Int. J. Communications, Network and System Sciences, 2011, 4, 266-285
doi:10.4236/ijcns.2011.44032 Published Online April 2011 (http://www.SciRP.org/journal/ijcns)

Adaptation in Stochastic Dynamic Systems—Survey and New Results II

Innokentiy V. Semushin
School of Mathematics and Information Technology, Ulyanovsk State University, Ulyanovsk, Russia
E-mail: *kentvsem@gmail.com*
Received March 7, 2011; *revised April* 2, 2011; *accepted April* 5, 2011

Abstract

This paper surveys the field of adaptation in stochastic systems as it has developed over the last four decades. The author's research in this field is summarized and a novel solution for fitting an adaptive model in state space (instead of response space) is given.

Keywords: Linear Stochastic Systems, Parameter Estimation, Model Identification, Identification for Control, Adaptive Control

1. Introduction

The history of adaptive control and identification is full of ups and downs, breakthroughs and setbacks [1-3] and [4]. In Ljung's opinion [1], the number of papers on identification-related problems published over the years must be close to 10^5. Combined with publications on adaptive control problems, this number is at least twice as large. This realm of research is clearly flourishing and attracting continuous attention.

At the same time, one can state that essentially different, landmark-size ideas or frameworks are not so many. The abundance of publications in this field signaled the need for some serious cleanup work in order to single out the truly independent concepts. According to Ljung, in system identification, the only two independent key concepts are the choice of a parametric model structure and the choice of identification criterion [5]. Indeed, these concepts are universal.

However, in the system identification community, understanding thereof is usually reduced to the impressive Prediction Error Framework (PEF) [6], as it can be found in literature: "All existing parameter identification methods could then be seen as particular cases of this prediction error framework" [2].

Time has shown that the importance of the Ljungian concepts is greater. Thus each of the five generalized principles of stochastic system adaptation [7] worked out in correlation with Mehra's ideas on adaptive filtering [8], can also be perceived from this point of view. Among these is the Performance Index-based Adaptive Model Principle of Adaptation. This name unites two of Ljung's principal concepts, *i. e*. Model Structure and Proximity Criterion (PC), which indicates erroneousness of a model.

It would be fortunate if someone managed to measure adequacy of a model in state space, considering that, as defined by Kalman [9], *state space* is a set of inner states of a system, which is rich enough to house all information about the system's prehistory necessary and sufficient to predict the effects of past history of the system on its future. For a dynamic data source, which in reality exists as a "black box," we have only its input and output, *i. e*. its state is beyond the reach of any practical methods. Impossibility to directly fit the adaptive model state to the Data Source state makes the block-diagram of **Figure 1** unrealistic. The question of how to overcome this *barrier* stimulates a search for novel approaches.

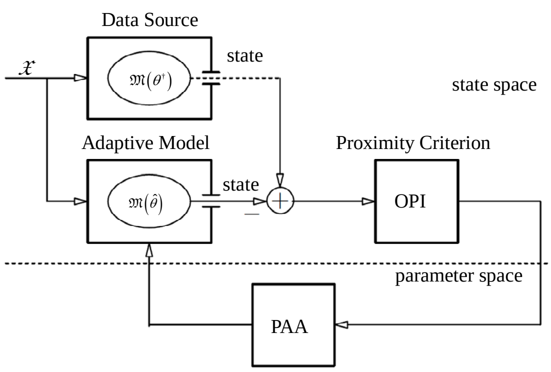

Figure 1. Unrealistic framework. *Legend*: $\mathfrak{X}$—experimental condition; OPI—Original Performance Index; PAA—Parameter Adaptation Algorithm.

Not putting the barrier on the agenda, existing PEF methods are used to fit the adaptive model to a data source in the response space instead of state space (**Figure 2**).

This position fits naturally into the presently accepted understanding of PE identification as an approximation [2], and can be expressed by the following maxima by G. E. P. Box: "All models are erroneous, but some of them can be useful." This leads to the assumption that a model set does not contain a "true" system, and so the concept of parameter error is meaningless since there are no "true parameters."

Overcoming the aforementioned barrier was perceived as a challenge in 1957 by Prouza [10] and then Šefl [11] and Gorsky [12]. The first solution to the problem was given by Semushin in 1968—70 [13-15]. The key was to exploit the complete observability property of Data Source under consideration in such a way that a finite batch of DS responses y_t to y_{t+s} would give access to the unreachable DS state (the dashed arrow in **Figure 3**) against the background of independent noise. In that event Auxiliary Performance Index (API) $\mathcal{J}^{a}$ equimodal

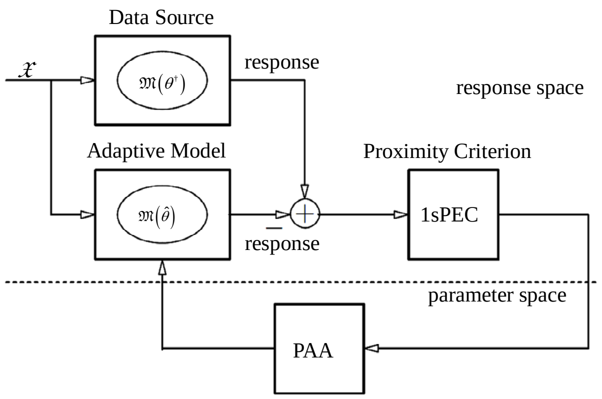

Figure 2. Minimum PE framework. *Legend*: 1sPEC— One-step Prediction Error Criterion.

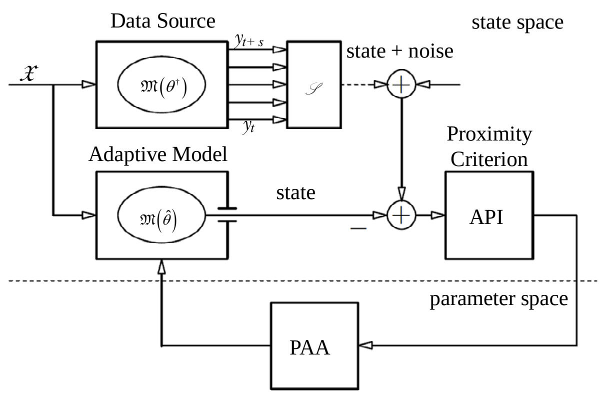

Figure 3. The API framework *Legend*: API—Auxiliary Performance Index.

to Original Performance Index $\mathcal{J}^{o}$ could be formed to suite the principal requirement $\mathcal{J}^{a} = \mathcal{J}^{o} + \text{const}$ as it was formulated in [7], and by doing so, to fit an adaptive model in state space.

System identification and adaptive control are deeply intertwined areas. Further still, the latter is unthinkable without the former being the central part of the triune "Classifier--Identifier--Modifier" whose presence in a system makes it adaptive, see **Figure 4** reproduced below from [7] where, keeping in mind adaptive nature of Identifier and the modern *identification for control* paradigm [2], the name Adaptor is used.

After this brief introduction (Section 1), which explains the essence of the author's approach, the paper (1) describes the adaptive control system structure in its two forms, *i.e*. Physical Data Model and Standard Observable Data Model (Section 2); (2) characterizes the innovation set of DMs (Section 3) and its levels of uncertainty (Section 4); (3) defines the ancillary matrix transformations, which are important for the approach (Section 5); (4) introduces a set of adaptive models and identifies five tasks at hand (Section 6). Then the paper solves four of the five tasks and gives an engineering example and simulation results. The author also offers a roadmap for further research to be reported in forthcoming papers.

2. Parameterized Data Models $\mathcal{D}(\theta)$

As assumed earlier [7], all data models $\mathcal{D}(\theta)$ forming a set $\mathcal{D}$ are parameterized by an l-component vector θ.

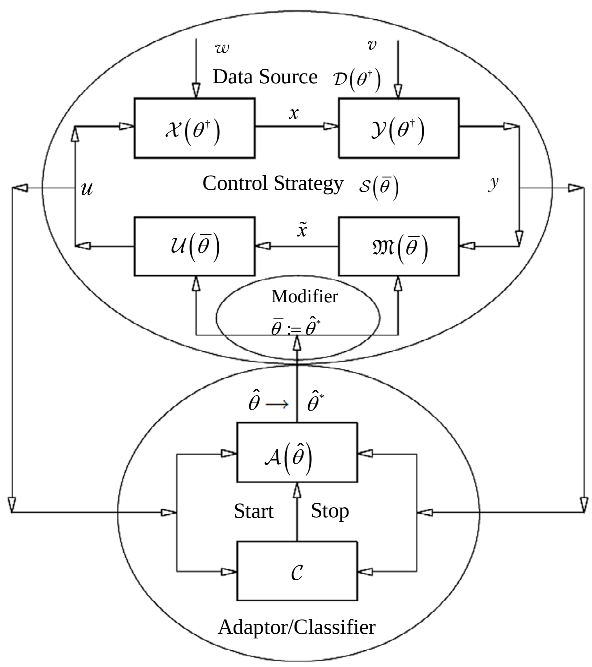

Figure 4. Adaptive stochastic control system structure.

Each particular value of θ (which does not depend on time) specifies a $\mathcal{D}(\theta)$. Within $\mathcal{D}$, a model switching mechanism exists. It is viewed as deterministic, and yet it is unknown to the observer (like controlled by an independent actor). Due to this mechanism, θ can switch over the compact subset Θ of $\mathbb{R}^l$ not frequently but rather abruptly with reference to the system dynamics; hence

$$\mathcal{D} = \left\{\mathcal{D}(\theta) \mid \theta \in \Theta \subset \mathbb{R}^l\right\}. \tag{1}$$

We write the time argument of signals as a lower index and omit the subscript θ for all the matrices describing a given physical data model (PhDM)

$$\mathcal{D}(\theta): \begin{aligned} x_{t+1} &= \mathbf{\Phi} x_t + \mathbf{\Psi} u_t + \mathbf{\Gamma} w_t, \quad t \in \mathbb{Z}_+ \\ y_t &= \boldsymbol{H} x_t + v_t, \quad t \in \mathbb{Z}_1 \end{aligned} \tag{2}$$

where $\mathbb{Z}_+$ denotes nonnegative integers, $\mathbb{Z}_1$ strictly positive integers (and $\mathbb{Z}$ all integers). Every model $\mathcal{D}(\theta)$ (2) is assumed to be acting between adjacent switches as long as it is sufficient for accepting as correct the *basic theoretical statement* (BTS) that all processes related to the $\mathcal{D}(\theta)$ are wide-sense stationary. This statement amounts to the following assumptions. The random x_0 with $\boldsymbol{E}\left\{\|x_0\|^2\right\} < \infty$ is orthogonal [16] to w_t and v_t, the zero-mean mutually orthogonal wide-sense stationary orthogonal sequences with $\boldsymbol{E}\left\{w_t w_t^{\mathrm{T}}\right\} = \boldsymbol{Q} \geq 0$ and $\boldsymbol{E}\left\{v_t v_t^{\mathrm{T}}\right\} = \boldsymbol{R} > 0$ for all $t \in \mathbb{Z}$; $\begin{bmatrix} w_t \\ v_t \end{bmatrix}$ is orthogonal to x_j and u_j for all $j \geq t$; u_t is a given signal; it is an "external input" when considering the open-loop case *or* a control strategy function

$$u_t = u_{\mathcal{X}}\left(t, y_1^t, u^{t-1}\right) \tag{3}$$

when considering the closed-loop setup (as in **Figure 4**).

Remark 1 *In **Figure** 4, we use the following nomenclatures: Signal Nomenclature: w—plant disturbance noise; v—sensor (observation) noise; x—(unknown) plant state (useful signal); $\tilde{x}$—suboptimally estimated plant state; u—(available) control signal; y—(available, measured) sensor output. Parameter Nomenclature: θ—generic name of the uncertainty parameter; $\theta^\dagger$—true value of θ in Data Source $\mathcal{D}(\cdot)$; $\bar{\theta}$—suboptimally estimated (preliminary designed) value of θ on which current Control Strategy $\mathcal{S}(\bar{\theta})$ is based; $\hat{\theta}$—current estimated value of θ; $\hat{\theta}^*$—final estimated value of θ resulting from the identification process in Adaptor $\mathcal{A}(\hat{\theta})$. Component Nomenclature: $\mathcal{X}(\cdot)$—Plant; $\mathcal{Y}(\cdot)$—Sensor; $\mathfrak{M}(\cdot)$—Model—based State Estimator (Kalman--like Filter); $\mathcal{U}(\cdot)$—Deterministic Controller; $\mathcal{A}(\cdot)$—Adaptor (Adaptive Parameter Identifier); $\mathcal{C}$—System Mode Classifier (abrupt change detector).*

Stackable vectors of previous values

$$\mathcal{X}: \begin{aligned} y_1^t &= (y_t, y_{t-1}, \cdots, y_1) \\ u^{t-1} &= (u_{t-1}, u_{t-2}, \cdots, u_0) \end{aligned} \tag{4}$$

constitute the experimental condition $\mathcal{X}$ (cf. Ljung [5]) on which both Adaptor and Classifier are based.

By assumption, $y_t \in \mathbb{R}^m$ is generated by the completely observable PhDM (2), so it is possible to move from the physical state variables $x \in \mathbb{R}^n$ in (2) to another x^* through the following similarity transformation $\boldsymbol{x}^* = \boldsymbol{W}_* x$. It is known [6] that matrices

$$\begin{aligned} \mathbf{\Phi}_* &= \boldsymbol{W}_* \mathbf{\Phi} \boldsymbol{W}_*^{-1}, & \mathbf{\Psi}_* &= \boldsymbol{W}_* \mathbf{\Psi} \\ \mathbf{\Gamma}_* &= \boldsymbol{W}_* \mathbf{\Gamma}, & \boldsymbol{H}_* &= \boldsymbol{H} \boldsymbol{W}_*^{-1} \end{aligned} \tag{5}$$

uniquely determine a new state representation

$$\mathcal{D}^*(\theta): \begin{aligned} x_{t+1}^* &= \mathbf{\Phi}_* x_t^* + \mathbf{\Psi}_* u_t + \mathbf{\Gamma}_* w_t, t \in \mathbb{Z}_+ \\ y_t &= \boldsymbol{H}_* x_t^* + v_t, t \in \mathbb{Z}_1 \end{aligned} \tag{6}$$

of the standard observable data model (SODM) with

$$\boldsymbol{H}_* = \begin{bmatrix} 1\,0\cdots0 & 0\,0\cdots0 & \cdots & 0\,0\cdots0 \\ 0\,0\cdots0 & 1\,0\cdots0 & \cdots & 0\,0\cdots0 \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ \underbrace{0\,0\cdots0}_{p_1} & \underbrace{0\,0\cdots0}_{p_2} & \cdots & \underbrace{1\,0\cdots0}_{p_m} \end{bmatrix} \tag{7}$$

$$\mathbf{\Phi}_* = \begin{array}{c} \left[\begin{array}{ccc|ccc|c|ccc} 0 & & & & & & & & & \\ \vdots & \boldsymbol{I} & & & 0 & & \cdots & & 0 & \\ 0 & & & & & & & & & \\ * & \ldots & * & * & \ldots & * & & * & \ldots & * \\ & & & 0 & & & & & & \\ & 0 & & \vdots & \boldsymbol{I} & & \cdots & & 0 & \\ & & & 0 & & & & & & \\ * & \ldots & * & * & \ldots & * & & * & \ldots & * \\ & \vdots & & & \vdots & & \ddots & & \vdots & \\ & & & & & & & 0 & & \\ & 0 & & & 0 & & \cdots & \vdots & \boldsymbol{I} & \\ & & & & & & & 0 & & \\ * & \ldots & * & * & \ldots & * & & * & \ldots & * \end{array}\right] \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} p_1 \\ \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} p_2 \\ \\ \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} p_m \end{array} \tag{8}$$

if some numbers r_j are chosen by the user so that

$$\begin{aligned} 0 &= r_0 < r_1 < \cdots < r_{m-1} < r_m = n \\ p_j &= r_j - r_{j-1}, \quad j = \overline{1, m} \end{aligned} \tag{9}$$

and the invertible $n \times n$ matrix $\boldsymbol{W}_*$ is determined by

$$\boldsymbol{W}_* = \left[\boldsymbol{h}_1^{\mathrm{T}} \cdots \left(\boldsymbol{h}_1 \mathbf{\Phi}^{p_1 - 1}\right)^{\mathrm{T}} \middle| \boldsymbol{h}_2^{\mathrm{T}} \cdots \left(\boldsymbol{h}_2 \mathbf{\Phi}^{p_2 - 1}\right)^{\mathrm{T}} \middle| \right. \\ \left. \cdots \middle| \boldsymbol{h}_m^{\mathrm{T}} \cdots \left(\boldsymbol{h}_m \mathbf{\Phi}^{p_m - 1}\right)^{\mathrm{T}} \right]^{\mathrm{T}} \tag{10}$$

Numbers $(p_1, p_2, \ldots, p_m)$ are known as the partial observability indices, and so W_* will be called the observability matrix. Benefits of this transformation will be seen later (at the end of Section 5).

Remark 2 *Since the eigenvalues of* $\mathcal{D}(\theta)$ (2) *remain unchanged in* $\mathcal{D}^*(\theta)$ (6), *the transformation* (5) *does not alter the dynamics of Data Source; it also has no effect on its inputs and outputs and so can be made at will*.

