A Closed-Loop Perspective on Fault Detection for Precision Motion Control: With Application to an Overactuated System

Koen Classens
dept. of Mechanical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands
K.H.J.Classens@tue.nl

W.P.M.H. (Maurice) Heemels
dept. of Mechanical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands
W.P.M.H.Heemels@tue.nl

Tom Oomen
dept. of Mechanical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands
T.A.E.Oomen@tue.nl

Abstract—Fault diagnosis is crucial in high-tech production equipment to minimize operational downtime and to facilitate targeted maintenance. Future high-tech systems have numerous complex closed-loop control systems and require compatible fault diagnosis systems. The aim of this paper is to develop a procedure for decentralized fault detection in the presence of additional feedback interconnections. The influence of the additional feedback interconnections on the fault diagnosis system is investigated by means of an illustrative experimental study that resembles a next generation flexible motion system.

I. INTRODUCTION

The movement towards predictive maintenance for mechatronic systems is driven by the high cost associated with unscheduled downtime. To minimize this cost, targeted maintenance is essential for future precision mechatronics, and is enabled by a combination of dedicated fault diagnosis systems and digital twins [1]–[7]. The field of fault diagnosis is broadly employed in safety-critical domains such as the chemical industry, aerospace and automotive. The present paper focuses on the domain of precision motion control.

As the high-tech industry is pushing for more throughput, motion systems are becoming increasingly more lightweight, and as a consequence, more flexible. To compensate for structural deformations introduced by this flexibility, motion systems are often equipped with additional actuators and sensors. This overactuation and oversensing allows to create additional closed-loop feedback interconnections, providing increased freedom to improve performance [8].

Fault diagnosis systems are usually based on parametric first principle models of the open-loop system. For instance, based on parametric models, nullspace-based fault detection and isolation (FDI) methods have been developed [9], [10] and enable fault diagnosis for large-scale complex systems. However, for precision mechatronics, data-driven modeling as opposed to first principles modeling, is fast, accurate, and inexpensive [11]–[13]. In addition, precision mechatronics operate in closed loop.

Although important progress has been made in fault detection for complex engineered systems, at present closed-loop aspects in fault diagnosis have not been clarified. The present research is driven by a lack of integral procedure for precision motion systems, and in particular, feedback controlled overactuated systems.

This paper is organized as follows. Section II addresses the problem formulation for flexible mode controlled systems and for closed-loop fault diagnosis. Moreover, an illustrative experimental setup is introduced. The procedure to decouple the flexible modes for control is presented in Section III, and the fault detection filter design is presented in Section IV. The proposed approach is applied to a relevant experimental prototype, described in Section V. Finally a conclusion is given in Section VI.

II. PROBLEM FORMULATION

Exploiting overactuation to compensate for structural deformations typically involves complex control configurations, introducing additional feedback interconnections. These additional loops have large consequences for FDI and system estimation for FDI [14]. This paper aims to address the implications on FDI by answering the following questions.

i) What is an overactuated and oversensed system and why is it relevant for next-generation motion systems? (Sections II-A and II-B)

ii) How to decouple an overactuated and oversensed system for flexible mode control? (Section III)

iii) How to augment an overactuated and oversensed system by a fault diagnosis system for decentralized fault detection? (Sections II-C and IV)

iv) Which model should be used to serve as basis for the fault diagnosis system design? In particular, where accurate models have to be obtained from data. (Sections II-C and IV)

v) What are the consequences of neglecting the additionally created control loops on the performance of the fault diagnosis system? (Section V)

The answers to these questions form important insights, crucial for fault diagnosis system design for precision motion systems.

A. Framework with additional actuators and sensors

In this subsection, the framework for adding additional actuators and sensors for control is introduced systematically. Consider the standard feedback interconnection as depicted in Figure 1, where the linear time-invariant (LTI) MIMO system
is described by the transfer function matrix (TFM) $G_u(s)$, $s \in \mathbb{C}$, and the controller by $C(s)$. The $k$ inputs to the plant are denoted by $u(t) \in \mathbb{R}^k$, $t \in \mathbb{R}_+$, and the $m$ outputs are denoted by $y(t) \in \mathbb{R}^m$, which means that $G_u : u \mapsto y$. Subtracting the output from the reference $r(t) \in \mathbb{R}^m$ forms the tracking error $e(t) := r(t) - y(t)$.

