

Adaptation in Stochastic Dynamic Systems —Survey and New Results I

Innokentiy V. Semushin

School of Mathematics and Information Technology, Ulyanovsk State University, Ulyanovsk, Russia

E-mail: kentvsem@gmail.com

Received December 16, 2010; revised December 30, 2010; accepted January 4, 2011

Abstract

This paper surveys the field of adaptation mechanism design for uncertainty parameter estimation as it has developed over the last four decades. The adaptation mechanism under consideration generally serves two tightly coupled functions: model identification and change point detection. After a brief introduction, the paper discusses the generalized principles of adaptation based both on the engineering and statistical literature. The stochastic multiinput multioutput (MIMO) system under consideration is mathematically described and the problem statement is given, followed by a definition of the active adaptation principle. The distinctive property of the principle as compared with the Minimum Prediction Error approach is outlined, and a plan for a more detailed exposition to be provided in forthcoming papers is given.

Keywords: Linear Stochastic Systems, Parameter Estimation, Model Identification, Change Point Detection, Learning and Adaptive Control

1. Introduction

In mathematical theory of data processing and control systems, the problem of overcoming a priori uncertainty has a long history. As from the first principles of automatic optimizing control systems by Draper and Li [1], the first approaches to designing such systems by Kalman [2], and from the fundamental works on dual control by Feldbaum [3-6], it has been the focus of attention for many professionals, giving rise a host of publications in journals, conference proceedings and books [7], and thus determining the front line research in developing a theory and design of systems with the highest capabilities. The solution to this problem is naturally and closely associated with the concept of adaptivity.

As can be seen from the available literature, *in an adaptive system the current information is perceived, analyzed, and used for the intentional change of the system*. Thus, the adaptivity is realized here as the triune:

1) Classification of data perceived by modes of operation: the nominal mode or one of faulty modes; the corresponding block can be called *Classifier*.

2) Identification of the process model and/or measurement model launched when the faulty mode point is detected; the corresponding block is naturally to call

Identifier.

3) System modification after the identification stopped with the new estimates for model parameters when the nominal mode point is detected; the corresponding block can be called *Modifier*.

These functions are not always broken into the three separate processes; in fact however, they are considered inherent to any adaptive system. Sometimes, identification is not separate from modification, and then no difference is made between these two processes. Such one-in-two process of changing the system is called, after Tsytkin [8], *adaptation* and the corresponding block *adaptor* (**Figure 1**), whereas making decision concerning modes of operation is sometimes called (for short) *control over* the system mode of operation. This term can imply emphasis on solving one of two (or both) subtasks: *change detection* and *change point detection* where “change” means switching from one mode of system operation to another. To equip the system with adaptivity, the crucial tasks to be solved effectively are the quickest change point detection and the unbiased estimation of the data source parameters, whose new unknown values may result from the change.

The quickest change point detection is of great importance for many applications and has extensive references

