identities (33a), (33b) and (33c) can easily be generalized to finite collections $(\mathcal{S}_i)_{1 \leq i \leq m}$ of bounded star-convex sets at 0 of $\mathbb{R}^n$ and whose interior contains 0:*

$$\mu_{\cup_{i=1}^m \mathcal{S}_i}(x) = \min_{1 \leq i \leq m} \mu_{\mathcal{S}_i}(x) \tag{35a}$$

$$\mu_{\cap_{i=1}^m \mathcal{S}_i}(x) = \max_{1 \leq i \leq m} \mu_{\mathcal{S}_i}(x) \tag{35b}$$

$$\mu_{\oplus_{i=1}^m \mathcal{S}_i}(x) = \inf_{x_1 + \cdots + x_m = x} \left\{ \max_{1 \leq i \leq m} \mu_{\mathcal{S}_i}(x_i) \right\} \tag{35c}$$

The identity (33a) can be extended to uncountable collections $(\mathcal{S}_j)_{j \in [0,1]}$ simply by replacing the $\min_{1 \leq i \leq m}$ by an $\inf_{j \in [0,1]}$. The existence of this inf *is easily deduced by the fact that the Minkowski functional is bounded from below by 0, and the deduction of this identity follows the same steps as (34), by leveraging $\inf \cup = \inf \inf$ in this context (using a similar proof by contradiction).*

Among the previous set operations, the Minkowski functional associated with the Minkowski sum (33c) is impractical for direct computation since it combines both an inf and a max in its expression. However, elementary upper- and lower-bounds can be retrieved respectively by evaluating the max at specific values of x_i (hence upper-bounding the inf), or by determining the inf of the average of all of the $\mu_{\mathcal{S}_i}(x_i)$ (hence lower-bounding the max). The authors also suggest in Property 4.4 some less straightforward inequalities to bound the value of (33c).

Property 4.4 (Minkowski sum of convex sets)**.** *Let $(\mathcal{S}_i)_{1 \leq i \leq m}$ be a collection of bounded star-convex sets at 0 of $\mathbb{R}^n$ whose interior contains 0. For all $x \in \mathbb{R}^n \setminus \{0\}$, the following holds:*

$$\frac{1}{m} \cdot \mu_{\text{hull}(\cup_{i=1}^m \mathcal{S}_i)}(x) \leq \mu_{\oplus_{i=1}^m \mathcal{S}_i}(x) \leq \frac{1}{\sum_{i=1}^m \frac{1}{\mu_{\mathcal{S}_i}(x)}} \tag{36}$$

which can be lower-bounded again using $\max_{1 \leq i \leq m}\{s_i \mu_{\mathcal{S}_i}(x)\} \leq \mu_{\text{hull}(\cup_{i=1}^m \mathcal{S}_i)}(x)$, where the $(s_i)_{1 \leq i \leq m} \in \mathbb{R}^m_{>0}$ are taken such that $\text{hull}(\cup_{i=1}^m \mathcal{S}_i) \subseteq \cap_{i=1}^m \frac{1}{s_i} \mathcal{S}_i$.

Proof. $\boxed{\mu_{\oplus} \leq \dots}$ Let $x \in \mathbb{R}^n \setminus \{0\}$. The $\mathcal{S}_i$ being star-convex at 0, their Minkowski sum is also star-convex at 0. Hence given $(t_i)_{1 \leq i \leq m} \in \mathbb{R}^m_{>0}$ taken such that $t_i x \in \mathcal{S}_i$, then for all $0 \leq \lambda \leq (\sum_{i=1}^m t_i)$, $\lambda x \in \oplus_{i=1}^m \mathcal{S}_i$ is verified by definition of the Minkowski sum. This provides the following lower-bound:

$$\sup \oplus_{i=1}^m \{t \in \mathbb{R}_{>0}, tx \in \mathcal{S}_i\} \leq \sup\{t \in \mathbb{R}_{>0}, tx \in \oplus_{i=1}^m \mathcal{S}_i\} \tag{37}$$

Note that $\sup \oplus = \sum \sup$ in this context [42], hence the expression above can be rewritten as [37]:

$$\sum_{i=1}^m \frac{1}{\inf\{t \in \mathbb{R}_{>0}, x \in t\mathcal{S}_i\}} \leq \frac{1}{\inf\{t \in \mathbb{R}_{>0}, x \in t \oplus_{i=1}^m \mathcal{S}_i\}} \tag{38}$$

Recognizing the Definition 4.1 of the Minkowski functional in the expression above provides the upper-bound in (36).

$\boxed{\cdots \leq \mu_{\oplus}}$ Since for all i, $\mathcal{S}_i \subseteq \text{hull}(\bigcup_{j=1}^m \mathcal{S}_j)$, the following upper-bound holds:

$$\sup\{t \in \mathbb{R}_{>0}, tx \in \oplus_{i=1}^m \mathcal{S}_i\} \leq \sup\{t \in \mathbb{R}_{>0}, tx \in \oplus_{i=1}^m \text{hull}(\cup_{i=1}^m \mathcal{S}_i)\} \tag{39}$$

Moreover, since $\oplus_{i=1}^m \text{hull}(\bigcup_{i=1}^m \mathcal{S}_i) = m \cdot \text{hull}(\bigcup_{i=1}^m \mathcal{S}_i)$, the following holds:

$$\frac{1}{\inf\{t \in \mathbb{R}_{>0}, x \in t \oplus_{i=1}^m \mathcal{S}_i\}} \leq \frac{m}{\inf\{t \in \mathbb{R}_{>0}, x \in t\, \text{hull}(\cup_{i=1}^m \mathcal{S}_i)\}} \tag{40}$$

Recognizing the Definition 4.1 of the Minkowski functional provides the lower-bound in (36). Finally, since all $\mathcal{S}_i$ contain 0 in their interior, for all i there exists $\varepsilon_i > 0$ such that $\varepsilon_i \mathcal{B}_2 \subseteq \mathcal{S}_i$, meaning there exists $s_i \in \mathbb{R}_{>0}$ such that $\text{hull}(\cup_{i=1}^m \mathcal{S}_i) \subseteq \frac{1}{s_i}\varepsilon_i \mathcal{B}_2 \subseteq \frac{1}{s_i}\mathcal{S}_i$. Given $\text{hull}(\cup_{i=1}^m \mathcal{S}_i) \subseteq \cap_{i=1}^m \frac{1}{s_i}\mathcal{S}_i$, the straightforward generalization of (33b) to m intersections combined with item 2 of Property 4.1 provides $\max_{1\leq i\leq m}\{s_i \mu_{\mathcal{S}_i}(x)\} \leq \mu_{\text{hull}(\cup_{i=1}^m \mathcal{S}_i)}(x)$. □

Although these inner- and outer-approximations may not always be practically handled, the Minkowski functional of a Minkowski sum can still be obtained for special classes of convex sets for which dedicated algorithms already exist (e.g. constrained zonotopes [55, 45]). Moreover, if the Minkowski sum is difficult to handle, the usage of the Minkowski functional can also lead to advantageous results for special classes of convex set, as demonstrated hereafter on ellipsoids.

Property 4.5 (Minkowski sum of two ellipsoids)**.** *Let $\mathcal{E}(Q_1) \triangleq \{x \in \mathbb{R}^n : x^\top Q_1 x \leq 1\}$ and $\mathcal{E}(Q_2)$ be defined similarly, with $Q_1, Q_2 \in \mathbb{S}_n^{++}(\mathbb{R})$. The eigenvalues of the matrix pencil (Q_2^{-1}, Q_1^{-1}), solutions of the equation*

$$\det(Q_1^{-1} - \lambda Q_2^{-1}) = 0 \tag{41}$$

are denoted as:

$$\underline{\lambda} = \lambda_1 \leq \cdots \leq \lambda_n = \overline{\lambda} \tag{42}$$

Let $\mathcal{I}$ be a finite subset of $[\underline{\lambda}, \overline{\lambda}]$. The following inequality holds for all $x \in \mathbb{R}^n$:

$$\max_{p\in\mathcal{I}} \mu_{\mathcal{E}(Q(p))}(x) \leq \mu_{\mathcal{E}(Q_1)\oplus\mathcal{E}(Q_2)}(x) \tag{43}$$

where

$$Q(p) \triangleq \left((1+p^{-1})Q_1^{-1} + (1+p)Q_2^{-1}\right)^{-1} \tag{44}$$

This outer-approximation can be made arbitrarily precise by increasing the number of points in $\mathcal{I}$.