Adding $l$ additional actuators $u^{\text{ext}}(t) \in \mathbb{R}^l$, and $n$ additional sensors $y^{\text{ext}}(t) \in \mathbb{R}^n$, gives the extended configuration depicted in Figure 2, where an additional feedback interconnection is created through the controller $C^{\text{ext}}(s)$.

**Definition 1** (Extended plant) Consider Figures 1 and 2, where the relation between the plant $G_u$ and the extended plant $G_u^{\text{ext}} : \begin{bmatrix} u & u^{\text{ext}} \end{bmatrix}^\top \mapsto \begin{bmatrix} y & y^{\text{ext}} \end{bmatrix}^\top$, can be described as

$$\begin{bmatrix} I & 0 \end{bmatrix} G_u^{\text{ext}} \begin{bmatrix} I \\ 0 \end{bmatrix} = G_u.$$  \hfill (1)

**Definition 2** (Equivalent plant) Consider Figure 2, where the equivalent plant $G_u^{\text{eq}} : u \mapsto y$ is formed by $F_1(G_u^{\text{ext}}, -C^{\text{ext}})$, the lower linear fractional transformation (LFT), i.e.,

$$G_u^{\text{eq}} = G_{u,11}^{\text{ext}} - G_{u,12}^{\text{ext}} (I + G_{u,22}^{\text{ext}})^{-1} G_{u,21}^{\text{ext}}.$$  \hfill (2)

**B. Motivational example for flexible mode control**

Conceptually, it is self-evident that increasing the number of control inputs and outputs provides additional freedom to improve performance. To illustrate this concept, consider the flexible two degree of freedom (DOF) system depicted in Figure 3, where the system is weakly connected to the fixed world via the springs with spring constant $k_1 = k_3 \ll k_2$ and dampers with damping constant $d_1 = d_3 \ll d_2$. In this motivational example, the system is excited at the first mass by force $F_1$, whereas performance is desired at the position of the second mass $x_2$. This behavior is described by the non-collocated plant $G_{u,\text{ncol}} : F_1 \mapsto x_2$, see Figure 5 (top left). Conventionally, the performance of these flexible systems is limited by the phase lag at the resonance frequency, see e.g. [8] for details. For flexible motion systems, the introduction of an additional actuator and sensor allows to achieve beyond conventional performance limits. Consider the extended flexible benchmark system shown in Figure 4, where the system is augmented with an additional sensor, able to measure $x_{\text{diff}} := x_2 - x_1$. An additional actuator allows to exert a force $F_{\text{diff}}$ in opposite direction to both masses. This results in the extended system

$$G_{u,\text{ncol}}^{\text{ext}} : \begin{bmatrix} u \\ u^{\text{ext}} \end{bmatrix} \mapsto \begin{bmatrix} y \\ y^{\text{ext}} \end{bmatrix} = \begin{bmatrix} F_1 \\ F_{\text{diff}} \end{bmatrix} \mapsto \begin{bmatrix} x_2 \\ x_{\text{diff}} \end{bmatrix}. \hfill (3)$$

Closing the feedback interconnection between $y^{\text{ext}}$ and $u^{\text{ext}}$, as in Figure 2, with the controller

$$C^{\text{ext}}(s) = k_p + k_d s,$$  \hfill (4)

allows to shift the effective resonance peak of the equivalent plant $G_u^{\text{eq}}$. Hence, the additional actuator and sensor allow to increase the bandwidth of the controlled system to a higher frequency and therefore achieve beyond conventional performance. In Figure 5, the open-loop system $G_{u,\text{ncol}}^{\text{ext}}(s)$...
is depicted, containing \( G_{u,ncol}(s) \) on its \((1,1)\)-entry. Additionally, the equivalent \( G_u^{eq}(s) \) is depicted which allows to increase the bandwidth of the controlled system to a higher frequency, and therefore improve performance.

**C. Framework for closed-loop fault detection**

Considering overactuated motion systems with additional feedback interconnections operating in the background, the following is assumed regarding the fault diagnosis system.

**Assumption 1** The additional sensor signals \( y^{ext} \) and additional actuator signals \( u^{ext} \) are not available to the fault diagnosis system.