3. Parameterized Innovations

The above data model of a time-invariant data source will be referred to as the *conventional model*, no matter whether it is PhDM (2) or SODM (6)-(8). Another commonly used representation is the time-invariant (due to BTS) *innovation model*

$$\mathfrak{M}(\theta): \begin{aligned} x_{t+1|t} &= \Phi x_{t|t-1} + \Psi u_t + \boldsymbol{G}\nu_{t|t-1} \\ y_t &= \boldsymbol{H}x_{t|t-1} + \nu_{t|t-1} \end{aligned} \tag{11}$$

with $t \in \mathbb{Z}_1$, the initial $x_{1|0} = \Phi\bar{x}_0 + \Psi u_0$, and $\bar{x}_0 = E\{x_0\}$, which is the well-known steady-state Kalman filter with the innovation process $\nu_{t|t-1}$, the optimal state predictor $x_{t+1|t}$, the gain $\boldsymbol{K} = \Sigma\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\Sigma\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}\right)^{-1}$, $\boldsymbol{G} = \Phi\boldsymbol{K}$, and Σ satisfying the algebraic Riccati equation (ARE) [17]

$$\Sigma = \Phi\left[\Sigma - \Sigma\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\Sigma\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}\right)^{-1}\boldsymbol{H}\Sigma\right]\Phi^{\mathrm{T}} + \Gamma\boldsymbol{Q}\Gamma^{\mathrm{T}}. \tag{12}$$

Concurrently, another form

$$\mathfrak{M}(\theta): \begin{aligned} x_{t+1|t} &= \Phi x_{t|t} + \Psi u_t, t \in \mathbb{Z}_+ \\ x_{t|t} &= x_{t|t-1} + \boldsymbol{K}\nu_{t|t-1}, t \in \mathbb{Z}_1 \\ y_t &= \boldsymbol{H}x_{t|t-1} + \nu_{t|t-1}, t \in \mathbb{Z}_1 \end{aligned} \tag{13}$$

with the initial $x_{0|0} = \bar{x}_0$, which is equivalent to (11), can be used where $x_{t|t}$ is the optimal "filtered" estimator for x_t based on experimental condition $\mathcal{X}$ (4). When θ ranges (or switches) over Θ as in (1), we obtain the set of Kalman filters

$$\mathfrak{M} = \left\{\mathfrak{M}(\theta) \middle| \theta \in \Theta \subset \mathbb{R}^l\right\}. \tag{14}$$

Theorem 1 (*Steady-state Kalman filter uniqueness* [18]).

Assume that in wide-sense stationary circumstances the following conditions hold:

1) $\boldsymbol{R}$ is positive definite $(\boldsymbol{R} > 0)$;
2) Matrix pair $\{\Phi, \boldsymbol{H}\}$ is detectable;
3) Matrix pair $\left\{\Phi, \left(\Gamma\boldsymbol{Q}\Gamma^{\mathrm{T}}\right)^{1/2}\right\}$ is stabilizable.

Then the following assertions are true:

a) Every solution $\boldsymbol{P}_{t|t-1}$ to the matrix discrete-time Riccati equation

$$\begin{aligned} \boldsymbol{P}_{t+1|t} &= \Phi\boldsymbol{P}_{t|t}\Phi^{\mathrm{T}} + \Gamma\boldsymbol{Q}\Gamma^{\mathrm{T}}, \quad t \in \mathbb{Z}_+ \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}\Big(\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}} \\ &\quad + \boldsymbol{R}\Big)^{-1}\boldsymbol{H}\boldsymbol{P}_{t|t-1}, \quad t \in \mathbb{Z}_1 \end{aligned} \tag{15}$$

with the initial condition $\boldsymbol{P}_{0|0} \geq 0$ has the limit

$$\lim_{t\to\infty}\boldsymbol{P}_{t|t-1} = \boldsymbol{P}_\infty \tag{16}$$

and this limit coincides with

$$\boldsymbol{\Sigma} = \lim_{t\to\infty} E\left\{\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)^T\right\} \triangleq \boldsymbol{P}_\infty$$

which is the unique non-negative definite solution to ARE (12) and defines the covariance of one-step prediction error

$$\tilde{x}_{t|t-1} \triangleq x_t - x_{t|t-1} \tag{17}$$

in the Wiener-Kolmogorov filter (as $t \to \infty$).

b) The Kalman gain

$$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}\right)^{-1} \tag{18}$$

in the Kalman filter (KF)

$$\begin{aligned} x_{t+1|t} &= \Phi x_{t|t} + \Psi u_t \\ x_{t|t} &= x_{t|t-1} + \boldsymbol{K}_t\nu_{t|t-1} \\ \nu_{t|t-1} &= y_t - \boldsymbol{H}x_{t|t-1} \end{aligned} \tag{19}$$

has the limit

$$\boldsymbol{K} = \lim_{t\to\infty}\boldsymbol{K}_t = \Sigma\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\Sigma\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}\right)^{-1} \tag{20}$$

such that the estimate transition matrix

$$\boldsymbol{A} = \Phi\left(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H}\right) \tag{21}$$

for $\mathfrak{M}(\theta)$ (13) is a stable limit.

c) Weighting function h_{t-k}^{WK} of the Wiener-Kolmogorov filter operating as the *l.s.* optimal one-step predictor

$$x_{t|t-1} = \sum_{k=-\infty}^{t-1} h_{t-k}^{\mathrm{WK}} y_k = \sum_{k=1}^{\infty} h_k^{\mathrm{WK}} y_{t-k}$$

coincides (asymptotically as $t \to \infty$ and uniformly in k) with the weighting function $h_k^{\mathrm{KF}} = \boldsymbol{A}^{k-1}(\boldsymbol{\Phi K})$ of the Kalman filter-predictor (15)-(21) computing

$$x_{t|t-1} = \sum_{k=1}^{\infty} h_k^{\mathrm{KF}} y_{t-k} = \sum_{k=1}^{\infty} \boldsymbol{A}^{k-1}\left(\boldsymbol{\Phi K}\right) y_{t-k}$$

Remark 3 *Theorem* 1 *is well-known, however this formulation is close to that given and completely proved in Fomin* [18]. *The basic notions used here* (*detectability, stabilizability and others*) *are known from literature, for example Wohnam* [19]. *Derivation of Kalman filter equations* (15)-(21) *can be found in Maybeck* [20] *or Anderson*

& Moore [21] *or in other known books*.

Varied criteria applicable for KF derivation are discussed in Meditch [22] based on the original work by Sherman [23]. Among them is the mean-square criterion

$$\mathcal{J}_t^{\mathrm{o}} \triangleq \frac{1}{2}\boldsymbol{E}\left\{\boldsymbol{e}_{t+1|t}^{\mathrm{T}}\boldsymbol{e}_{t+1|t}\right\} \tag{22}$$

defined for a one-step predictor $\hat{g}_{t+1|t}$ through its error $e_{t+1|t} = x_{t+1} - \hat{g}_{t+1|t}$, which has the form of $\tilde{x}_{t+1|t}$ (17) in the Kalman filter. Thus in the basis forming the state-space, $\mathfrak{M}(\theta)$ (13) is the unique steady-state model minimizing the Original Performance Index (OPI) $\mathcal{J}_t^{\mathrm{o}}$ (22) at any t, which is large enough for BTS to hold, so that writing t or $t+1$ or any other finitely shifted time in (22) makes no difference.

Remark 4 *The set of SODM* $\mathcal{D}^*(\theta)$ (6)-(8) *when used for Theorem* 1 *leads to the isomorphic set* $\mathfrak{M}^* = \left\{\mathfrak{M}^*(\theta)\middle|\theta \in \Theta \subset \mathbb{R}^l\right\}$ *of Kalman filters* $\mathfrak{M}^*(\theta)$ *in the form of Equations* (11)-(13) (*steady-state version*) *or* (15)-(21) (*temporal version*) *where matrices* $\boldsymbol{H} = \boldsymbol{H}_*$ *and* $\Phi = \Phi_*$ *are* (7) *and* (8). *When there is no need to distinguish between* $\mathfrak{M}^*(\theta)$ *and* $\mathfrak{M}(\theta)$, *as well as between* $\mathcal{D}^*(\theta)$ *and* $\mathcal{D}(\theta)$, *as in* ***Figure* 4**, *we omit the asterisk mark thus implying that symbols* $\mathcal{D}$ *and/or* $\mathfrak{M}$ *may be taken to mean* $\mathcal{D}^*$ *and/or* $\mathfrak{M}^*$, *as it is the case in the following two Remarks*.

Remark 5 *It might be well to point out that* $\mathcal{D}$ *in* (1) *and* $\mathfrak{M}$ *in* (14) *are two different and yet equivalent representations of one and the same Set of Data Sources. When* θ *takes the true value* $\theta^\dagger \in \Theta$, *one can choose either the true Data Model* $\mathcal{D}(\theta^\dagger)$ *or the optimal innovation model* $\mathfrak{M}(\theta^\dagger)$ *to be sought as a hidden object within the set. As we need Adaptor* $\mathcal{A}(\hat{\theta})$ *for control* (*cf.* ***Figure* 4**) *able to serve the purpose of feedback filter optimization, our priority now is the seeking of* $\mathfrak{M}(\theta)$ *instead of* $\mathcal{D}(\theta)$. (*However we do not rule out the seeking of* $\mathcal{D}(\theta)$ *instead of* $\mathfrak{M}(\theta)$ *as another possibility* (*see Section* 9) *for further research*.) *This suggests that we need to have developed the unbiased* $\mathfrak{M}(\theta^\dagger)$ *identification methods. Had OPI* (22) *been accessible, minimizing the OPI by a numerical optimization method would produce the desired result. However, this is not the case, and this creates the problem as stated in* [7].

Remark 6 *One more point needs to be made: Packing the* θ *with elements in models* (11) *and* (13) *will differ from that in model* (2) *because some* θ *elements of model* $\mathcal{D}(\theta)$ (2) *appear in* $\boldsymbol{K}$ (*or* $\boldsymbol{G}$) *of* $\mathfrak{M}(\theta)$ *not directly but through the solution* Σ *of equation* (12).

Remark 6 leads to the four levels of uncertainty to be included into the subsequent consideration.

4. Uncertainty Parameterization

Let symbol $\gtrdot$ read: "*enters as a parameter into the elements of*". By reference to this binary relation, the levels of uncertainty inherent in $\mathcal{D}(\theta)$ and as a consequence in $\mathfrak{M}(\theta)$, are as follows:

Level 1 The θ-dependent in $\mathcal{D}(\theta)$ are matrices Γ, $\boldsymbol{Q}$, and $\boldsymbol{R}$, only. In $\mathfrak{M}(\theta)$, it can be treated as $\theta \gtrdot \{\boldsymbol{K}\}$.

Level 2 The θ-dependent in $\mathcal{D}(\theta)$ are matrices Φ, Γ, $\boldsymbol{Q}$, and $\boldsymbol{R}$, only. For $\mathfrak{M}(\theta)$, it can be visualized as $\theta \gtrdot \{\Phi, \boldsymbol{K}\}$.

Level 3 The θ-dependent in $\mathcal{D}(\theta)$ are matrices Ψ, Φ, Γ, $\boldsymbol{Q}$, and $\boldsymbol{R}$, only. In $\mathfrak{M}(\theta)$, it can be conceived as $\theta \gtrdot \{\Psi, \Phi, \boldsymbol{K}\}$.

Level 4 The θ-dependent in $\mathcal{D}(\theta)$ are matrices $\boldsymbol{H}$, Ψ, Φ, Γ, $\boldsymbol{Q}$, and $\boldsymbol{R}$, only. For $\mathfrak{M}(\theta)$, it can be thought of as $\theta \gtrdot \{\boldsymbol{H}, \Psi, \Phi, \boldsymbol{K}\}$.

Remark 7 *Level* 4 *takes place for PhDM and not for SODM because matrix* $\boldsymbol{H}$ *in the latter case is equal to* $\boldsymbol{H}_*$ (7). *Following inclusions are valid:* $L1 \subset L2 \subset L3 \subset L4$.

5. Ancillary Matrix Transformations

Several system related transformations will be needed in the sequel. Let the observability index s of Data Sources be defined as the greatest of partial indices (9):

$$s = \max(p_1, \ldots, p_m) \le p_1 + \cdots + p_m = n. \tag{23}$$

Introduce the following matrices

$$\boldsymbol{W}(\boldsymbol{H}, \Phi) = \left[\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\Phi)^{\mathrm{T}} \cdots \left(\boldsymbol{H}\Phi^{s-1}\right)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{24}$$

$$\boldsymbol{F}_0(\boldsymbol{H}, \Phi, \Psi) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \boldsymbol{H}\Psi & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \boldsymbol{H}\Phi^{s-2}\Psi & \boldsymbol{H}\Phi^{s-3}\Psi & \cdots & 0 \end{bmatrix} \tag{25}$$

$$\boldsymbol{F}_1(\boldsymbol{H}, \Phi, \Psi) = \begin{bmatrix} \boldsymbol{H}\Psi & 0 & \cdots & 0 \\ \boldsymbol{H}\Phi\Psi & \boldsymbol{H}\Psi & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \boldsymbol{H}\Phi^{s-1}\Psi & \boldsymbol{H}\Phi^{s-2}\Psi & \cdots & \boldsymbol{H}\Psi \end{bmatrix} \tag{26}$$

$$\boldsymbol{S}_0(\boldsymbol{H}, \Phi, \boldsymbol{G}) = \begin{bmatrix} \boldsymbol{I} & 0 & \cdots & 0 \\ \boldsymbol{H}\boldsymbol{G} & \boldsymbol{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \boldsymbol{H}\Phi^{s-2}\boldsymbol{G} & \boldsymbol{H}\Phi^{s-3}\boldsymbol{G} & \cdots & \boldsymbol{I} \end{bmatrix} \tag{27}$$

$$\boldsymbol{S}_1(\boldsymbol{H}, \Phi, \boldsymbol{D}) = \begin{bmatrix} \boldsymbol{I} & 0 & \cdots & 0 \\ \boldsymbol{H}\Phi\boldsymbol{D} & \boldsymbol{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \boldsymbol{H}\Phi^{s-1}\boldsymbol{D} & \boldsymbol{H}\Phi^{s-2}\boldsymbol{D} & \cdots & \boldsymbol{I} \end{bmatrix} \tag{28}$$

with $n \times m$ matrix $\boldsymbol{D}$. Consider an arbitrary $r \times q$

matrix $\boldsymbol{A}$ with $r = sm$ as composed of s submatrices $\boldsymbol{A}_i$, each $\boldsymbol{A}_i$ of m rows and some $q \in \mathbb{Z}_1$ columns:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 \\ \cdots \\ \boldsymbol{A}_s \end{bmatrix}; \quad \boldsymbol{A}_i = \begin{bmatrix} a_i^1 \\ \cdots \\ a_i^m \end{bmatrix}; \quad i = \overline{1, s} \tag{29}$$

where a_i^j is the j-th row of the i-th submatrix A_i.

Definition 1 *Rearrangement of matrix* $\boldsymbol{A}$ (29) *to the following* $s \times mq$ *matrix* $\boldsymbol{A}_T$ *is called the* T*-transform of* $\boldsymbol{A}$, *i. e.* (***Figure*** **5**)

$$\boldsymbol{A}_T = \begin{bmatrix} a_1^1 & \dots & a_1^m \\ \vdots & \ddots & \vdots \\ a_s^1 & \dots & a_s^m \end{bmatrix}. \tag{30}$$

Definition 2 $\mathscr{S}$*-transform of matrix* $\boldsymbol{A}$ (29) *is the* $n \times q$ *matrix* $\mathscr{S}(\boldsymbol{A})$ *whose* n *rows are obtained by taking the elements* a_i^j *from* $\boldsymbol{A}_T$ (30) *and placing them into* $\mathscr{S}(\boldsymbol{A})$ *as rows in the following order* (*cf.* ***Figure 5***)*:*

$$\left\{a_1^1, a_2^1, \cdots, a_{p_1}^1, a_1^2, a_2^2, \cdots, a_{p_2}^2, \cdots, a_1^m, a_2^m, \cdots, a_{p_m}^m\right\}.$$

Definition 3 $\mathscr{F}$-, $\mathscr{B}$-, $\mathscr{P}$-, *and* $\mathscr{N}$*-transforms of matrices* $\boldsymbol{\Psi}$, $\boldsymbol{\Psi}$, $\boldsymbol{B}$, *and (correspondingly)* $\boldsymbol{D}$ *are the following matrices*

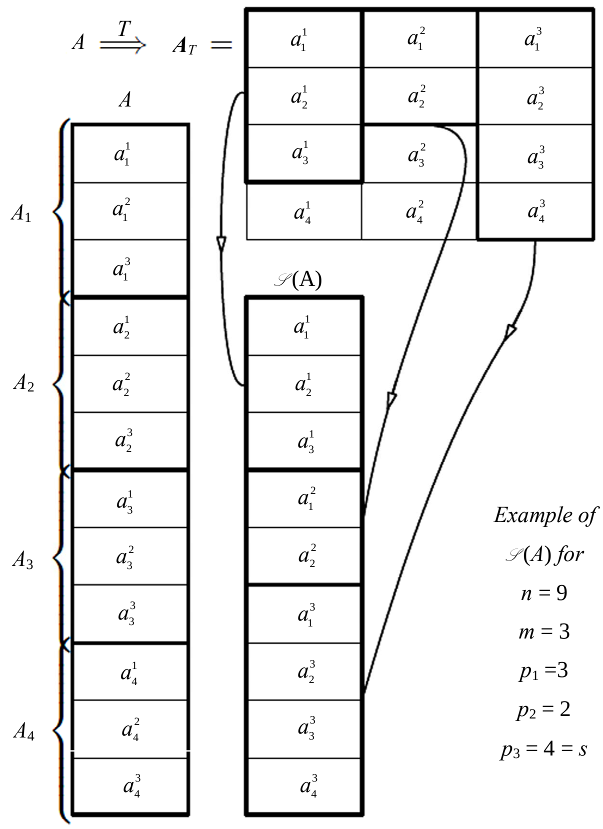

Figure 5. An illustration for Definitions 1 and 2.

$$\begin{aligned} \mathscr{F}(\Psi) &= \mathscr{S}\left(\boldsymbol{F}_0\left(\boldsymbol{H}_*, \Phi_*, \Psi\right)\right) \\ \mathscr{B}(\Psi) &= \mathscr{S}\left(\boldsymbol{F}_1\left(\boldsymbol{H}_*, \Phi_*, \Psi\right)\right) \\ \mathscr{P}(\boldsymbol{B}) &= \mathscr{S}\left(\boldsymbol{S}_0\left(\boldsymbol{H}_*, \Phi_*, \boldsymbol{B}\right)\right) \\ \mathscr{N}(\boldsymbol{D}) &= \mathscr{S}\left(\boldsymbol{S}_1\left(\boldsymbol{H}_*, \Phi_*, \boldsymbol{D}\right)\right) \end{aligned} \tag{31}$$

where $\boldsymbol{F}_0(\cdot,\cdot,\cdot)$, $\boldsymbol{F}_1(\cdot,\cdot,\cdot)$, $\boldsymbol{S}_0(\cdot,\cdot,\cdot)$, and $\boldsymbol{S}_1(\cdot,\cdot,\cdot)$ are matrices whose structure is defined by (25) to (28); $\boldsymbol{H}_*$ and $\boldsymbol{\Phi}_*$ are matrices whose structure is defined by (7) and (8); $\boldsymbol{\Psi}$ is an arbitrary $n \times q$ matrix; $\boldsymbol{B}$ and $\boldsymbol{D}$ are arbitrary $n \times m$ matrices, so that matrices (31) are of dimensions $(n \times sq)$, $(n \times sq)$, $(n \times sm)$, $(n \times sm)$ correspondingly.

