Robust Detection and Identification of Simultaneous Sensor and Actuator Faults

Hamza El-Kebir
Dept. of Aerospace Engr.
University of Illinois Urbana-Champaign
Urbana, IL 61801, USA
elkebir2@illinois.edu

Melkior Ornik
Dept. of Aerospace Engr,
University of Illinois Urbana-Champaign
Urbana, IL 68101, USA
mornik@illinois.edu

Yashwanth Kumar Nakka
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109, USA
yashwanth.kumar.nakka@jpl.nasa.gov

Changrak Choi
Jet Propulsion Laboratory
California Institute of Technologyy
Pasadena, CA 91109, USA
changrak.choi@jpl.nasa.gov

Amir Rahmani
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109, USA
amir.rahmani@jpl.nasa.gov

***Abstract*—In this work, we address the problem of fault detection, identification, and recovery (FDIR) for simultaneous actuator and sensor degradation. Prior studies have considered sensor degradation to some extent, but results for systems that do not have sensor redundancy while experiencing simultaneous actuator degradation are lacking. We present a novel method for robustly detecting and identifying sensor degradation in the presence of potential bounded actuator degradation modes under the influence of process and output noise. Subsequently, we develop novel theory for guaranteed forward reachability in the case of stochastic differential equations, as well as theory for robust fault detection for dynamical systems. Our approach enables in robust state estimation under sensor degradation, namely in the framework of zonotopic–Gaussian Kalman filters (ZGKF), providing an end-to-end FDIR framework for simultaneous sensor and actuator faults in a multi-agent setting. We apply our approach, which can be run in real time, to a realistic model of rigid-body satellite attitude dynamics in the presence gyroscope bias and gain changes, as well as changes in the thruster efficacy. We also present a second examples based on a multi-agent rover mission with limited periodic information sharing.**

***Index Terms*—fault tolerance, sensor/actuator degradation, real-time reconfiguration, guaranteed reachability, multi-agent fault detection**

I. Introduction

Fault detection, identification, and recovery (FDIR) has been a problem that attracts constant attention from the fault-tolerant control community [1]. FDIR is often based on nominal system models and/or redundant sensors to detect changes in the system dynamics (fault detection), characterize these changes (fault identification), and recover system performance despite these changes (fault recovery). We focus in this work on sensor and actuator faults. In prior work, sensor and actuator faults are treated in isolation, an idealization that may not hold in practice in the case of, e.g., collocated actuator-sensor clusters. In this paper, we focus our attention exactly to FDIR for simultaneous actuator and sensor degradation.

Actuator degradation has been studied extensively in the past, as shown in work by Wang et al. [2], who considered control input map degradation and actuator saturation in discrete-time linear systems, where a fault-tolerant control is developed by solving a constrained optimization problem. A similar approach was developed by Si et al. [3], where system reliability was assessed using an event-based Monte Carlo simulation approach, wherein potential degradation modes are simulated *en masse*, further limiting the applicability of this method due to the intractable number of potential failure modes that may be encountered in practice, which would demand a very large number of Monte Carlo simulations.

Sensor degradation, on the other hand, has only received limited attention in prior work, with results lacking for systems experiencing simultaneous actuator degradation. Xu et al. [4] studied a class of fault-tolerant controllers based on constrained model predictive control (MPC) in the presence of sensor faults. Their method relies on tightening of MPC state constraints due to faulty sensor measurements; actuator changes are not accounted for in this work. Efimov et al. [5] studied the problem of fault detection and compensation by treating actuator faults as unknown inputs, but their approach does not account for sensor degradation. Their method is applicable to discrete-time nonlinear systems, but relies heavily on higher-order numerical differentiation; importantly, actuator disturbance structure is not leveraged in this case, which makes the approach ill-suited for use in predictive fault mitigation.

We present a novel method for robustly detecting and identifying sensor degradation in the presence of potential bounded actuator degradation modes, under the influence of process and output noise. Our approach is structured as follows:

1) Assuming bounded propagation error rates due to actuator degradation, as well as interval-bounds on output mea-

This work was supported by NASA grant no. 80NSSC22M0070, AFOSR grant no. FA9550-23-1-0131, and ONR grant no. N00014-23-1-2505.

surements, we use interval-based ordinary least squares regression to fit a potentially biased affine sensor degradation map from the expected output to the sensed output.
2) We account for inherent stochastic process and sensor noise by introducing bounds on the sensor degradation map parameter estimates to obtain confidence intervals that contain the original sensor readings at a specified level of confidence, which we term the *confidently reached output set*.
3) The estimated sensor degradation map is used to robustly test for the presence of sensor degradation, which we do by studying a parsimonious model of the sensor as opposed to raw high-dimensional output time series.
4) We leverage guaranteed reachable output sets to account for allowable model uncertainty, enabling us to detect the presence of actuator degradation by monitoring the excursion of the guaranteed reachable output set from the confidently reached output set.
5) We reconstruct the full state of the potentially impaired system by leveraging the theory of zonotopic-Gaussian Kalman filters (ZGKF). Here, the zonotopic disturbance is reconstructed directly from the observed sensor data, allowing us to obtain output-based set-valued estimates of the system state, which can in turn be used in planning.

The proposed framework is largely model-agnostic, and only requires a means of simulating the nominal system's trajectory as well as rudimentary propagation error rate bounds. Despite these lax requirements, we are capable of reconstructing a range of sensor degradation modes, as well as narrowing down the set of potential actuator degradation modes over time. In addition, our approach allows for robust state estimation, providing an end-to-end FDIR framework for simultaneous sensor and actuator faults, with extensions to multi-agent systems with periodic information sharing. In this multi-agent extension, we consider periodic availability of potentially corrupted external output information, which we use to facilitate faster FDIR on the ego system.

We apply our approach, which can be run in real time, to a realistic example based on rigid-body satellite attitude dynamics in the presence gyroscope bias and gain changes, as well as changes in the thruster efficacy. Additionally, we present an example on a multi-agent ground rover system with limited information exchange.

II. Preliminaries

A. Notation

Let $\mathscr{S}_\tau(X,Y)$ denote the set of simple functions whose defining disjoint sets have measure greater than $\tau > 0$, i.e., all $f \in \mathscr{S}_\tau(X,Y)$ are such that $f(x) := \sum_k \alpha_k \chi_{A_k}(x)$, such that $\mu(A_k) \geq \tau$ for all k. Here, χ_A is the indicator function, where $\chi_A(x) = 1$ if $x \in A$, and 0 otherwise; μ is the Lebesgue measure. By $\mathscr{S}_\infty(X,Y)$ we refer to the set of constant functions. By calligraphic lower-case letters we denote intervals or hyperrectangles, e.g., $\mathscr{x} \in \mathscr{K}(\mathbb{R}^m)$, where $\underline{\mathscr{x}}$ and $\overline{\mathscr{x}}$ denote the lower and upper corner, respectively; $\mathscr{K}(\mathbb{R}^m)$ denotes the set of compact hyperrectangles in $\mathbb{R}^m$. We denote a closed ball in $\mathbb{R}^m$ centered around the origin with radius $r > 0$ as $\mathscr{B}_r^m$.