Proof. Theorem 2.2.3 at page 118 of [31] states that:

$$\mathcal{E}(Q_1) \oplus \mathcal{E}(Q_2) = \bigcap_{p\in[\underline{\lambda},\overline{\lambda}]} \mathcal{E}(Q(p)) \tag{45}$$

hence:

$$\mathcal{E}(Q_1) \oplus \mathcal{E}(Q_2) \subseteq \bigcap_{p\in\mathcal{I}} \mathcal{E}(Q(p)) \tag{46}$$

The result is then obtained by applying (35b) to the right-hand side of the inclusion (46). □

Finally, the Minkowski functional of a linearly transformed set is given below.

Property 4.6 (Linear transformation)**.** *Given M a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ in matrix form and $\mathcal{S}$ a non-empty set of $\mathbb{R}^n$, if M has a left inverse, then for all $x \in \text{Im}(M)$:*

$$\mu_{M\mathcal{S}}(x) = \mu_{\mathcal{S}}(M^\dagger x) \tag{47}$$

with $M^\dagger$ the Moore–Penrose inverse of M, such that $M^\dagger M = I_n$.

Proof. The linear transformation M is injective, hence $M^\dagger$ is the inverse of M on $\text{Im}(M)$. Moreover, for all $t \in \mathbb{R}_{\geq 0}$, $tM\mathcal{S} \subseteq \text{Im}(M)$, hence (47) holds for all $x \in \text{Im}(M)$. □

Other practical characterizations of the linear transformation of a convex set are obtained thereafter. This manuscript focuses on the non-invertible linear case since, to the authors' knowledge, practical characterizations of $\mu_{M\mathcal{S}}$ are not found in the literature of the Minkowski functional.

4.3. Linear transformation of smooth convex sets

In order to generalize Property 4.6, this paper provides two characterizations for the Minkowski functional of a linearly transformed smooth compact and convex set $\mathcal{S}$ of $\mathbb{R}^n$ whose interior contains the origin. The first characterization is obtained for strictly convex sets with the help of the Legendre transform, which is a common tool in convex analysis. The second characterization is obtained using a result on the orthogonal projection of convex set onto the linear subspaces of $\mathbb{R}^n$ [4]. From now on, M is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, k is defined by $k \triangleq \dim(\mathrm{Im}(M)) = \dim(\mathrm{Ker}(M)^\perp)$ with the assumption that $k \geq 1$ (indeed, $M\mathcal{S} = \{0\}$ if $k = 0$, which is not an interesting case to consider).

Definition 4.2 (Legendre transform [53])**.** *Given a closed and convex function $f : \mathbb{R}^n \to \tilde{\mathbb{R}}$, the Legendre transform of f is given by the convex function $f^* : \mathbb{R}^n \to \tilde{\mathbb{R}}$ such that:*

$$f^*(y) \triangleq \sup_{x \in \mathbb{R}^n} \left(\langle y|x\rangle - f(x)\right) \tag{48}$$

Theorem 4.1 (First characterization)**.** *Let $\mathcal{S}$ be a smooth compact and strictly convex set of $\mathbb{R}^n$ whose interior contains 0, and $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a bijective strictly increasing and strictly convex function such that $g \circ \mu_{\mathcal{S}}$ is differentiable over $\mathbb{R}^n$. Let $f_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $f_2 : \mathbb{R}^m \to \mathbb{R}^m$ be two functions defined by*

$$f_1 = \frac{\partial}{\partial x}[g \circ \mu_{\mathcal{S}}] \tag{49a}$$

$$f_2 = M \circ \frac{\partial}{\partial x}[(g \circ \mu_{\mathcal{S}})^*] \circ M^\top \tag{49b}$$

Then, f_1 is invertible on $f_1(\mathbb{R}^n)$, and for all $y \in \frac{\partial}{\partial x} f_1(\mathbb{R}^n)$

$$(g \circ \mu_{\mathcal{S}})^*(y) = \left\langle f_1^{-1}(y)|y\right\rangle - g \circ \mu_{\mathcal{S}} \circ f_1^{-1}(y) \tag{50}$$

From here, if f_2 is invertible on $f_2(\mathbb{R}^m)$, it is possible to retrieve the expression of $\mu_{M\mathcal{S}}$ by computing

$$\mu_{M\mathcal{S}}(y) = g^{-1}(\left\langle f_2^{-1}(y)|y\right\rangle - (g \circ \mu_{\mathcal{S}})^* \circ M^\top \circ f_2^{-1}(y)) \tag{51}$$

which offers a characterization of the linear transformation of $\mathcal{S}$ by M that - in spite of its apparent complexity - can sometimes be treated algebraically. The problem is reduced to finding a suitable g and an inverse for f_1 and f_2.

Proof. First it is proven that $g \circ \mu_{\mathcal{S}}$ is a strictly convex function. Let $x_1, x_2 \in \mathbb{R}^n$ such that $x_1 \neq x_2$ and $t \in (0, 1)$. On one hand, if $\mu_{\mathcal{S}}(x_1) = \mu_{\mathcal{S}}(x_2) = \lambda$, then necessarily $x_1, x_2 \neq 0$, and by strict convexity of $\mathcal{S}$:

$$\begin{aligned} tx_1 + (1-t)x_2 \in \mathrm{intr}(\lambda\mathcal{S}) &\Leftrightarrow \mu_{\lambda\mathcal{S}}(tx_1 + (1-t)x_2) < 1 \\ &\Leftrightarrow \mu_{\mathcal{S}}(tx_1 + (1-t)x_2) < \lambda \end{aligned} \tag{52}$$

Since g is a strictly increasing function, $(g \circ \mu_{\mathcal{S}})(tx_1 + (1-t)x_2) < g(\lambda) = tg(\lambda) + (1-t)g(\lambda)$, hence

$$(g \circ \mu_{\mathcal{S}})(tx_1 + (1-t)x_2) < t(g \circ \mu_{\mathcal{S}})(x_1) + (1-t)(g \circ \mu_{\mathcal{S}})(x_2) \tag{53}$$

On the other hand, if $\mu_{\mathcal{S}}(x_1) \neq \mu_{\mathcal{S}}(x_2)$, then by convexity of $\mu_{\mathcal{S}}$ and from the fact that g is increasing, the following holds: $(g \circ \mu_{\mathcal{S}})(tx_1 + (1-t)x_2) \leq g(t\mu_{\mathcal{S}}(x_1) + (1-t)\mu_{\mathcal{S}}(x_2))$. Finally by strict convexity of g, it follows:

$$(g \circ \mu_{\mathcal{S}})(tx_1 + (1-t)x_2) < t(g \circ \mu_{\mathcal{S}})(x_1) + (1-t)(g \circ \mu_{\mathcal{S}})(x_2) \tag{54}$$

Hence $g \circ \mu_{\mathcal{S}}$ is a strictly convex function. The rest of the proof is based on explicitly calculating the right-hand side of the following identity (the Legendre transform of the identity found at page 142-143 of [50]):

$$g \circ \mu_{M\mathcal{S}} = ((g \circ \mu_{\mathcal{S}})^* \circ M^\top)^* \tag{55}$$

Given $f : \mathbb{R}^n \to \mathbb{R}$ a strictly convex and differentiable function, $\frac{\partial}{\partial x} f$ is invertible on $\frac{\partial}{\partial x} f(\mathbb{R}^n)$, and for all $y \in \frac{\partial}{\partial x} f(\mathbb{R}^n)$, the Legendre transform of f at y is given by (see Remark 1.6.18 of [53]):

$$f^*(y) = \left\langle (\frac{\partial}{\partial x} f)^{-1}(y) | y \right\rangle - f \circ (\frac{\partial}{\partial x} f)^{-1}(y) \tag{56}$$

