A behavioral approach to direct data-driven fault detection

Ivan Markovsky, Alba Muixí, Sergio Zlotnik, and Pedro Diez

***Abstract*— Efficient and reliable fault detection methods are needed for monitoring and evaluation of processes, *e.g.*, in structural health assessment. Most existing methods, however, rely on a priori given model. Thus, these methods require a model suitable for fault detection. Obtaining such a model is nontrivial and is often the bottleneck in applications. Direct data-driven methods were recently developed in signal processing and control. These methods avoid the model identification step and were shown to outperform state-of-the-art model-based methods in practical applications. In this paper, we propose a direct data-driven method for fault detection. The monitored process is modeled as a linear time-invariant system with unobserved deterministic disturbance. We use the behavioral approach to systems theory in order to define a representation invariant measure for the distance between data and model. The main contribution of the paper is computing the distance directly from offline and online data without parametric model identification. A second contribution of the paper is a direct data-driven method for input estimation.**

***Index Terms*— fault detection, input estimation, direct data-driven methods, behavioral approach.**

I. Introduction

Fault detection is a real-time monitoring problem aimed to determine based on data of a dynamical system if the system is in a "healthy" mode of operation [1], [2]. The healthy mode of operation, as well as possible "faulty" modes of operation, are specified by parametric models or data collected offline when the system in the corresponding mode.

There are two fundamentally different approaches for fault detection. The first one is based on parametric model identification [3]. The existence of a fault is detected by monitoring the estimated model parameters. The rationale for this approach is that a fault causes a change in the model parameters. The second approach is based on a signal, call a *residual*, that measures the discrepancy between the data and the model. The occurrence of a fault is detected by thresholding the residual signal. The rationale for this approach is that the residual is "small" (ideally zero) when the system is in the corresponding mode of operation and "large" otherwise. Residuals are defined in terms of representations of the system. For linear of time-invariant systems, the representation could be impulse response, transfer function, or state-space, and possible residuals are the prediction error, output error, or equation error.

I. Markovsky is with Catalan Institution for Research and Advanced Studies and Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE) (e-mail: imarkovsky@cimne.upc.edu), A. Muixí, S. Zlotnik, and P. Diez are with CIMNE and the Universitat Politècnica de Catalunya, LaCaN, DECA, (e-mails: alba.muixi@upc.edu, sergio.zlotnik@upc.edu, pedro.diez@upc.edu).

The approach in this paper is reminiscent to the residual thresholding approach, however, it uses a data–model discrepancy measure that is not based on a representation of the system and can be computed directly from data. The new *representation invariant* distance measure, is defined in the behavioral setting, where dynamical systems are viewed as sets of trajectories [4]. This view is naturally suited for data-driven analysis and design. In contemporary machine learning language, the behavioral setting is non-parametric and unsupervised since the data does not have to be labeled into inputs and outputs.

In the behavioral setting, a natural choice for the discrepancy, called *misfit*, between the system and a signal is the orthogonal distance from the signal to the system, *i.e.*, the projection of the signal on the system. This operation is equivalent to errors-in-variables Kalman smoothing [5]. It is statistically optimal under the assumption that the signal is generated in the errors-in-variables setting, *i.e.*, it is a trajectory of the system corrupted by measurement error, which is assumed to be a zero mean, white, Gaussian process. The misfit attributes the lack of fit between the data and the model to measurement error. An alternative data–model discrepancy measure, called *latency*, attributes the lack of fit between the data and the model to an unobserved latent signal, referred to as *disturbance* [6]. In the latency setting, the model describes the joint dynamics of the disturbance and the observed variables. The latency is defined as the size of the smallest disturbance that makes the observations compatible with the model. The latency computation corresponds to Kalman smoothing [7]. As the misfit, the latency is also statistically optimal under suitable stochastic assumptions about the disturbance signal.