B. Problem Formulation

Consider the following *nominal* affine-in-control switched stochastic differential equation (SDE):

$$\begin{aligned} \dot{\bar{x}}(t,\sigma(t)) &= f(\bar{x}(t)) + g(\bar{x}(t))P_{\sigma(t)}u(t), \\ \bar{y}(t,\sigma(t)) &= h(\bar{x}(t,\sigma(t))), \end{aligned} \tag{1}$$

where $W(t)$ is a Wiener process, and $\nu(t) \sim \mathscr{N}(0,\sigma_\nu^2)$. We denote the system's state $x(t) \in \mathbb{R}^n$, control input $u(t) \in \mathbb{R}^m$ and output $y(t) \in \mathbb{R}^d$. Let $y_*(t) := h(x(t))$. Let $\bar{x}(t) := \bar{x}(t,1)$. Define the following *off-nominal* system:

$$\begin{aligned} \mathrm{d}x(t) &= f(x(t))\mathrm{d}t + g(x(t))P_{\sigma(t)}u(t)\mathrm{d}t + \mathrm{d}W(t), \\ y(t) &= q_{\sigma(t)} + Q_{\sigma(t)}(h(x(t)) + \nu(t)). \end{aligned} \tag{2}$$

Here, $P_{\sigma(t)} \in C_0(\mathbb{R}^m,\mathbb{R}^m)$ is called the *control degradation mode* or *map* (CDM), and $q_{\sigma(t)} \in \mathbb{R}^d$ and $Q_{\sigma(t)} \in \mathbb{R}^{d\times d}$ define an affine *sensor degradation mode* or *map* (SDM).

We make the following assumptions about the switching signal:

Assumption 1. For switching signal $\sigma : \mathbb{R}_+ \to \Sigma$, assume that $\#\Sigma \in \mathbb{N}$, and let $\sigma = 1$ represent the *nominal operational mode*, where $P_1 = I$ and $(p_1, Q_1) = (0, I)$.

Furthermore, assume that admissible switching signals σ originate from the set $\mathscr{S}_\tau(\mathbb{R}_+,\Sigma)$ for some known dwell time $\tau > 0$.

We define the guaranteed (forward) reachable set (GRS) as follows:

Definition II.1 (Guaranteed Reachable Set)**.** Assuming that sensor degradation mode $P_{\sigma(t)}$ satisfies $P_{\sigma(t)} \in \mathscr{P} \subseteq \mathscr{B}_{\delta_P}(I)$ for all values of $\sigma(t) \in \Sigma$, the *guaranteed reachable set* (GRS) is defined as some set $X^{\rightarrow}(t, X_0, u)$ satisfying:

$$\begin{aligned} &X^{\rightarrow}(t, X_0, u) \\ &\supseteq \{\varphi(t; x_0, u, P) : x_0 \in X_0, P \in \mathscr{S}_\tau([0,t],\mathscr{P})\}, \end{aligned} \tag{3}$$

where $\varphi(t; x_0, u, P)$ denotes the trajectory of system (1) with initial state x_0, control input signal u, and CDM P.

The *guaranteed reachable output set* (GRO) is defined as some set $Y^{\rightarrow}(t, X_0, u)$ such that

$$\begin{aligned} &Y^{\rightarrow}(t, X_0, u) \\ &\supseteq \{h(\varphi(t; x_0, u, P(t))) : x_0 \in X_0, P \in \mathscr{S}_\tau([0,t],\mathscr{P})\}. \end{aligned} \tag{4}$$

This set is simply a superset of $h(X^{\rightarrow}(t, X_0, u))$.

Any set that satisfies the above condition (4) and is described by a nominal output trajectory $\bar{y}(t)$ as $\bar{y}(t) + \mathscr{B}_{\rho(t)}$, where $\rho : [0,T] \to \mathbb{R}_+$ defines a *funnel*, is called a *confidently reached output set* (CRO).

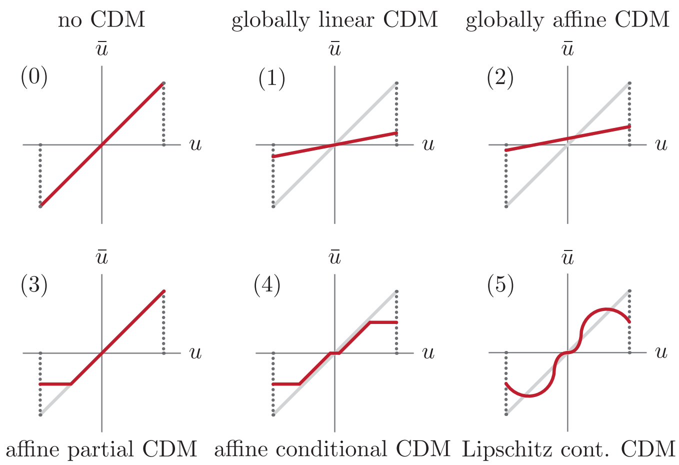


Fig. 1: Comparison between various classes of common degradation modes. Our approach focuses on (2) where sensor degrades with changes in gain and offset. This allows our method to recover the sensor readings and separate out actuator failures even in the presence of sensor degradation.

C. Sensor Degradation Modes

We briefly consider a practical classification of common degradation modes, illustrated in Fig. 1. The following categories for sensor faults have been identified in the literature [6]:

1) **Bias**: Constant drift $(y(t) = y_*(t) + \delta)$.
2) **Drift**: Time-varying $(y(t) = y_*(t) + \delta(t))$.
3) **Noise**: 0-mean random $(y(t) = \nu(t))$.
4) **Scaling/Gain**: Scaled magnitude $(y(t) = \alpha(t)y_*(t))$.
5) **Hard faults**: Lost/stuck sensor $(y(t) = C_{\sigma(t)} + \nu(t))$.
6) ***Intermittents***: Intermittent loss of sensor signal $(y(t) = [\![\sigma(t) = 1]\!]y_*(t) + C_{\sigma(t)} + \nu(t))$.

Faults 1–3 are often referred to as 'tame' faults, whereas faults 4–6 are called severe faults [6]. Before we proceed to described methods for robust detection of sensor degradation, it bears mentioning what difficulties underly this problem.

In general, it is hard to ascribe anomalous sensor readings either to changes in the dynamics or to sensor failure, since these may overlap. Consider as an example an aircraft reporting a constant roll rate; such a report may be due to a stuck gyroscope reading, or could be due to an unresponsive aileron. Discriminating between these two failure modes requires active probing, i.e., altering the control signal, in cases where one does not have sensor redundancy, as is the case here.

Since we would like to develop a passive detection algorithm, we consider cases in which sensors experience gain changes or constant biases that could be accounted for, as opposed to completely ignoring these readings. In particular, we focus on *affine degradation maps* (Fig. 1.2). We refer to these sensor failure modes as *reversible* or *recoverable failure modes*, since the affine transformation $y = ax + b$ can be inverted if $a \neq 0$, giving $x = (y - b)/a$. A key motivation here is that disregarding recoverable sensor readings may often result in much greater problems in systems that lack sensor redundancy, with the possibility of loss of mission.

We now proceed by posing the problem of robust detection of sensor degradation.

III. Robust Detection of Sensor Degradation

We proceed by posing two problems regarding sensor degradation mode detection. We distinguish between the absence and presence of control authority degradation modes. It bears noting here that we assume that sensor degradation always precedes actuator degradation; we comment on the implications of the opposite case later in this work.

Problem 1 (Robust SDM Detection)**.** Given system (2), assuming that $\sigma \in \mathscr{S}_\infty(\mathbb{R}_+, \Sigma \setminus \{1\})$, determine in time T with confidence $\gamma \in (0, 1)$ whether $(q, Q) \neq (0, I)$ given input–output pair $(y(t), u(t))_{t=0}^T$ and knowledge of the GRS of (1) under the influence of CDMs in $\mathscr{P}$.

In this first problem, we aim to determine whether or the acting sensor degradation mode is not equal to the identity transform $(0, I)$ with confidence γ. This amounts to robust detection of a sensor degradation mode. We extend this problem by also demanding that the presence of simultaneous actuator degradation be detected:

Problem 2 (Robust CDM–SDM Detection)**.** Given system (2), assuming that $\sigma \in \mathscr{S}_\infty(\mathbb{R}_+, \Sigma \setminus \{1\})$, determine in time T with confidence $\gamma \in (0, 1)$ whether $(q, Q) \neq (0, I)$ and $P \neq I$ given input–output pair $(y(t), u(t))_{t=0}^T$ and knowledge of nominal trajectories and the GRS of (1) under the influence of CDMs in $\mathscr{P}$.