Applying the previous result to $g \circ \mu_{\mathcal{S}}$ provides:

$$(g \circ \mu_{\mathcal{S}})^*(y) = \left\langle f_1^{-1}(y) | y \right\rangle - g \circ \mu_{\mathcal{S}} \circ f_1^{-1}(y) \tag{57}$$

The result can be applied again, this time to $(g \circ \mu_{\mathcal{S}})^* \circ M^\top$. Under the assumption that $(g \circ \mu_{\mathcal{S}})^* \circ M^\top$ is strictly convex and differentiable, then $f_2 = \frac{\partial}{\partial x}((g \circ \mu_{\mathcal{S}})^* \circ M^\top) = M \circ \frac{\partial}{\partial x}[(g \circ \mu_{\mathcal{S}})^*] \circ M^\top$ is invertible on $f_2(\mathbb{R}^m)$, and the identity (55) can be written:

$$\begin{aligned} g \circ \mu_{M\mathcal{S}} &= ((g \circ \mu_{\mathcal{S}})^* \circ M^\top)^* \\ &= \left\langle (\frac{\partial}{\partial x}[(g \circ \mu_{\mathcal{S}})^* \circ M^\top])^{-1}(y) | y \right\rangle - (g \circ \mu_{\mathcal{S}})^* \circ M^\top \circ (\frac{\partial}{\partial x}[(g \circ \mu_{\mathcal{S}})^* \circ M^\top])^{-1}(y) \\ &= \left\langle (M \circ \frac{\partial}{\partial x}[g \circ \mu_{\mathcal{S}}]^* \circ M^\top)^{-1}(y) | y \right\rangle - (g \circ \mu_{\mathcal{S}})^* \circ M^\top \circ (M \circ \frac{\partial}{\partial x}[g \circ \mu_{\mathcal{S}}]^* \circ M^\top)^{-1}(y) \end{aligned} \tag{58}$$

which concludes the proof. □

Remark 4.7. *This first result is a broad generalization of the well-known result on ellipsoids which states that* $M\mathcal{E}(P) = \mathcal{E}((MP^{-1}M^\top)^{-1})$*, where* $\mathcal{E}(P) \triangleq \{x \in \mathbb{R}^n : x^\top P x \leq 1\}$ *and* M *is a full row rank matrix.*

Example 4.6. *Consider the Minkowski functional associated to* $\mathcal{B}_{4/3}$*, the unit ball of norm* $4/3$ *of* $\mathbb{R}^3$*:*

$$\mu_{\mathcal{B}_{4/3}}(x, y, z) \triangleq \left(x^{4/3} + y^{4/3} + z^{4/3} \right)^{3/4} \tag{59}$$

and let us compute the Minkowski functional associated to $M\mathcal{B}_{4/3}$ *where* M *is defined by:*

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \tag{60}$$

Since $\mathcal{B}_{4/3}$ *is a smooth compact and strictly convex set, Theorem 4.1 is applied.*

The function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *defined by* $g(s) = (3/4)s^{4/3}$ *is introduced. It is easily verified that* g *is bijective (with* $g^{-1}(s) = \left(2/\sqrt{3}\right)^{3/2} s^{3/4}$*), strictly increasing, strictly convex, and that* $g \circ \mu_{\mathcal{B}_{4/3}}$ *is smooth over* $\mathbb{R}^3$*. Following Theorem 4.1,* $f_1 = \frac{\partial}{\partial x}[g \circ \mu_{\mathcal{B}_{4/3}}]$ *and its inverse are given by:*

$$f_1(x, y, z) = \begin{pmatrix} x^{1/3} & y^{1/3} & z^{1/3} \end{pmatrix}^\top \tag{61a}$$

$$f_1^{-1}(x, y, z) = \begin{pmatrix} x^3 & y^3 & z^3 \end{pmatrix}^\top \tag{61b}$$

This provides the Legendre transform of $g \circ \mu_{\mathcal{B}_{4/3}}$*:*

$$(g \circ \mu_{\mathcal{B}_{4/3}})^*(x, y, z) = \frac{1}{4}(x^4 + y^4 + z^4) \tag{62}$$

Again, following Theorem 4.1, $f_2 = M \circ \frac{\partial}{\partial x}[(g \circ \mu_{\mathcal{B}_{4/3}})^*] \circ M^\top$ *and its inverse are given by:*

$$f_2(x, y) = \begin{pmatrix} 2(x+y)^3 + (x-y)^3 \\ 2(x+y)^3 - (x-y)^3 \end{pmatrix} \tag{63a}$$

$$f_2^{-1}(x, y) = \begin{pmatrix} 2^{-4/3} \left(2^{-1/3}(x+y)^{1/3} + (x-y)^{1/3}\right) \\ 2^{-4/3} \left(2^{-1/3}(x+y)^{1/3} - (x-y)^{1/3}\right) \end{pmatrix} \tag{63b}$$

Finally, the Minkowski functional associated to $M\mathcal{B}_{4/3}$ is given by:

$$\mu_{M\mathcal{B}_{4/3}}(x,y) = \left(\frac{2}{\sqrt{3}}\right)^{3/2} \left(\frac{x}{2^{4/3}}\left(\frac{(x+y)^{1/3}}{2^{1/3}} + (x-y)^{1/3}\right) + \frac{y}{2^{4/3}}\left(\frac{(x+y)^{1/3}}{2^{1/3}} - (x-y)^{1/3}\right) - \frac{1}{2^{10/3}}\left(\frac{(x+y)^{4/3}}{2^{1/3}} + (x-y)^{4/3}\right)\right)^{3/4} \tag{64}$$

Let $(v_1, \dots, v_k)$ denotes a basis of $\mathrm{Ker}(M)^\perp$ and $(v_{k+1}, \dots, v_n)$ denotes a basis of $\mathrm{Ker}(M)$. $V_{\mathrm{Ker}^\perp}$ is the matrix whose columns are formed by $(v_1, \dots, v_k)$ and V_{Ker} is the matrix whose columns are formed by $(v_{k+1}, \dots, v_n)$. The invertible matrix $V \in \mathbb{R}^{n\times n}$ is also defined by the matrix whose columns are formed by $(v_1, \dots, v_n)$, i.e. such that:

$$V = \begin{pmatrix} V_{\mathrm{Ker}^\perp} & V_{\mathrm{Ker}} \end{pmatrix} \tag{65}$$

Theorem 4.2 (Second characterization)**.** *Let $\mathcal{S}$ be a smooth compact and convex set of $\mathbb{R}^n$ whose interior contains 0. Then the following identity holds:*

$$\mu_{M\mathcal{S}}(x) = \begin{cases} \mu_{p_{\mathrm{Ker}^\perp}(\mathcal{S})}((MV_{\mathrm{Ker}^\perp})^\dagger x) & \text{if } x \in \mathrm{Im}(M) \\ +\infty & \text{else} \end{cases} \tag{66}$$

where $p_{\mathrm{Ker}^\perp}(\mathcal{S})$ denotes the orthogonal projection of $\mathcal{S}$ onto $\mathrm{Ker}(M)^\perp$. The Minkowski functional on the right-hand side can be computed using the following system:

$$\mu_{p_{\mathrm{Ker}^\perp}(\mathcal{S})}(y_{\mathrm{Ker}^\perp}) = \inf\left\{ t \in \mathbb{R}_+^* \,:\, \exists y_{\mathrm{Ker}} \in \mathbb{R}^{n-k} \mid \begin{cases} \eta_{\mathcal{S}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) \le t \\ y_{\mathrm{Ker}^\perp} + y_{\mathrm{Ker}} \neq 0 \Rightarrow \frac{\partial \eta_{\mathcal{S}}}{\partial y_{\mathrm{Ker}}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) = 0 \end{cases} \right\} \tag{67}$$

where $y_{\mathrm{Ker}^\perp} \in \mathbb{R}^k$ and $y_{\mathrm{Ker}} \in \mathbb{R}^{n-k}$ are expressed in the $(v_1, \dots, v_k)$ and $(v_{k+1}, \dots, v_n)$ basis respectively, with:

$$\eta_{\mathcal{S}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) \triangleq \mu_{\mathcal{S}}\left(V \begin{pmatrix} y_{\mathrm{Ker}^\perp} \\ y_{\mathrm{Ker}} \end{pmatrix}\right) \tag{68}$$

The right-hand side of (67) can sometimes be treated algebraically, providing a characterization of the linear transformation of $\mathcal{S}$ by M. The problem is reduced to solving the equation $\frac{\partial \eta_{\mathcal{S}}}{\partial y_{\mathrm{Ker}}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) = 0$ with y_{Ker} unknown.