It is clear from Definition 2 that $\boldsymbol{W}_* = \mathscr{S}\left(\boldsymbol{W}\left(\boldsymbol{H}, \Phi\right)\right)$ where $\boldsymbol{W}_*$ and $\boldsymbol{W}(\boldsymbol{H}, \Phi)$ are given by relations (10) and (24). Also, it is straightforward to check the identity

$$\mathscr{S}\left(\boldsymbol{W}\left(\boldsymbol{H}_*, \Phi_*\right)\right) = \boldsymbol{I} \tag{32}$$

and derive the following rule for computing matrices (31).

Algorithm 1 *Cycle through the following nested items*:

for $k = 1, 2, \cdots, m$ **do**
  for $j = 1, 2, \cdots, p_k$ **do**
    begin

$$i = \left(p_1 + \cdots + p_{k-1}\right) + j$$

$$\mathscr{F}_i(\Psi) = \begin{cases} \boldsymbol{O}, & j = 1, \\ \Psi_{i-1} \odot \mathscr{F}_{i-1}(\Psi), & j \neq 1, \end{cases}$$

$$\mathscr{B}_i(\Psi) = \begin{cases} \Psi_i \odot \boldsymbol{O}, & j = 1, \\ \Psi_i \odot \mathscr{B}_{i-1}(\Psi), & j \neq 1, \end{cases}$$

$$\mathscr{P}_i(\boldsymbol{B}) = \begin{cases} \boldsymbol{I}_k, & j = 1, \\ \boldsymbol{B}_{i-1} \odot \mathscr{P}_{i-1}(\boldsymbol{B}), & j \neq 1, \end{cases}$$

$$\mathscr{N}_i(\boldsymbol{D}) = \begin{cases} \boldsymbol{I}_k, & j = 1, \\ \boldsymbol{D}_i \odot \mathscr{N}_{i-1}(\boldsymbol{D}), & j \neq 1, \end{cases}$$

    end
  end
end

where

- $\boldsymbol{O}$ is a row of zeros of an appropriate length,
- $\boldsymbol{I}_k$ is the k-th row of the identity matrix,
- $(\cdot)_i$ is the i-th row of any matrix $(\cdot)$,
- $\odot$ denotes the concatenation of two rows into one row whose length is limited on the right to the required number of elements, viz. sq for $\mathscr{F}_i(\Psi)$ and $\mathscr{B}_i(\Psi)$, and sm for $\mathscr{P}_i(\boldsymbol{B})$ and $\mathscr{N}_i(\boldsymbol{D})$.

Thus matrices (31) do not depend on the state transition matrix $\boldsymbol{\Phi}$ when $\boldsymbol{\Phi}$ is represented in the block-companion form (8) and observation matrix $\boldsymbol{H}$ in the form of (7), i. e. in the context of SODM. This is the heart of our approach to constructing the auxiliary performance indices (APIs, $\mathcal{J}_t^{\mathrm{a}}$) as was announced in [7] and is considered below.

Remark 8 *The above matrix transformations as stated here were first introduced and used in* [24].

6. The Set $\mathcal{A}$ of Adaptive Models $\mathfrak{M}\left(\hat{\theta}\right)$

Let us define the set of adaptive models

$$\mathcal{A} = \left\{\mathfrak{M}\left(\hat{\theta}\right) \middle| \hat{\theta} \in \Theta \subset \mathbb{R}^l\right\} \tag{33}$$

By this notation, we emphasize the fact that we construct adaptive models in the same class as $\mathfrak{M}$ belongs to with the only difference that the unknown parameter θ in $\mathfrak{M}(\theta)$ is replaced by $\hat{\theta}$ to obtain $\mathfrak{M}\left(\hat{\theta}\right)$. In so doing, each particular value of $\hat{\theta}$, an estimate of θ, leads to a fixed model $\mathfrak{M}\left(\hat{\theta}\right)$. In accordance with The Active Principle of Adaptation (APA) [7], only when $\hat{\theta}$ ranges over Θ in search of $\theta^\dagger$ for *the goal* $\mathfrak{M}^*\left(\theta^\dagger\right)$ or $\mathfrak{M}\left(\theta^\dagger\right)$ as governed by a smart, unsupervised helmsman equipped by a vision of *the goal in state space* (cf. **Figure 3**) and able to pursue it, we obtain an adaptive model $\mathfrak{M}\left(\hat{\theta}\right)$ of *active type* within the set $\mathcal{A}$ (33). In this case, $\hat{\theta}$ will act as a self-tuned model parameter and so should be labeled by τ, the time instant of model's inner clock, in order to get thereby the emphasized notations $\hat{\theta}_\tau$ and $\mathfrak{M}\left(\hat{\theta}_\tau\right)$ in describing parameter adaptation algorithms (PAAs) to be developed. From this point on $\mathfrak{M}\left(\hat{\theta}_\tau\right)$ becomes an *adaptive estimator*.

Remark 9 *Note in passing that pace of* τ *may differ from that of* t: *e.g.* $\tau = kt$ *for* $k > 0$. *In general, there exist three time scales in adaptive systems*, *as stated by Anderson* [3]: *time scale for underlying plant dynamics, time scale for identifying plant, and time scale of plant parameters variation. We shall need discriminate between* τ *and* t *later when developing a PAA.*

Remark 10 *If we work in the context of SODM*, *the set*

$$\mathcal{A}^* = \left\{\mathfrak{M}^*\left(\hat{\theta}\right) \middle| \hat{\theta} \in \Theta \subset \mathbb{R}^l\right\} \tag{34}$$

instead of (33) should be used.

At this junction, we identify the following tasks as pending:

1) Express $\mathfrak{M}^*\left(\hat{\theta}\right)$ or $\mathfrak{M}\left(\hat{\theta}\right)$ in explicit form.
2) Build up APIs that could offer vision of the goal.
3) Examine APIs' capacity to visualize the goal.
4) Put forward feasible schemata for APIs computation.
5) Develop a PAA that could help pursueing the goal.

Consider here the first four points consecutively.

6.1. Parameterized Adaptive Models

Reasoning from (11), (13), we set the adaptive model

$$\mathfrak{M}\left(\hat{\theta}\right): \begin{aligned} \hat{g}_{t+1|t} &= \boldsymbol{A}\hat{g}_{t|t-1} + \boldsymbol{F}u_t + \boldsymbol{B}\eta_{t|t-1} \\ y_t &= \boldsymbol{C}\hat{g}_{t|t-1} + \eta_{t|t-1} \end{aligned} \tag{35}$$

or equivalently (due to $\boldsymbol{B} = \boldsymbol{AD}$) the model

$$\mathfrak{M}\left(\hat{\theta}\right): \begin{aligned} \hat{g}_{t+1|t} &= \boldsymbol{A}\hat{g}_{t|t} + \boldsymbol{F}u_t \\ \hat{g}_{t|t} &= \hat{g}_{t|t-1} + \boldsymbol{D}\eta_{t|t-1} \\ y_t &= \boldsymbol{C}\hat{g}_{t|t-1} + \eta_{t|t-1} \end{aligned} \tag{36}$$

as a member of $\mathcal{A}$ (33). Here $\hat{\theta} \succ \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{F}\}$ is the self-tuned parameter intended to estimate (in one-to-one correspondence) parameter $\theta \succ \{\boldsymbol{\Phi}, \boldsymbol{G}, \boldsymbol{H}, \boldsymbol{K}, \boldsymbol{\Psi}\}$. In parallel, reasoning from $\mathfrak{M}^*(\theta)$ (cf. Remark 4), we build the adaptive model

$$\mathfrak{M}^*\left(\hat{\theta}\right): \begin{aligned} \hat{g}_{t+1|t} &= \boldsymbol{A}\hat{g}_{t|t-1} + \boldsymbol{F}u_t + \boldsymbol{B}\eta_{t|t-1} \\ y_t &= \boldsymbol{H}_*\hat{g}_{t|t-1} + \eta_{t|t-1} \end{aligned} \tag{37}$$

or equivalently (due to $\boldsymbol{B} = \boldsymbol{AD}$) the model

$$\mathfrak{M}^*\left(\hat{\theta}\right): \begin{aligned} \hat{g}_{t+1|t} &= \boldsymbol{A}\hat{g}_{t|t} + \boldsymbol{F}u_t \\ \hat{g}_{t|t} &= \hat{g}_{t|t-1} + \boldsymbol{D}\eta_{t|t-1} \\ y_t &= \boldsymbol{H}_*\hat{g}_{t|t-1} + \eta_{t|t-1} \end{aligned} \tag{38}$$

with $\boldsymbol{H}_*$ and $\boldsymbol{A} = \boldsymbol{A}_*$ taken in the form of (7) and (8).

Adaptor $\mathcal{A}\left(\hat{\theta}\right)$ using (35)-(36) (or alternatively, $\mathcal{A}^*\left(\hat{\theta}\right)$ using (37)-(38)) is supposed to contain a PAA to offer the prospect of convergence. As viewed in **Figures 1** to **3**, convergence can take place in three spaces:

- in response space; this is model response convergence,
- in state space; this is model state convergence, and
- in parameter space; this is model parameter convergence.

For convergence of the last-named type, we anticipate almost surely (a.s.) convergence, as it is the case for MPE identification methods [16]. It actuates either or both of the two other types of convergence. The type of convergence in state space, as well as in response space, is induced by the type of Proximity Criterion, PC (cf. **Figures 1-3**). As seen from (22), we are oriented to the PC, which is quadratic in error; this being so, it would appear reasonable that these covergences would be in mean square (m.s.). Thus we anticipate the following properties of our estimators:

$$\left.\begin{matrix} \mathfrak{M}\left(\hat{\theta}_\tau\right) \xrightarrow{\text{m.s.}} \mathfrak{M}\left(\theta^\dagger\right) \\ \mathfrak{M}^*\left(\hat{\theta}_\tau\right) \xrightarrow{\text{m.s.}} \mathfrak{M}^*\left(\theta^\dagger\right) \end{matrix} : \begin{matrix} \hat{\theta}_\tau & \xrightarrow{\text{a.s.}} \theta^\dagger \\ \hat{g}_{t+1|t} & \xrightarrow{\text{m.s.}} x_{t+1|t} \\ \hat{g}_{t|t} & \xrightarrow{\text{m.s.}} x_{t|t} \end{matrix} \right|_{t\to\infty} \tag{39}$$

With the understanding that errors for PC

$$\begin{aligned} e_{t+1|t} &\triangleq x_{t+1} - \hat{g}_{t+1|t}, \quad & e_{t|t} &\triangleq x_t - \hat{g}_{t|t} \\ r_{t+1|t} &\triangleq x_{t+1|t} - \hat{g}_{t+1|t}, \quad & r_{t|t} &\triangleq x_{t|t} - \hat{g}_{t|t} \end{aligned} \tag{40}$$

are fundamentally unmeasurable, we search for a function

$$\varepsilon_t(\hat{\theta}) = f\left(y_{t+1}^{t+s} - \hat{y}_{t+1|t}^{t+s|t}\right) \in \mathbb{R}^n \quad (41)$$

of the difference in two terms: outputs y_{t+1}^{t+s} generated by Data Source described in any appropriate form (2), (6), (11), or (13), and their estimates $\hat{y}_{t+1|t}^{t+s|t}$ generated by the adaptive model $\mathfrak{M}(\hat{\theta})$ (or $\mathfrak{M}^*(\hat{\theta})$). For $\varepsilon_t(\hat{\theta})$ in (41), we will also use notations $\varepsilon_{t+1|t}$ or $\varepsilon_{t|t}$, thus bringing them into correlation with $e_{t+1|t}$ or $e_{t|t}$ (correspondingly, with $r_{t+1|t}$ or $r_{t|t}$) from (40). Then

$$\mathcal{J}_t^{\mathrm{a}} = \mathcal{J}_t^{\mathrm{a}}(\hat{\theta}) \triangleq \frac{1}{2} E\left\{\boldsymbol{\varepsilon}_t(\hat{\theta})^{\mathrm{T}} \boldsymbol{\varepsilon}_t(\hat{\theta})\right\} \quad (42)$$

will be taken as the PC and determined with the key aim:

True (Unbiased) System Identifiability
$\min_{\hat{\theta}} J_t^{\mathrm{a}}(\hat{\theta}) \Leftrightarrow \mathfrak{M}(\hat{\theta}) \equiv \mathfrak{M}(\theta^\dagger) \vee \mathfrak{M}^*(\hat{\theta}) \equiv \mathfrak{M}^*(\theta^\dagger)$

Here, the equivalence symbol $\equiv$ needs clarification. Its sense correlates with the above concept of convergence (39). Necessary refinements will be done (in Theorem 2).

6.2. API Generalized Residual

Since (41) requires many-step-predicted value $\hat{y}_{t+1|t}^{t+s|t}$ as a stackable vector

$$\hat{y}_{t+1|t}^{t+s|t} = \left[\hat{y}_{t+1|t}^{\mathrm{T}} \mid \hat{y}_{t+2|t}^{\mathrm{T}} \mid \cdots \mid \hat{y}_{t+s|t}^{\mathrm{T}}\right]^{\mathrm{T}} \quad (43)$$

we supplement our model (35), (36) with the k-step ahead predictors

$$\begin{aligned} \hat{g}_{t+k|t} &= \boldsymbol{A}\hat{g}_{t+k-1|t} + \boldsymbol{F}u_{t+k-1} \\ \hat{y}_{t+k|t} &= \boldsymbol{C}\hat{g}_{t+k|t} \end{aligned} \quad k = \overline{1,s} \quad (44)$$

and call the residual in (41) by Generalized Residual

$$\mathrm{GR}: \eta_{t+1|t}^{t+s|t} = y_{t+1}^{t+s} - \hat{y}_{t+1|t}^{t+s|t} \quad (45)$$

Recursively using (2) with notations (23)-(28) yields

$$\begin{aligned} y_{t+1}^{t+s} = & \ \boldsymbol{W}(\boldsymbol{H},\boldsymbol{\Phi})x_{t+1} + \boldsymbol{F}_0(\boldsymbol{H},\boldsymbol{\Phi},\boldsymbol{\Psi})u_{t+1}^{t+s} \\ & + \boldsymbol{F}_0(\boldsymbol{H},\Phi,\Gamma)w_{t+1}^{t+s} + v_{t+1}^{t+s} \\ y_{t+1}^{t+s} = & \ \boldsymbol{W}(\boldsymbol{H},\Phi)\Phi x_t + \boldsymbol{F}_1(\boldsymbol{H},\Phi,\boldsymbol{\Psi})u_t^{t+s-1} \\ & + \boldsymbol{F}_1(\boldsymbol{H},\Phi,\Gamma)w_t^{t+s-1} + v_{t+1}^{t+s} \end{aligned} \quad (46)$$

From (11)--(13) by the same technique, we obtain

$$\begin{aligned} y_{t+1}^{t+s} = & \ \boldsymbol{W}(\boldsymbol{H},\Phi)x_{t+1|t} + \boldsymbol{F}_0(\boldsymbol{H},\Phi,\boldsymbol{\Psi})u_{t+1}^{t+s} \\ & + \boldsymbol{S}_0(\boldsymbol{H},\Phi,\boldsymbol{G})\nu_{t+1|t}^{t+s|t+s-1} \\ y_{t+1}^{t+s} = & \ \boldsymbol{W}(\boldsymbol{H},\Phi)\Phi x_{t|t} + \boldsymbol{F}_1(\boldsymbol{H},\Phi,\boldsymbol{\Psi})u_t^{t+s-1} \\ & + \boldsymbol{S}_0(\boldsymbol{H},\Phi,\boldsymbol{G})\nu_{t+1|t}^{t+s|t+s-1} \end{aligned} \quad (47)$$

Applying the same technique to (35), (36) together with (44) yields

$$\begin{aligned} \hat{y}_{t+1|t}^{t+s|t} &= \boldsymbol{W}(\boldsymbol{C},\boldsymbol{A})\hat{g}_{t+1|t} + \boldsymbol{F}_0(\boldsymbol{C},\boldsymbol{A},\boldsymbol{F})u_{t+1}^{t+s} \\ \hat{y}_{t+1|t}^{t+s|t} &= \boldsymbol{W}(\boldsymbol{C},\boldsymbol{A})\boldsymbol{A}\hat{g}_{t|t} + \boldsymbol{F}_1(\boldsymbol{C},\boldsymbol{A},\boldsymbol{F})u_t^{t+s-1} \end{aligned} \quad (48)$$

and

$$y_{t+1}^{t+s} = \hat{y}_{t+1|t}^{t+s|t} + \boldsymbol{S}_0(\boldsymbol{C},\boldsymbol{A},\boldsymbol{B})\eta_{t+1|t}^{t+s|t+s-1} \quad (49)$$

Remark 11 *The following diagram* (***Figure 6***) *shows that so far we have obtained formulae* (46) *to* (49) *only for the PhDM. We now turn to the case of SODM* (*by application of* (5)).