We also state the robust SDM identification problem:

Problem 3 (Robust SDM Identification)**.** Given System 2, assuming that $\sigma \in \mathscr{S}_\infty(\mathbb{R}_+, \Sigma \setminus \{1\})$, construct in time T a compact set K in which (q, Q) resides with confidence $\gamma \in (0, 1)$, given input–output pair $(y(t), u(t))_{t=0}^T$ and knowledge of the GRS of (1) under the influence of CDMs in $\mathscr{P}$.

In this problem, our goal is not only to detect the SDM as in Problem 1, but also identify it. We briefly consider some alternative approaches to solving Problem 3, noting shortcomings of these approaches prior to presenting our own method:

Problem 3 can be formulated as an *unknown input observer* (UIO) synthesis problem for control-affine stochastic differential equations. Consider the following system:

$$\begin{aligned} \mathrm{d}x(t) &= f(x(t))\mathrm{d}t + g(x(t))v_{:m}(t)\mathrm{d}t + \mathrm{d}W(t), \\ y(t) &= v_{m:}(t), \end{aligned} \tag{5}$$

where $v_{:m}(t) = P_{\sigma(t)}u(t)$ and $v_{m:}(t) = q_{\sigma(t)} + Q_{\sigma(t)}(h(x(t)) + \nu(t))$. This problem has been addressed in the linear parameter variance discrete-time case in [7]. One major drawback here is the fact that much of the problem structure is discarded, making it prohibitive to tractably reconstruct $v(t)$.

Another possibility is to frame the problem as a state estimation problem with partially unknown initial state [8]:

$$\begin{aligned} \mathrm{d}x_{:n}(t) &= f(x_{:n}(t))\mathrm{d}t + g'(x(t))u(t)\mathrm{d}t + \mathrm{d}W(t), \\ \dot{x}_{n:}(t) &= 0 \\ y(t) &= h'(x(t)+\nu(t)) = q + Q(x_{:n}(t)+\nu(t)), \\ x_{n:}(0) &= \mathrm{param}(P, q, Q), \\ g'(x(t)) &= g(x_{:n}(t))P, \end{aligned} \tag{6}$$

where only $x_{:n}(0)$ is known and $\mathrm{param}(P, q, Q)$ describes P, q, and Q as a vector of parameters. Note that here P is assumed to be a linear map. Both problems have not been addressed in full generality, necessitating a novel approach such as ours even to address the fault detection problem, let alone the identification problem.

Before presenting our approach, we proceed by making some simplifying assumptions:

Assumption 2. Noise ν is normally distributed with zero mean and variance σ^2. Additionally, there exists known $r > 0$ such that $\|y(t)\| \leq r$ and $\|\mathrm{bias}(y(t))\| = \|\mathbb{E}[y_*(t)|y(t)] - y(t)^\mathsf{T}\beta\| \leq a$ almost surely for all finite-time trajectories generated by system (2). Here, β is the vector of regressors.

The assumption of normally distributed output noise is a common one in the literature. On the other hand, the almost-sure boundedness of the output signal is a given in practical systems with finite sensor output [9].

We now proceed by introducing a method for SDM *identification*, which allows us to directly detect the presence of SDMs at the same time.

A. Robust Identification of SDMs

Our approach to solving Problem 1 is based on the idea of separating deterministic and stochastic uncertainty. We leverage the funnel radius $\rho(t)$ to account for small model imperfections so as to obtain a confidently reached output (CRO) set; doing so accounts for deterministic uncertainties. On the other hand, the effects of the process noise $W(t)$ and output noise $\nu(t)$ are harder to account for. We propose to fit our affine sensor degradation model to the noisy output data and the CRO, and account for the effect of stochastic noise errors on the parameter estimations by producing a guaranteed set of parameters for some given confidence $0 < \gamma < 1$.

Since we wish to find a relationship between the CRO, which is a set, and the measured (possibly corrupted) sensor output, we proposed to adopt a method known as *interval-based least squares* [10]. We consider the following model:

$$\mathcal{y}(t) = Q\mathcal{y}_*(t) + \mathcal{q}, \tag{7}$$

where $\mathcal{y}, \mathcal{y}_*, \mathcal{q} \in \mathscr{K}(\mathbb{R}^m)$ are intervals and $Q \in \mathrm{diag}(\mathbb{R}^m)$ is a diagonal matrix. Our goal is to find Q and $\mathcal{q}$ such that the Euclidean norm error between the left- and right-hand sides of the equation is minimized. After some manipulation, this problem reduces to an ordinary least squares problem of the following form:

$$\min_\beta \|(b_e + \tilde{b}_e) - (A_e + \tilde{A}_e)\beta\|, \tag{8}$$

where $b_e, \beta \in \mathbb{R}^{3m}$. Here, $A_e + \tilde{A}_e$ is known as a *random design matrix*, where A_e is the known mean and $\tilde{A}_e$ is a zero-mean Gaussian matrix with known variance per Assumption 2. Similarly, the vector of outputs $b_e + \tilde{b}_e$ is normally distributed with known variance bounds and known mean.

It is possible to obtain a set in which the coefficient vector β, which describes Q and $\mathcal{q}$, is guaranteed to reside with confidence $0 < \gamma < 1$. Such a result was provided by Hsu et al. [9], who studied the effect of random design matrix ordinary least squares on the model coefficients β:

Proposition 1 ([9, Thm. 1, p. 8])**.** *Let Assumption 2 hold. Then, for any fixed* $\delta \in (0, 1)$*, if* N *is strictly greater than*

$$N_{2,\delta} := 4\rho^2_{2,\mathrm{cov}} d \log(d/\delta), \tag{9}$$

the following bound holds with probability at least $1 - 2\delta$*:*

$$\begin{aligned} &\|\hat{\beta}_{\mathrm{OLS}} - \beta\|^2_\Sigma \\ &\leq K_{2,\delta,N} \frac{\sigma^2_{\mathrm{noise}}(d + 2\sqrt{d\log(1/\delta)} + 2\log(1/\delta))}{N}, \end{aligned} \tag{10}$$

where

$$K_{2,\delta,N} := \left(1 - \sqrt{\frac{2\rho^2_{2,\mathrm{cov}} d \log(d/\delta)}{N}}\right)^{-1}. \tag{11}$$

The bound in (10), which is known as the *excess loss*, can directly be used to find intervals for the SDM coefficients at a desired confidence level, namely $(\mathcal{q}, \mathcal{Q})$. Using this result, we may now solve a hypothesis testing problem which will yield whether or not sensor degradation is present at a desired confidence level γ.

B. Robust Hypothesis Testing for SDM Detection

We now proceed by posing a solution to Problem 1 based on hypothesis testing on the parameters that describe the SDM, (q, Q). Using the approach described in Proposition 1, we obtain two intervals $(\mathcal{q}, \mathcal{Q})$. These intervals can be subject to the following two tests to determine whether or not an SDM is present with confidence γ.

a) Presence of SDM with confidence γ*:* If $(0, I) \notin (\mathcal{q}, \mathcal{Q})$, we know with confidence γ that an *SDM is indeed active*.

b) Absence of SDM with confidence γ*:* On the other hand, detecting if a sensor is operating nominally requires us to consider confidence $1 - \gamma$ and obtain the same interval $(\mathcal{q}, \mathcal{Q})$. If $(0, I)$ lies inside this interval, then there is *no active off-nominal SDM* with confidence γ.

This approach solves the sensor degradation mode detection problem at a desired confidence γ. As a matter of fact, we immediately obtain a family of sensor degradation modes which is guaranteed to contain the true SDM with confidence γ. We may now develop an approach to detecting actuator degradation.