We define a new distance measure that combines misfit and latency with disturbance signal that is an unknown deterministic signal. Relaxing the assumptions about the disturbance allows us to apply the new distance in applications where there is no prior information about the disturbance or either of the zero-mean, whiteness, and Gaussianity assumptions is not satisfied. The new measure reduces to the misfit when the system has no disturbance. Also, the new distance measure is zero when there is no measurement error, *i.e.*, the signal is exact. As a byproduct of its computation, the disturbance signal is estimated. Thus, an independent contribution is a new input estimation method.

Based on the representation invariant distance measure, we propose a fault detection method. In addition to being representation invariant, the new distance measure and resulting fault detection method allow for direct computation from offline data of the system, bypassing the parametric model identification required by model-based methods. The approach of using data as a representation of the system is called *direct data-driven* and is successfully used in signal processing and control, where it is shown to have advantages over alternative model-based methods [8], [9], [10]. The direct data-driven method proposed in the paper applies to data obtained from a transient response, forced response due to observed excitation signal, as well as forced response due to unobserved excitation signal. It can be computed efficiently in real-time and is validated on simulated data.

Section II introduces the terminology, notation, and results from the behavioral systems theory that are used in the paper. Section III defines the concepts of misfit, latency, and the new distance measure that unifies them. Section IV presents the direct data-driven fault detection method based on the new distance measure. Section V validates the method on simulated data.

II. Preliminaries and notation

We use the behavioral approach [4], [11], [9]. A real-valued q-variate signal w with time axis $\mathscr{T} \subset \mathbb{R}$ is a map from $\mathscr{T}$ to $\mathbb{R}^q$. The set of signals $w : \mathscr{T} \to \mathbb{R}^q$ with q variables is denoted by $(\mathbb{R}^q)^{\mathscr{T}}$. In this paper, the signals are discrete-time and $\mathscr{T} = \mathbb{N}$ — the time axis is the set of natural numbers. The *unit shift operator* is $(\sigma w)(t) := w(t+1)$.

In the behavioral setting, a *dynamical system* $\mathscr{B}$ with q variables is defined as a subset of the set of signals $(\mathbb{R}^q)^{\mathscr{T}}$. A system $\mathscr{B}$ is *linear* if $\mathscr{B}$ is a linear subspace of $(\mathbb{R}^q)^{\mathscr{T}}$ and *time-invariant* if $\mathscr{B}$ is invariant to the action of the shift operator, *i.e.*, $\sigma\mathscr{B} = \mathscr{B}$. The set of *linear time-invariant systems* with q variables is denoted by $\mathscr{L}^q$.

The *restriction* of a signal $w \in (\mathbb{R}^q)^{\mathbb{N}}$ and a system $\mathscr{B} \subset (\mathbb{R}^q)^{\mathbb{N}}$ to the interval $1,\dots,T$ is denoted by $w|_T$ and $\mathscr{B}|_T$, respectively. The *restricted behavior* $\mathscr{B}|_T$ is a subspace of the set $(\mathbb{R}^q)^T$. When $\mathscr{B}$ is linear time-invariant,

$$\dim \mathscr{B}|_T = mT + n, \qquad \text{for all } T \geq \ell,$$

where m, ℓ, and n are natural numbers that are properties of the system $\mathscr{B}$:

- m is the *number of inputs* (in an input/output representation of the system),
- ℓ, called the *lag* of $\mathscr{B}$, is the minimal degree of a difference equation representation of $\mathscr{B}$, and
- n, called the *order* of $\mathscr{B}$, is the minimal total degree of a difference equation representation of $\mathscr{B}$.

The triple (m,ℓ,n) characterizes the *complexity* of $\mathscr{B} \in \mathscr{L}^q$. The set of *linear time-invariant systems* with q variables and complexity bounded by (m,ℓ,n) is denoted by $\mathscr{L}^q_{(m,\ell,n)}$.