In this case, substituting $\boldsymbol{H}_*$ from (7) for $\boldsymbol{H}$ and $\boldsymbol{C}$ into (46)-(48) as well as taking there $\Phi = \Phi_*$ and $\boldsymbol{A} = \boldsymbol{A}_*$ in the form of (8) and using transformations (31), we obtain:

a) for the conventional form —

$$\begin{aligned} \mathscr{S}\left(y_{t+1}^{t+s}\right) = & \ x_{t+1} + \mathscr{T}(\boldsymbol{\Psi})u_{t+1}^{t+s} \\ & + \mathscr{T}(\Gamma)w_{t+1}^{t+s} + \mathscr{S}\left(v_{t+1}^{t+s}\right) \\ \mathscr{S}\left(y_{t+1}^{t+s}\right) = & \ \Phi x_t + \mathscr{B}(\boldsymbol{\Psi})u_t^{t+s-1} \\ & + \mathscr{B}(\Gamma)w_t^{t+s-1} + \mathscr{S}\left(v_{t+1}^{t+s}\right) \end{aligned} \quad (50)$$

b) for the innovation form —

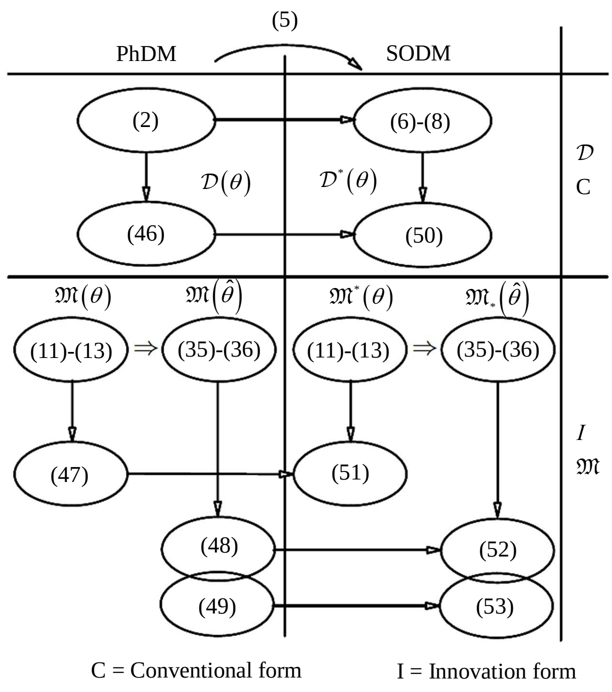

Figure 6. Taxonomy of models. The arrows indicate the consequtive order in which the numbered formulae are obtained.

$$\begin{aligned}\mathscr{S}\left(y_{t+1}^{t+s}\right) = &\ x_{t+1|t} + \mathscr{F}\left(\Psi\right)u_{t+1}^{t+s} \\ &+ \mathscr{P}\left(\boldsymbol{G}\right)v_{t+1|t}^{t+s|t+s-1} \\ \mathscr{S}\left(y_{t+1}^{t+s}\right) = &\ \Phi x_{t|t} + \mathscr{B}\left(\Psi\right)u_{t}^{t+s-1} \\ &+ \mathscr{P}\left(\boldsymbol{G}\right)v_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{51}$$

c) for the predicted values —

$$\begin{aligned}\mathscr{S}\left(\hat{y}_{t+1|t}^{t+s|t}\right) &= \hat{g}_{t+1|t} + \mathscr{F}\left(\boldsymbol{F}\right)u_{t+1}^{t+s} \\ \mathscr{S}\left(\hat{y}_{t+1|t}^{t+s|t}\right) &= \boldsymbol{A}\hat{g}_{t|t} + \mathscr{B}\left(\boldsymbol{F}\right)u_{t}^{t+s-1}\end{aligned} \tag{52}$$

and using $\mathscr{S}$ -transform in (49) yields

$$\mathscr{S}\left(y_{t+1}^{t+s}\right) = \mathscr{S}\left(\hat{y}_{t+1|t}^{t+s|t}\right) + \mathscr{P}\left(\boldsymbol{B}\right)\eta_{t+1|t}^{t+s|t+s-1} \tag{53}$$

Now that we have written formulae (46) to (53), many of them arranged in pairs, we find all possible representations for the GR by substituting these equations into (45):

1) for the PhDM with the predicted output $\hat{g}_{t+1|t}$ —

the 1st of (46) *minus* the 1st of (48) *gives* (54)

$$\begin{aligned}\eta_{t+1|t}^{t+s|t} = &\left[\boldsymbol{W}\left(\boldsymbol{H},\Phi\right)x_{t+1} - \boldsymbol{W}\left(\boldsymbol{C},\boldsymbol{A}\right)\hat{g}_{t+1|t}\right] \\ &+\left[\boldsymbol{F}_0\left(\boldsymbol{H},\Phi,\Psi\right) - \boldsymbol{F}_0\left(\boldsymbol{C},\boldsymbol{A},\boldsymbol{F}\right)\right]u_{t+1}^{t+s} \\ &+\left[\boldsymbol{F}_0\left(\boldsymbol{H},\Phi,\Gamma\right)w_{t+1}^{t+s} + v_{t+1}^{t+s}\right]\end{aligned} \tag{54}$$

the 1st of (47) *minus* the 1st of (48) *gives* (55)

$$\begin{aligned}\eta_{t+1|t}^{t+s|t} = &\left[\boldsymbol{W}\left(\boldsymbol{H},\Phi\right)x_{t+1|t} - \boldsymbol{W}\left(\boldsymbol{C},\boldsymbol{A}\right)\hat{g}_{t+1|t}\right] \\ &+\left[\boldsymbol{F}_0\left(\boldsymbol{H},\Phi,\Psi\right) - \boldsymbol{F}_0\left(\boldsymbol{C},\boldsymbol{A},\boldsymbol{F}\right)\right]u_{t+1}^{t+s} \\ &+\boldsymbol{S}_0\left(\boldsymbol{H},\Phi,\Gamma\right)v_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{55}$$

2) for the PhDM with the filtered output $\hat{g}_{t|t}$ —
the 2nd of (46) *minus* the 2nd of (48) *gives* (56)

$$\begin{aligned}\eta_{t+1|t}^{t+s|t} = &\left[\boldsymbol{W}\left(\boldsymbol{H},\Phi\right)\Phi x_{t} - \boldsymbol{W}\left(\boldsymbol{C},\boldsymbol{A}\right)\boldsymbol{A}\hat{g}_{t|t}\right] \\ &+\left[\boldsymbol{F}_1\left(\boldsymbol{H},\Phi,\Psi\right) - \boldsymbol{F}_1\left(\boldsymbol{C},\boldsymbol{A},\boldsymbol{F}\right)\right]u_{t}^{t+s-1} \\ &+\left[\boldsymbol{F}_1\left(\boldsymbol{H},\Phi,\Gamma\right)w_{t}^{t+s-1} + v_{t+1}^{t+s}\right]\end{aligned} \tag{56}$$

the 2nd of (47) *minus* the 2nd of (48) *gives* (57)

$$\begin{aligned}\eta_{t+1|t}^{t+s|t} = &\left[\boldsymbol{W}\left(\boldsymbol{H},\Phi\right)\Phi x_{t} - \boldsymbol{W}\left(\boldsymbol{C},\boldsymbol{A}\right)\boldsymbol{A}\hat{g}_{t|t}\right] \\ &+\left[\boldsymbol{F}_1\left(\boldsymbol{H},\Phi,\Psi\right) - \boldsymbol{F}_1\left(\boldsymbol{C},\boldsymbol{A},\boldsymbol{F}\right)\right]u_{t}^{t+s-1} \\ &+\boldsymbol{S}_0\left(\boldsymbol{H},\Phi,\boldsymbol{G}\right)v_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{57}$$

and from (49) (taking into account (27)-(28) and relation $\boldsymbol{B} = \boldsymbol{AD}$), we have

$$\begin{aligned}\eta_{t+1|t}^{t+s|t} = &\ \boldsymbol{S}_0\left(\boldsymbol{C},\boldsymbol{A},\boldsymbol{B}\right)\eta_{t+1|t}^{t+s|t+s-1} \\ = &\ \boldsymbol{S}_1\left(\boldsymbol{C},\boldsymbol{A},\boldsymbol{D}\right)\eta_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{58}$$

3) for the SODM with the predicted output $\hat{g}_{t+1|t}$ —
the 1st of (50) *minus* the 1st of (52) *gives* (59)

$$\begin{aligned}\mathscr{S}\left(\eta_{t+1|t}^{t+s|t}\right) = &\left(x_{t+1} - \hat{g}_{t+1|t}\right) \\ &+\left[\mathscr{F}\left(\Psi\right) - \mathscr{F}\left(\boldsymbol{F}\right)\right]u_{t+1}^{t+s} \\ &+\left[\mathscr{F}\left(\Gamma\right)w_{t+1}^{t+s} + \mathscr{S}\left(v_{t+1}^{t+s}\right)\right]\end{aligned} \tag{59}$$

the 1st of (51) *minus* the 1st of (52) *gives* (60)

$$\begin{aligned}\mathscr{S}\left(\eta_{t+1|t}^{t+s|t}\right) = &\left(x_{t+1|t} - \hat{g}_{t+1|t}\right) \\ &+\left[\mathscr{F}\left(\Psi\right) - \mathscr{F}\left(\boldsymbol{F}\right)\right]u_{t+1}^{t+s} \\ &+\mathscr{P}\left(\boldsymbol{G}\right)v_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{60}$$

4) for the SODM with the filtered output $\hat{g}_{t|t}$ —
the 2nd of (50) *minus* the 2nd of (52) *gives* (61)

$$\begin{aligned}\mathscr{S}\left(\eta_{t+1|t}^{t+s|t}\right) = &\left(\Phi x_{t} - \boldsymbol{A}\hat{g}_{t|t}\right) \\ &+\left[\mathscr{B}\left(\Psi\right) - \mathscr{B}\left(\boldsymbol{F}\right)\right]u_{t}^{t+s-1} \\ &+\left[\mathscr{B}\left(\Gamma\right)w_{t}^{t+s-1} + \mathscr{S}\left(v_{t+1}^{t+s}\right)\right]\end{aligned} \tag{61}$$

the 2nd of (51) *minus* the 2nd of (52) *gives* (62)

$$\begin{aligned}\mathscr{S}\left(\eta_{t+1|t}^{t+s|t}\right) = &\left(\Phi x_{t|t} - \boldsymbol{A}\hat{g}_{t|t}\right) \\ &+\left[\mathscr{B}\left(\Psi\right) - \mathscr{B}\left(\boldsymbol{F}\right)\right]u_{t}^{t+s-1} \\ &+\mathscr{P}\left(\boldsymbol{G}\right)v_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{62}$$

and from (58) (taking into account (31)), we have

$$\mathscr{S}\left(\eta_{t+1|t}^{t+s|t}\right) = \mathscr{P}\left(\boldsymbol{B}\right)\eta_{t+1|t}^{t+s|t+s-1} = \mathscr{N}\left(\boldsymbol{D}\right)\eta_{t+1|t}^{t+s|t+s-1} \tag{63}$$

Generalized Residual, as introduced in (45), is important in allowing the user to extract all possible amount of information from experimental condition $\mathscr{X}$ (4) concerning the unmeasurable errors (40). The foregoing equations (54) to (63) reveal the features of GR from the practical standpoint as a possible tool for system identification under different levels of uncertainty and with different data models used: PhDM or SODM.

6.3. API Identifiability of $\mathfrak{M}^*\left(\theta^\dagger\right)$

Let the auxiliary process (41) for the API (42) be built as

$$\begin{aligned}\varepsilon_{t+1|t} = &\ \mathscr{S}\left(y_{t+1}^{t+s}\right) - \hat{g}_{t+1|t} - \mathscr{F}\left(\boldsymbol{F}\right)u_{t+1}^{t+s} \\ = &\ \mathscr{P}\left(\boldsymbol{B}\right)\eta_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{64}$$

(**Figures 7** and **8**) or, equivalently, as

$$\begin{aligned}\varepsilon_{t|t} = &\ \mathscr{S}\left(y_{t+1}^{t+s}\right) - \boldsymbol{A}\hat{g}_{t|t} - \mathscr{B}\left(\boldsymbol{F}\right)u_{t}^{t+s-1} \\ = &\ \mathscr{N}\left(\boldsymbol{D}\right)\eta_{t+1|t}^{t+s|t+s-1}\end{aligned} \tag{65}$$

(**Figures 9** and **10**).

Theorem 2 *Let* $\varepsilon_t\left(\hat{\theta}\right)$ (41) *be a vector-valued* n*-component function of* (45). *If* $\varepsilon_t\left(\hat{\theta}\right)$ *is defined by* (64) *or (equivalently)* (65) *in order to form the API* (42), *then minimum in* $\hat{\theta}$ *of the API fixed out at any instant* t *is*

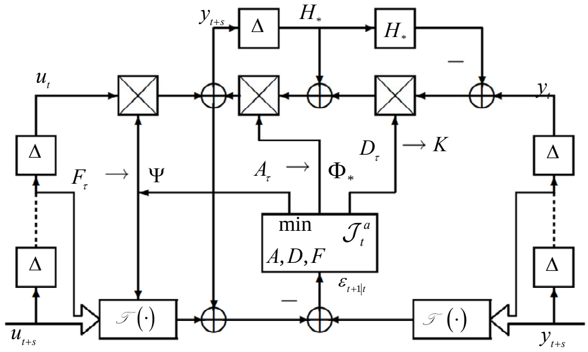

Figure 7. Adaptor based on $\varepsilon_{t+1|t}$, the first equality in (64). Δ denotes the unitary delay operator.

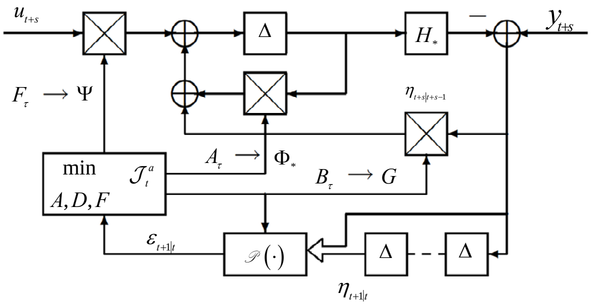

Figure 8. Adaptor based on $\varepsilon_{t+1|t}$, the second equality in (64).

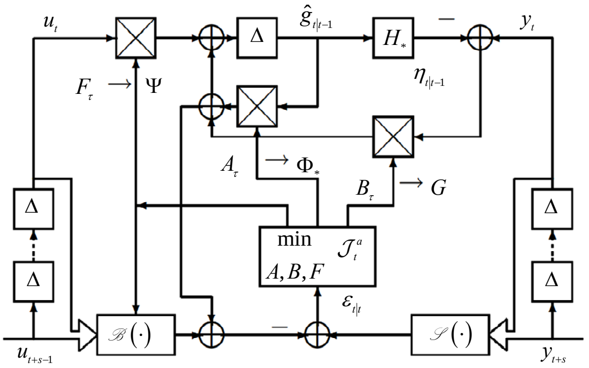

Figure 9. Adaptor based on $\varepsilon_{t|t}$, the first equality in (65).

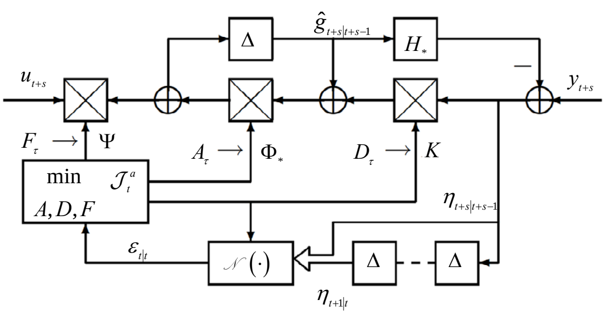

Figure 10. Adaptor based on $\varepsilon_{t|t}$, the second equality in (65).

the necessary and sufficient condition for adaptive model $\mathfrak{M}^*(\hat{\theta})$ to be consistent estimator of $\mathfrak{M}^*(\theta^\dagger)$ in mean square, $\forall t \in \mathbb{Z}_+ : \mathfrak{M}^*(\hat{\theta}) \overset{\text{m.s.}}{\equiv} \mathfrak{M}^*(\theta^\dagger)$, that is

True (Unbiased) *m.s.* System Identifiability

$$\min_{\hat{\theta}} \mathcal{J}_t^{a}(\hat{\theta}) \Leftrightarrow \boldsymbol{E}\left\{\left\| x_{t+1|t} - \hat{g}_{t+1|t} \right\|^2\right\} = 0$$

in the following three setups:

Setup 1 (Random Control Input) $\{u(t)\}$ is a preassigned zero-mean orthogonal wide-sence stationary process orthogonal to $\{w(t), v(t)\}$ but in contrast to $\{w(t)\}$ and $\{v(t)\}$, known and serving as a testing signal;

Setup 2 (Pure Filtering) $\forall t \in \mathbb{Z}_+ : u(t) = 0$, *and*

Setup 3 (Close-loop Control) with known $\boldsymbol{F} = \boldsymbol{\Psi}$.