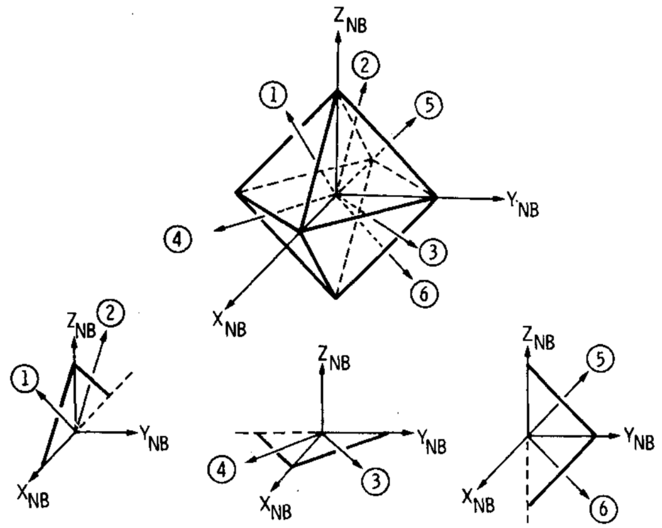


Fig. 2: Illustration of the octahedral gyroscope setup [11]. This configuration nominally allows for error detection using parity relations, where three or four sensors can be used to detect failure of another sensor.

C. Robust Detection of Sensor & Actuator Degradation

In detecting sensor degradation modes, our method directly produces a set of identified SDMs that is guaranteed to contain the true SDM at a specified confidence level.

Having shown how to detect and identify sensor degradation modes, we may know consider detecting actuator degradation. Our proposed approach uses the confidently reachable output set (CRO), which we assumed to be available as $h(x_*(t)+\mathscr{B}_{\rho(t)})$. Given the identified set of SDMs, we directly reconstruct a set of candidate outputs from the corrupted sensor output by inverting the set of affine maps. If the CRO transgresses the reconstructed output set, we can say with confidence γ that there is in fact actuator degradation. We illustrate this method by way of example in the following section.

IV. Single-agent Application

Consider the following rigid-body satellite attitude dynamics [12]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\boldsymbol{\omega}\\ \mathbf{q}\\ q_4\\ \mathbf{h}\end{bmatrix} = \begin{bmatrix} J^{-1}\left(\boldsymbol{\omega}^\times J\boldsymbol{\omega} - \boldsymbol{\omega}^\times \mathbf{h} + \mathbf{u} + \mathbf{d}\right)\\ \frac{1}{2}\boldsymbol{\omega}^\times \mathbf{q} + \frac{1}{2}q_4\boldsymbol{\omega}\\ -\frac{1}{2}\boldsymbol{\omega}^\mathsf{T}\mathbf{q}\\ -\mathbf{u}\end{bmatrix}, \tag{12}$$

where $\boldsymbol{\omega}$ denotes the body angular rate vector, $\mathbf{q}$ denotes the vector part of the quaternion, q_4 is the scalar part of the quaternion, and $\mathbf{h}$ represents the angular momentum of the spacecraft. J denotes the intertia tensor, which is assumed to be invertible, $\mathbf{u}$ denotes the body axis-aligned torque input, and $\mathbf{d}$ denotes a disturbance signal. Finally, $\boldsymbol{\omega}^\times$ is a skew symmetric matrix:

$$\boldsymbol{\omega}^\times := \begin{bmatrix} 0 & -\omega_z & \omega_y\\ \omega_z & 0 & -\omega_x\\ -\omega_y & \omega_x & 0\end{bmatrix}.$$

We consider in this work the following inertia tensor from [13]:

$$J = \begin{bmatrix} 20 & 0 & 0.9\\ 0 & 17 & 0\\ 0.9 & 9 & 15\end{bmatrix}.$$

A. Output Model

We consider the following octahedral two-degree-of-freedom IMU configuration [11], as illustrated in Fig. 2:

$$y_*(t) := h(x(t)) = H\boldsymbol{\omega}, \tag{13}$$

where

$$H := \begin{bmatrix} a & 0 & a\\ -a & 0 & a\\ a & a & 0\\ a & -a & 0\\ 0 & a & a\\ 0 & a & -a\end{bmatrix}, \tag{14}$$

with $a = \cos(\pi/4) = \sqrt{2}/2$. This configuration is very common in application because it enables so-called *parity relations* to be established; given four sensor readings, it is possible to rule out failure of a single axis through a linear combination of output signals that should equal 0 in the nominal case [11].

B. Gyroscope Failure Modes

We consider the following sensor degradation modes, which were chosen arbitrarily to cover a number of different failures:

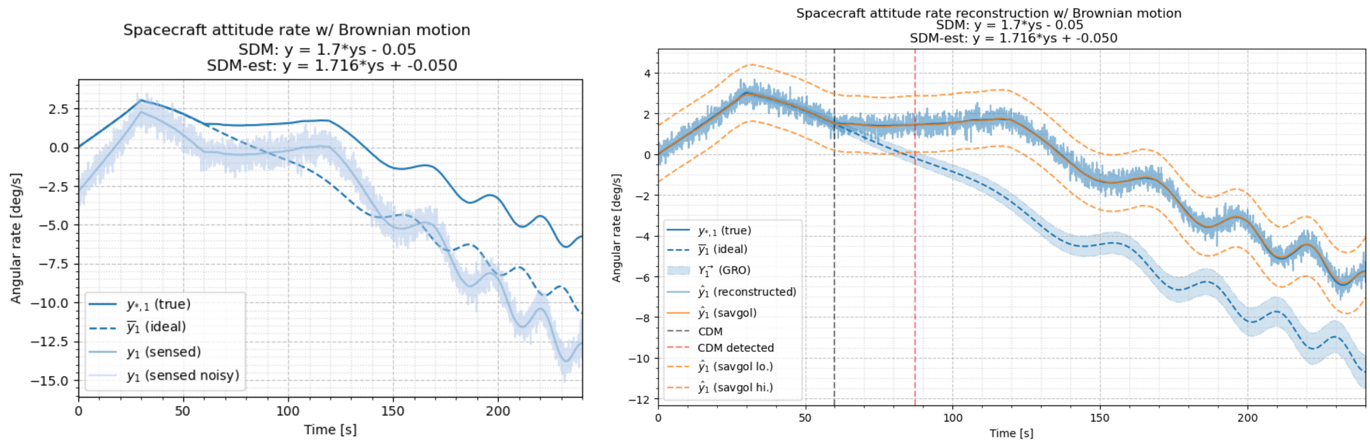
$$\begin{aligned} y_1 &= 1.7y_{*,1} - 0.05,\\ y_2 &= y_{*,2},\\ y_3 &= 10^{-3}y_{*,3} + 0.2,\\ y_4 &= y_{*,4},\\ y_5 &= 0.4y_{*,5} + 0.1,\\ y_6 &= 10^{-3}y_{*,6} + 0.6, \end{aligned} \tag{15}$$

Here, the second and fourth sensor do not experience any sensor degradation (other than inherent noise). Sensors #3 and #6, on the other hand, experience an almost complete loss of signal, being replaced by 0.2 and 0.6 instead. Finally, sensors one and five experience affine maps that include a reasonable gain change and bias.

We also introduce a fault in one of the thruster pairs; at 60 seconds, the first thruster pair produces zero thrust for a duration of 60 seconds, after which control authority is regained. We apply our approach with a 95% confidence bound.

C. Results

We demonstrate the efficacy of our approach by considering the results on the first sensor, as shown in Fig. 3. Fig. 3a shows the true (unknown) sensor output, as well as the GRO (ideal output), and the sensed (corrupted) output. Applying our approach, we can reconstruct the original sensor output based on the corrupted sensed output, as shown in Fig. 3b. Our method directly produces confidence bounds, as shown by



(a) Corrupted sensor readings. The solid line represents the true output, whereas the dashed line shows the ideal output based on the guaranteed reachable output set; this reading starts to deviate at 60 seconds due to actuator failure. Only the noise corrupted output is available for processing.