The variables w can be partitioned into *inputs* u (free variable) and *outputs* y (dependent variable) via a permutation matrix $\Pi \in \mathbb{R}^{q\times q}$, *i.e.*, $w = \Pi \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$. The inputs u can be chosen freely while the outputs y are uniquely defined by the model, the given inputs u, and the initial conditions. As shown in [12, Lemma 1], the *initial conditions* for a trajectory $w = \big(w(1),\dots,w(T)\big)$ can be specified by $T_{\text{ini}} \geq \ell$ "past" samples $w_{\text{ini}} = \big(w(-T_{\text{ini}}+1),\dots,w(0)\big)$. A partitioning of the variables into inputs and outputs is not unique. In the context of fault detection, we use an input/output partitioning in order to model user defined excitation signals and disturbances.

The restricted behavior $\mathscr{B}|_T$ of a linear time-invariant system $\mathscr{B} \in \mathscr{L}^q$ is an $r := \dim \mathscr{B}|_T = Tm + n$ dimensional subspace and, therefore, it can be represented by a basis

$$\mathscr{B}|_T = \operatorname{image} B, \quad \text{where } B := \begin{bmatrix} b^1 & \cdots & b^r \end{bmatrix} \in \mathbb{R}^{qT\times r}. \quad (B)$$

We refer to the matrix B of the basis vectors as the basis. The representation (B) of $\mathscr{B}|_T$ is *nonparameteric.* A trajectory $w \in \mathscr{B}|_T$ is specified using the data-driven representation (B) via the equation $w = Bg$, where $g \in \mathbb{R}^r$. For $T \geq \boldsymbol{\ell}(\mathscr{B}) + 1$, the basis B for the finite-horizon behavior $\mathscr{B}|_T$ uniquely defines the system $\mathscr{B}$ [13, Lemma 13].

Consider a finite trajectory $w_{\text{d}} \in \mathscr{B}|_{T_{\text{d}}}$ (the subscript index "d" stands for "data") of a bounded complexity linear time-invariant system $\mathscr{B} \in \mathscr{L}^q_{(m,\ell,n)}$. The following result from [13, Theorem 17], is used for obtaining a basis B for $\mathscr{B}|_T$ from the data w_{d}. Define the Hankel matrix $\mathscr{H}_L(w_{\text{d}})$ with depth L

$$\mathscr{H}_L(w_{\text{d}}) := \begin{bmatrix} w|_L & (\sigma w)|_L & \cdots & (\sigma^{T_{\text{d}}-L} w)|_L \end{bmatrix} \in \mathbb{R}^{qL\times(T_{\text{d}}-L+1)}.$$

For any $L \geq \ell$, $\mathscr{B}|_L = \operatorname{image} \mathscr{H}_L(w_{\text{d}})$ if and only if

$$\operatorname{rank} \mathscr{H}_L(w_{\text{d}}) = mL + n. \qquad \text{(GPE)}$$

The condition (GPE), called *generalized persistency of excitation*, is verifiable from the data w_{d} and the model's complexity (m,ℓ,n). The result of [13, Theorem 17] is a generalization of the *fundamental lemma* [14]. For detailed discussion on the similarities and differences between the generalized persistency of excitation and the conditions of the fundamental lemma, see [15].

Linear time-invariant systems can be represented in different ways by equations. The most popular ones—convolution, transfer function, and state-space—assume a priori given input/output partitioning of the variables. In this section, we do not review parametric representations because they are not used in the paper. For more details on the behavioral approach to systems theory, it's relation to the classical input/output approach, and its relevance to direct data-driven signal processing and control, we refer the reader to [9].

III. Data–model discrepancy measures

Stochastic system identification problems and methods can be classified into errors-in-variables [16] and auto-regressive moving-average exogenous (ARMAX). Deterministic counterparts of the likelihood functions in the errors-in-variables and the ARMAX settings are, respectively, the misfit and the latency. In this section, we propose a new distance measure that combines and generalizes the misfit and the latency.

We model the disturbance as an unknown deterministic input, *i.e.*, no prior about it is imposed. Under conditions on the system however the disturbance can be inferred from the observed variables. In this section, we assume that the system is given. In the next section, we show how the new distance measure can be estimated directly from data.