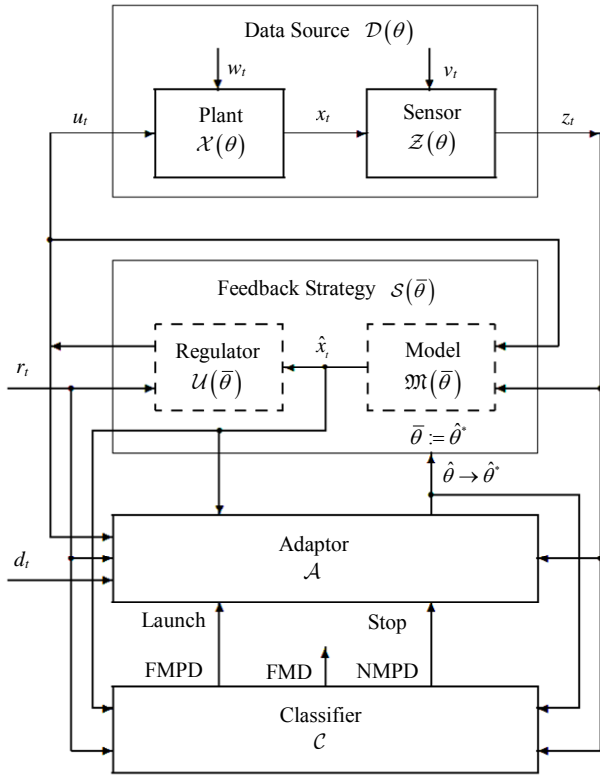


Figure 1. The generalized block-diagram of an adaptive system. FMPD—Faulty Mode Point Detected; NMPD—Nominal Mode Point Detected; FMD—Faulty Mode Detected. *Legend:* r_t —external “reference” signal; d_t —external “desired” signal; θ —uncertainty parameter; $\bar{\theta}$ —suboptimal estimated parameter on which Feedback is based; $\hat{\theta}$ —current estimated parameter; $\hat{\theta}^*$ —final value of $\hat{\theta}$ resulting from identification.

(e.g. [9] and many others). Parameter identification methods have also become the topic of a large body of research. However, little attention is still being given to the study of bias in parameter estimates. The bias in $\hat{\theta}^*$, the final result of identification, with respect to the true value θ^* of the uncertainty parameter θ is the most difficult to cope with when based on the incomplete noisy observations over a closed-loop stochastic control [10,11]. As this takes place, the existence of bias is not so annoying as its dependence on (unknown) experimental conditions [12].

2. Generalized Principles of Adaptation

The comprehensive and systematic analysis of existing system adaptation methods shows that design methodologies of adaptive filtering (state estimation) and stochastic control can be grouped into five large cate-

gories as follows.

2.1. Bayesian Adaptive Model (BAM)

The whole vector space of unknown model parameters is, firstly, partitioned and secondly, approximated by a set of design vectors. Based on each of those vectors, the optimal system is built, and the conditional (a’posteriori) probability density for each design vector is calculated to be used as a weighting factor for the system output. The approximation for the optimal estimate is computed as the weighted mean of outputs over the set of designed systems (**Figure 2**).

2.2 Extended Adaptive Model (EAM)

The state variable vector is augmented with the constant subvector of unknown parameters. Formed in this fashion, the extended model state equation is used as the basis for the optimal system design (**Figure 3**).

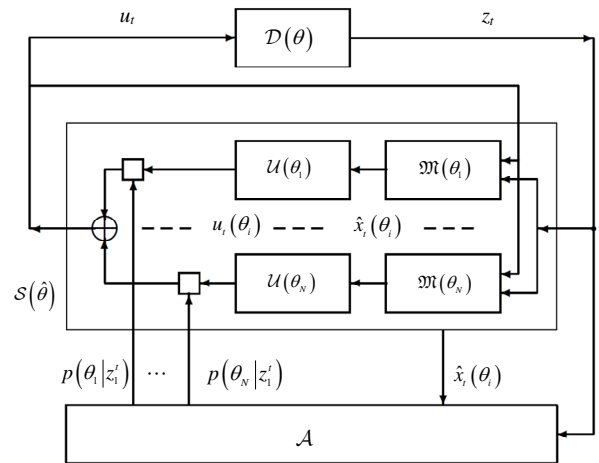


Figure 2. Bayesian (partitioned) Adaptive Model (BAM) principle: $p(\theta_i | z_t^i)$ —a’posteriori probability densities, $i = 1, 2, \dots, N$. (The small square-wise blocks denote multiplication.)

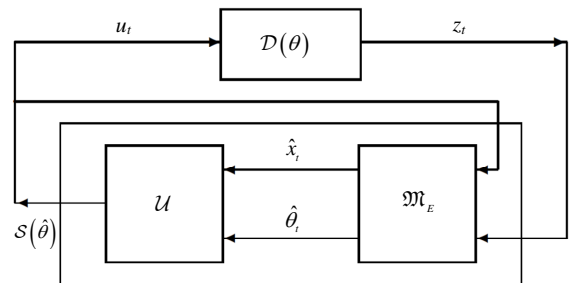


Figure 3. Extended (state based) Adaptive Model (EAM) principle: $(\hat{x}_t, \hat{\theta}_t)$ —extended model \mathfrak{M}_E state vector.