Proof: See the Appendix.

Corollary 1 *Under the assumptions of Theorem* 2, *minimum in $\hat{\theta}$ of the API fixed out at any $t \in \mathbb{Z}_+$ is the necessary and sufficient condition for adaptive model $\mathfrak{M}^*(\hat{\theta})$ to be consistent estimator of $\mathfrak{M}^*(\theta^\dagger)$ in mean square, $\forall t \in \mathbb{Z}_+ : \mathfrak{M}^*(\hat{\theta}) \overset{\text{m.s.}}{\equiv} \mathfrak{M}^*(\theta^\dagger)$, up to the equality*

$$x_{t+1|t} + \mathscr{T}(\Psi) u_{t+1}^{t+s} = \hat{g}_{t+1|t} + \mathscr{T}(\boldsymbol{F}) u_{t+1}^{t+s} \tag{66}$$

or, what is equivalent, equality

$$\Phi x_{t|t} + \mathscr{B}(\Psi) u_t^{t+s-1} = \boldsymbol{A}\, \hat{g}_{t|t} + \mathscr{B}(\boldsymbol{F}) u_t^{t+s-1} \tag{67}$$

Corollary 2 *Under the assumptions of Theorem* 2, *if upper s rows of $\boldsymbol{\Psi}$ and $\boldsymbol{F}$ are zero, then in Corollary* 1 *equalities* (66)-(67) *are replaced by equalities*

$$\forall t \in \mathbb{Z}_+ : x_{t+1|t}^{(j)} = \hat{g}_{t+1|t}^{(j)}; \quad \Phi x_{t|t}^{(i)} = \boldsymbol{A}\, \hat{g}_{t|t}^{(i)} \tag{68}$$

where $j = \overline{1,(s+1)}$ and $i = \overline{1,s}$ are numbers of vector components. If additionally $u_t \equiv 0$ or if $\{u_t\}$ is orthogonal to $\{w_t\}$ and $\{v_t\}$ for models (11), (13), (35) and (36), then equalities (68) hold for all vector components, and the following equalities

$$\boldsymbol{A} = \Phi, \quad \boldsymbol{B} = \boldsymbol{G}, \quad \boldsymbol{D} = \boldsymbol{K}, \quad \boldsymbol{F} = \boldsymbol{\Psi} \tag{69}$$

are added to (68) thus assuring that $\mathfrak{M}^*(\hat{\theta}) \overset{\text{m.s.}}{\equiv} \mathfrak{M}^*(\theta^\dagger)$.

Proof of (68) and $\boldsymbol{F} = \boldsymbol{\Psi}$ leans upon Algorithm 1 for computing matrix $\mathscr{T}$ (see Section 5). After that, equalities $\boldsymbol{A} = \Phi$, $\boldsymbol{B} = \boldsymbol{G}$ and $\boldsymbol{D} = \boldsymbol{K}$ follow from uniqueness of Kalman filter equations.

Corollary 3 *If* Φ, Γ, $\boldsymbol{Q}$, *and* $\boldsymbol{R}$ *are unknown and* $\boldsymbol{H}$ (7) *and* $\boldsymbol{\Psi}$ *known, thereby allowing for situation when $\hat{\theta} \succ (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{D})$, then in Corollary* 1 *equalities* (66)-(67) *are replaced by equalities*

$$\begin{gathered} \forall t \in \mathbb{Z}_+ : \quad x_{t+1|t} = \hat{g}_{t+1|t}, \quad x_{t|t} = \hat{g}_{t|t} \\ \boldsymbol{A} = \Phi, \quad \boldsymbol{B} = \boldsymbol{G}, \quad \boldsymbol{D} = \boldsymbol{K} \end{gathered} \tag{70}$$

thus assuring that $\mathfrak{M}^*(\hat{\theta}) \overset{\text{m.s.}}{\equiv} \mathfrak{M}^*(\theta^\dagger)$.

6.4. API Identifiability of $\mathfrak{M}\left(\theta^{\dagger}\right)$

When transformation of PhDM, $\mathfrak{M}(\theta)$ to SODM, $\mathfrak{M}^{*}(\theta)$ is troublesome or objectionable, we have to work with the given PhDM. In this case, we take, as $\varepsilon_t\left(\hat{\theta}\right)$ for (42), either of the two processes:

$$\varepsilon_{t+1|t} = \mathbf{W}_1^{+}\eta_{t+1|t}^{t+s|t}; \quad \varepsilon_{t|t} = \Phi_1^{-1}\mathbf{W}_1^{+}\eta_{t+1|t}^{t+s|t}, \tag{71}$$

where relation (58) can be used. In (71), we denote: Φ_1 to be an ($n\times n$)-matrix; H_1 an $(m\times n)$-matrix, and $W_1 = W\left(H_1,\Phi_1\right)$ with $W(\cdot,\cdot)$ defined by (24). We assign matrices H_1, Φ_1 rather arbitrarily choosing them as substitutes for unknown matrices H and Φ from the conditions: $\left|\Phi_1\right|\neq 0$ and $\mathrm{r}ank\, W_1 = n$. By the latter condition, the pseudo-inverse matrix W_1^{+} is found as $\left(W_1^{\mathrm{T}}W_1\right)^{-1}W_1^{\mathrm{T}}$, and the following $(n\times n)$-matrices

$$\begin{aligned} &T \triangleq W_1^{+}W(H,\Phi); && \hat{T} \triangleq W_1^{+}W(C,A)\\ &T_0 \triangleq \Phi_1^{-1}T\Phi; && \hat{T}_0 \triangleq \Phi^{-1}\hat{T}A \end{aligned} \tag{72}$$

are non-singular if the addional conditions

$$\begin{aligned} |\Phi| &\neq 0; \quad \mathrm{rank}\, W(H,\Phi) = n\\ |A| &\neq 0; \quad \mathrm{rank}\, W(C,A) = n \end{aligned} \tag{73}$$

hold.

The last-added term in (55), as well as in (57), does not depend on the model parameter $\hat{\theta} \succ (A,B,C,D,F)$ and is formed by the innovation process $\nu_{t+1|t}$, which is an orthogonal process. In addition, the set $\mathfrak{M}$ of models (35), or equivalently (36) contains the optimal model (11), (13). On this grounds, we come to the following result.

Theorem 3 *Let* $\varepsilon_t\left(\hat{\theta}\right)$ (41) *be a vector-valued* n*-component function of* (45). *If* $\varepsilon_t\left(\hat{\theta}\right)$ *is taken from* (71) *in order to form the API* (42), *then minimum in* $\hat{\theta} \succ (A,B,C,D,F)$ *of the API fixed out at any instant* t *is the necessary and sufficient condition for adaptive model* $\mathfrak{M}\left(\hat{\theta}\right)$ *to be consistent estimator of* $\mathfrak{M}\left(\theta^{\dagger}\right)$ *in mean square,* $\forall t\in\mathbb{Z}_+ : \mathfrak{M}\left(\hat{\theta}\right) \overset{\text{m.s.}}{\equiv} \mathfrak{M}\left(\theta^{\dagger}\right)$ *up to the equations*

$$\begin{aligned} &x_{t+1|t} - \hat{T}\hat{g}_{t+1|t}\\ &= W_1^{+}\left[F_0(C,A,F) - F_0(H,\Phi,\Psi)\right]u_{t+1}^{t+s} \end{aligned} \tag{74}$$

in case of the first process of (71), and

$$\begin{aligned} &T_0 x_{t|t} - \hat{T}_0\ \hat{g}_{t|t}\\ &= \Phi_1^{-1}W_1^{+}\left[F_1(C,A,F) - F_1(H,\Phi,\Psi)\right]u_t^{t+s-1} \end{aligned} \tag{75}$$

in case of the second process of (71) taken to form the API.

Proof: Given in [25].

Corollary 4 *On the assumption that* Γ, Q *and* R *are unknown, but* Φ, Ψ *and* H *known, i. e. allowing for situation when* $\hat{\theta} \succ (B,D)$ *and* $H_1 = H$, $\Phi_1 = \Phi$ *in* (71), *equalities* (74) *and* (75) *in Theorem* 3 *are replaced by equalities*

$$\begin{aligned} \forall t\in\mathbb{Z}_+ : \quad & x_{t+1|t} = \hat{g}_{t+1|t}; \quad x_{t|t} = \hat{g}_{t|t}\\ & B = G, \quad D = K \end{aligned} \tag{76}$$

This follows from the right sides of (74)-(75) being zero and from $T = \hat{T} = T_0 = \hat{T}_0 = I$ under the assumption of this corollary.

Remark 12 *The similar result however relating to the second process of* (71) *only and in the following form of Learning Criterion* (LC)

$$\forall t\in\mathbb{Z}_+ :$$
$$E\left\{\left[\Phi^{-1}W^{+}\left(y_{t+1}^{t+s} - W\Phi\hat{g}_{t|t-1}\right) - D\eta_{t|t-1}\right]\eta_{t|t-1}^{\mathrm{T}}\right\} = 0$$

with $W = W(H,\Phi)$, and matrix D being the only adjustable parameter of filter (36) in the event that $A=\Phi$, $C = H$, and $u_t \equiv 0$, was also obtained [26] and restated in a different way [27] by Hampton where the problem of minimizing LC in the form of

$$\left\|\Phi^{-1}W^{+}\left(y_{t+1}^{t+s} - W\Phi\hat{g}_{t|t-1}\right) - D\eta_{t|t-1}\right\|^2$$

was formulated. The convergence properties of Hampton's solution were studied by Perriot-Mathonna [28].

Corollary 5 *On the assumption that* Φ, Γ, Q *and* R *are unknown and* $u_t \equiv 0$, *i. e. allowing for situation when* $\hat{\theta} \succ (A,B,C,D)$, *equalities* (74) *and* (75) *in Theorem* 3 *are replaced by equalities*

$$\begin{aligned} &\forall t\in\mathbb{Z}_+ : \ \hat{g}_{t+1|t} = Sx_{t+1|t}; \quad \hat{g}_{t|t} = Sx_{t|t}\\ &A = S\Phi S^{-1};\ B = S\mathrm{G};\ C = HS^{-1};\ D = SK \end{aligned} \tag{77}$$

where S *is an arbitrary non-singular* $(n\times n)$*-matrix.*

The proof is obtained by setting the right sides of (74)-(75) to zero if definitions (72) are accounted for. In so doing, we consider $S = \hat{T}^{-1}T$ (or $S = \hat{T}_0^{-1}T$) an unknown matrix of similarity transformation. A pure algebraic proof of the result is also available due to [25,29]. Analyzing conditions of Corollary 5 in more detail, state the following results.

Corollary 6 *On the assumption that* Φ, Γ, Q *and* R *are unknown and* H *known and* $u_t \equiv 0$, *i. e. allowing for situation when* $\hat{\theta} \succ (A,B,D)$ *and* $C = H$, *equalities* (74)-(75) *in Theorem* 3 *are replaced by equalities*

$$\begin{aligned} &\forall t\in\mathbb{Z}_+ : \quad \hat{g}_{t+1|t} = Sx_{t+1|t}; \quad \hat{g}_{t|t} = Sx_{t|t}\\ &A = S\Phi S^{-1};\ B = SG;\ D = SK \end{aligned} \tag{78}$$

where S is a non-singular $(n\times n)$-matrix subject to relation $H = HS$.

Corollary 7 *On the assumption that* Γ, Q, H *and* R *are unknown and* Φ *known and* $u_t \equiv 0$, *i. e. allowing for situation when* $\hat{\theta} \succ (B,C,D)$ *and* $A = \Phi$, *equalities* (74)-(75) *in Theorem* 3 *are replaced by equalities*

$$\forall t \in \mathbb{Z}_+ : \quad \hat{g}_{t+1|t} = \boldsymbol{S} x_{t+1|t}; \quad \hat{g}_{t|t} = \boldsymbol{S} x_{t|t}$$
$$\boldsymbol{B} = \boldsymbol{SG}; \quad \boldsymbol{C} = \boldsymbol{HS}^{-1}; \quad \boldsymbol{D} = \boldsymbol{SK} \tag{79}$$

where $\boldsymbol{S}$ is a non-singular $(n \times n)$-matrix subject to relation $\Phi \boldsymbol{S} = \boldsymbol{S} \Phi$.

Corollary 8 *On the assumption that* Γ, $\boldsymbol{Q}$, *and* $\boldsymbol{R}$ *are unknown and* Φ $\boldsymbol{H}$ *known and* $u_t \equiv 0$, *i. e. allowing for situation when* $\hat{\theta} \succ (\boldsymbol{B}, \boldsymbol{D})$, $\boldsymbol{A} = \Phi$ *and* $\boldsymbol{C} = \boldsymbol{H}$, *equalities* (74)-(75) *in Theorem* 3 *are replaced by equalities*

$$\forall t \in \mathbb{Z}_+ : \quad \hat{g}_{t+1|t} = x_{t+1|t}; \quad \hat{g}_{t|t} = x_{t|t}$$
$$\boldsymbol{B} = \boldsymbol{G}, \quad \boldsymbol{D} = \boldsymbol{K} \tag{80}$$

This is a special case of Corollary 4 at $u_t \equiv 0$.

Corollary 9 *If under the assumptions of Corollary* 6 r*ank* $\boldsymbol{H} = n$, *then* $\boldsymbol{S} = \boldsymbol{I}$ *(cf. theorems of* [30]).

Corollary 10 *If under the assumptions of Corollary* 6 *matrix* $\boldsymbol{H}$ *is given by (7) and* $\boldsymbol{A}$ *is sought in the form of* (8), *then* $\boldsymbol{S} = \boldsymbol{W}_*$ *when* $\boldsymbol{W}_*$ *is matrix* (10). *If in addition* Φ *has the form of* (8), *then* $\boldsymbol{S} = \boldsymbol{I}$.

This establishes the association of this case with the identifiability of matrices Φ, $\boldsymbol{G}$ and $\boldsymbol{K}$ for SODM — cf. Corollary 3.

Corollary 11 *For the adaptive model* (35)-(36), *which is optimal in structure,* $(n \times n)$ *-matrix* $\boldsymbol{H}$ *cannot be identified with the APIs under consideration*, *i. e. it must be known. As this takes place, the whole amount of identifiable entries of matrix* Φ *is* $(m \times n)$.

This follows from Corollary 7 where we have the Frobenius Problem of finding matrices $\boldsymbol{S}$ commutative with the given matrix Φ. It is known that the total number of linear independent solutions to the problem is not less than n. The second part of Corollary 11 is obtained from similarity (isomorphism) of any system (2) to its standard observable form (6)-(8).

6.5. Main Conceptual Novelty

The goal of an identification method is to find a model, whose “behavior” best approximates that of the system under consideration. However, what meaning may be attributed to the term “behavior”? In the context of APA, the *inner state of a dynamical system* is emphasized, whereas classical MPE methods imply the *output behavior*. Theorems 2-3 and corollaries solve the task by an indirect minimization of either the errors (40) in prediction or mean square estimation of the inner state, whereas classical MPE methods do so by a direct minimization of a “prediction error criterion,” which expresses the one-step “prediction performance” of the model on the given input-output experimental condition (4). The difference is illustrated graphically in **Figures 2** and **3**. At the same time, both approaches share a common trait of having a proximity criterion to be numerically minimized. Subspace-based identification methods [31] also put emphasis on *state of a dynamical system*, but by doing the following: combining the past input-output data and future inputs linearly to predict future outputs; minimizing the error of prediction measured in the Frobenius norm; obtaining the KF state sequence by using the robust Singular Value Decomposition; and finally, estimating system matrices with Least Squares techniques.

Theorem 2 and its corollaries establish the point that generally, the API approach is rather useful as it helps us identify unknown parameters of optimal discrete time filters used either independently or as a part of a control strategy. It also indicates the levels of uncertainty (see Section 4), within which the approach still remains practicable. General characteristic values of these levels are defined by three setups stated in the theorem (and realistically reproduced in *Proof*). Corollary 1 is an accurate generalization of Theorem 2 for the case where there is no point in specifying Setups 1, 2 or 3. A few details of the levels are stated by Corollaries 2 and 3. Feasible schemata for APIs computation visually support these identification results associated with the standard observable data model, SODM (**Figures 7-10**), and the physical data model, PhDM (see below **Figures 11-14**).