The configuration for closed-loop fault detection is illustrated in Figure 6, where additive faults \( f(t) \in \mathbb{R}^p \) affect the output through the TFM \( G_f(s) \) and a disturbance \( d(t) \in \mathbb{R}^p \) affects the output through the TFM \( G_d(s) \). Hence, the output is given by

\[
y = (I + G^{eq}_u C)^{-1} G_u r + (I + G^{eq}_u C)^{-1} G_f f + (I + G^{eq}_u C)^{-1} G_d d,
\]

and the input is given by

\[
u = (I + CG_u^{eq})^{-1} r - (I + CG_u^{eq})^{-1} CG_f f - (I + CG_u^{eq})^{-1} CG_d d.
\]

The system is augmented by a residual generator formed by a proper and stable TFM \( Q := [Q_y \ Q_u] \). The residual \( \varepsilon \), used for fault detection, is equal to

\[
\varepsilon = Q_u u + Q_y y.
\]

Loosely speaking, the fault detection goal is that the residual \( \varepsilon \neq 0 \), i.e., is sufficiently larger than zero in the presence of faults and the residual \( \varepsilon \approx 0 \), i.e., is sufficiently small in the absence of faults.

**Remark 1** Note that by setting \( C = C^{ext} = 0 \) and considering \( r = u \), the standard open-loop fault detection problem is obtained [10], where \( y \) is given by

\[
y = G_u u + G_f f + G_d d.
\]
III. MODAL DECOUPLING FOR CONTROL

To suppress flexible modes, it is beneficial to first apply a sensor and actuator transformation to decouple the mode-shapes of the system. In addition, the modal framework is cast in the control framework presented in Section II-A.

Consider a physical model of the system $G_u(s)$ with ordered inputs $\bar{u} := \begin{bmatrix} u_1 & \ldots & u_{k+1} \end{bmatrix}^\top$ and outputs $\bar{y} := \begin{bmatrix} y_1 & \ldots & y_{m+n} \end{bmatrix}^\top$. A sensor transformation and an actuator transformation are employed to transform the system into the modal form, given by

$$G_u^{\text{mod}}(s) = T_y G_u^{\text{ext}}(s) T_u,$$

(9)

where $T_u = (\Phi^{-1})^\top$ and $T_y = \Phi^{-1}$ define the transformation to the modal form. The transformation to modal inputs $u_m$ and model outputs $y_m$ is achieved with the relations $u_m = T_u^{-1} \bar{u}$ and $y_m = T_y \bar{y}$. The matrix $\Phi$ can be found by solving the generalized eigenvalue problem of the undamped equations of motion, i.e., $\Phi$ contains the eigenvectors resembling the various modes.

Remark 4 For the motivational example, described in Section II-B, $\Phi = [\phi_1 \ \phi_2]$, where $\phi_1 = [1 \ 1]^\top$ corresponds to the rigid body mode and $\phi_2 = [-1 \ 1]^\top$ represents the flexible mode.

In order to retain the structure presented in Section II-A, condition (1) must be fulfilled. I.e., after modal decoupling, the equivalent model $G_u^{\text{ext}}(s)$ must have the same input-output pair as the original open-loop system $G_u(s)$. To that end, the transformation matrices $T_u$ and $T_y$ are partitioned as

$$T_y = \begin{bmatrix} T_{y_1,1} \\ T_{y_2,1} \end{bmatrix}, \quad T_u = \begin{bmatrix} T_{u_1,1} & T_{u_2,1} \end{bmatrix}.$$

The first part of these transformation matrices is replaced by $\tilde{T}_{y,1}$ and $\tilde{T}_{u,1}$, giving

$$\tilde{T}_y = \begin{bmatrix} \tilde{T}_{y_1,1} \\ \tilde{T}_{y_2,1} \end{bmatrix}, \quad \tilde{T}_u = \begin{bmatrix} \tilde{T}_{u_1,1} & \tilde{T}_{u_2,1} \end{bmatrix},$$

such that condition (1) is fulfilled after the transformation

$$G_u^{\text{ext}}(s) = \tilde{T}_y G_u^{\text{ext}}(s) \tilde{T}_u.$$

(10)

Hence, $\tilde{T}_{y,1}$ and $\tilde{T}_{u,1}$ are chosen such that the result after transformation is cast into the framework of Section II-A, and thus $G_u : u \mapsto y$.