2.3. Analytical Relations Based Adaptive Model (AAM)

The unknown parameters of the optimal system are expressed analytically or computationally through statistics of one or more chosen processes within the suboptimal system structure. These statistics are calculated through time averaging over a single sample representing the ensemble of samples for the chosen process. The analytical relations are then solved for the unknown parameters of the optimal system (Figure 4).

2.4. Performance Index Based Adaptive Model (PIAM)

In PIAM, the specifications are given in terms of the residual between an explicit reference signal and a model (predicted) signal. The model parameters are adjusted so that the model output would be close to that of real system taken as the reference. The mutual proximity of these outputs is taken as a performance index (a functional) to be minimized (Figure 5). Minimum seeking methods are designed to provide convergence of the adaptive model parameters to the values minimizing the performance index. It means that PIAM principle provides a closed loop identification scheme, i. e. the scheme with the feedback in performance index. The approach unites many methods: Minimum Prediction Error (MPE) method, Least Squares method, Maximum Likelihood method and many others [7].

2.5. Characteristic Matching Adaptive Model (CMAM)

An indirect and simplified performance index is intro-

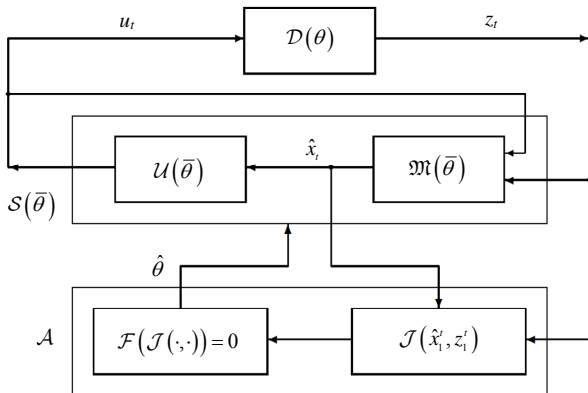


Figure 4. Analytical (optimality equation based) Adaptive Model (AAM) principle: $\mathcal{J}(\cdot, \cdot)$ —a statistic obtained from the sample (\hat{x}_t, z_t) ; $\mathcal{F}(\mathcal{J}(\cdot, \cdot))$ —the model optimality equation.

duced instead of the original design performance index. Usually, it is defined as a deflection of a theoretical value of the inner system process characteristic from the corresponding actual value. The adaptation mechanism is designed to minimize this deflection (Figure 6). One possible technique for that is introducing a fictitious noise into adaptive model equations [13,14].

3. Dynamic System and Problem Statement

Consider a linear time-invariant state-space stochastic MIMO control system

$$x(t_{i+1}) = \Phi_{\theta}x(t_i) + \Psi_{\theta}u(t_i) + w(t_i), \quad x \in \mathbb{R}^n \quad (1)$$

$$y(t_i) = Hx(t_i) + v(t_i), \quad y \in \mathbb{R}^m \quad (2)$$

$$\tilde{x}(t_{i+1}^-) = \bar{\Phi}\tilde{x}(t_i^+) + \bar{\Psi}u(t_i), \quad \tilde{x} \in \mathbb{R}^n \quad (3)$$

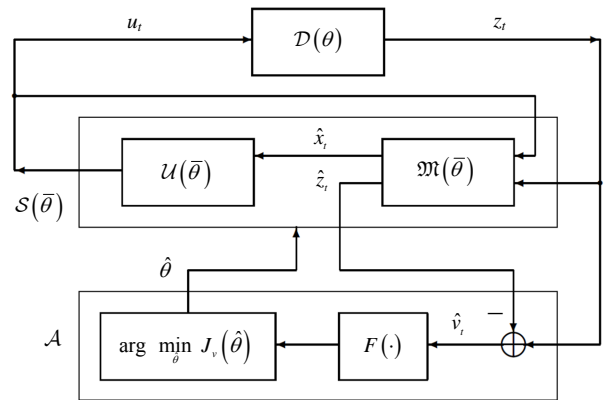


Figure 5. Performance Index based Adaptive Model (PIAM) principle: $F(\cdot)$ —an even loss function; $\mathcal{J}_v(\cdot)$ —model performance index.