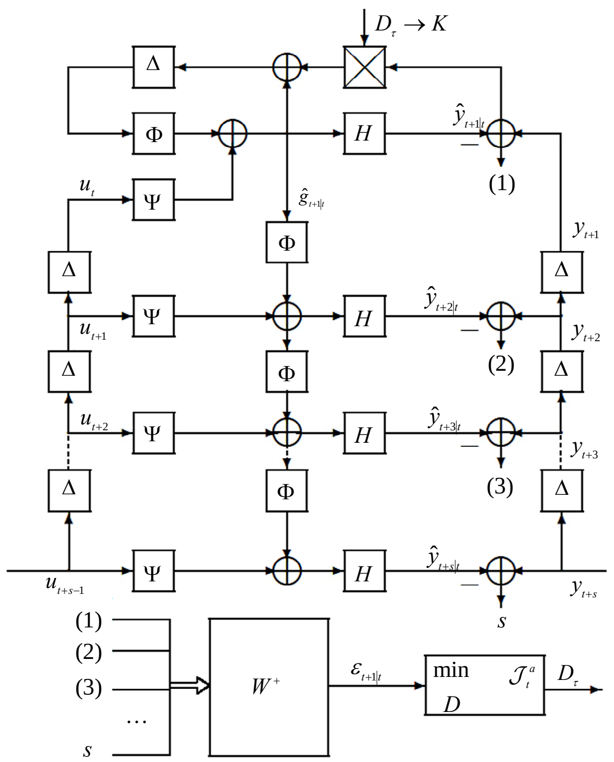

Figure 11. Adaptor based on process $\varepsilon_{t+1|t}$ **from (71) and relation (48) under assumptions of Corollary 4. Here** $\hat{\theta} \succ \{\mathbf{D}\}$**.**

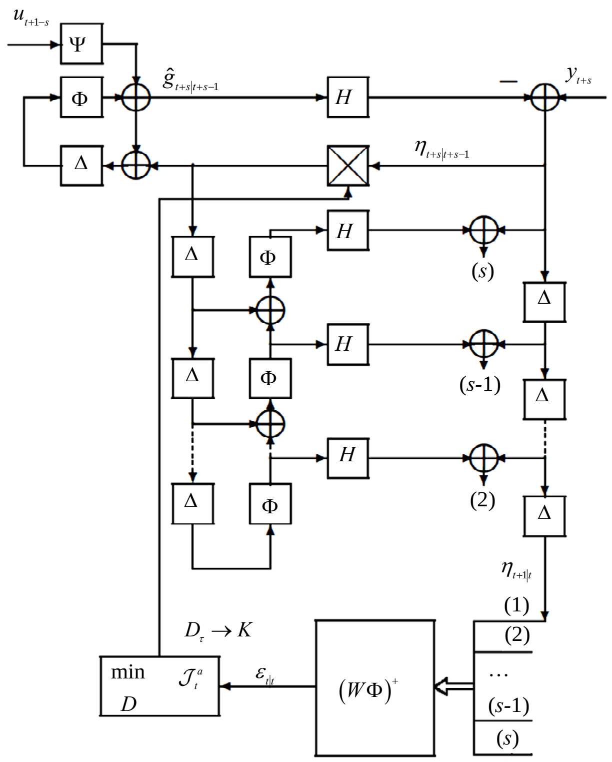

Figure 12. Adaptor based on process $\varepsilon_{t|t}$ from (71) and relation (58) under assumptions of Corollary 4. Here $\hat{\theta} \succ \{\mathbf{D}\}$.

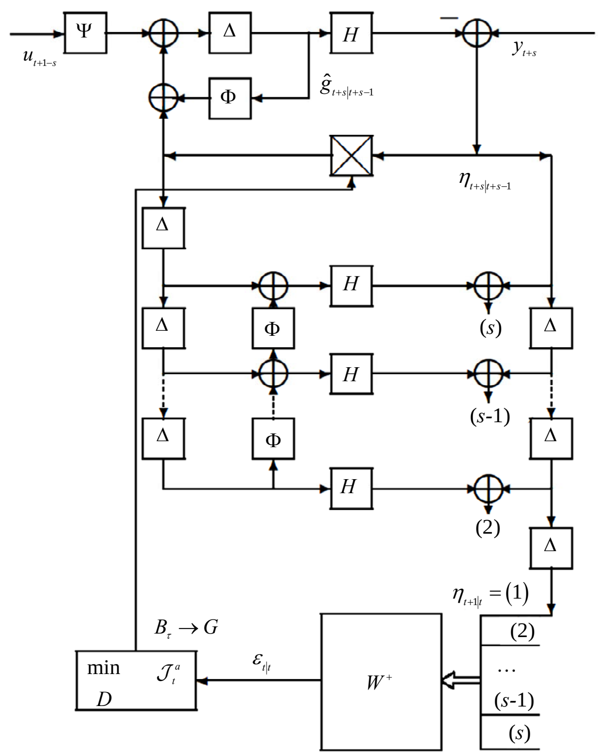

Figure 13. Adaptor based on process $\varepsilon_{t+1|t}$ from (71) and relation (58) under assumptions of Corollary 4. Here $\hat{\theta} \succ \{\mathbf{B}\}$.

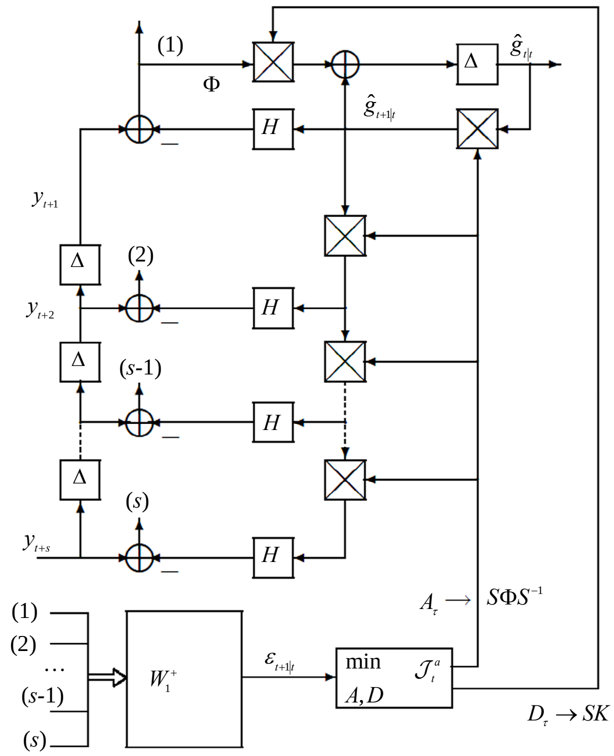

Figure 14. Identification of Φ and K as parameters of optimal steady-state filter. Here $\hat{\theta} \succ \{\mathbf{A}, \mathbf{D}\}$ and $u_t \equiv 0$.

6.6. API Adaptor Forms

We have stated the uniqueness of identification under conditions of corollaries 2 to 4 and 8 to 10. This is expressed by equalities (69), (70), (76), and (80). Identification is accomplished non-uniquely—up to arbitrary similarity transformation if conditions for corollaries 5 to 7 hold. This situation is expressed by equalities (77)-(79). Uniquely accomplished identification is possible in some particular cases, as stated in corollary 6 and shown in **Figure 14**. In this figure, PhDM is used where $\boldsymbol{W}_1 = \boldsymbol{W}(\boldsymbol{H}_1, \Phi_1)$. Matrices Φ and $\boldsymbol{K}$ are identifiable when $\boldsymbol{H}$ is known and if equations $\boldsymbol{HS} = \boldsymbol{H}$ and $\Phi'\boldsymbol{S} = \boldsymbol{S}\Phi$ imply $\boldsymbol{S} = \boldsymbol{I}$ where Φ' is a matrix differing from Φ only in that unknown parameters are denoted differently (cf. Corollary 8). On this condition, $\hat{g}_{t+1|t} \xrightarrow{m.s.} x_{t+1|t}$ and $\hat{g}_{t|t} \xrightarrow{m.s.} x_{t|t}$.

When **Figures 7-10** are compared with **Figures 11-14**, it is apparent that Adaptor for SODM is much simpler than Adaptor for PhDM. These benefits can be realized only if the transition from (2) to (6) has been preliminary performed. If it is the case, the estimates in terms of PhDM can be obtained according to the following statement.

Corollary 12 *Let* $\boldsymbol{W}_*$ (10) *be defined analytically as a known function* $\boldsymbol{W}_* = \boldsymbol{W}_*(\alpha_0)$ *of the unknown parameters* $\alpha_0 = (\alpha_1^0, \ldots, \alpha_{N_1}^0)$ *entering* $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\alpha_0)$ *of* (2) *and to be identified by minimization of criterion* (42), *and* $N_1 \le mn$ *as stated by Corollary* 11. *Let matrix* $\boldsymbol{\Phi}_* = \boldsymbol{W}_* \boldsymbol{\Phi} \boldsymbol{W}_*^{-1}$ *be found in the companion form* (8) *as a known function of* α_0, *namely* $\boldsymbol{\Phi}_* = \boldsymbol{\Phi}_*(\varphi_*)$ *where vector* $\varphi_* = (\varphi_1^*, \cdots, \varphi_{N_2}^*)$ *is composed of* N_2 *non-trivial (i.e. non-zero ot non-unit) unknown entries of* $\boldsymbol{\Phi}_*$ ($0 < N_1 \le N_2 \le mn$), *and in so doing the type of function* $\varphi_* = f(\alpha_0)$ *be determined as continuous function having its inversion* $\alpha_0 = f^{-1}(\varphi_*)$ (*the last-named hypothesis can be true as* $N_1 \le N_2$).

Then minimization of (42) *under conditions of Corollary* 3 *ensures*, *as necessary and sufficient condition*, *attaining the following limits*:

$$
\begin{aligned}
&\boldsymbol{W}_*^{-1}\left(f^{-1}(a)\right)\hat{g}_{t+1|t} \xrightarrow{\text{m.s.}} x_{t+1|t} \\
&\boldsymbol{W}_*^{-1}\left(f^{-1}(a)\right)\hat{g}_{t|t} \xrightarrow{\text{m.s.}} x_{t|t} \\
&\boldsymbol{W}_*^{-1}\left(f^{-1}(a)\right)\boldsymbol{A}\boldsymbol{W}_*\left(f^{-1}(a)\right) \xrightarrow{\text{a.s.}} \boldsymbol{\Phi} \\
&\boldsymbol{W}_*^{-1}\left(f^{-1}(a)\right)\boldsymbol{D} \xrightarrow{\text{a.s.}} \boldsymbol{K}
\end{aligned}
\qquad (81)
$$

where a *are the adjustable parameters of matrix* $\boldsymbol{A}$, $\boldsymbol{A}$ *taken in the form* (8), *if* $a \xrightarrow{\text{a.s.}} \varphi_*$ *by a PAA*.

Proof: Done by inverting equalities (78), see Corollaries 6 and 10.

Thus the corresponding block-diagram (**Figure 15**) differs from the preceding block-diagram (**Figure 14**) by including the operations turning back from SODM to PhDM by relations (81). However, some questions remain open: "Must of necessity the transition from PhDM to SODM and back be performed? What benefits are harboured by the transition if not performed ?"

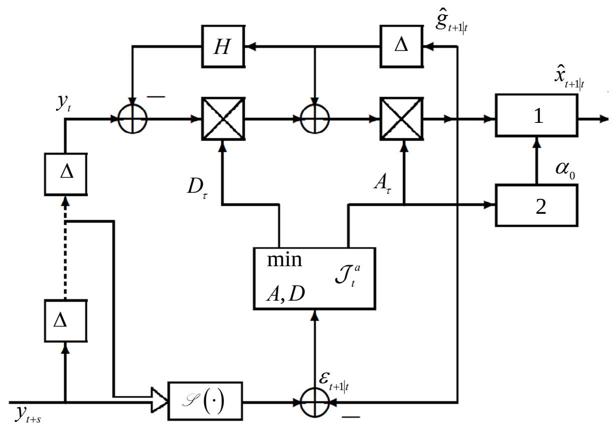

Figure 15. Using $\mathfrak{M}^*(\hat{\theta})$ **(34) for** $D(\theta)$ **(2) with the property** $A_\tau \to W_* \Phi W_*^{-1}$ **and** $D_\tau \to W_* K$. ***Legend*: 1—operation** $W_*^{-1}(\alpha_0)$**; 2---operation** $f^{-1}(a)$.

7. Engineering Illustration & a Rule

Consider a simplified version of the application problem from aeronautical equipment engineering [32] whose complete statement is given in [33].

The simplified version is the instrument error model for one channel of the Inertial Navigation System (INS) of semi-analytical type, which looks as follows:

$$
\begin{bmatrix} \Delta v_x \\ \beta \\ m_{Ax} \\ n_{Gy} \end{bmatrix}_{t+1} = \begin{bmatrix} 1 & -\tau g & \tau & 0 \\ \tau/a & 1 & 0 & \tau \\ 0 & 0 & b_1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \Delta v_x \\ \beta \\ m_{Ax} \\ n_{Gy} \end{bmatrix}_t + \begin{bmatrix} 0 \\ 0 \\ a_1 \\ 0 \end{bmatrix} w_t \qquad (82)
$$
$$
y_t = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} x_t + v_t
$$

where subscripts $_x$, $_y$, $_A$, and $_G$ stand for "axis Ox", "axis Oy", "Accelerometer", and "Gyro", correspondingly.[1] State vector $\boldsymbol{x} = \left[\Delta v_x, \beta, m_{Ax}, n_{Gy}\right]^{\mathrm{T}}$ consists of: Δv_x, random error in reading velocity along axis Ox of a gyro-stabled platform (GSP); β, angular error in determining the local vertical; m_{Ax}, the accelerometer reading random error; and n_{Gy}, the gyro constant drift rate. Parameters $a_1 = H_1\sqrt{1-b_1^2} \simeq H_1\sqrt{2\gamma_1\tau}$ and $b_1 = \exp(-\gamma_1\tau) \simeq 1-\gamma_1\tau$ are obtained from the correlation function

$$
R_{m_{Ax}}(t) = H_1^2 \exp(-\gamma_1 |t|) \qquad (83)
$$

describing m_{Ax} after transition from continuous time t in (83) to the discrete—time index $_t$ in (82) with the sampling period τ.

Let parameters H_1 and γ_1 be unknown. Rewriting (82) in terms of $p_{Ax} = m_{Ax}/a_1$ results in that parameters a_1 and b_1 move into $\boldsymbol{\Phi}$ of equations (2) with

$$
\boldsymbol{\Phi} = \begin{bmatrix} 1 & -\tau g & \tau a_1 & 0 \\ \tau/a & 1 & 0 & \tau \\ 0 & 0 & b_1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Gamma} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad (84)
$$

and $\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$. Using (10) in (5) gives (6) with $\boldsymbol{H}_* = \boldsymbol{H}$ and

[1]These subscripts are a tribute to the engineering tradition: $_x$ must not be confused with x denoting the state vector.

$$
\begin{aligned}
\Gamma_* &= \begin{bmatrix} 0 \\ \tau a_1 \\ \tau\left(1+b_1\right)a_1 \\ \tau\left(1-\rho+b_1+b_1^2\right)a_1 \end{bmatrix} \\
\rho &= \tau^2 g/a \\
\Phi_* &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_4 & -\alpha_3 & -\alpha_2 & -\alpha_1 \end{bmatrix} \\
\alpha_4 &= b_1\left(1+\rho\right) \\
\alpha_3 &= -b_1\left(3+\rho\right)-\left(1+\rho\right) \\
\alpha_2 &= 3b_1+\left(3+\rho\right) \\
\alpha_1 &= -b_1-3
\end{aligned}
\tag{85}
$$

As seen from (85), parameter a_1 vanishes in Φ_*. This is quite reasonable, because it belongs, by its nature, to Γ of (82) (and Γ_* of (85)), that is a_1 has entered Φ of (84) artificially. However it is very difficult to reveal the presence of such non-identifiable parameters in Φ by $\Phi's$ visual appearance. Transition to SODM reveals such problem parameters, as it is the case for (85), and this is its benefit although this may prove to be a difficult algebra.

The question, that ended Section 6.6, can be reformulated: "Is it always necessary to move from PhDM to SODM in order to reveal non-identifiable parameters in matrix Φ?" The answer is: "No," and it is given by Corollary 6 (Section 6.4) from where we obtain the following

General Rule: (Algebraic Identifiability Criterion)

1) Take Φ as it is up to the unknown parameters.
2) Take Φ in the same form as Φ, but using other designations for the unknown parameters.
3) Take an arbitrary $n\times n$ matrix $\boldsymbol{S}$, $\det \boldsymbol{S} \neq 0$ satisfying equation $\boldsymbol{HS}=\boldsymbol{H}$ (matrix $\boldsymbol{H}$ must be known).
4) Write $\Phi'\boldsymbol{S}=\boldsymbol{S}\Phi$ in the component-wise form.
5) If $\boldsymbol{S}=\boldsymbol{I}$ is the only solution, the unknown parameters of Φ are identifiable; you need not do the transition to SODM.
6) If $\boldsymbol{S}=\boldsymbol{I}$ is not the only solution, find the constraints needed to have the solution $\boldsymbol{S}=\boldsymbol{I}$ as unique.
7) Those parameters that require to maintain the found constraints may be non-identifiable; they become identifiable only if the constrains are fulfilled.