(b) Reconstructed sensor readings. Using the sensed noisy output reading and the ideal output, our approach can reconstruct the true output (orange solid line) along with guaranteed confidence bounds (orange dashed line). This enables us to autonomously detect actuator faults in a timely fashion, as shown by the vertical red dashed line.

Fig. 3: True, ideal, sensed, and reconstructed output readings for sensor 1 based on the single spacecraft attitude dynamics.

the dashed lines, in this case at a 95% confidence level; these confidence bounds are tight enough to allow for reasonable fast failure detection, *without* demanding sensor redundancy. Note that we have smoothened the upper and lower bounds using a Savitzky–Golay filter for ease of exposition.

We detect the thruster malfunction 27 seconds after failure using the data obtained by this sensor. Sensor 2 data gives us the same result after 32 seconds and sensor 4 takes 37 seconds. This data may be seen in Figs. 8 and 9.

The approach presented here allows for all of the above to be achieved with marginal knowledge, requiring only the nominal trajectory to be known. In addition to being able to reconstruct the original sensor output from its corrupted counterpart, we also allow for robust detection of actuator faults, without a need to know for a list of potentially failure modes. Unlike past methods, we do not require a hand-crafted filter bank to detect sensor failure, allowing for the present method to be adopted easily for a wide variety of systems without requiring additional design effort.

We now continue by discussing a multi-agent extension to fault detection based on limited periodic information sharing.

V. Multi-agent Fault Detection

In the case of networked multi-agent systems capable of (limited) information sharing, it is possible to periodically introduce additional sensed outputs to a set-based state estimation process, thereby tightening the full-state set estimate.

The approach is straightforward: we compute an F-radius minimizing Luenberger gain matrix L_a based on the new augmented input matrix C_a, allowing us to robustly incorporate both the internal sensed outputs and the new external ones. For the external outputs, we can easily introduce additional noise bounds in the F matrix based on the trustworthiness of the external data.

Consider now the case in which we are only allowed to request external data for at most $N_r \in \mathbb{N}$ times. We are interested in minimizing the time-to-detection (TTD) of any fault since its initial occurrence, without imposing control input constraints:

Problem 4 (Cooperative Passive SDM Time-to-Detection Minimization)**.** Given System 2, assume that $\sigma \in \mathcal{S}_\infty(\mathbb{R}_+, \Sigma \setminus \{1\})$ for all $t \in \mathbb{R}_+$, and assume $N_r \in \mathbb{N}$ τ-length external data requests (τ-EDRs) are allowed. Minimize the time-to-detection T within which, with confidence $\gamma \in (0,1)$, we can detect $(q, Q) \neq (0, I)$ given input–output pair $(y(t), u(t))_{t=0}^T$ and knowledge of the GRS of (1) under the influence of CDMs in $\mathscr{P}$.

We shall refer to this problem as the *TTD-minimizing EDR allocation problem*. We now introduce some theory on set-propagation-based observers.

a) Set-propagation Observer in a Multi-agent Setting: Consider following the discrete-time linear time-invariant system:

$$x[k+1] = Ax[k] + Bu[k] + w[k],$$
$$y[k] = Cx[k] + v[k], \tag{16}$$

where $w[k] \in W$ and $v[k] \in V \subseteq \times_{i=1}^m [-\sigma_i, \sigma_i]$. We shall use in this work an F-radius (Frobenius norm) minimizing Luenberger observer [14, §3.2]:

$$X[k+1] = (A - LC)X[k] + Bu[k] + Ly[k] + (-L)V + W, \tag{17}$$

where

$$L = AGG^\mathsf{T} C^\mathsf{T} [CGG^\mathsf{T} C^\mathsf{T} + FF^\mathsf{T}]^{-1}, \tag{18}$$

for a zonotope $X[k] = c + G\mathcal{B}_1^r$, $W = E\mathcal{B}_1^{n_w}$ and $V = F\mathcal{B}_1^{n_v}$.

This choice of Luenberger gain minimizes the F-radius of G [14, §3.2], which is defined as:

$$\|G\|_{\mathrm{F}} = \sqrt{\mathrm{trace}(G^\mathsf{T} G)}. \tag{19}$$

If V is a symmetric hyperrectangle defined by a vector $\delta_v \in \mathbb{R}^d_+$, then we have $F = \mathrm{diag}(\delta_v)$; the same holds for W such that $E = \mathrm{diag}(\delta_w)$.

We shall consider the introduction of external outputs in the observer, with the goal of choosing a external source that minimizes state uncertainty in a desired direction. This desired direction will be the one in which there is most state uncertainty, which will enable us to most accurately determine a set of candidate SDM models as in Proposition 1. This solves Problem 4. We pose the following problem on data source selection:

Problem 5 (Data Source Selection)**.** For system (16), where output map C can be chosen from a finite set $\mathscr{C} \in \mathscr{K}(\mathbb{R}^{n\times m})$, choose $C \in \mathscr{C}$ such that $X[k+1]$ has minimal width along the direction $\lambda \in \partial\mathscr{B}^n_1$.

Since we can only influence the $(A - LC)G[k]$ and $-LF$ terms of $G[k+1]$, we shall exclusively consider the effects of the output map $C \in \mathscr{C}$:

$$(A - LC)G = AG - AGG^\mathsf{T} C^\mathsf{T}[CGG^\mathsf{T} C^\mathsf{T} + FF^\mathsf{T}]^{-1} CG.$$

Here, we would like to minimize the following quantity:

$$\begin{aligned}\min_{C\in\mathscr{C}} \|\lambda^\mathsf{T} AG(I - G^\mathsf{T} C^\mathsf{T}[CGG^\mathsf{T} C^\mathsf{T} + FF^\mathsf{T}]^{-1}CG) + \\ \lambda^\mathsf{T} AGG^\mathsf{T} C^\mathsf{T}[CGG^\mathsf{T} C^\mathsf{T} + FF^\mathsf{T}]^{-1}F\|.\end{aligned} \tag{20}$$

Since $\mathscr{C}$ is finite, we can directly solve for a desired external information source. In this case, the matrix F is extended based on the output uncertainty of the external data source. This relates directly to the SDMs $(\mathscr{Q}, \mathscr{G})_i$ identified for each external agent, based on the reconstructed output funnels described in Sec. III-B.

The remainder of the approach remains the same; with the introduction of periodic external data, we obtain tighter reconstruct output sets, which will allow for faster detection of sensor and actuator degradation, as shown in Sec. VI.

We proceed with an example dealing with a multi-agent system based on a ground rover.

VI. Multi-agent Application

For our multi-agent application, we consider the case were an agent runs the proposed single-agent SDM–CDM detection algorithm, but has the capability of requesting information about its own state from stationary beacons periodically.

We consider the following dynamics of a rover:

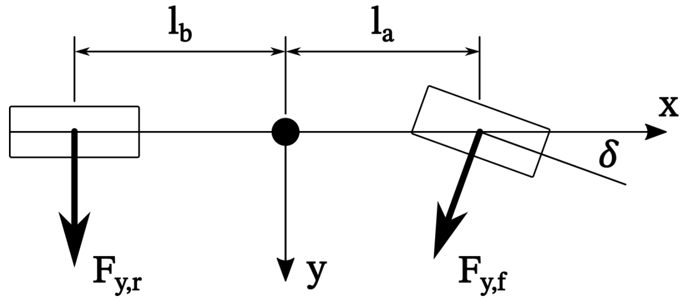


Fig. 4: Illustration of the rover geometry considered in this work.