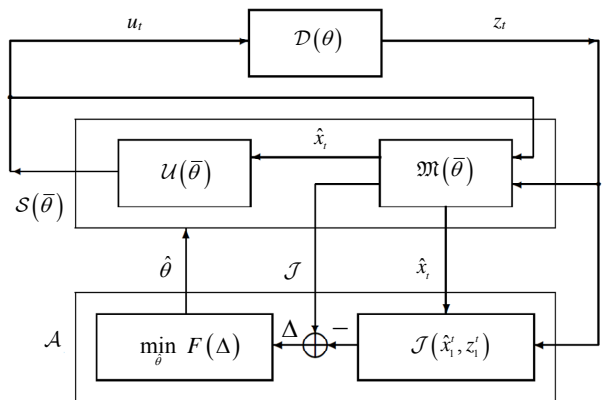


Figure 6. Characteristic Matching based Adaptive Model (CMAM) principle: \mathcal{J} —theoretical (predicted) value of a characteristic; $\mathcal{J}(\cdot, \cdot)$ —experimental (virtual) value of the characteristic; $F(\cdot)$ —an even function of residual Δ .

$$\tilde{x}(t_i^+) = \tilde{x}(t_i^-) + \bar{K} \left[y(t_i) - H\tilde{x}(t_i^-) \right] \quad (4)$$

$$u(t_i) = f_{\text{FR}} \left[\tilde{x}(t_i^+) \right] = -\bar{F}\tilde{x}(t_i^+), \quad u \in \mathbb{R}^q \quad (5)$$

where $i \in \mathbb{Z}$, composed of a plant P, Equation (1), a sensor S, Equation (2), and a feedback controller FC, Equations (3)-(5), realizing Feedback Strategy $S(\bar{\theta})$ as shown in **Figure 1**.

According to the separation theorem, FC is composed of a feedback filter FF, Equations (3)-(4), cascaded with a feedback regulator FR, Equation (5). Based on the Kalman theory, FF is the one-step predictor, Equation (3), coupled with the estimator, Equation (4). Note that in signal processing expositions, the process $x(t_i)$ in Equation (1) is called the *useful signal*, the process $Hx(t_i)$ in Equation (2) the *observation signal* and $v(t_i)$ the *observation noise*. In our case, the available data are complemented by control input $u(t_i)$, and so the *observation data* are $z(t_i) = \begin{bmatrix} y(t_i) \\ u(t_{i-1}) \end{bmatrix}$ at every instant $t_i \in \mathbb{R}$.

The main assumptions are quite standard as follows. The random initial state $x(t_0)$ is given at some $t_0 \in \mathbb{R}$ with $E \left\{ \|x(t_0)\|^2 \right\} < \infty$ where $E \left\{ \|\cdot\| \right\}$ denotes the expectation of Euclidean norm $\|\cdot\|$; $w(t_i)$ and $v(t_i)$ are zero-mean mutually orthogonal wide-sense stationary orthogonal processes such that

$$E \left\{ w(t_i) w(t_i)^T \right\} = Q_\theta \geq 0 \quad \text{and} \quad E \left\{ v(t_i) v(t_i)^T \right\} = R_\theta > 0$$

for all $t_i \in \mathbb{R}$ with $\begin{bmatrix} w(t_i) \\ v(t_i) \end{bmatrix}$ orthogonal to $x(t_i)$ and $u(t_j)$ for all $j \geq i$; $u(t_i)$ given as a function $f_{\text{FR}}[\cdot]$ of $\tilde{x}(t_i^+)$ is wide-sense stationary and $E \left\{ \|u(t_i)\|^2 \right\} < \infty$ for all $t_i \in \mathbb{R}$.