Following this Rule in the example yields

$$
\begin{array}{cc}
\Phi & \Phi' \\
\begin{bmatrix} 1 & \alpha & x & 0 \\ \beta & 1 & 0 & \tau \\ 0 & 0 & y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, &
\begin{bmatrix} 1 & \alpha & x_1 & 0 \\ \beta & 1 & 0 & \tau \\ 0 & 0 & y_1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
\text{ and }
\begin{array}{c}
\boldsymbol{S} \\
\begin{bmatrix} 1 & 0 & 0 & 0 \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{bmatrix}
\end{array}
$$

where the following known constants are non-zero: α, β, and τ; the following entries are unknown: $x\neq 0$, $x_1\neq 0$, $y\neq 1$, $y_1\neq 1$; and $|\boldsymbol{S}|\neq 0$. Comparing left and right sides of $\Phi'\boldsymbol{S}=\boldsymbol{S}\Phi$ (this is much more simpler than the transition to SDOM) yields the unique solution $y_1=y$ and $\boldsymbol{S}=\mathrm{diag}[1,1,s_{33},1]$ where $s_{33}=x/x_1$. Element y (that equals b_1 in (84)) is thus seen to be identifiable even if x (that equals τa_1 in (84)) is in error. However x is not identifiable, that is it should be estimated by some other methods. Notice that in active type identification, there is no need to identify $x=\tau a_1$ because optimal gain $\boldsymbol{K}$ and matrix Φ are being estimated directly—avoiding estimation of Γ, $\boldsymbol{Q}$ and $\boldsymbol{R}$ (this is the general result).

8. Simulation Example

E1 Second order system with unknown covariances $\boldsymbol{Q}$ and $\boldsymbol{R}$ of the noises w_t and v_t is given by

$$
\begin{aligned}
x_{t+1} &= \begin{bmatrix} 0 & 1 \\ f_1 & f_2 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ \beta \end{bmatrix} u_t + \begin{bmatrix} 0 \\ \alpha \end{bmatrix} w_t \\
y_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_t + v_t
\end{aligned}
$$

$\alpha=0.4$, $\beta=1.0$, $f_1=-0.8$ and $f_2=0.1$. Kalman gain $\boldsymbol{K}=[k_1\,|\,k_2]^{\mathrm{T}}$ should be estimated by adaptive model gain $\boldsymbol{D}=[d_1\,|\,d_2]^{\mathrm{T}}$, $\hat\theta=(d_1,d_2)$.

E2 The same system as in E1. Unknowns are $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\left(f_1,f_2\right)$ of matrix Φ. Adaptive model parameter is the four-component vector: $\hat\theta=\left(\hat f_1,\hat f_2,d_1,d_2\right)$.

We analyze the adaptive model behavior with the Integral Percent Error (IPE) defined by

$$
\rho_{\mathrm{IPE}} = \|\hat\theta(\tau)-\hat\theta_{\mathrm{OPT}}\|_\infty / \|\hat\theta_{\mathrm{OPT}}\|_\infty \cdot 100\% \tag{86}
$$

with respect to $\hat\theta_{\mathrm{OPT}}$, the optimal value of $\hat\theta$ for different levels of *signal-to-noise ratio*, $SNR=\boldsymbol{Q}/\boldsymbol{R}$.

Results of **Figures 16-17** are obtained in conext of **Figure 10** using the simulation toolbox developed by Gorokhov [34]. The results of this and other simulation experiments confirm applicability of the presented method.

9. Conclusions

This paper develops The Active Principle of Adaptation for linear time-invariant state-space stochastic MIMO filter systems included into the feedback or considered

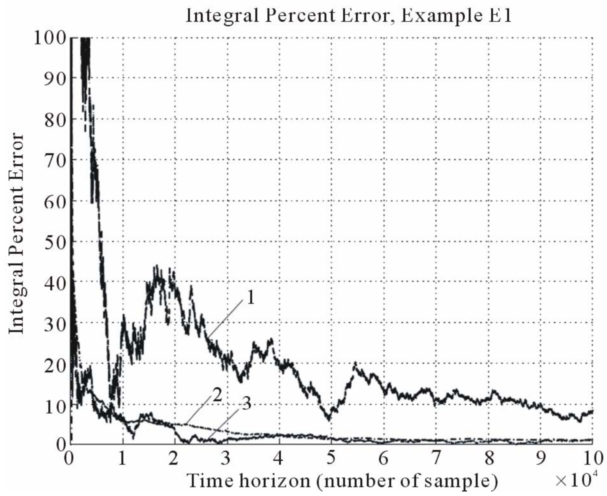

Figure 16. Integral percent error, ρ_{IPE} Equation (86), versus number of signal samples. *Example E*1: (1) for $SNR = 0.10$, (2) for $SNR = 1.00$, and (3) for $SNR = 10.0$.

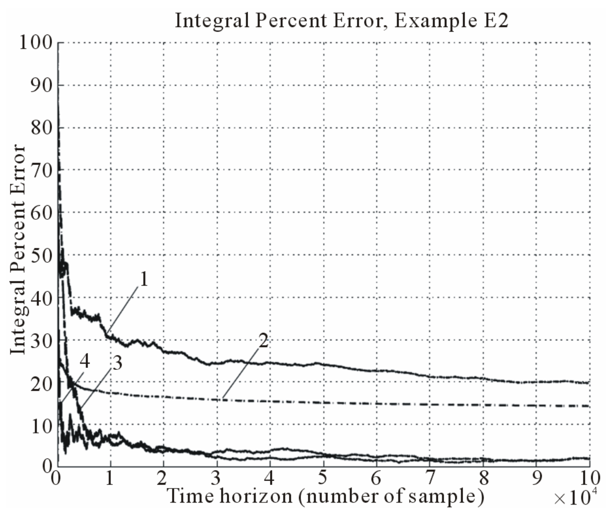

Figure 17. Integral percent error, ρ_{IPE} Equation (86), versus number of signal samples. *Example E*2: (1) for $SNR = 0.01$, (2) for $SNR = 0.10$, (3) for $SNR = 1.00$, and (4) for $SNR = 10.0$.

independently. The Principle, as well as its defining term "*active*" is conceptually different from that which is used in the collective monograph [35] where the authors associate this term with the problem of optimal input design for system identification and where they follow the solutions of Mehra [36].

Our approach is addressed to filters as the *state* estimators, whose original performance index is fundamentally inaccessible, in actual practice of a priori parameter uncertainty and unpredictable abrupt changeability. The problem lies in constructing an auxiliary performance index (API), which would have the following two properties:

- Accessibility for direct use in adaptation algorithms;
- Equimodality with the original performance index.

The present paper gives a comprehensive solution to the problem. We have solved the following tasks:

1) Clearly conveyed the adaptive model. Just as $\mathfrak{M}^*\left(\hat{\theta}\right)$, a replica of the standard observable data model, has been specified, $\mathfrak{M}\left(\hat{\theta}\right)$ has been patterned after the physical data model.
2) Introduced the notion of Generalized Residual as the *multi*-step (s-step) prediction error. In so doing, we exploited the system's complete observability as its key property, and used s, the system observability index.
3) Constructed the API that could offer ways of gaining indirect access to the data source *state* or to the Kalman filter *state*.
4) Examined API's capacity to "visualize" the *state* with respect to different levels of uncertainty.
5) Put forward feasible schemata for API computation.
6) Illustrated the theoretical identifiability by a real life example from inertial navigation.
7) Verified the theory by a numerical experimental testing of the approach.

Our further research is aimed at obtaining solutions to the following issues:

- Using the modern computational techniques in Kalman filtering for computer implementation of the approach.
- Seeking minimum of $J_t^a\left(\hat{\theta}\right)$ in parameters $\hat{\theta}$ of $\mathcal{D}\left(\hat{\theta}\right)$ or $\mathcal{D}^*\left(\hat{\theta}\right)$ instead of $\mathfrak{M}\left(\hat{\theta}\right)$ or $\mathfrak{M}^*\left(\hat{\theta}\right)$.
- Economic feasibility, numeric stability and convergence reliability of each proposed parameter adaptation algorithm.
- Numerical testing of the approach and determining the scope of its appropriate use in real life problems.

10. Acknowledgements

I would like to thank Prof. B. Verkhovsky of CS Department, New Jersey Institute of Technology for his most helpful input, Dr. A. Murgu of Electrical Engineering Department, University of Cape Town for his useful comments, and an anonymous reviewer for a number of corrections that improved the style of this paper.

11. References

[1] L. Ljung, "Perspectives on System Identification," *Annual Reviews in Control*, Vol. 34, No. 1, 2010, pp. 1-12. doi:10.1016/j.arcontrol.2009.12.001

[2] M. Gevers, "System Identification without Lennart Ljung:

What Would Have Been Different?" In: T. Glad and G. Hendeby, Eds., *Forever Ljung in System Identification*, Studentlitteratur AB, Norrtalje, 2006.

[3] B. D. O. Anderson, "Two Decades of Adaptive Control Pitfalls," *Proceedings of the* 8*th International Conference on Control*, *Automation*, *Robotics and Vision*, Kunming, December 2004.

[4] V. V. Terekhov and I. Y. Tyukin, "Adaptive Control Systems: Issues and Tendencies," *Control and Information Technology*, St. Petersburg University of Information Technology, St. Petersburg, 2003, pp. 146-154.

[5] L. Ljung, "Convergence Analysis of Parametric Identification Methods," *IEEE Transactions on Automatic Control*, Vol. AC-23, No. 5, 1978, pp. 770-783. doi:10.1109/TAC.1978.1101840

[6] L. Ljung, "System Identification: Theory for the User," Prentice-Hall, Inc., Englewood Cliffs, 1987.

[7] I. V. Semushin, "Adaptation in Stochastic Dynamic Systems—Survey and New Results I," *International Journal of Communications*, *Network and System Sciences*, Vol. 4, No. 1, 2011, pp. 17-23. doi:10.4236/ijcns.2011.41002

[8] R. K. Mehra, "Approaches to Adaptive Filtering," *IEEE Transactions on Automatic Control*, Vol. AC-17, No. 5, 1972, pp. 693-698. doi:10.1109/TAC.1972.1100100

[9] R. E. Kalman, P. L. Falb and M. A. Arbib, "Topics in Mathematical System Theory," McGraw Hill, Boston, 1969.

[10] L. Prouza, "Bemerkung Zur Linearen Prediktoren Mittels Eines Lernenden Filters," *Transactions of the* 1*st Prague Conference on the Information Theory and Statistical Decision Functions*, Prague, 28-30 November 1956, pp. 37-41.

[11] O. Šefl, "Filters and Predictors Which Adapt Their Values to Unknown Parameters of the Input Process," *Transactions of the* 2*nd Prague Conference on the Information Theory and Statistical Decision Functions*, Prague, 1961, pp. 597-608.

[12] A. A. Gorsky, "Automatic Optimal Filtering," *Transactions on Power Engineering and Automation*, USSR Academy of Sciences, Engineering Division, Moscow, No. 4, 1962, pp. 3-30.

[13] I. V. Semushin, "Use of Active Principle for Filtering Nonstationary Random Processes," *Proceedings of the* 3*rd Science and Technology Conference—Novgorod LETI Branch*, St. Petersburg, 1968, p. 64.

[14] I. V. Semushin, "Multi-Channel Adaptive Filter of Active Type," *Transactions on Instrument Making*, USSR University, Moscow, Vol. 12, No. 10, 1969, pp. 47-50.

[15] I. V. Semushin, "Closed Loop Adaptive Filters Investigation," Ph.D. Thesis, V. I. Ulyanov (Lenin) Saint Petersburg Electrotechnical University, St. Petersburg, 1970.

[16] P. E. Caines, "Linear Stochastic Systems," John Wiley and Sons, Inc., Hoboken, 1988.

[17] P. Lancaster and L. Rodman, "Algebraic Riccati Equations," Oxford University Press, Inc., Oxford, 1995.

[18] V. N. Fomin, "Reccurent Estimation and Adaptive Filtering," Nauka, Moscow, 1984.

[19] W. M. Wohnam, "Linear Multivariable Control—A Geomentic Approach," Springer-Verlag, Berlin, 1974.

[20] P. S. Maybeck, "Stochastic Models, Estimation, and Control," Vol. 1, Academic Press, Cambridge, 1979.

[21] B. D. O. Anderson and J. V. Moore, "Optimal Filtering," Prentice-Hall, Inc., Englewood Cliffs, 1979.

[22] J. S. Meditch, "Stochastic Optimal Linear Filtering and Control," McGraw Hill, Boston, 1969.

[23] S. Sherman, "Non-Mean-Square Error Criteria," *IRE Transactions on Information Theory*, Vol. IT-4, 1958, p. 125. doi:10.1109/TIT.1958.1057451

[24] I. V. Semushin, "Active Methods of Adaptation and Control for Discrete-Time Systems," D.Sc. Dissertation, Leningrad State University of Aerospace Instrumentation "LIAP", St. Petersburg, April 1987.

[25] I. V. Semushin, "Identification of Linear Stochastic Plants from the Incomplete Noisy Measurements of the State Vector," *Automation and Remote Control*, USSR Academy of Sciences, Moscow, Vol. 45, No. 8, 1985, pp. 61-71.

[26] R. L. T. Hampton, "Unsupervised Learning of the Kalman Filter," *Electronics Letters*, Vol. 9, No. 17, 1971, pp. 383-384.

[27] R. L. T. Hampton, "On Unknown State-Dependent Noise, Modeling Errors and Adaptive Filtering," *Electronics Letters*, Vol. 2, No. 2-3, 1975, pp. 195-201.

[28] D. Perriot-Mathonna, "The Use of Ljung's Results for Studuing the Convergence Properties of Hampton's Adaptive Filter," *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 6, 1980, pp. 1165-1169.

[29] I. V. Semushin, "Adaptive Identification and Fault Detection Methods in Random Signal Processing," Saratov University Publishers, Saratov, 1985.

[30] I. V. Semushin, "Active Adaptation of the Optimal Discrete Filters," *Transactions on Engineering Cybernetics*, USSR Academy of Sciences, Moscow, No. 5, 1975, pp. 192-198.

[31] P. Van Overschee and B. De Moor, "Subspace Identification for Linear Systems: Theory—Implementation—Applications," Kluwer Academic Publishers, Norwell, 1996.

[32] C. Broxmeyer, "Inertial Navigation Systems," McGraw-Hill Book Company, Boston, 1956.

[33] I. V. Semushin, "Identifying Parameters of Linear Stochastic Differential Equations from Incomplete Noisy Measurements," *Recent Developments in Theories* & *Numerics—International Conference on Inverse Problems*, Hong Kong, January 2002, pp. 281-290.

[34] O. Yu. Gorokhov and I. V. Semushin , "Developing a Simulation Toolbox in MATLAB and Using It for Non-linear Adaptive Filtering Investigation," *Lecture Notes in Computer Science*, Vol. 2658, 2003, pp. 436-445. doi:10.1007/3-540-44862-4_46

[35] V. I. Denisov, *et al.*, "Active Parameter Identification of Stochastic Linear Systems," NSTU Publisher, Novosibirsk, 2009.

[36] R. K. Mehra, "Optimal Input Signals for Parameter Estimation in Dynamic Systems—Survey and New Results," *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 6, 1974, pp. 753-768. doi:10.1109/TAC.1974.1100701

Appendix

Proof of Theorem 2. Processes $\varepsilon_{t+1|t}$ and $\varepsilon_{t|t}$ defined by (64) and (65) are equal to each other and also to $\mathscr{S}\left(\eta_{t+1|t}^{t+s|t}\right)$ in equations (59)-(60) and (61)-(62) correspondingly. Let $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ denote pro-tem the first, the second and the third summands in (59) so that

$$\begin{aligned} \boldsymbol{a} &\triangleq \left(x_{t+1} - \hat{g}_{t+1|t}\right) \\ \boldsymbol{b} &\triangleq \left[\mathscr{F}(\Psi) - \mathscr{F}(\mathbf{F})\right] u_{t+1}^{t+s} \\ \boldsymbol{c} &\triangleq \left[\mathscr{F}(\Gamma) w_{t+1}^{t+s} + \mathscr{S}\left(v_{t+1}^{t+s}\right)\right] \\ \boldsymbol{d} &\triangleq \boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c} = \boldsymbol{\varepsilon}_{t+1|t} \end{aligned} \tag{87}$$

Hence for Euclidean vector norms, it follows that

$$\|\boldsymbol{d}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 + \|\boldsymbol{c}\|^2 + 2\left(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{c} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{c}\right) \tag{88}$$

We examine the conditions under which all cross-terms in (88) (under sign $\boldsymbol{E}\{\cdot\}$ of expectation) could vanish. For such a consequence to ensue, we should provide orthogonality of $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ of (87) to each other. To do this, we restrict our consideration to three setups as follows.

Setup 1 (*Random Control Input*). This is an open--loop mode of operation in which $\{u(t)\}$ is a preassigned external zero-mean (and say, unit covariance) orthogonal wide- sence stationary process orthogonal to both $\{w(t)\}$ and $\{v(t)\}$ but in contrast to the last-named noises, such $\{u(t)\}$ is applied intentionally and meant to serve as an independent testing signal.

Setup 2 (*Pure Filtering*). This is the control-free (and hence open-loop) mode of operation, in which $\forall t : u(t) = 0$.

Setup 3 (*Close-loop Control*). For this mode, recall that we assume model based certainty equivalence optimal con- trol design [7]. Hence Control Strategy $S\left(\overline{\theta}\right)$ (cf. **Figure 4**) is linear in the experimental condition $\mathscr{X}$ (4).

Remark 13 *As in* [16], *denote* $\overline{\mathrm{Sp}}\{\cdots\}$ *to be the space* L_2 *of square integrable linear combinations of a process* $\{\cdots\}$ *with all limits in quadratic mean of all such combinations adjoined. Then* $\boldsymbol{H}_{-\infty,t}^{p} \triangleq \overline{\mathrm{Sp}}\{\ldots, p_{t-1}, p_t\}$ *is the subspace of a Hilbert space* $\boldsymbol{H}$ *spanned by a process* $\{p_t\}$ *from infinitely remote past up to* t.

Consider the above setups consecutively.

Setup 1 *Control Input*.