The misfit between a signal $w \in (\mathbb{R}^q)^T$ and a system $\mathscr{B} \subset \mathscr{L}^q$ is defined as the minimum norm perturbation of w that makes the perturbed signal $\widehat{w}$ consistent with the system $\mathscr{B}$,

$$\text{misfit}(w,\mathscr{B}) := \min_{\widehat{w}\in\mathscr{B}|_T} \|w-\widehat{w}\|. \qquad \text{(M)}$$

The misfit$(w,\mathscr{B})$ is the likelihood of w given $\mathscr{B}$ in the *errors-in-variables setting:* $w = \overline{w}+\widetilde{w}$, where $\overline{w} \in \mathscr{B}|_T$ and $\widetilde{w}$ is a zero-mean, white, Gaussian process [17].

Lemma 1. *Assuming that the data w is generated in the errors-in-variables setting, misfit$(w,\mathscr{B}) \leq \|\widetilde{w}\|$.*

Proof: For $w = \overline{w}+\widetilde{w}$, where $\overline{w} \in \mathscr{B}|_T$, we have that $\text{misfit}(w,\mathscr{B}) = \min_{\Delta w\in\mathscr{B}|_T} \|\widetilde{w}-\Delta w\|$. Since $\Delta w = 0 \in \mathscr{B}|_T$, $\text{misfit}(w,\mathscr{B}) \leq \|\widetilde{w}\|$. □

An alternative way of measuring the discrepancy between data and model is an unobserved signal $e \in (\mathbb{R}^{n_e})^{\mathbb{N}}$ acting on the system. In this case, the system $\mathscr{B}$ describes the extended signal $(e,w) \in (\mathbb{R}^{n_e+q})^{\mathbb{N}}$. The *latency* of $w \in (\mathbb{R}^q)^T$, given $\mathscr{B} \in \mathscr{L}^{n_e+q}$ is defined as

$$\text{latency}(w,\mathscr{B}) = \min_{(\widehat{e},w)\in\mathscr{B}|_T} \|\widehat{e}\|. \qquad \text{(L)}$$

For latency$(w,\mathscr{B})$ to be well-defined, problem (L) should have a unique solution. A necessary and sufficient condition for existence and uniqueness of solution is that w is compatible with $\mathscr{B}$, *i.e.*, $w \in \Pi_w\mathscr{B}|_T$, where Π_w is the projection of (e,w) on the w component. The condition is satisfied when 1) e is an input of $\mathscr{B}$ and 2) $n_e = p$, where p is the number of outputs of $\mathscr{B}$. These assumptions are standard in the ARMAX setting [3]: $(\overline{e},w) \in \mathscr{B}|_T$, where the disturbance $\overline{e}$ is a zero-mean, white, Gaussian process. The latency is the likelihood of w given $\mathscr{B}$ in the ARMAX setting. The following statement follows directly from the assumption that $(\overline{e},w) \in \mathscr{B}|_T$, and the definition of latency$(w,\mathscr{B})$.

Lemma 2. *Assuming that the data w is generated in the ARMAX setting, latency$(w,\mathscr{B}) \leq \|\overline{e}\|$.*

Consider next the missing data estimation problem: Given a system $\mathscr{B} \in \mathscr{L}^{n_e+q}$ with variables partitioned as (e,w) and a signal $w \in (\mathbb{R}^q)^T$, find e, such that $(e,w) \in \mathscr{B}|_T$. The goal is to ensure *exact* recovery of e from w, *i.e.*, there should be a unique signal $e \in (\mathbb{R}^{n_e})^T$ that is compatible with the data w and the model $\mathscr{B}$. The following result from [18] gives necessary and sufficient conditions for exact recovery.

Lemma 3. *Consider a system $\mathscr{B} \in \mathscr{L}^{n_e+q}_{(m,\ell,n)}$, let $w \in \Pi_w\mathscr{B}|_T$, and let B be a basis for $\mathscr{B}|_T$. There is a unique $e \in (\mathbb{R}^{n_e})^T$, such that $(e,w) \in \mathscr{B}|_T$ if and only if rank$\,\Pi_w B = mT+n$.*

A necessary condition for unique recovery is $p > n_e$, where p denotes the number of outputs of $\mathscr{B}$. Note that in the ARMAX setting unique recovery of e is not possible.