As we summarize the specific (but not restrictive) theoretical assumptions, we have the following:

A1 An infinitely long sample of observations $z(t_i)$ is a realization of a stochastic process on a probability space (Ω, \mathcal{B}, P) where the measure P on the space Ω is such that a form of ergodicity holds. For the purposes of this paper, it is sufficient to require the ergodicity in the mean for the inner processes of the system that are supposed to be wide-sense stationary. According to [15], the ergodicity theorem for wide-sense stationary processes states that the sample mean computed on increasing finite segments

$z_1^N = \left[z^T(t_1) \cdots z^T(t_N) \right]^T$ of the data will converge in *quadratic means (q.m.)* to the exact mean given in terms of P .

A2 Plant Equation (1) and sensor Equation (2) are given by the *standard observable model*, SOM.

A3 Four matrices in the system description, namely the state transition matrix Φ , actuator matrix Ψ , and noise covariances Q and R in the Equations (1)-(2) depend on an uncertainty l -component vector $\theta \in \Theta$, as they are marked by the subscript θ . Each particular value of θ specifies a *mode*.

Remark 1 The resulting parameterization of x , y , \tilde{x} and u by θ has been suppressed for simplicity of notation. For the same reason, henceforth we shall write Φ , Ψ , Q and R without subscript θ .

A4 The charts of the atlas covering the parameter manifold Θ constitute either a finite, countable or a continuous set. Thus the system itself can operate in several modes (*finitely* or *infinitely* many).

A5 Θ constitutes an *identifiable* SOM parameterization [7], i.e., for all z and all $\theta, \theta' \in \Theta$, we have

$$P(z_1^N; \theta) = P(z_1^N; \theta') \Leftrightarrow \theta = \theta'$$

A6 The system is designed to operate with a minimum expected control cost

$$J_c = \lim_{t_0 \rightarrow -\infty} E \left\{ x^T(t_{N+1}) X_f x(t_{N+1}) + \sum_{i=0}^N \left[x^T(t_i) X x(t_i) + u^T(t_i) U u(t_i) \right] \right\} \quad (6)$$

for each $N > 0$ and some $X_f \geq 0$, $U > 0$, and $X \geq 0$.

A7 $(\Phi, Q^{1/2})$ is stabilizable, $(\Phi, X^{1/2})$ is detectable, (Φ, H) is completely observable, and (Φ, Ψ) is completely controllable.

A8 For any $\theta \in \Theta$ the designed system holds its property to be δ -mean exponentially stable [7].

A9 A mode switching mechanism is viewed as deterministic, and yet it is unknown to the observer (like controlled by an independent actor) but allows for enough time between successive switches to leave no doubt that **A1** practically holds.

Assumption **A2** is made for simplicity and may be omitted if one includes a similarity transformation into the inference (thus defining new basis vectors for SOM in the state space). Assumptions **A3-A5** bring the system to the realm of *hybrid* or *multi-mode* systems. One value of $\theta \in \Theta$, denoted by θ_0 , designates the so called *adopted operating mode*, which is recognized—not obligatorily albeit possibly—as a *nominal operating mode*, NOM. The other values $\theta \in \Theta \setminus \theta_0$ can be viewed as specifying some *alternative operating modes* recognized—not obligatorily albeit possibly—as *faulty operating modes*, FOMs, $\Theta_{\text{FOM}} = \Theta \setminus \theta_0$. Assumption **A7** guarantees the existence of the optimal steady-state parameters $\bar{\Phi}$, $\bar{\Psi}$, \bar{K} , \bar{F} in Equations (3)-(5) for any mode. Algorithms to compute these parameters are as follows where subscript “ $_0$ ” corresponds to θ_0 .

$$\begin{array}{c}
\boxed{\bar{\Phi} := \Phi_0 \quad \bar{Q} := Q_0 \quad \bar{R} := R_0} \quad \boxed{\bar{K} := K} \\
\downarrow \quad \quad \quad \uparrow \\
\boxed{M = \bar{\Phi}[M - MH^T(HMH^T + \bar{R})^{-1}HM]\bar{\Phi}^T + \bar{Q}} \\
\boxed{K = MH^T(HMH^T + \bar{R})^{-1}}
\end{array}$$