Taking three external inputs u, w, v as one composite process p in the above notation $\boldsymbol{H}_{-\infty,t}^{p}$ yields

$$\boldsymbol{H}_{-\infty,t}^{u,w,v} \triangleq \overline{\mathrm{Sp}}\left\{\ldots, \begin{bmatrix} u_{t-1} \\ w_{t-1} \\ v_{t-1} \end{bmatrix}, \begin{bmatrix} u_t \\ w_t \\ v_t \end{bmatrix}\right\}$$

For (87), we have $a \in \boldsymbol{H}_{-\infty,t}^{u,w,v}$, $b \in \boldsymbol{H}_{t+1,t+s}^{u}$, and $c \in \boldsymbol{H}_{t+1,t+s}^{w,v}$. By virtue of the fact that

$$\left\{\cdots, \begin{bmatrix} u_{t-1} \\ w_{t-1} \\ v_{t-1} \end{bmatrix}, \begin{bmatrix} u_t \\ w_t \\ v_t \end{bmatrix}\right\} \text{ and } \left\{\begin{bmatrix} u_{t+1} \\ w_{t+1} \\ v_{t+1} \end{bmatrix}, \cdots, \begin{bmatrix} u_{t+s} \\ w_{t+s} \\ v_{t+s} \end{bmatrix}\right\}$$

are two separate portions of the wide-sense stationary orthogonal process $p = (u, w, v)$, the following asserons (three orthogonalities) are true:

$$\begin{aligned} \boldsymbol{H}_{-\infty,t}^{u,w,v} &\perp \boldsymbol{H}_{t+1,t+s}^{u} \Rightarrow \boldsymbol{E}\left\{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}\right\} = 0 \\ \boldsymbol{H}_{-\infty,t}^{u,w,v} &\perp \boldsymbol{H}_{t+1,t+s}^{w,v} \Rightarrow \boldsymbol{E}\left\{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}\right\} = 0 \\ \boldsymbol{H}_{t+1,t+s}^{u} &\perp \boldsymbol{H}_{t+1,t+s}^{w,v} \Rightarrow \boldsymbol{E}\left\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{c}\right\} = 0 \end{aligned} \tag{89}$$

Remark 14 *The assertions are true for discrete-time systems only because only they have a finite sampling interval. This fact is crucial for our development.*

Hence

$$\begin{aligned} \boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} &= \boldsymbol{E}\left\{\|\boldsymbol{a}+\boldsymbol{b}\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{c}\|^2\right\} \\ \boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} &= \boldsymbol{E}\left\{\|\boldsymbol{a}\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{c}\|^2\right\} \\ \boldsymbol{E}\left\{\|\boldsymbol{a}+\boldsymbol{b}\|^2\right\} &= \boldsymbol{E}\left\{\|\boldsymbol{a}\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\} \end{aligned}$$

Define

$$\mathcal{J}_t^{\mathrm{a}}\left(\hat{\theta}\right) \triangleq (1/2)\,\boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} \tag{90}$$

to be the API, as it was made in (42), and denote

$$\mathcal{J}_t^{\mathrm{o}}\left(\hat{\theta}\right) \triangleq (1/2)\,\boldsymbol{E}\left\{\|\boldsymbol{a}\|^2\right\} \tag{91}$$

to be the Original Performance Index, OPI in line with (22). Use $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ and $\mathcal{P}_4$ to stand for the following statements:

$\mathcal{P}_1 \triangleq$ "$\mathcal{J}_t^{\mathrm{a}}\left(\hat{\theta}\right)$ attains its minimum in $\hat{\theta}$."

$\mathcal{P}_2 \triangleq$ "$\boldsymbol{E}\left\{\|\boldsymbol{a}+\boldsymbol{b}\|^2\right\}$ attains its minimum in $\hat{\theta}$."

$\mathcal{P}_3 \triangleq$ "$\boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\}$ attains its minimum in $\hat{\theta}$, that is $\boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\} = 0$."

$\mathcal{P}_4 \triangleq$ "$\mathcal{J}_t^{\mathrm{o}}\left(\hat{\theta}\right)$ attains its minimum in $\hat{\theta}$ at $\hat{\theta} = \theta^{\dagger}$, *i.e.* $\min_{\hat{\theta}} \mathcal{J}_t^{\mathrm{o}}\left(\hat{\theta}\right) = J_t^{\mathrm{o}}\left(\theta^{\dagger}\right) = (1/2)\,tr\{\Sigma\}$."

Our argument is as follows:

$$\mathcal{P}_1 \equiv \mathcal{P}_2, \mathcal{P}_2 \equiv \mathcal{P}_3 \wedge \mathcal{P}_4, \mathcal{P}_4 \to \mathcal{P}_3, \ \therefore \mathcal{P}_1 \equiv \mathcal{P}_4$$

The argument is valid because one can verify that

$$\left(\left[\mathcal{P}_1 \equiv \mathcal{P}_2\right] \wedge \left[\mathcal{P}_2 \equiv \mathcal{P}_3 \wedge \mathcal{P}_4\right] \wedge \left[\mathcal{P}_4 \to \mathcal{P}_3\right]\right) \to \left(\mathcal{P}_1 \equiv \mathcal{P}_4\right)$$

is a tautology. Three premises (they are in brackets) are always *TRUE*. The first premise, $\mathcal{P}_1 \equiv \mathcal{P}_2$ is *TRUE* because $\boldsymbol{E}\left\{\|c\|^2\right\}$ is *constant* in $\hat{\theta}$. The second premise, $\mathcal{P}_2 \equiv \mathcal{P}_3 \wedge \mathcal{P}_4$ follows from the properties of norms, and the third premise, $\mathcal{P}_4 \to \mathcal{P}_3$ is *TRUE* due to uniqueness of Kalman filter parameters when optimized in terms of criterion (22).

Since the conclusion $\mathcal{P}_1 \equiv \mathcal{P}_4$ is *TRUE*, this completes the proof for **Setup 1**: criteria (90) and (91) are equimodal (have the same minimizing arguments).

Setup 2 *Pure Filtering*.

This mode eliminates $\boldsymbol{b}$ from (87). For terms of (87), we have $\boldsymbol{a} \in \boldsymbol{H}^{w,v}_{-\infty,t}$, $\boldsymbol{b} = \boldsymbol{0}$, and $\boldsymbol{c} \in \boldsymbol{H}^{w,v}_{t+1,t+s}$. By virtue of the fact that $\boldsymbol{H}^{w,v}_{-\infty,t} \perp \boldsymbol{H}^{w,v}_{t+1,t+s}$, we obtain $E\left\{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}\right\} = 0$ and $\boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} = \boldsymbol{E}\left\{\|\boldsymbol{a}\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{c}\|^2\right\}$. It means, together with definitions (40), (90) and (91) that

$$\begin{aligned} \mathcal{J}^{\mathrm{a}}_t\left(\hat{\theta}\right) &= \mathcal{J}^{\mathrm{o}}_t\left(\hat{\theta}\right) + \text{const} \\ \text{const} &= (1/2)\boldsymbol{E}\left\{\|\mathbf{c}\|^2\right\} \end{aligned} \tag{92}$$

and again, we are done.

Setup 3 *Close-loop Control*.

Consider Level 2 of uncertainty as acting in this setup. This is tantamount to stating that identification of $\boldsymbol{\Psi}$ is not needed. As in Setup 2, we observe that $\boldsymbol{a} \in \boldsymbol{H}^{w,v}_{-\infty,t}$, $\boldsymbol{b} = \boldsymbol{0}$, and $\boldsymbol{c} \in \boldsymbol{H}^{w,v}_{t+1,t+s}$. The orthogonality $\boldsymbol{H}^{w,v}_{-\infty,t} \perp \boldsymbol{H}^{w,v}_{t+1,t+s}$ (and as a result, $\boldsymbol{E}\left\{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}\right\} = 0$) is true since

$$\left\{\cdots, \begin{bmatrix} w_{t-1} \\ v_{t-1} \end{bmatrix}, \begin{bmatrix} w_t \\ v_t \end{bmatrix}\right\} \text{ and } \left\{\begin{bmatrix} w_{t+1} \\ v_{t+1} \end{bmatrix}, \cdots, \begin{bmatrix} w_{t+s} \\ v_{t+s} \end{bmatrix}\right\}$$

are two separate portions of the zero-mean orthogonal wide-sence stationaly process (w, v).

In order to get a deeper insight into the basic relation (92) and the conclusion "$\mathcal{P}_1 \equiv \mathcal{P}_4$ is *TRUE*", let us look at them from some other point of view. Let $\boldsymbol{a}'$, $\boldsymbol{b}$, and $\boldsymbol{c}'$ denote pro-tem the first, the second and the third summands in (60) so that

$$\begin{aligned} \boldsymbol{a}' &\triangleq \left(x_{t+1|t} - \hat{g}_{t+1|t}\right) \\ \boldsymbol{b} &\triangleq [\mathscr{T}\left(\boldsymbol{\Psi}\right) - \mathscr{T}\left(\boldsymbol{F}\right)]u^{t+s}_{t+1} \\ \boldsymbol{c}' &\triangleq \mathscr{P}\left(\boldsymbol{G}\right)v^{t+s|t+s-1}_{t+1|t} \\ \boldsymbol{d} &\triangleq \boldsymbol{a}' + \boldsymbol{b} + \boldsymbol{c}' = \boldsymbol{\varepsilon}_{t+1|t} \end{aligned} \tag{93}$$

We are interested in $\boldsymbol{b}$ to vanish and in eliminating all cross-terms of (88) and its innovation analogue

$$\|\boldsymbol{d}\|^2 = \|\boldsymbol{a}'\|^2 + \|\boldsymbol{b}\|^2 + \|\boldsymbol{c}'\|^2 + 2\left(\boldsymbol{a}'^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{a}'^{\mathrm{T}}\boldsymbol{c}' + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{c}'\right) \tag{94}$$

Innovation form (93) reinforces the fact that innovation version of the system (only if not in Setup 1) has the only "external" (better to say a *hidden* external) input. This is the innovation process $\left\{\nu_{t|t-1}\right\}$, which is linear, orthogonal and wide-sense stationary with the $\overline{\mathrm{Sp}}\left\{\cdots, \nu_{t-1|t-2}, \nu_{t|t-1}\right\}$ forming $\boldsymbol{H}^{\nu}_{-\infty,t}$. The above outlined additional input $\{u_t\}$ appears in Setup 1.

Starting from (93), we revisit the same setups:

- Setup 1 *Random Control Input*. We observe that $\boldsymbol{a}' \in \boldsymbol{H}^{u,\nu}_{-\infty,t}$, $\boldsymbol{b} \in \boldsymbol{H}^{u}_{t+1,t+s}$, $\boldsymbol{c}' \in \boldsymbol{H}^{\nu}_{t+1,t+s}$ where $\boldsymbol{H}^{u,\nu}_{-\infty,t} \triangleq \overline{\mathrm{Sp}}\left\{\cdots, \begin{bmatrix} u_{t-1} \\ \nu_{t-1|t-2} \end{bmatrix}, \begin{bmatrix} u_t \\ \nu_{t|t-1} \end{bmatrix}\right\}$. By this, the following assertions are true:

$$\begin{aligned} \boldsymbol{H}^{u,\nu}_{-\infty,t} &\perp \boldsymbol{H}^{u}_{t+1,t+s} \Rightarrow \boldsymbol{E}\left\{\boldsymbol{a}'^{T}\boldsymbol{b}\right\} = 0 \\ \boldsymbol{H}^{u,\nu}_{-\infty,t} &\perp \boldsymbol{H}^{\nu}_{t+1,t+s} \Rightarrow \boldsymbol{E}\left\{\boldsymbol{a}'^{\mathrm{T}}\boldsymbol{c}'\right\} = 0 \\ \boldsymbol{H}^{u}_{t+1,t+s} &\perp \boldsymbol{H}^{\nu}_{t+1,t+s} \Rightarrow \boldsymbol{E}\left\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{c}'\right\} = 0 \end{aligned} \tag{95}$$

Hence

$$\begin{aligned} \boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} &= \boldsymbol{E}\left\{\|\boldsymbol{a}' + \boldsymbol{b}\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{c}'\|^2\right\} \\ \boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} &= \boldsymbol{E}\left\{\|\boldsymbol{a}'\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{c}'\|^2\right\} \\ \boldsymbol{E}\left\{\|\boldsymbol{a}' + \boldsymbol{b}\|^2\right\} &= \boldsymbol{E}\left\{\|\boldsymbol{a}'\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\} \end{aligned}$$

Denote

$$\mathcal{J}^{\mathrm{o}'}_t\left(\hat{\theta}\right) \triangleq (1/2)E\left\{\|\boldsymbol{a}'\|^2\right\} \tag{96}$$

to be another Original Performance Index, OPI$'$ and add $\mathcal{P}'_2$ and $\mathcal{P}'_4$ to stand for the following statements:

$$\mathcal{P}'_2 \triangleq \text{"}\boldsymbol{E}\left\{\|\boldsymbol{a}' + \boldsymbol{b}\|^2\right\} \text{ attains its minimum in } \hat{\theta}.\text{"}$$

$$\mathcal{P}'_4 \triangleq \text{"}\mathcal{J}^{\mathrm{o}'}_t\left(\hat{\theta}\right) \text{ attains its minimum in } \hat{\theta} \text{ at } \hat{\theta} = \theta^{\dagger},$$

$$\textit{i.e.}\ \min_{\hat{\theta}} \mathcal{J}^{\mathrm{o}'}_t\left(\hat{\theta}\right) = \mathcal{J}^{\mathrm{o}'}_t\left(\theta^{\dagger}\right) = 0.\text{"}$$

Consider them together with the above statements

$$\mathcal{P}'_1 \triangleq \text{"}\mathcal{J}^{a}_t\left(\hat{\theta}\right) \text{attains its minimum in } \hat{\theta}.\text{"}$$

$$\mathcal{P}'_3 \triangleq \text{"}\boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\} \text{ attains its minimum in } \hat{\theta}, \text{ that is } \boldsymbol{E}\left\{\|\boldsymbol{b}\|^2\right\} = 0.\text{"}$$

The following argument

$$\mathcal{P}_1 \equiv \mathcal{P}'_2, \mathcal{P}'_2 \equiv \mathcal{P}_3 \wedge \mathcal{P}'_4, \mathcal{P}'_4 \to \mathcal{P}_3, \therefore \mathcal{P}_1 \equiv \mathcal{P}'_4$$

is valid because

$$\left(\left[\mathcal{P}_1 \equiv \mathcal{P}'_2\right] \wedge \left[\mathcal{P}'_2 \equiv \mathcal{P}_3 \wedge \mathcal{P}'_4\right] \wedge \left[\mathcal{P}'_4 \to \mathcal{P}_3\right]\right) \to \left(\mathcal{P}_1 \equiv \mathcal{P}'_4\right)$$

is a tautology. All premises (placed in brackets) are always *TRUE*. The premise $\mathcal{P}_1 \equiv \mathcal{P}'_2$ is *TRUE* because

$\boldsymbol{E}\left\{\|\boldsymbol{c}'\|^2\right\}$ is *constant* in $\hat{\theta}$. The premise $\mathcal{P}_2' \equiv \mathcal{P}_3 \wedge \mathcal{P}_4'$ follows from the properties of norms. The third premise, $\mathcal{P}_4' \rightarrow \mathcal{P}_3$ is *TRUE* due to uniqueness of Kalman filter parameters when optimized in terms of criterion (22).

By the conclusion $\mathcal{P}_1 \equiv \mathcal{P}_4'$ being *TRUE*, the proof for **Setup 1** is completed: criteria (90) and (96) have the same minimizing argument: $\hat{\theta} = \theta^{\dagger}$.

- Setup 2 *Pure Filtering*. In this case, $\boldsymbol{a}' \in \boldsymbol{H}^{\nu}_{-\infty,t}$, $\boldsymbol{b} = \boldsymbol{0}$, $\boldsymbol{c}' \in \boldsymbol{H}^{\nu}_{t+1,t+s}$. By virtue of the fact that $\boldsymbol{H}^{\nu}_{-\infty,t} \perp \boldsymbol{H}^{\nu}_{t+1,t+s}$, we obtain $\boldsymbol{E}\left\{\boldsymbol{a}'^{\mathrm{T}}\boldsymbol{c}'\right\} = 0$ and $\boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} = \boldsymbol{E}\left\{\|\boldsymbol{d}\|^2\right\} = \boldsymbol{E}\left\{\|\boldsymbol{a}'\|^2\right\} + \boldsymbol{E}\left\{\|\boldsymbol{c}'\|^2\right\}$. It means, together with definitions (40), (90) and (96) that

$$
\begin{aligned}
\mathcal{J}_t^{\mathrm{a}}\left(\hat{\theta}\right) &= \mathcal{J}_t^{o'}\left(\hat{\theta}\right) + \mathrm{const}' \\
\mathrm{const}' &= \left(1/2\right)\boldsymbol{E}\left\{\|\boldsymbol{c}'\|^2\right\}
\end{aligned}
\tag{97}
$$

Again, the conclusion $\mathcal{P}_1 \equiv \mathcal{P}_4'$ is *TRUE* for Setup 2.

- Setup 3 *Closed-loop Control*. In this case we are in the same situation: $\boldsymbol{a}' \in \boldsymbol{H}^{\nu}_{-\infty,t}$, $\boldsymbol{b} = \boldsymbol{0}$, $\boldsymbol{c}' \in \boldsymbol{H}^{\nu}_{t+1,t+s}$. The conclusion $\mathcal{P}_1 \equiv \mathcal{P}_4'$ is *TRUE* for Setup 3, as well.

Thus Theorem 2 is true. □