Proof. By definition of $\mathrm{Ker}(M)$ and $\mathrm{Ker}(M)^\perp$, the following equality holds:

$$M = MV_{\mathrm{Ker}^\perp} \begin{pmatrix} I_k & 0 \end{pmatrix} V^{-1} \tag{69}$$

where moreover, $MV_{\mathrm{Ker}^\perp}$ is left-invertible, hence by Property 4.6, the following holds:

$$\mu_{M\mathcal{S}}(x) = \mu_{\begin{pmatrix} I_k & 0 \end{pmatrix} V^{-1}\mathcal{S}}((MV_{\mathrm{Ker}^\perp})^\dagger x) \tag{70}$$

From here, it can be noticed that $\begin{pmatrix} I_k & 0 \end{pmatrix} V^{-1}\mathcal{S}$ is the orthogonal projection of $\mathcal{S}$ onto $\mathrm{Ker}(M)^\perp$ expressed in the $(v_1, \dots, v_k)$ basis. Finally, the results from [4] provides the expression (66) and (67) when $\mathcal{S}$ is a smooth compact and convex set of $\mathbb{R}^n$ whose interior contains 0, concluding the proof. □

Example 4.7. *Consider the Minkowski functional associated to $\mathcal{B}_4$, the unit ball of norm 4 of $\mathbb{R}^2$:*

$$\mu_{\mathcal{B}_4}(x,y) \triangleq \left(x^4 + y^4\right)^{1/4} \tag{71}$$

and let us compute the Minkowski functional associated to $M\mathcal{B}_4$ where M is defined by:

$$M = \begin{pmatrix} 1 & 2 \end{pmatrix} \tag{72}$$

Since $\mathcal{B}_4$ is a smooth compact and convex set, Theorem 4.2 is applied.

The matrix V associated with M can be taken to be $V = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$. Hence:

$$\eta_{\mathcal{B}_4}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) \triangleq \left(\left(y_{\mathrm{Ker}^\perp} + 2y_{\mathrm{Ker}}\right)^4 + \left(2y_{\mathrm{Ker}^\perp} - y_{\mathrm{Ker}}\right)^4\right)^{1/4} \tag{73}$$

and:

$$\frac{\partial \eta_{\mathcal{B}_4}}{\partial y_{\mathrm{Ker}}}\left(y_{\mathrm{Ker}^\perp}, y_{\mathrm{Ker}}\right) = 0 \Leftrightarrow 17y_{\mathrm{Ker}}^3 + 18y_{\mathrm{Ker}}^2 y_{\mathrm{Ker}^\perp} + 24y_{\mathrm{Ker}} y_{\mathrm{Ker}^\perp}^2 - 6y_{\mathrm{Ker}^\perp}^3 = 0 \tag{74}$$

the equation on the right possesses a single real solution given by $y_{\mathrm{Ker}} = \frac{1}{17}(-6 - 5 \cdot 2^{1/3} + 10 \cdot 2^{2/3}) y_{\mathrm{Ker}^\perp}$, moreover $(MV_{\mathrm{Ker}^\perp})^\dagger = 1/5$. Finally, the Minkowski functional associated to $M\mathcal{B}_4$ is given by:

$$\mu_{M\mathcal{B}_4}(x) = \mu_{p_{\mathrm{Ker}^\perp}(\mathcal{B}_4)}(x/5) = \eta_{\mathcal{B}_4}\left(x/5, x(-6 - 5 \cdot 2^{1/3} + 10 \cdot 2^{2/3})/85\right) \tag{75}$$

4.4. Analytic fault isolation

The flexibility and generality of the Minkowski functional for set-membership approaches is illustrated here by applying its properties to the fault isolation scheme of Theorem 3.2. In Theorem 4.3, the problem of calculating the DIT set (15) is reduced to determining analytical thresholds under the form of scalar inequalities. In Theorem 4.4, a condition on the minimal fault magnitude is given to guarantee isolability by the DIT. In both theorems, for all i, the sets $\mathcal{W}_i$ and $\mathcal{F}_i$ of Definition 3.1 are assumed to be bounded star-convex sets at 0 whose interior contains 0.

Theorem 4.3 (Analytic Direct Image Test)**.** *For all $\mathcal{I} \in \mathbf{P}_{n_f}$, the Minkowski functional associated with the feasible sets $\mathcal{R}_{\theta,\mathcal{I}}$ (13) introduced in Definition 3.2 can be computed by:*

$$\mu_{\theta,\mathcal{I}}(r) = \inf_{\substack{\Sigma r_i = r \\ i \in \mathcal{J}}} \max\left\{ \max_{1 \le i \le n_w} \mu_{W_i(\theta)\mathcal{W}_i}(r_i), \max_{i \in \mathcal{I}} \mu_{F_i(\theta)\mathcal{F}_i}(r_{n_w+i}) \right\} \tag{76}$$

with $\mathcal{J} = [\![1, n_w]\!] \cup (\{n_w\} \oplus \mathcal{I})$. Given $\underline{\mu}_{\theta,\overline{\mathcal{I}}}$ and $\overline{\mu}_{\theta,\overline{\mathcal{I}}}$ resp. a lower-bound and an upper-bound to $\mu_{\theta,\overline{\mathcal{I}}}$, the DIT set (15) can be approximated by:

$$\left\{\mathcal{I} \in \mathbf{P}_{n_f} \,:\, \sup_{\theta \in \Theta} \underline{\mu}_{\theta,\overline{\mathcal{I}}}(r(\theta)) > 1\right\} \subseteq DIT \subseteq \left\{\mathcal{I} \in \mathbf{P}_{n_f} \,:\, \sup_{\theta \in \Theta} \overline{\mu}_{\theta,\overline{\mathcal{I}}}(r(\theta)) > 1\right\} \tag{77}$$

Proof. This is simply a rewriting of Theorem 3.2 obtained by leveraging the Minkowski functional identities of Property 4.1 and Remark 4.6. □

Corollary 4.1 (Analytic Robust Direct Image Test)**.** *Similarly, the robust DIT set (18) can be approximated by:*

$$\left\{\mathcal{I} \in \mathbf{P}_{n_f} \,:\, \sup_{\theta_2 \in \Theta} \inf_{\theta_1 \in \Theta} \underline{\mu}_{\theta_1,\overline{\mathcal{I}}}(r(\theta_2)) > 1\right\} \subseteq RDIT \subseteq \left\{\mathcal{I} \in \mathbf{P}_{n_f} \,:\, \sup_{\theta_2 \in \Theta} \inf_{\theta_1 \in \Theta} \overline{\mu}_{\theta_1,\overline{\mathcal{I}}}(r(\theta_2)) > 1\right\} \tag{78}$$

As noted in Remark 3.5 and Corollary 3.1, the tests performed to obtain these two sets do not require to know which uncertainty θ is responsible for each residual in R.