IV. FAULT DETECTION FILTER DESIGN

To formalize the closed-loop fault detection problem, presented in Section II-C, the closed-loop fault detection approach presented in [11] is employed. Substitution of (5) and (6) into (7) results in the residual

$$\varepsilon = (Q_u + Q_y G_u^{\text{ext}}) r + (Q_y G_f) f + (Q_y G_d) d.$$

(11)

The residual generator $Q$ can always be parameterized such that the residual $\varepsilon$ is decoupled from the reference $r$ by the parameterization $Q_u = -Q_y G_u^{\text{eq}}$. The effects of the noise $d$ can usually not be decoupled from $\varepsilon$. Hence, $Q_y$ should be designed to achieve that the residual $\varepsilon$ is significantly influenced by all fault entries $f_i$, where $i = 1, \ldots, q$, and the influence of the noise signal $d$ is negligible.

Specifically, the objective is to maximize the gap between the fault detectability and noise attenuation. An optimization-based approach is employed following [16, Chapter 5]. Given $\gamma \geq 0$, a stable and proper optimal fault detection filter $Q$ and corresponding optimal fault sensitivity level $\beta > 0$ need to be determined such that

$$\beta = \max_Q \left\{ \|G_{\varepsilon f}\|_{\infty}^{-1} \|G_{\varepsilon d}\|_{\infty} \leq \gamma \right\},$$

(12)

where the index

$$\|G_{\varepsilon f}\|_{\infty} := \min_{1 \leq i \leq q} \|G_{\varepsilon f_i}\|_{\infty},$$

(13)

is used as a sensitivity measure, covering globally all fault inputs.

V. APPLICATION TO FLEXIBLE BEAM SETUP

This case study demonstrates the procedure proposed in this paper. First, the flexible beam setup, introduced in Section II-D, is decoupled with the aim for modal control. Subsequently, a data-driven model is estimated and parameterized on which the fault diagnosis system is based. In addition, it is shown that accurate models are a necessity for fault diagnosis design and that it is a common misconception that the open-loop plant model can be used. In fact, it severely compromises the fault diagnosis system.
A. Modal decoupling and control

Consider $\hat{G}_u(s) : [u_1\ u_2\ u_3]^\top \mapsto [y_1\ y_2\ y_3]^\top$ and the modeshapes depicted in Figure 9. From these modeshapes is concluded that

$$\Phi^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & -1 \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix}. \tag{14}$$

Employing transformation (9) allows to identify the modal decoupled system $\hat{G}_u^{mod} : [u_{trans}\ u_{rot}\ u_{flex}]^\top \mapsto [y_{trans}\ y_{rot}\ y_{flex}]^\top$. To this end, open-loop experiments are performed, exciting one modal input at a time, using random phase multisines up to a frequency of 1000 Hz [13]. The identified model is depicted in Figure 10.

Examination of the relative gain array, $\Lambda(G) = G \circ (G)^{-\top} \approx I$, shows that the modal system is indeed decoupled. Due to the decoupling, a decentralized control solution may be employed in order to control the modeshapes. As described in Section II-D, the aim is to achieve performance at $y = y_2$ with the actuators $u = \frac{1}{2}(u_1 + u_3)$. Hence choosing $\hat{T}_{u,1} = [0 \ 1 \ 0]$ and $\hat{T}_{u,3} = [\frac{1}{2} \ 0 \ \frac{1}{2}]^\top$ results in the extended plant $\hat{G}_u^{eq}$, c.f. (10), of the form

$$\begin{bmatrix} y_2 \\ y_{rot} \\ y_{flex} \end{bmatrix} = \hat{G}_u^{eq} \begin{bmatrix} \frac{1}{2} (u_1 + u_3) \\ u_{rot} \\ u_{flex} \end{bmatrix}. \tag{15}$$

First, the flexible mode is controlled with the control law $u_{flex} = -C_{flex} y_{flex}$. A gain and a lead-lag filter are used as controller in order to reduce the effective resonance peak and shift it to a higher frequency, see Figure 11. The limiting factor, preventing to move the resonance to an even higher frequency, is the time-delay, estimated to be equal to $T_d = 9.25 \cdot 10^{-4}$ s. By means of a sequential loop-closing procedure, the rotational free-body mode is suppressed by means of $u_{rot} = -C_{rot} y_{rot}$, where $C_{rot}$ consists of a gain and a lead-lag filter.