In the missing data estimation problem, the data w may be corrupted by measurement noise. Then, generically, there is no $e \in (\mathbb{R}^{n_e})^T$, such that $(e,w) \in \mathscr{B}|_T$. In this case, under the assumptions of Lemma 3, we choose the signal $\widehat{e}$ that achieves the best in the least-squares sense fit to the data w:

$$\text{dist}(w,\mathscr{B}) := \min_{(\widehat{e},\widehat{w})\in\mathscr{B}|_T} \|w-\widehat{w}\|. \qquad \text{(dist)}$$

Problem (dist) is a generalization of (M). Indeed, under the assumptions of Lemma 3, when there is no latent input, (dist) coincides with (M). Moreover, dist$(w,\mathscr{B})$ is the likelihood of w, given $\mathscr{B}$, when

$$w = \overline{w}+\widetilde{w}, \quad \text{where } (\overline{e},\overline{w}) \in \mathscr{B}|_T, \qquad \text{(EIV-ARMAX)}$$

for some $\overline{e} \in (\mathbb{R}^{n_e})^T$ and measurement noise $\widetilde{w}$ that is a zero-mean, white, Gaussian process.

Lemma 4. *Under the assumptions of Lemma 3, if the data w is generated in the (EIV-ARMAX) setting,* $\text{dist}(w,\mathscr{B}) \leq \|\widetilde{w}\|$.

Proof: By Lemma 3, we have that

$$\text{dist}(w,\mathscr{B}) = \text{misfit}(w,\Pi_w\mathscr{B}). \qquad (\text{dist} \leftrightarrow \text{misfit})$$

Lemma 4 follows then from Lemma 1. □

Next, we show how misfit$(w_{\text{d}},\mathscr{B})$ and dist$(w,\mathscr{B})$ can be computed in practice and derive a direct data-driven fault detection method based on dist$(w,\mathscr{B})$.

IV. Direct data-driven fault detection

In the data-driven fault detection problem considered, the monitored process is a bounded complexity linear time-invariant system with variables (e,w). Its nominal behavior $\mathscr{B}^0$ is implicitly specified by offline data $(e^0_{\text{d}},{w_{\text{d}}}^0) \in \mathscr{B}^0|_{T_{\text{d}}}$, that satisfies the generalized persistency of excitation condition (GPE). The fault detection problem aims to check if an observed signal $w \in (\mathbb{R}^q)^T$, generated in the (EIV-ARMAX) setting, is compatible with $\mathscr{B}^0$. This is done by computing and comparing the distance measures $d_i := \text{dist}(w,\mathscr{B}^i)$ to the nominal behavior $\mathscr{B}^0$ and possible faulty behaviors $\mathscr{B}^i$, $i = 1,\dots,N$, also specified by offline data $(e^i_{\text{d}},{w_{\text{d}}}^i) \in \mathscr{B}^i|_{T_{\text{d}}}$ satisfying the generalized persistency of excitation condition.

The method proposed has two phases:

1) using the offline data $(e^i_{\text{d}},{w_{\text{d}}}^i)$ and the complexity specification c, find orthonormal bases B^i for $\mathscr{B}^i|_T$,
2) using B^i, compute $d_i := \text{dist}(w,\mathscr{B}^i)$.

In phase 1, the bases B^i can be computed by low-rank approximation, *i.e.*, truncation of the singular value decomposition of the Hankel matrices $\mathscr{H}_T({w_{\text{d}}}^i)$ to the theoretical rank $mT+n$. Although this method is cheap and easy to compute, it does not preserve the shift-invariant structure and is suboptimal [19]. Alternatively, a Hankel structured low-rank approximation can be used [20], [21], however, it leads to a nonconvex optimization problem.

Problem (M) is a projection of w on the subspace $\mathscr{B}|_T$. Thus, with B being an orthonormal basis for $\mathscr{B}|_T$, we have

$$\text{misfit}(w,\mathscr{B}) = \min_g \|w-Bg\| = \sqrt{w^\top(I_{qT}-BB^\top)w},$$

where I_{qT} is the identity matrix of size qT.