Remark 4.8. *By definition, the sets on the left-hand side of (77) and (78) avoid false fault detection. Similarly, the sets on the right-hand side can guarantee the absence of faults.*

The inf in (76) comes from Minkowski sums, and can be handled either by leveraging the inequalities described before and in Property 4.4, or by using dedicated algorithms which compute Minkowski sums for special classes of convex shapes. Similarly, computing the linear transformations of the sets $\mathcal{W}_i$ and

$\mathcal{F}_i$ can be achieved either by using dedicated algorithms, or by leveraging Theorem 4.1 or 4.2. All of these approaches allow to find upper- and lower-bounds to $\mu_{\theta,\mathcal{I}}$, here denoted $\overline{\mu}_{\theta,\mathcal{I}}$ and $\underline{\mu}_{\theta,\mathcal{I}}$, and which respectively correspond to an inner- and outer-approximation of the feasible set $\mathcal{R}_{\theta,\mathcal{I}}$. These upper- and lower-bounds provide closed-form inequalities for residual thresholding, and their real-time evaluation can be viewed as the computation of *meta-residuals* synthesizing *how close is the system from the boundary of* $\mathcal{R}_{\theta,\mathcal{I}}$. The authors believe that the name *Minkowski signals* is appropriate to refer the real-time evaluation of these functions in the context of system diagnosis.

Theorem 4.4 (Minimal fault magnitude for isolation)**.** *Let* $Q_{\mathcal{I}}$ *with* $\mathcal{I} \triangleq \{i_1, \ldots, i_p\} \in \mathbf{P}_{n_f}$ *denote the following statement:*

$$Q_{\mathcal{I}} : \text{“ At least one of the } i_1\text{-th, } \ldots, \ i_p\text{-th vector fault is active”} \tag{79}$$

This statement is guaranteed to be verified (or isolated) using the DIT set (15) *if the faults* $(f_i)_{i\in\mathcal{I}}$ *satisfy the following inequality:*

$$\sup_{\theta\in\Theta} \mu_{\mathcal{S}_\theta}\left(\sum_{i\in\mathcal{I}} F_i(\theta) f_i\right) > 1 \text{ with } \mathcal{S}_\theta \triangleq \mathcal{R}_{\theta,\overline{\mathcal{I}}} \oplus \left(-\mathcal{R}_{\theta,\overline{\mathcal{I}}}\right) \tag{80}$$

Similarly, this statement is guaranteed to be verified (or isolated) using the RDIT set (18) *if the faults* $(f_i)_{i\in\mathcal{I}}$ *satisfy the following inequality:*

$$\inf_{\theta\in\Theta} \mu_{\mathcal{S}_\Theta}\left(\sum_{i\in\mathcal{I}} F_i(\theta) f_i\right) > 1 \text{ with } \mathcal{S}_\Theta \triangleq \mathcal{R}_{\Theta,\overline{\mathcal{I}}} \oplus \left(-\mathcal{R}_{\Theta,\overline{\mathcal{I}}}\right) \tag{81}$$

Proof. Since the operations of linear transformation, union and Minkowski sum preserve star-convexity at 0, $\mathcal{S}_\theta$ is star-convex at 0. Moreover, the inequality (80) implies that there exists $\theta \in \Theta$:

$$\sum_{i\in\mathcal{I}} F_i(\theta) f_i \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}} \oplus \left(-\mathcal{R}_{\theta,\overline{\mathcal{I}}}\right) \tag{82}$$

meaning there exists $\theta \in \Theta$, such that for all $w_i \in \mathcal{W}_i$ (with $i \in [\![1, n_w]\!]$) and $f_i \in \mathcal{F}_i$ (with $i \in \overline{\mathcal{I}}$),

$$\sum_{i\in\mathcal{I}} F_i(\theta) f_i + \sum_{i=1}^{n_w} W_i(\theta) w_i + \sum_{i\in\overline{\mathcal{I}}} F_i(\theta) f_i \notin \mathcal{R}_{\theta,\overline{\mathcal{I}}} \tag{83}$$

hence, by Theorem 3.2, (83) guarantees that $\mathcal{I} \in \text{DIT}$, which in turn provides the statement $Q_{\mathcal{I}}$. The proof follows the same principle in the robust case. □

5. Application

Consider the following academic example of residuals $r(t)$ following the uncertain linear internal structure of Definition 3.1:

$$\begin{pmatrix} r_1(t) \\ r_2(t) \end{pmatrix} = \begin{pmatrix} 1 & \theta_1(t) & 0 \\ 0 & 1 & \theta_2(t) \end{pmatrix} \begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \end{pmatrix} + \begin{pmatrix} 1 & \theta_1(t) \\ 0 & \varepsilon \end{pmatrix} \begin{pmatrix} f_{1,1}(t) \\ f_{1,2}(t) \end{pmatrix} + \begin{pmatrix} \varepsilon \\ 1 \end{pmatrix} f_2(t) \tag{84a}$$

$$\text{i.e. } r(t) = W(\theta(t))w(t) + F_1(\theta(t))f_1(t) + F_2(\theta(t))f_2(t) \tag{84b}$$

where $r(t)$ are the residuals of the system, $\theta(t) \in \Theta$ are time-varying parametric uncertainties, $f_1(t) \in \mathcal{F}_1$ and $f_2(t) \in \mathcal{F}_2$ are potential faults to detect, and $w(t) \in \mathcal{W}$ are bounded noises affecting the residuals. This

example is chosen with matrices in $\mathbb{R}^{2\times 3}$, $\mathbb{R}^{2\times 2}$ and $\mathbb{R}^{2\times 1}$ to cover as many cases of matrix multiplication as possible. According to the set-membership approach developed in Section 3, the three following sets should be determined or approximated in order to threshold the residuals of the system:

$$\mathcal{R}_{\theta,\emptyset} \triangleq W(\theta)\mathcal{W} \tag{85a}$$
$$\mathcal{R}_{\theta,\{1\}} \triangleq W(\theta)\mathcal{W} \oplus F_1(\theta)\mathcal{F}_1 \tag{85b}$$
$$\mathcal{R}_{\theta,\{2\}} \triangleq W(\theta)\mathcal{W} \oplus F_2(\theta)\mathcal{F}_2 \tag{85c}$$

These sets can be approximated using any sort of set-membership methodologies already available in the literature, e.g. by leveraging results on orthotopes, ellipsoids, parallelotopes, zonotopes, constrained zonotopes, hybrid zonotopes, etc; or (not exclusively) by using the analytical results exposed in this paper. From here, following the analytic RDIT (Corollary 4.1) methodology of this paper, the problem consists in obtaining or approximating the following Minkowski signals:

$$\mu_{\Theta,\emptyset}(r) \triangleq \mu_{\mathcal{R}_{\Theta,\emptyset}}(r) = \inf_{\theta\in\Theta} \mu_{\mathcal{R}_{\theta,\emptyset}}(r) \qquad \text{with } \mathcal{R}_{\Theta,\emptyset} \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_{\theta,\emptyset} \tag{86a}$$
$$\mu_{\Theta,\{1\}}(r) \triangleq \mu_{\mathcal{R}_{\Theta,\{2\}}}(r) = \inf_{\theta\in\Theta} \mu_{\mathcal{R}_{\theta,\{2\}}}(r) \qquad \text{with } \mathcal{R}_{\Theta,\{2\}} \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_{\theta,\{2\}} \tag{86b}$$
$$\mu_{\Theta,\{2\}}(r) \triangleq \mu_{\mathcal{R}_{\Theta,\{1\}}}(r) = \inf_{\theta\in\Theta} \mu_{\mathcal{R}_{\theta,\{1\}}}(r) \qquad \text{with } \mathcal{R}_{\Theta,\{1\}} \triangleq \bigcup_{\theta\in\Theta} \mathcal{R}_{\theta,\{1\}} \tag{86c}$$

where:

- $\mu_{\Theta,\emptyset}(r) > 1$ guarantees that at least one of the vector faults f_1 or f_2 is active (detection);
- $\mu_{\Theta,\{1\}}(r) > 1$ guarantees that the vector fault f_1 is active (isolation);
- $\mu_{\Theta,\{2\}}(r) > 1$ guarantees that the fault f_2 is active (isolation).