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} 0\,0\,0 & 0 & 0 \\ 0\,0\,0 & 1 & 0 \\ 0\,0\,0 & 0 & 1 \\ 0\,0\,0 & (-C_f - C_r)/(mu) & (-C_f l_a + C_r l_b)/(mu) - u \\ 0\,0\,0 & (-C_f l_a + C_r l_b)/(Iu) & (-C_f l_a^2 - C_r l_b^2)/(Iu) \end{bmatrix} x(t) \\ &\quad + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & C_f/m \\ 0 & C_f l_a/I \end{bmatrix} \begin{bmatrix} v(t) \\ r(t) \end{bmatrix} \\ y(t) &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} x(t).\end{aligned} \tag{21}$$

Here, $C_r = 0.2$ and $C_f = 0.1$ are tire turning friction coefficients, $l_a = 1$ and $l_b = 1$ are the front and rear tire axel lengths, $m = 10$ is the mass of the vehicle, $I = 1$ the rotational inertia, and $u = 0.5$ the longitudinal velocity. Fig. 4 shows an overview of the rover geometry.

We consider five randomly placed beacons, which are of types 'A', 'B', and 'C', as shown in Fig. 5. Beacon 'A' provides high quality data for state 1 and mediocre data for states 2 and 3, beacons 'B' does so for state 2, and beacon 'C' for state 3. For each 120 time samples, a 20 sample window is provided during which the agent can communicate with any one of the beacons. The uncertainty associated with each of the beacon readings is attenuated with distance d as $e^{0.02d}$, to account for ranging errors with distance.

We simulated the system described above, prioritizing tightness of state 1 zonotope bounds. As shown in Fig. 6, external data greatly tightens the uncertainty bounds, allowing for timely detection of anomalies. To described the data selection results, during the first 20 samples, data source 4, a C-type beacon in the middle of the traveled path, is selected. Immediately thereafter, source 3, a B-type beacon at the end of the course, is selected for the remainder of the run.

The effects of data sharing on the other zonotope bounds are shown in Fig. 7. One can notice that there is a certain 'grace period' after data sharing during which the zonotope bounds slowly grow back. This phenomenon will be the subject of future work, as it could be leveraged to request data at different rates depending on this deterioration rate.

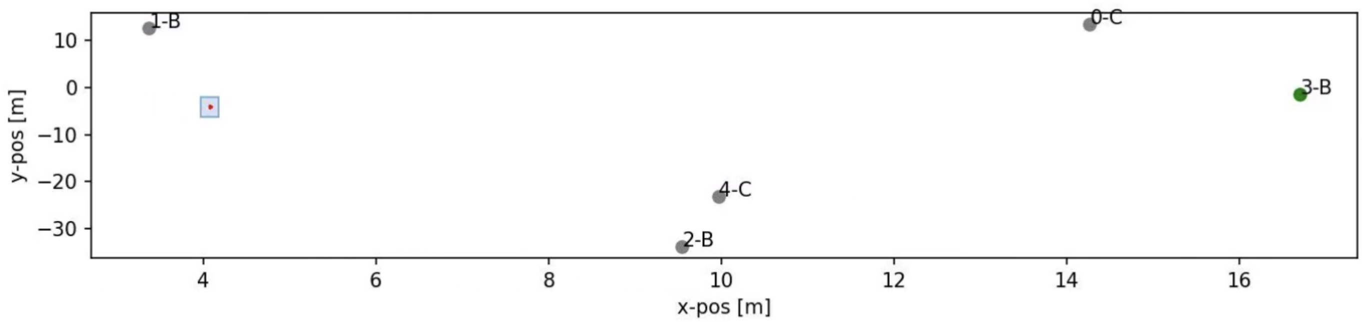
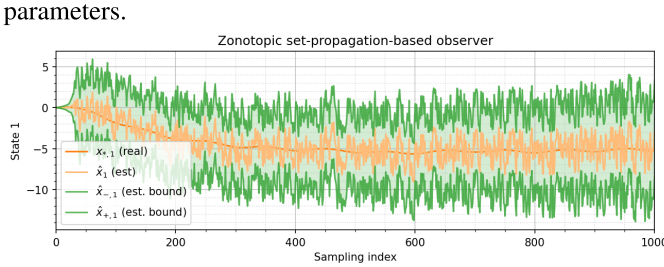
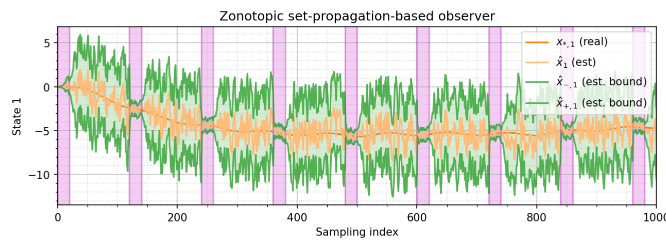


Fig. 5: External ranging beacon types and locations considered in the multi-agent rover example. In addition to experiencing attention due to the distance to a beacon, we also consider different noise levels for each supported state based on the beacon type. Beacon 'A' report high-fidelity readings for the first state, beacon 'B' for the second state, and 'C' for the third state. Using our approach, an agent is capable of autonomously choosing the desired information source depending on the uncertainty parameters.

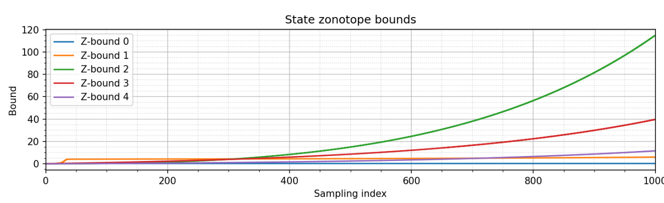


(a) State 1 estimate without external data. State uncertainty bounds remain large throughout the entire system run, making this data ill-suited for fault detection.

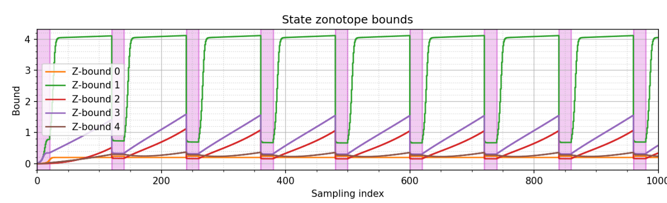


(b) State 1 estimate with external data. During periods of information exchange, the state uncertainty decreases, allowing for detection of potential faults.

Fig. 6: Comparison between state 1 estimate based on internal sensors and external information.



(a) Zonotope bounds without external data. Note the exponential growth in the zonotopic state uncertainty bounds. Without an external ground truth or periodic information exchange, these bounds quickly grow to an extent where fault detection is no longer tenable.



(b) Zonotope bounds with external data. Note the gradual increase of the error bounds after a period of information exchange. Such a 'grace period' is instrumental in allowing for fault detection even when no external data is available.

Fig. 7: Comparison between state zonotope bounds based on internal sensors and external information.

VII. Conclusion

In this work, we have presented a novel method for fault detection and identification (FDI) in the case of simultaneous sensor and actuator failure. Our method is model-agnostic, and can be applied in real time. We have applied our approach to satellite attitude dynamics in the case of reaction wheel and gyro degradation, showing that we can adequately reconstruct corrupted sensor signals, as well as detect actuator failure in a timely fashion despite a lack of sensor redundancy. We have also presented a new approach for guaranteed state estimation in case of multiple information sources, which we applied to a moving rover example, resulting in optimal selection of information sources in data-constrained environments.

For future work, we aim to address the case of actuator-to-sensor degradation. This problem is prohibitive to solve at a fundamental level, since it is not certain when or if sensor data gets corrupted in the case of uncertain actuator degradation. Given the assumption of an oracle that tells us if we are dealing with sensor-to-actuator or actuator-to-sensor degradation, it may be possible to tighten detection error bounds. The existence of such an oracle could be justified by considering low-cost diagnostics sensors that provide limited information about sensor or actuator health. In the case of multi-agent systems, a method of finding an opportunity cost for requesting external information is an open problem; such an opportunity cost will allow an agent to optimally schedule when to request information from external sources.