Next, we consider the computation of the new measure dist$(w,\mathscr{B})$. Note that by the assumption of Lemma 3 $B_w :=$

ΠB is a basis for $\Pi_w \mathscr{B}|_T$. (However, it need not be an orthonormal basis.) Using (dist $\leftrightarrow$ misfit), we have

$$\operatorname{dist}(w, \mathscr{B}) = \min_g \|w - B_w g\| = \sqrt{w^\top (I_{qT} - B_w B_w^+) w},$$

where B_w^+ is the pseudo-inverse of B_w.

V. Empirical validation

For the validation of the method, we use a mechanical system consisting of three masses interconnected via springs and dampers as shown in Figure 1. The system is excited by an external force u, applied on the first mass, and a disturbance e acting on the third mass. Both u and e are generated as random independent and uniformly distributed in the interval $[0,1]$ processes. The nominal system $\mathscr{B}^0$ has parameters $m_1 = m_2 = m_3 = 10$, $k_1 = k_2 = k_3 = 1$, $b_1 = b_2 = b_3 = 0.5$. A scenario of a fault $\mathscr{B}^1$ is considered, where k_3 is changed by 5%, *i.e.*, $k_3 = 1.05$ with the other parameters being the same as in the nominal case.

The observed variables w are the external force and the positions of the three masses. Data $w^0 \in (\mathbb{R}^4)^{100}$, $w^1 \in (\mathbb{R}^4)^{100}$ is collected from the nominal and faulty behaviors, respectively, in the (EIV-ARMAX) setups. Using the data, the four distance measures $d_i^j := \operatorname{dist}(w^j, \mathscr{B}^i)$, $i = 0,1$, $j = 0,1$ are computed for increasing measurement noise $\widetilde{w}$ variance. Figure 2 shows the distance measures d_i^j, averaged over 100 Monte-Carlo repetitions of the experiment.

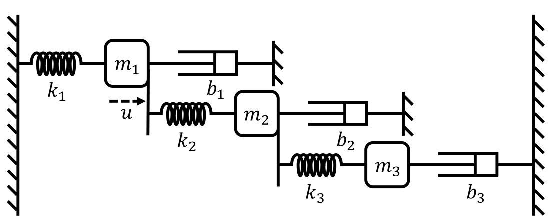

Fig. 1. The empirical validation is done on an interconnected mass-spring-damper system with external force u applied on mass m_1 and disturbance e acting on mass m_3. The observed signals are the force u and the positions of the masses. The fault is a change of the elasticity coefficient k_3 by 5%.

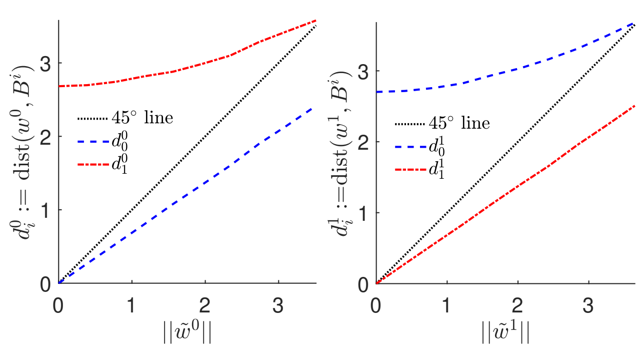

Fig. 2. The distances from the data to the corresponding behaviors, plotted in the figure as a function of the norms of the measurement noises, show that $d_i^i \leq \|\widetilde{w}^i\|$ (Lemma 4) and fault detection can be done by comparison of the distances to the nominal and faulty behaviors.