Finally, a FRF of $G_u$ as well as the equivalent plant $\hat{G}_u^{eq}$ are estimated by means of open-loop random phase multisine experiments [13], resulting in the FRFs depicted in Figure 12.

Based on $\hat{G}_u^{eq}$, the last controller $C(s)$ is designed, consisting of a gain, a lead-lag filter, two notch filters, a second order low-pass and an integrator. $\hat{G}_u(s)$ and $\hat{G}_u^{eq}(s)$ are parameterized by means of the instrumental variable approach. Overall, the deviations between $G_u$ and $\hat{G}_u^{eq}$ are marginal, but the newly introduced feedback interconnections have a significant impact on the fault diagnosis design, as illustrated next.

B. Actuator fault detection

Based on the identified parametric models, two fault detection filters are designed, following the procedure in Section IV.

Assumption 2 It is assumed that only the first and third actuator are sensitive to faults. Hence, it is assumed that only faults enter the system in $u$, shown in Figure 6.

Assumption 3 It is assumed that the disturbance $d$ affects the output mainly at higher frequencies, i.e. through the filter $G_d(s) = \frac{250s + 7.9 \cdot 10^4}{50s + 7.9 \cdot 10^4}$.

Two residual generators are compared on the basis of the following assumptions.

1. Including interaction, i.e., based on $\hat{G}_u^{eq}(s)$, $G_f(s) = \hat{G}_u^{eq}(s)$ and $G_d(s) = \frac{250s + 7.9 \cdot 10^4}{50s + 7.9 \cdot 10^4}$.

2. Neglecting interaction, i.e., based on $\hat{G}_u(s)$, $G_f(s) = \hat{G}_u(s)$ and $G_d(s) = \frac{250s + 7.9 \cdot 10^4}{50s + 7.9 \cdot 10^4}$.
Fig. 13. Actuator fault detection on flexible beam setup, with scaled actuator fault signal (—). The residual signals corresponding to the filters based on $\hat{G}_u(s)$ (—) and $G_u(s)$ (—) are depicted as well as their moving averages over a period of 1 s, indicated by (−−) and (−−) respectively. Both residual signals increase in the presence of the fault, however, the signal based on $\hat{G}_u(s)$ shows significant traces of the reference signal.

Fig. 14. Actuator fault detection on flexible beam setup, with scaled actuator fault signal (—). The residual signals corresponding to the filters based on $\hat{G}_u^\text{eq}(s)$ (—) and $G_u(s)$ (—) are depicted as well as their moving averages over a period of 1 s, indicated by (−−) and (−−) respectively. The residual signal based on $\hat{G}_u(s)$ shows significant traces of the reference signal.

The first case, see Figure 13, follows the reference $r(t) = 5 \cdot 10^{-4} \sin(\pi t)$ and has an additive drift as actuator fault. Both residual generators give an interpretable residual, enabling to detect the fault. However, the signal based on $\hat{G}_u(s)$ has larger perturbations during the fault as well as in the non-faulty regions, complicating fault detection.

The second case, see Figure 14, attempts to track the reference $r(t) = 5 \cdot 10^{-4} (\sin(2\pi t) + \sin(208t))$ and has a step as additive actuator fault. The residual generator based on $\hat{G}_u(s)$ gives a poorly interpretable signal because the reference has a significant frequency contribution near the resonance peaks, where the effect of neglecting interaction becomes apparent, see (11) where $G_{rr} \neq 0$. This illustrative example shows that using a model with a marginal mismatch for fault diagnosis system design, e.g. by neglecting interaction, results in a poorly interpretable residual signal due to large traces of the reference signal.

VI. CONCLUSION

The presented approach addresses ambiguity in fault detection for closed-loop systems. By means of an experimental setup, exhibiting flexible behavior similar to next-generation positioning systems, it is shown that neglecting interaction between submodules results in a severely compromised fault diagnosis system. A solution is proposed where submodules are identified and used as basis for the fault diagnosis system. The presented procedure shows that, provided appropriate system boundary, the framework for fault detection can serve as the basis for digital twins for complex mechatronic systems.

ACKNOWLEDGEMENT

This work is supported by Topconsortia voor Kennis en Innovatie (TKI), and is supported by ASML Research, Veldhoven, The Netherlands.

REFERENCES