ACKNOWLEDGEMENTS

This work was supported by the Jet Propulsion Laboratory's Research and Technology Development (R&TD) program. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. Berlin, Germany: Springer Berlin Heidelberg, 2006.

[2] Z. Wang, M. Rodrigues, D. Theilliol, and Y. Shen, "Fault-tolerant control for discrete linear systems with consideration of actuator saturation and performance degradation," in *19th IFAC World Congress*, Cape Town, South Africa, 2014, pp. 499–504.

[3] X. Si, Z. Ren, X. Hu, C. Hu, and Q. Shi, "A novel degradation modeling and prognostic framework for closed-loop systems with degrading actuator," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 11, pp. 9635–9647, 2020.

[4] F. Xu, S. Olaru, V. Puig, C. Ocampo-Martinez, and S.-I. Niculescu, "Sensor-fault tolerance using robust MPC with set-based state estimation and active fault isolation," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 8, pp. 1260–1283, 2017.

[5] D. Efimov, A. Zolghadri, and T. Raïssi, "Actuator fault detection and compensation under feedback control," *Automatica*, vol. 47, no. 8, pp. 1699–1705, 2011.

[6] E. Balaban, A. Saxena, P. Bansal, K. F. Goebel, and S. Curran, "Modeling, Detection, and Disambiguation of Sensor Faults for Aerospace Applications," *IEEE Sensors Journal*, vol. 9, no. 12, pp. 1907–1917, 2009.

[7] F. Xu, J. Tan, X. Wang, V. Puig, B. Liang, B. Yuan, and H. Liu, "Generalized set-theoretic unknown input observer for LPV systems with application to state estimation and robust fault detection," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 17, pp. 3812–3832, 2017.

[8] A. Bobtsov, R. Ortega, B. Yi, and N. Nikolaev, "Adaptive state estimation of state-affine systems with unknown time-varying parameters," *International Journal of Control*, vol. 95, no. 9, pp. 2460–2472, 2022.

[9] D. Hsu, S. M. Kakade, and T. Zhang, "An analysis of random design linear regression," 2011, arXiv:1106.2363v1 [cs, math, stat].

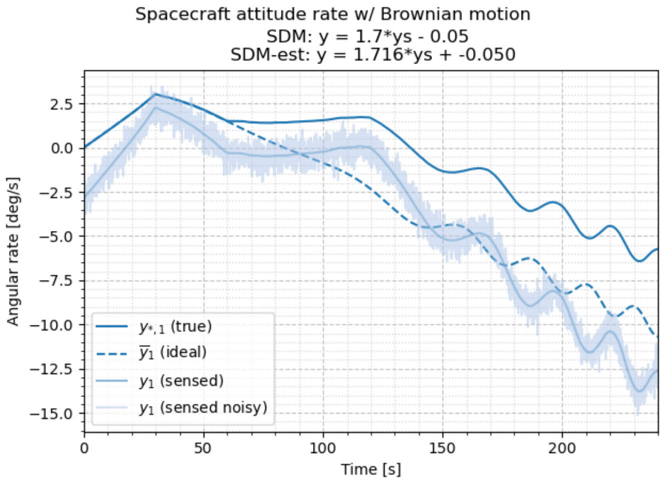
[10] P. Diamond, "Least squares fitting of compact set-valued data," *Journal of Mathematical Analysis and Applications*, vol. 147, no. 2, pp. 351–362, 1990.

[11] A. L. Satin and R. L. Gates, "Evaluation of parity equations for gyro failure detection and isolation," *Journal of Guidance and Control*, vol. 1, no. 1, pp. 14–20, 1978.

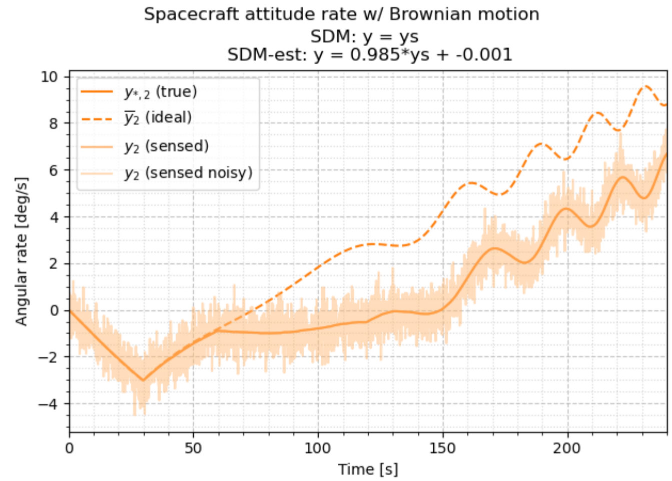
[12] M. Blanke and M. B. Larsen, *Satellite Dynamics and Control in a Quaternion Formulation*, 2nd ed. Technical University of Denmark, Department of Electrical Engineering, 2010.

[13] Z. Chen and J. Huang, "Attitude tracking and disturbance rejection of rigid spacecraft by adaptive control," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 600–605, 2009.

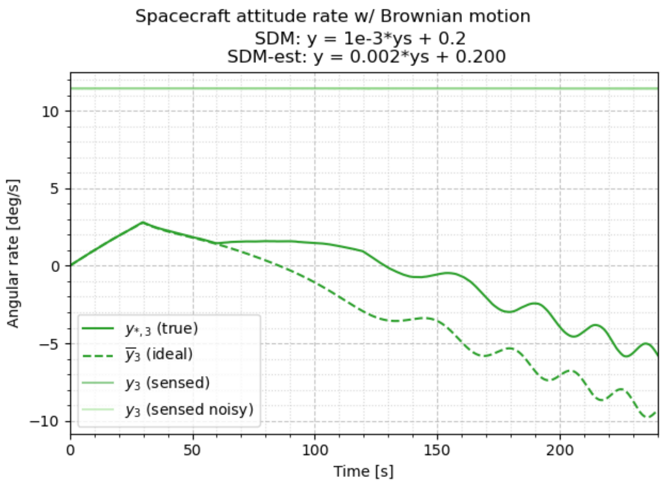
[14] M. Althoff and J. J. Rath, "Comparison of guaranteed state estimators for linear time-invariant systems," *Automatica*, vol. 130, p. 109662, 2021.