Algorithm 1

$$\begin{array}{c}
\boxed{\bar{\Phi} := \Phi_0 \quad \bar{\Psi} := \Psi_0} \quad \boxed{\bar{F} := F} \\
\downarrow \quad \quad \quad \uparrow \\
\boxed{\Pi = \bar{\Phi}^T[\Pi - \Pi\bar{\Psi}(\bar{\Psi}^T\Pi\bar{\Psi} + U)^{-1}\bar{\Psi}^T\Pi]\bar{\Phi} + X} \\
\boxed{F = (\bar{\Psi}^T\Pi\bar{\Psi} + U)^{-1}\bar{\Psi}^T\Pi\bar{\Phi}}
\end{array}$$

Algorithm 2

These algorithms hold when the FR function, $f_{FR}[\cdot]$, is chosen to satisfy the second equality in Equation (5). In this case, the assumption **A8** holds if

$$\begin{bmatrix} \Phi - \Psi\bar{F}\bar{K}H & -\Psi\bar{F}(I - \bar{K}H) \\ (\bar{\Phi} - \bar{\Psi}\bar{F})\bar{K}H & (\bar{\Phi} - \bar{\Psi}\bar{F})(I - \bar{K}H) \end{bmatrix}$$

being the transition matrix for the stackable state vector $\left[x^T(t_i) \mid \tilde{x}^T(t_i^-) \right]^T$ of the given system Equations (1)-(5), is such that its eigenvalues lie in the open unit disc.

The problem is to identify the optimal Steady State Kalman Filter (SSKF) for any system mode that appears after a mode switch, in order to use the identification result $\hat{\theta}^*$ as a substitute for FF ($\bar{\theta} := \hat{\theta}^*$ as in **Figures 4, 5 and 6** is assumed).

4. Active Principle of Adaptation

Analyzing possibilities to apply any of the adaptation principles considered above in Section 2, we put in the forefront the behavioral aspect of the adaptive part of a system. In so doing, we notice that there are two more general principles of (or two approaches to) adaptation to be contrasted to each other:

- *Passive Principle*: Adaptation based on following some foremade prescriptions and so free of any tracking (and, as a sequel, without any guarantee of the desired quality of system and its small deflection from the point of optimum, and
- *Active Principle*: Adaptation based on tracking (and, as a sequel, with certain guarantee of) the desired quality of system and its small deflection from the point of optimum.

Viewed from this perspective, BAM, EAM and AAM are passive, whereas PIAM and CMAM are active. Naturally, there may exist combinations of these behavioral features in one adaptive system; however we focus on the

active type of adaptation as using PIAM of Section 2 for MIMO control systems described in Section 3.

Original performance indices: Let $\hat{x}_{EST}(t_i^-)$ and $\hat{x}_{EST}(t_i^+)$ be a one-step predicted and a filtered estimate of $x(t_i)$, correspondingly. Evidently, $e(t_i^-)$ and $e(t_i^+)$ being the errors committed by those estimators,

$$e(t_i^-) = x(t_i) - \hat{x}_{EST}(t_i^-) \quad (7)$$

$$e(t_i^+) = x(t_i) - \hat{x}_{EST}(t_i^+) \quad (8)$$

are not available in practice. They are useful in theory only when the corresponding original performance indices (OPI) given as the limiting mean square error cost in one of two equivalent forms

$$J_o^1 = \lim_{j \rightarrow -\infty} E \left\{ e^T(t_i^-) W e(t_i^-) \mid Y(t_j, t_{i-1}) \right\} \quad (9)$$

or

$$J_o^2 = \lim_{j \rightarrow -\infty} E \left\{ e^T(t_i^+) W e(t_i^+) \mid Y(t_j, t_i) \right\} \quad (10)$$

whose domain includes all time-invariant one-step predicting estimators $\hat{x}_{EST}(t_i^-)$ or filtering estimators $\hat{x}_{EST}(t_i^+)$ of $x(t_i)$, conditioned on the knowledge of all preceding measurements extracted from the above closed-loop system, are minimized, where

$$Y(t_j, t_i) = \left[y(t_j) \mid y(t_{j+1}) \mid \dots \mid y(t_i) \right]^T \quad (11)$$

is the standard notation for a stackable vector composed of column vectors $y(t_j)$ through $y(t_i)$, $i > j$, and W is a symmetric semi-positive definite ($W \geq 0$) weighting matrix.