The sets $\mathcal{W}$, $\mathcal{F}_1$ and $\mathcal{F}_2$ are modeled by:

$$\mathcal{W} \triangleq \mathcal{E}(Q_{w,1}) \cup G_{w,2}\mathcal{B}_\infty \tag{87a}$$
$$\mathcal{F}_1 \triangleq \mathcal{E}(Q_{f_1}) \tag{87b}$$
$$\mathcal{F}_2 \triangleq \mathcal{I}_{f_2} = [-1/\varepsilon, 1/\varepsilon] \tag{87c}$$

In order to be as illustrative as possible, these sets are selected to exhibit many of the properties developed in this paper. However, any other choice could have been made at this point. A more involved modeling of these sets would lead to better detection capabilities, but also to more elaborate expressions for the Minkowski signals, whereas simpler sets would be more easily handled, but less accurate in their detection capabilities. The expression of $\mu_{\Theta,\emptyset}(r)$ is explicitly obtained as:

$$\mu_{\Theta,\emptyset}(r) = \inf_{\theta\in\Theta} \min \left\{ \sqrt{r^\top \left(W(\theta)Q_{w,1}^{-1}W^\top(\theta)\right)^{-1} r}, \left\| L_{w,2}(\theta) r^\top \right\|_\infty \right\} \tag{88}$$

where Theorem 4.1 is leveraged to handle the linear transformation of the ellipsoid, and $L_{w,2}$ is obtained from $W(\theta)G_{w,2}$ using the results of Example 4.3.

In order to obtain an expression for $\mu_{\Theta,\{1\}}(r)$, $\mathcal{E}(Q_{w,1})$ is approximated by the zonotope $G_{w,1}\mathcal{B}_\infty$, leading to an hybrid zonotope defined by:

$$\begin{aligned} W(\theta)\mathcal{W} \oplus F_2(\theta)\mathcal{F}_2 &= W(\theta)\left(G_{w,1}\mathcal{B}_\infty \cup G_{w,2}\mathcal{B}_\infty\right) \oplus F_2(\theta)\mathcal{I}_{f_2} \\ &= \left(W(\theta)G_{w,1}\mathcal{B}_\infty \oplus F_2(\theta)\mathcal{I}_{f_2}\right) \cup \left(W(\theta)G_{w,2}\mathcal{B}_\infty \oplus F_2(\theta)\mathcal{I}_{f_2}\right) \\ &= \begin{pmatrix} W(\theta)G_{w,1} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix} \mathcal{B}_\infty \cup \begin{pmatrix} W(\theta)G_{w,2} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix} \mathcal{B}_\infty \end{aligned} \tag{89}$$

The Minkowski functional associated with this hybrid zonotope is given by:

$$\mu_{\Theta,\{1\}}(r) = \inf_{\theta\in\Theta} \min\left\{\left\|\mathcal{L}_1(\theta)r\right\|_\infty, \left\|\mathcal{L}_2(\theta)r\right\|_\infty\right\} \tag{90}$$

where $\mathcal{L}_1(\theta)$ and $\mathcal{L}_2(\theta)$ are resp. obtained from $\begin{pmatrix} W(\theta)G_{w,1} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix}$ and $\begin{pmatrix} W(\theta)G_{w,2} & \frac{1}{\varepsilon}F_2(\theta) \end{pmatrix}$ using the results of Example 4.3. By leveraging the central symmetry of the set $\mathcal{R}_{\Theta,\{2\}}$, it can be noted that Theorem 4.4 guarantees that the magnitude of the vector fault f_1 is sufficient to ensure a fault detection if the following inequality holds:

$$\inf_{\theta\in\Theta} \mu_{\Theta,\{1\}}(F_1(\theta)f_1) > 2 \tag{91}$$

Finally, in order to obtain an analytical expression to $\mu_{\Theta,\{2\}}(r)$, $G_{w,2}\mathcal{B}_\infty$ is approximated by the ellipsoid $\mathcal{E}(Q_{w,2})$, hence:

$$\begin{aligned} W(\theta)\mathcal{W} \oplus F_1(\theta)\mathcal{F}_1 &= W(\theta)\left(\mathcal{E}(Q_{w,1}) \cup \mathcal{E}(Q_{w,2})\right) \oplus F_1(\theta)\mathcal{E}(Q_{f_1}) \\ &= \left(W(\theta)\mathcal{E}(Q_{w,1}) \oplus F_1(\theta)\mathcal{E}(Q_{f_1})\right) \cup \left(W(\theta)\mathcal{E}(Q_{w,2}) \oplus F_1(\theta)\mathcal{E}(Q_{f_1})\right) \end{aligned} \tag{92}$$

Although only ellipsoids are leveraged in the representation of the initial sets, $\mathcal{R}_{\Theta,\{1\}}$ is obtained here as a generic convex shape whose Minkowski functional can be approximated arbitrarily closely by:

$$\mu_{\Theta,\{2\}}(r) = \inf_{\theta\in\Theta} \min\left\{\max_{p\in\mathcal{I}_1} \sqrt{r^\top \mathcal{Q}_1(\theta,p)r}, \max_{p\in\mathcal{I}_2} \sqrt{r^\top \mathcal{Q}_2(\theta,p)r}\right\} \tag{93}$$

where $\mathcal{Q}_1(\theta,p)$, $\mathcal{Q}_2(\theta,p)$ and $\mathcal{I}_1$ and $\mathcal{I}_2$ are obtained with Property 4.5 on the Minkowski sum of two ellipsoids.

$$\mathcal{Q}_1(\theta,p) = \left((1+p^{-1})W(\theta)Q_{w,1}^{-1}W^\top(\theta) + (1+p)F_1(\theta)Q_{f_1}^{-1}F_1(\theta)^\top\right)^{-1} \tag{94a}$$

$$\mathcal{Q}_2(\theta,p) = \left((1+p^{-1})W(\theta)Q_{w,2}^{-1}W^\top(\theta) + (1+p)F_1(\theta)Q_{f_1}^{-1}F_1(\theta)^\top\right)^{-1} \tag{94b}$$

Again, by leveraging the central symmetry of the set $\mathcal{R}_{\Theta,\{1\}}$, it can be noted that Theorem 4.4 guarantees that the magnitude of the fault f_2 is sufficient to ensure a fault detection if the following inequality holds:

$$\inf_{\theta\in\Theta} \mu_{\Theta,\{2\}}(F_2(\theta)f_2) > 2 \tag{95}$$

The simulations have been carried out with the following numerical values:

$$\begin{aligned} &\Theta = [-1,1]^2, && \theta(t) = \begin{pmatrix} \cos(t) & \sin(t) \end{pmatrix}^\top, \\ &Q_{w,1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.3 \end{pmatrix}, && Q_{w,2} = \begin{pmatrix} 0.4 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ &G_{w,1} = Q_{w,1}^{-1/2}, && G_{w,2} = Q_{w,2}^{-1/2}, \\ &Q_{f_1} = \begin{pmatrix} 0.025 & 0.01 \\ 0.01 & 1 \end{pmatrix}, && \varepsilon = 0.2, \end{aligned} \tag{96}$$

The results of the simulation can be found in Figure 6 of this document. The vector fault f_1 is active for $t \in [1,3] \cup [6,8]$, and the fault f_2 is active for $t \in [2,4] \cup [5,7]$. The Minkowski signal $\mu_{\Theta,\emptyset}(r)$ detects faults for $t \in [1,4] \cup [5,8]$ almost perfectly (with the exception of a few points). Moreover, the Minkowski signal $\mu_{\Theta,\{1\}}(r)$ perfectly isolates the fault f_1 for $t \in [1,3]$, and isolates, although less precisely, its activity in the interval $[6,8]$. Finally, the Minkowski signal $\mu_{\Theta,\{2\}}(r)$ almost perfectly isolates the fault f_2 for $t \in [2,4] \cup [5,7]$, again with the exception of a few points that could be easily dealt with by classical filtering techniques. The quality of the detection demonstrates the quality of the feasible sets underlying the Minkowski signals, which were computed by combining both the properties of this document and some existing results from the set-membership literature. Overall, for each $\theta \in \Theta$, the Minkowski signals are computed explicitly, while:

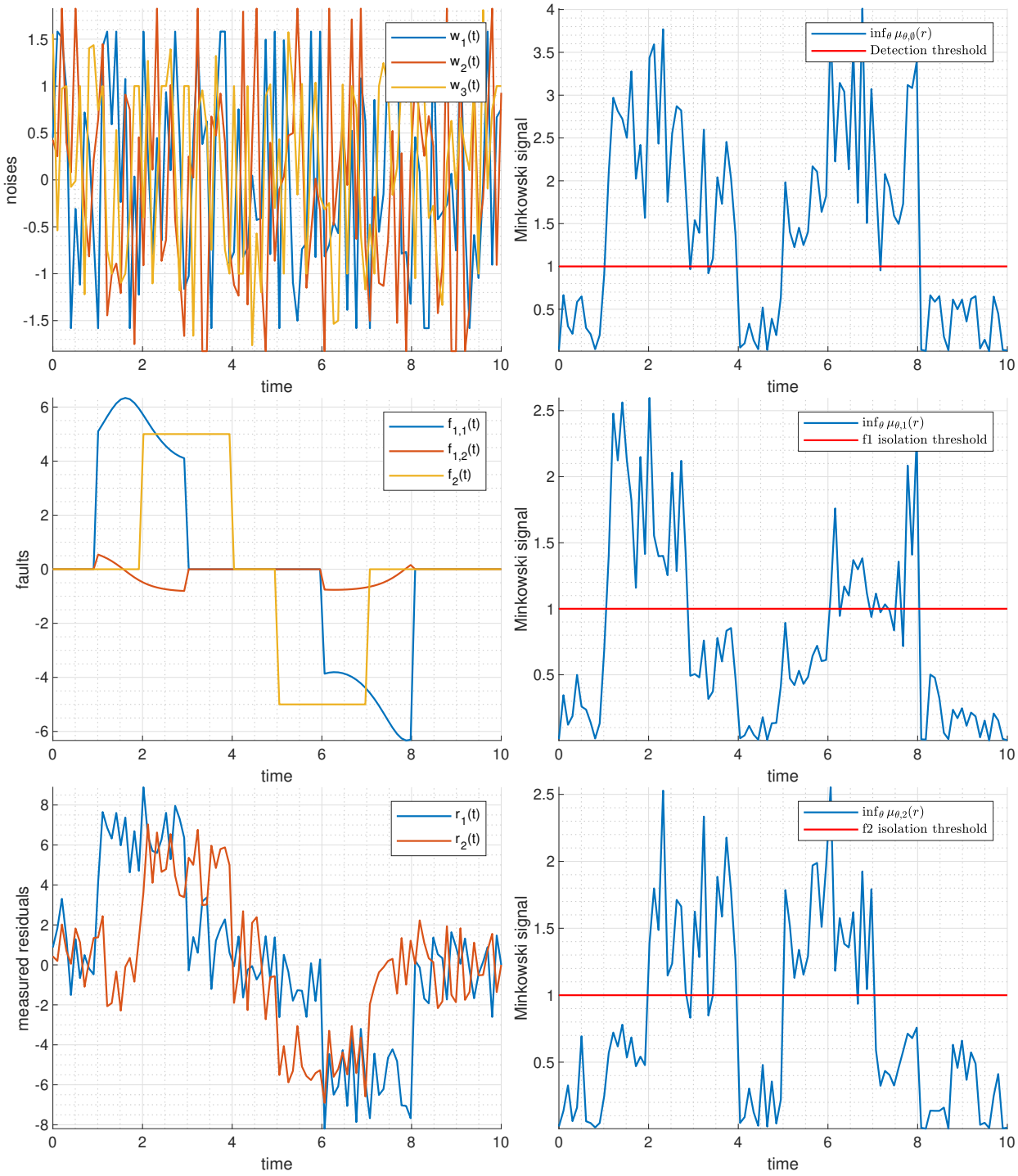

Figure 6: Computation of the Minkowski signals to perform the robust direct image.

– $\mathcal{R}_{\theta,\emptyset}$ is considered to be the union of an ellipse and a zonotope;
– $\mathcal{R}_{\theta,\{1\}}$ is considered to be the union of two zonotopes (or an hybrid zonotope);
– $\mathcal{R}_{\theta,\{2\}}$ is considered to be the Minkowski sum of two ellipsoids.

This ultimately demonstrates the unifying nature of the proposed thresholding framework.

6. Conclusions and perspectives

This paper first stated a generic set-theoretic FDI scheme to evaluate residuals following an uncertain linear internal structure. This scheme is obtained using results from order theory, which remain to be generalized to fault-trees for multi-component systems. Then, the Minkowski functional has been introduced as a novel and unified analytical thresholding tool for model-based diagnosis. Using the properties of the Minkowksi functional, including two characterizations of linear transformations of smooth convex sets, the threshold computation problem of the previously introduced fault isolation scheme has been stated in an analytical way. This led to the introduction of Minkowski signals, an intuitive measurement of the extend to which a set-membership relation is verified. An analytical expression of a minimal fault magnitude guaranteeing fault isolation has also been provided. The fault isolation scheme described in the paper has finally been illustrated by an academic example. A promising direction for future research is the integration of the Minkowski results proposed in this paper with the established set-based methodologies, such as set-based distinguishability [59]. Simplifying the expression of the Minkowski functional of a convex set, particularly after it has been subject to several sets operation, also remains an open challenge which needs to be tackled in future works.

References

[1] O. Adrot, D. Maquin, and J. Ragot. Bounding approach to fault detection of uncertain dynamic systems. *IFAC Proceedings Volumes*, 33(11):1131–1136, June 2000.
[2] C. D. Aliprantis and K. C. Border. *Topological vector spaces*, page 161–236. Springer Berlin Heidelberg, 1999.
[3] M. Althoff, O. Stursberg, and M. Buss. Computing reachable sets of hybrid systems using a combination of zonotopes and polytopes. *Nonlinear Analysis: Hybrid Systems*, 4(2):233–249, May 2010.
[4] G. Bainier, B. Marx, and J.-C. Ponsart. Orthogonal projection of convex sets with a differentiable boundary. *arXiv*, 2024.
[5] A. Benavoli and D. Piga. A probabilistic interpretation of set-membership filtering: Application to polynomial systems through polytopic bounding. *Automatica*, 70:158–172, Aug. 2016.
[6] D. Bertsekas and I. Rhodes. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, 16(2):117–128, Apr. 1971.
[7] T. J. Bird, H. C. Pangborn, N. Jain, and J. P. Koeln. Hybrid zonotopes: A new set representation for reachability analysis of mixed logical dynamical systems. *Automatica*, 154:111107, Aug. 2023.
[8] F. Blanchini and S. Miani. Convex sets and their representation. In *Set-Theoretic Methods in Control*, pages 93–119. Springer International Publishing, 2015.
[9] J. Blesa, V. Puig, and J. Saludes. Robust fault detection using polytope-based set-membership consistency test. In *2010 Conference on Control and Fault-Tolerant Systems (SysTol)*, page 726–731. IEEE, Oct. 2010.
[10] J. Blesa, V. Puig, and J. Saludes. Identification for passive robust fault detection using zonotope-based set-membership approaches. *International Journal of Adaptive Control and Signal Processing*, 25(9):788–812, Apr. 2011.
[11] J. Blesa, V. Puig, J. Saludes, and R. M. Fernández-Cantí. Set-membership parity space approach for fault detection in linear uncertain dynamic systems. *International Journal of Adaptive Control and Signal Processing*, 30(2):186–205, Mar. 2014.
[12] A. Bregon. *Integration of FDI and DX Techniques within Consistency-based Diagnosis with Possible Conflicts*. PhD thesis, Universidad de Valladolid, 05 2010.
[13] J. Chen and R. J. Patton. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer US, 1999.
[14] E. Chow and A. Willsky. Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control*, 29(7):603–614, July 1984.
[15] C. Combastel. An extended zonotopic and Gaussian Kalman filter (EZGKF) merging set-membership and stochastic paradigms: Toward non-linear filtering and fault detection. *Annual Reviews in Control*, 42:232–243, 2016.
[16] C. Combastel. Functional sets with typed symbols: Mixed zonotopes and polynotopes for hybrid nonlinear reachability and filtering. *Automatica*, 143:110457, Sept. 2022.
[17] C. Combastel, V. Puig, T. Raïssi, and T. Alamo. Set-membership methods applied to FDI and FTC. *International Journal of Adaptive Control and Signal Processing*, 30(2):150–153, Jan. 2016.
[18] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Apr. 2002.