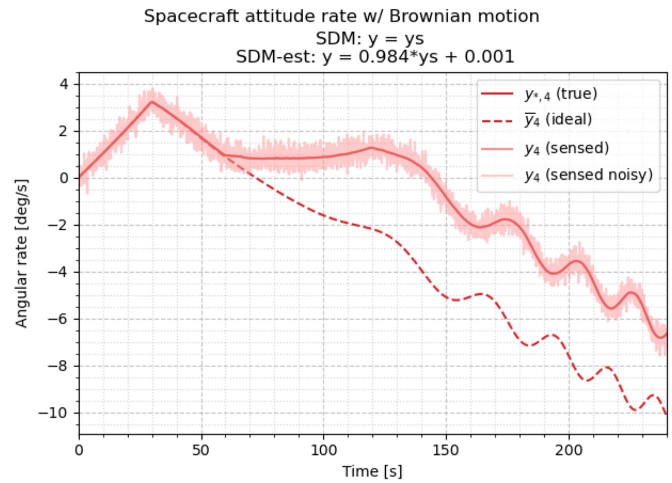
(a) Sensor 1



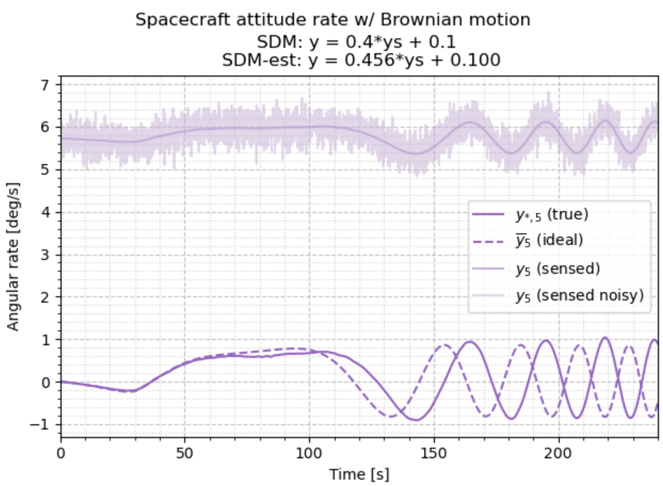
(b) Sensor 2



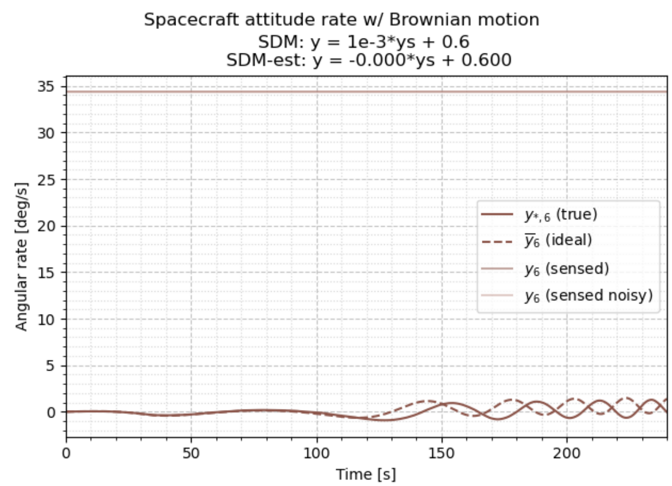
(c) Sensor 3



(d) Sensor 4

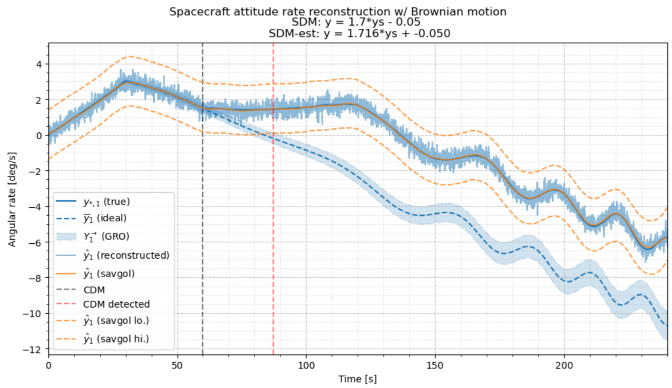


(e) Sensor 5

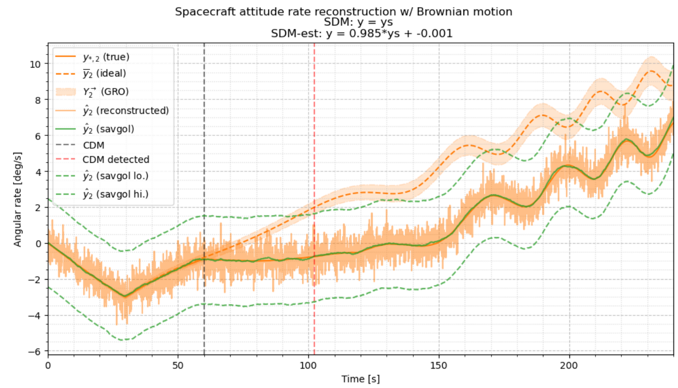


(f) Sensor 6

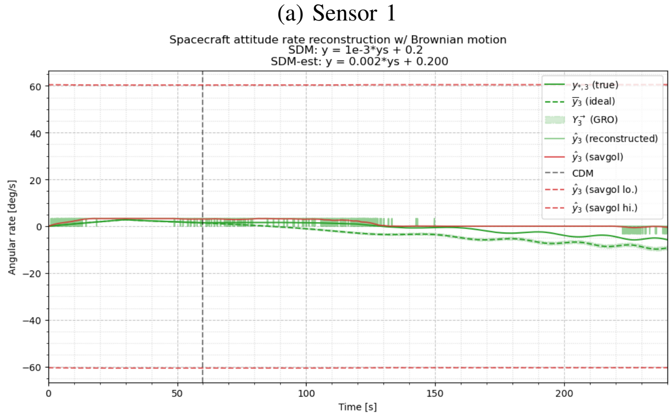
Fig. 8: True, ideal, and sensed output readings for the single spacecraft attitude dynamics.



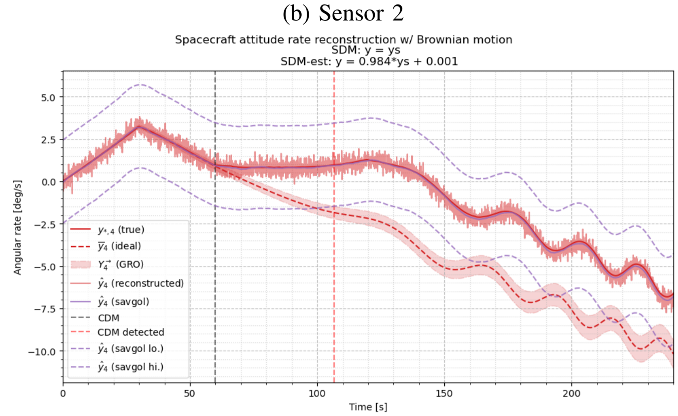
(a) Sensor 1



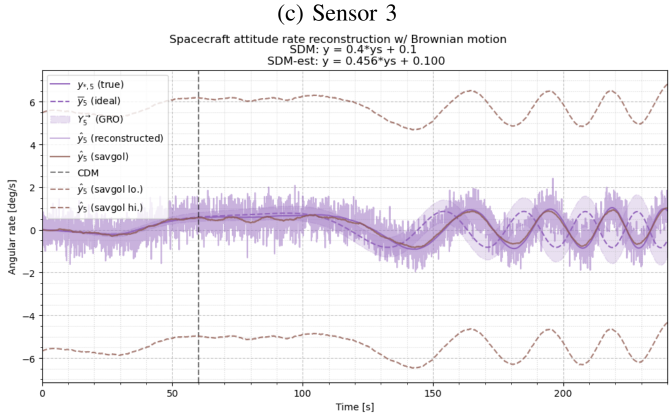
(b) Sensor 2



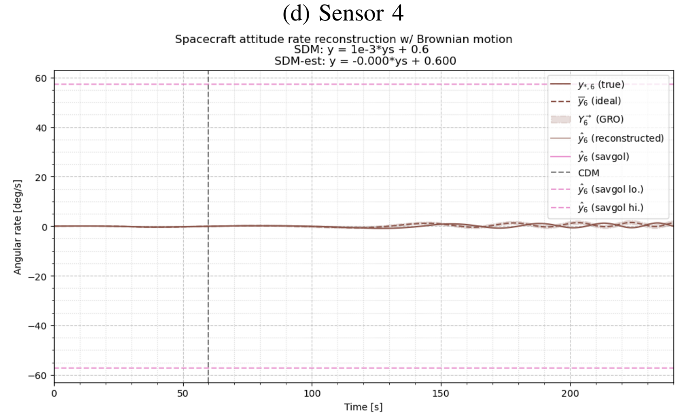
(c) Sensor 3



(d) Sensor 4



(e) Sensor 5



(f) Sensor 6

Fig. 9: True, ideal, and reconstructed output readings with 95% uncertainty bounds for the single spacecraft attitude dynamics.