Problem of active principle: Given the original performance index as (9) or (10). The minimum of (9) or, equivalently, (10) is achieved for all $W \geq 0$ if and only if $\hat{x}_{EST}(t_i^-)$ is identical to $\hat{x}(t_i^-)$ and $\hat{x}_{EST}(t_i^+)$ is identical to $\hat{x}(t_i^+)$ where $\hat{x}(t_i^-)$ and $\hat{x}(t_i^+)$ are estimates generated by the steady-state optimal (Kalman) filter

$$\begin{aligned}
\hat{x}(t_{i+1}^-) &= \Phi \hat{x}(t_i^+) + \Psi u(t_i) \\
&= \Phi \hat{x}(t_i^-) + \Psi u(t_i) + G_f v(t_i)
\end{aligned} \quad (12)$$

$$y(t_i) = H \hat{x}(t_i^-) + v(t_i), \quad \hat{x}(t_i^+) = \hat{x}(t_i^-) + K v(t_i) \quad (13)$$

In (12) and (13), $G_f = \Phi K$, $K = \Sigma H^T (H \Sigma H^T + R)^{-1}$ and Σ satisfies the algebraic Riccati equation

$$\Sigma = \Phi \left[\Sigma - \Sigma H^T (H \Sigma H^T + R)^{-1} H \Sigma \right] \Phi^T + Q. \quad (14)$$

However, (9) or (10) are unfit for use in minimum seeking methods (for example, in stochastic approxi-

mation or any others used in PIAM) by virtue of (7) and (8) being unavailable. The core of our method is to form an auxiliary performance index (API), using the available values $u(t_i)$ and $y(t_i)$ only, so that necessary and sufficient conditions for its minimum are equivalent to those for minimum of (9) or (10).

Note that MPE methods use a performance index to be minimized, too, in the process of system (better to say, model) adaptation. However, they predict not the exact state vector $x(t_i)$ as $\hat{x}(t_i)$ is inaccessible, but the measurement vector $y(t_i)$, and this is the difference between our approach and the MPE approach.

An intermediate discussion: In the Equations (7) and (8), the first operand, $x(t_i)$, can be called the *hidden* reference model output, while the second operand is the explicit adaptive model output to be fitted to $x(t_i)$. The name “hidden” comes from the fact that the “reference model output” $x(t_i)$, which is subject to Equation (1), is unavailable.

Alternative forms of OPI: If we make use of the well-known innovation representation (12)-(13) of Data Source (1)-(2), then we can consider the following resi- duals

$$r(t_i^-) = \hat{x}(t_i^-) - \hat{x}_{EST}(t_i^-) \quad (15)$$

$$r(t_i^+) = \hat{x}(t_i^+) - \hat{x}_{EST}(t_i^+) \quad (16)$$

instead of errors (7)-(8). In (15)-(16), the first operand again represents the *hidden* reference model, in this case described by (12)-(13). Correspondingly, one can take the original performance indices in the altertative forms

$$J_o^3 = \lim_{j \rightarrow -\infty} E \left\{ r^T(t_i^-) W r(t_i^-) \middle| Y(t_j, t_{i-1}) \right\} \quad (17)$$

or

$$J_o^4 = \lim_{j \rightarrow -\infty} E \left\{ r^T(t_i^+) W r(t_i^+) \middle| Y(t_j, t_i) \right\} \quad (18)$$

equivalent to (9)-(10) in the sense that minimization of any of these performance indices leads to the optimal parameter values of the adaptive model labeled as an estimator by subscript EST .

5. Conclusions

In this paper, the author formulates a new approach called The Active Principle of Adaptation. It is intended for one class of systems, whose original performance index can be used only as a theoretical one, namely for linear time-invariant state-space stochastic MIMO filter systems, possibly included into Feedback Strategy of stochastic control or considered independently, for example, in communication systems where filtering or detecting

signals is a priority.