The empirical results confirm that $d_i^i \leq \|\widetilde{w}^i\|$ (Lemma 4), $d_0^0 < d_1^0$, and $d_1^1 < d_0^1$. The margin between the distances $d_0 := \operatorname{dist}(w, \mathscr{B}^0)$, $d_1 := \operatorname{dist}(w, \mathscr{B}^1)$ to the nominal and faulty behaviors allows us to do reliable fault detection. Indeed. $d_0 < d_1$ implies no fault, while $d_1 < d_0$ implies a fault. Note that the test does not require a thresholding hyper-parameter.

VI. Acknowledgments

The research leading to these results received funding from: Catalan Institution for Research and Advanced Studies (ICREA), Fond for Scientific Research Vlaanderen (FWO) project G033822N, Spanish Ministry of Science MCIU/AEI/10.13039/501100011033/ FEDER, UE, grants PID2023-148952OB-I00, PID2023-153082OB-I00, and Generalitat de Catalunya, grant SGR 01049. AM is Fellow of the Serra Húnter Programme of the Generalitat de Catalunya.

References

[1] J. Chen, P. M. Frank, M. Kinnaert, J. Lunze, and R. J. Patton, *Fault Detection and Isolation*. Springer London, 2001, pp. 191–207.

[2] J. Antoni, K. Kestel, C. Peeters, Q. Leclère, F. Girardin, T. Ooijevaar, and J. Helsen, "On the design of optimal health indicators for early fault detection and their statistical thresholds," *Mechanical Systems and Signal Processing*, vol. 218, p. 111518, 2024.

[3] L. Ljung, *System Identification: Theory for the User*. Prentice-Hall, 1999.

[4] J. C. Willems, "The behavioral approach to open and interconnected systems: Modeling by tearing, zooming, and linking," *IEEE Control Systems Magazine*, vol. 27, pp. 46–99, 2007.

[5] I. Markovsky and B. De Moor, "Linear dynamic filtering with noisy input and output," *Automatica*, vol. 41, no. 1, pp. 167–171, 2005.

[6] P. Lemmerling and B. De Moor, "Misfit versus latency," *Automatica*, vol. 37, pp. 2057–2067, 2001.

[7] J. C. Willems, "Deterministic least squares filtering," *J. Econometrics*, vol. 118, pp. 341–373, 2004.

[8] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Annual Reviews in Control*, vol. 52, pp. 42–64, 2021.

[9] I. Markovsky, L. Huang, and F. Dörfler, "Data-driven control based on behavioral approach: From theory to applications in power systems," *IEEE Control Systems Magazine*, vol. 43, pp. 28–68, 2023.

[10] I. Markovsky, "Comparison of adaptive and model-free methods for dynamic measurement," *IEEE Signal Proc. Lett.*, vol. 22, no. 8, pp. 1094–1097, 2015.

[11] J. Polderman and J. C. Willems, *Introduction to Mathematical Systems Theory*. Springer-Verlag, 1998.

[12] I. Markovsky and P. Rapisarda, "Data-driven simulation and control," *Int. J. Control*, vol. 81, no. 12, pp. 1946–1959, 2008.

[13] I. Markovsky and F. Dörfler, "Identifiability in the behavioral setting," *IEEE Trans. Automat. Contr.*, vol. 68, pp. 1667–1677, 2023.

[14] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor, "A note on persistency of excitation," *Systems & Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.

[15] I. Markovsky, E. Prieto-Araujo, and F. Dörfler, "On the persistency of excitation," *Automatica*, p. 110657, 2023.

[16] T. Söderström, *Errors-in-Variables Methods in System Identification*. Springer, 2018.

[17] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[18] I. Markovsky and F. Dörfler, "Data-driven dynamic interpolation and approximation," *Automatica*, vol. 135, p. 110008, 2022.

[19] J.-M. Papy, L. De Lathauwer, and S. Van Huffel, "A shift invariance-based order-selection technique for exponential data modelling," *IEEE Signal Proc. Letters*, vol. 14, no. 7, pp. 473–476, 2007.

[20] I. Markovsky, *Low-Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2019.

[21] R. Arablouei and K. Dogancay, "Linearly-constrained recursive total least-squares algorithm," *IEEE Signal Proc. Letters*, vol. 19, no. 12, pp. 821–824, 2012.