[19] J. de Kleer, A. K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56(2–3):197–222, Aug. 1992.
[20] J. Deckert, M. Desai, J. Deyst, and A. Willsky. F-8 DFBW sensor failure identification using analytic redundancy. *IEEE Transactions on Automatic Control*, 22(5):795–803, Oct. 1977.
[21] S. X. Ding. *Model-Based Fault Diagnosis Techniques*. Springer London, 2013.
[22] D. Efimov and T. Raïssi. Design of interval observers for uncertain dynamical systems. *Automation and Remote Control*, 77(2):191–225, Feb. 2016.
[23] M. Fiacchini, C. Prieur, and S. Tarbouriech. On the computation of set-induced control Lyapunov functions for continuous-time systems. *SIAM Journal on Control and Optimization*, 53(3):1305–1327, Jan. 2015.
[24] P. M. Frank. Fault diagnosis in dynamic systems via state estimation - a survey. In S. G. Tzafestas S, Singh M, editor, *System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches*, page 35–98. Springer Netherlands, 1987.
[25] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. *Automatica*, 26(3):459–474, May 1990.
[26] S. Gentil, J. Montmain, and C. Combastel. Combining FDI and AI approaches within causal-model-based diagnosis. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 34(5):2207–2221, Oct. 2004.
[27] J. Gertler. Analytical redundancy methods in fault detection and isolation - survey and synthesis. *IFAC Proceedings Volumes*, 24(6):9–21, Sept. 1991.
[28] J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. CRC Press, Nov. 2017.
[29] S. Hafstein and P. Giesl. Review on computational methods for Lyapunov functions. *Discrete and Continuous Dynamical Systems - Series B*, 20(8):2291–2331, Aug. 2015.
[30] H. Hogbe-Nlend. *Bornologies and functional analysis*. North-Holland Mathematics Studies. North-Holland, Jan. 1977.
[31] A. B. Kurzhanskii and I. Valyi. *Ellipsoidal calculus for estimation and control*. IIASA Birkhauser Boston, Laxenburg, Austria Boston, 1997.
[32] C. Le Guernic and A. Girard. Reachability analysis of linear systems using support functions. *Nonlinear Analysis: Hybrid Systems*, 4(2):250–262, May 2010.
[33] D. Luenberger. *Optimization by vector space methods*. Wiley, New York, 1968.
[34] R. S. Mah, G. M. Stanley, and D. M. Downing. Reconcillation and rectification of process flow and inventory data. *Industrial & Engineering Chemistry Process Design and Development*, 15(1):175–183, Jan. 1976.
[35] F. Mazenc, T. N. Dinh, and S. I. Niculescu. Interval observers for discrete-time systems: interval observers for discrete-time systems. *International Journal of Robust and Nonlinear Control*, 24(17):2867–2890, June 2013.
[36] M. Milanese, J. Norton, H. Piet-Lahanier, and É. Walter, editors. *Bounding Approaches to System Identification*. Springer US, 1996.
[37] L. Narici and E. Beckenstein. *Topological Vector Spaces*. Chapman and Hall/CRC, July 2010.
[38] B. Noack, F. Pfaff, and U. D. Hanebeck. Optimal Kalman gains for combined stochastic and set-membership state estimation. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, page 4035–4040. IEEE, Dec. 2012.
[39] R. Patton and J. Chen. A review of parity space approaches to fault diagnosis. *IFAC Proceedings Volumes*, 24(6):65–81, Sept. 1991.
[40] S. Ploix and O. Adrot. Parity relations for linear uncertain dynamic systems. *Automatica*, 42(9):1553–1562, Sept. 2006.
[41] D. Poole. Representing diagnosis knowledge. *Annals of Mathematics and Artificial Intelligence*, 11(1–4):33–50, Mar. 1994.
[42] ProofWiki. Supremum of sum equals sum of suprema. *proofwiki.org/wiki/Supremum_of_Sum_equals_Sum_of_Suprema*.
[43] V. Puig. Fault diagnosis and fault tolerant control using set-membership approaches: Application to real case studies. *International Journal of Applied Mathematics and Computer Science*, 20(4):619–635, Dec. 2010.
[44] D. Qu, Z. Huang, Y. Zhao, G. Song, K. Yi, and X. Zhao. Nonlinear state estimation by extended parallelotope set-membership filter. *ISA Transactions*, 128:414–423, Sept. 2022.
[45] V. Raghuraman and J. P. Koeln. Set operations and order reductions for constrained zonotopes. *Automatica*, 139:110204, May 2022.
[46] S.-A. Raka and C. Combastel. Fault detection based on robust adaptive thresholds: A dynamic interval approach. *Annual Reviews in Control*, 37(1):119–128, Apr. 2013.
[47] S. V. Raković. Control Minkowski-Lyapunov functions. *Automatica*, 128:109598, June 2021.
[48] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987.
[49] V. Reppa and A. Tzes. Fault-detection relying on set-membership techniques for an atomic force microscope. *IFAC Proceedings Volumes*, 42(8):1186–1191, 2009.
[50] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Dec. 1970.
[51] H. H. Schaefer and M. P. Wolff. Locally convex topological vector spaces. In *Topological Vector Spaces*, pages 36–72. Springer New York, 1999.
[52] E. Schechter. Convexity. In *Handbook of Analysis and Its Foundations*, pages 302–325. Elsevier, 1997.
[53] R. Schneider. Basic convexity. In *Convex Bodies The Brunn-Minkowski Theory*, pages 1–73. Cambridge University Press, Oct. 2013.
[54] F. Schweppe. Recursive state estimation: Unknown but bounded errors and system inputs. *IEEE Transactions on Automatic Control*, 13(1):22–28, Feb. 1968.
[55] J. K. Scott, D. M. Raimondo, G. R. Marseglia, and R. D. Braatz. Constrained zonotopes: A new tool for set-based estimation and fault detection. *Automatica*, 69:126–136, July 2016.
[56] P. Seifried. Fault detection and diagnosis in chemical and petrochemical processes. *Chemie Ingenieur Technik*, 51(7):766–766, July 1979.
[57] D. Silvestre. Constrained convex generators: A tool suitable for set-based estimation with range and bearing measurements.

IEEE Control Systems Letters, 6:1610–1615, 2022.

[58] D. Silvestre. Exact set-valued estimation using constrained convex generators for uncertain linear systems. *IFAC-PapersOnLine*, 56(2):9461–9466, 2023.

[59] D. Silvestre, P. Rosa, and C. Silvestre. Distinguishability of discrete-time linear systems. *International Journal of Robust and Nonlinear Control*, 31(5):1452–1478, Dec. 2020.

[60] W. A. Sutherland. *Introduction to metric and topological space*. Oxford University Press, London, England, Aug. 1975.

[61] L. Travé-Massuyès and T. Escobet. *Bridge: Matching Model-Based Diagnosis from FDI and DX Perspectives*. Springer International Publishing, 2019.

[62] S. Zhang, V. Puig, and S. Ifqir. Robust fault detection using set-based approaches for LPV systems: Application to autonomous vehicles. *IFAC-PapersOnLine*, 55(6):31–36, 2022.