Annual Reviews in Control 47 (2019) 35–50

Contents lists available at ScienceDirect

Annual Reviews in Control

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

Review

Input design for active fault diagnosis

Tor Aksel N. Heirung, Ali Mesbah*

Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 13 January 2019
Revised 5 March 2019
Accepted 8 March 2019
Available online 27 March 2019

Keywords:
Active fault diagnosis
Optimal input design
Model discrimination
Bayesian decision theory
System uncertainty

ABSTRACT

Reliably diagnosing faults and malfunctions has become increasingly challenging in modern technical systems because of their growing complexity as well as increasingly stringent requirements on safety, availability, and high-performance operation. Traditional methods for fault detection and diagnosis rely on nominal input–output data, which can contain insufficient information to support reliable conclusions. Recent years have witnessed a growing interest in active fault diagnosis, which addresses this issue by injecting input signals specifically designed to reveal the fault status of the system. This paper provides an overview of state-of-the-art methods for input design for active fault diagnosis and discusses the primary considerations in the formulation and solution of the input-design problem. We also discuss the primary challenges and suggest avenues for future research in this rapidly evolving field.

Contents

---

* Corresponding author.
E-mail address: mesbah@berkeley.edu (A. Mesbah).

https://doi.org/10.1016/j.arcontrol.2019.03.002
1367-5788/

1. Introduction

Faults and malfunctions can happen in any modern technical system, with potentially detrimental effects on safety, performance, reliability, environmental footprint, and economics. In 2013, every Boeing 787 Dreamliner was grounded indefinitely after battery failures had occurred in two planes, with enormous consequences for the finances and reputations of the affected airlines, the manufacturer, and its suppliers (Williard, He, Hendricks, & Pecht, 2013). The 2005 series of explosions and fires at the BP refinery in Texas City, in part caused by an overflowing isomerization column, resulted in 15 fatalities and 180 injuries (Khan & Amyotte, 2007; Manca & Brambilla, 2012). Before losing control of El Al Flight 1862 in the 1992 accident, in which both engines on the starboard side detached because of material fatigue, the pilots were able to keep the plane in the air for almost fifteen minutes. Had the fault been diagnosed during this time, followed by appropriate action, the disaster could have been averted (Alwi, Edwards, Stroosma, & Mulder, 2008; Maciejowski & Jones, 2003). Reliable and timely diagnosis of faults is not only critical to safety, reliability, availability, and maintainability of a system, it is also essential in ensuring a system's ability to function as designed (Isermann, 2006). However, the growing complexity and strict performance requirements of modern technical systems have made reliable fault diagnosis increasingly challenging.

Fault diagnosis is generally a multi-step process, commonly including fault *detection, isolation, identification*, and *estimation*. Informally, these terms in turn refer to: determining whether or not the system is fault free; if not, which part of the system is faulty; the type of fault that has occurred in that part; and the magnitude of the fault (e.g., Blanke, Kinnaert, Lunze, & Staroswiecki, 2006). This paper deals with fault diagnosis in its entirety, rather than treating these activities individually. In particular, we focus on the problem of enhancing fault diagnosis through the design of system inputs, which is known as *active fault diagnosis*, or AFD. The remainder of this section gives an overview of some common types of faults, contrasts the active and passive approaches to fault diagnosis, discusses advantages of the active approach, highlights some connections to related branches of the control literature, and states the objective of the paper.

For clarity, we conform to the common practice of distinguishing between faults and failures, since these two terms are sometimes conflated in the literature. While a fault may cause a reduction in a system's ability to perform the tasks for which it is designed, a failure is generally understood as an event that renders the system inoperable. The two terms can thus be defined as follows (after Blanke et al., 2006; Isermann, 2006; Varga, 2017).

Definition 1 (Fault)**.** A *fault* in a dynamic system is an anomalous variation in a characteristic system property that causes an unacceptable deviation from the specified limits of normal operation.

Definition 2 (Failure)**.** A *failure* is generally an irrecoverable event that renders the system incapable of operating such that it fulfills its purpose.

Hence, a failure is more critical than a fault, and a fault may lead to a failure unless diagnosed and managed appropriately.

Much of the literature makes a distinction between faults that arise from structural and gradual changes in the system. *Structural* changes are discrete events, such as actuators that are stuck in some position, the complete loss of a sensor, or a system component that breaks entirely. Faults arising from structural changes are often *abrupt*. Conversely, faults that stem form *gradual* changes can increase in severity or magnitude over time; examples include actuators that become slower to respond because of wear, sensor biases, and system components that suffer from issues like leaks or changing material characteristics. *Incipient* faults are in their earliest stages, primarily of the gradual type. Finally, a structural fault about to happen is *impending*.

1.1. Active versus passive approaches to fault diagnosis

The growing complexity of modern technical systems has made faults possibly more frequent and harder to diagnose. Generally, an important consideration in the design of technical systems is the potential occurrence of faults and failures to ensure some level of inherent robustness to such anomalies through the system design. For example, sensor and actuator redundancy can enable graceful degradation of system performance in the event of certain faults. Nonetheless, systematically accounting for all potential faults in the system design stage is impractical or impossible. This has motivated the use of fault diagnostics during operation, which are typically developed once the system is designed. However, a complex design, as well as feedback control and system uncertainties, can significantly limit the ability to diagnose faults (Sampath, Lafortune, & Teneketzis, 1998). Therefore, there has been a growing interest in the development of methods for faster and more reliable fault diagnosis during operation (Campbell & Nikoukhah, 2004; Zhang, 1989).

Fault diagnosis approaches are commonly classified as *active* or *passive*. The latter approach, also known as non-invasive, generally relies on comparing recorded input–output data to some reference data, which can be historical or generated through simulation. Importantly, the system is not perturbed to investigate its fault status. Comprehensive survey papers (e.g., Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003, Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003) and a growing number of textbooks, such as Chen and Patton (1999), Chiang, Russell, and Braatz (2001), Blanke et al. (2006), Isermann (2006), Gonzalez, Qi, and Huang (2016), and Varga (2017), discuss passive methods in detail. Algorithms for passive fault diagnosis are broadly classified as data or model based (Venkatasubramanian, Rengaswamy, Yin, et al., 2003; Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003). Model-based methods generally require a model for every fault. These models are often based on first principles, but can also be identified from data (e.g., see Ljung, 1999). In contrast, data-based methods rely less on domain knowledge about the system, with a stronger focus on analysis of large historical data sets to characterize fault-free and different types of faulty operation.

A shortcoming of passive approaches arises from the potential lack of diagnostically relevant information in the input–output data generated while the system is operated under the assumption that no fault has occurred. We refer to this as *normal* operation, and say the system then generates *nominal* input–output data. Note that normal operation does not imply that no fault has occurred, and is thus distinct from *fault-free operation*.[1] That is, passive approaches do not account for the fact that nominal operating data may not be sufficiently informative for reliable fault diagnosis. Common reasons of this lack of diagnostically relevant information include system uncertainties and the presence of feedback controllers. Incomplete knowledge, or uncertainty, about the system and its state can result from inadequate measurements (including issues such as low signal-to-noise ratio, which lowers the information content in the measurements) and system disturbances that may not be readily distinguishable from faults through analysis of nominal operating data. Similarly, feedback controllers, the purpose of which

---

[1] Some authors use *nominal* as an antonym for faulty in the context of operation and models, a convention we do not follow here.