The approach is focused on constructing an auxiliary performance index which would have two properties:

- accessibility for direct use in adaptation algorithms;
- equimodality with the original performance index.

The latter notion means that API and OPI have the same minimizing arguments. This is necessary and sufficient to guarantee (theoretically) nonbiasedness of parameter estimates. The direct way for these properties to hold is as follows:

- to form an available auxiliary process $\varepsilon(t_i)$ as a substitute for any process (7), (8), (15), (16), and then
- to form the API as

$$J_a = \lim_{j \rightarrow -\infty} E \left\{ \varepsilon^T(t_i) W \varepsilon(t_i) \middle| Y(t_j, t_{i-1}) \right\}$$

$$J_a = J_o + \text{const}$$

with J_o taken as J_o^k , $k=1,2,3$, or 4 from (9), (10), (17), (18).

Such a novel set-up provides a new, broad look at this field of reasearch. Our plan for further research consists of solving the following tasks:

- 1) Forming the auxiliary performance index for MIMO control systems desribed in Sect. 3.
- 2) Studying some particular cases with different and progressively increasing levels of uncertainty, in order to determine the level, for which the API approach still remains workable.
- 3) Numerical experimental testing the approach and demonstrating its applicability to different applications.
- 4) Elucidation of the numerical properties of this approach and using the modern efficient computational techniques for its computer implementation.

6. Acknowledgements

I would like to thank Prof. B. Verkhovsky for his most helpful input, Dr. A. Murgu for his useful comments, and an anonymous reviewer for a number of corrections that improved the style of this paper.

7. References

- [1] C. S. Draper, “Principles of Optimizing Control Systems and an Application to the Inertial Combustion Engine,” ASME Publications, Mawson, 1951.
- [2] R. E. Kalman, “Design of a Self-Optimizing Control Systems,” *Transactions of ACME*, Vol. 80, 1958, pp. 468-478.
- [3] A. A. Feldbaum, “The Theory of Dual Control I,” *Automation and Remote Control*, Vol. 21, No. 9, 1961, pp. 874-883.

- [4] A. A. Feldbaum, "The Theory of Dual Control II," *Automation and Remote Control*, Vol. 21, No. 11, 1961, pp. 1033-1039.
- [5] A. A. Feldbaum, "The Theory of Dual Control III," *Automation and Remote Control*, Vol. 22, No. 1, 1962, pp. 1-12.
- [6] A. A. Feldbaum, "The Theory of Dual Control IV," *Automation and Remote Control*, Vol. 22, No. 2, 1962, pp. 109-121.
- [7] P. E. Caines, "Linear Stochastic Systems," John Wiley and Sons, Inc., Hoboken, 1988.
- [8] Y. Z. Tsypkin, "Adaptation and Learning in Automatic Systems," Nauka Publications, Moscow, 1968.
- [9] T. L. Lai, "Sequential Change-point Detection in Quality Control and Dynamical Systems," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 57, No. 4, 1995, pp. 613-658.
- [10] P. Ansay, M. Gevers and V. Wertz, "Closed-Loop or Open-Loop Models in Identification for Control?" In: P. Frank, Ed., *CD-ROM Proceedings of 5th European Control Conference*, Karlsruhe, 31 August-3 September 1999, File F0544.
- [11] O. Grospeaud, T. Pointot and J. C. Trigeassou, "Unbiased Identification in Closed-Loop by an Output Error Technique," In: P. Frank, Ed., *CD-ROM Proceedings of 5th European Control Conference*, Karlsruhe, 31 August-3 September 1999, File F0792.
- [12] 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
- [13] A. H. Jazwinski, "Stochastic Processes and Filtering Theory," Academic Press, Cambridge, 1970.
- [14] H. Kaufman and D. Beaulier, "Adaptive Parameter Identification," *IEEE Transactions on Automatic Control*, Vol. AC-17, No. 5, 1972, pp. 729-731. doi:10.1109/TAC.1972.1100111
- [15] H. Cramér and M. R. Leadbetter, "Stationary and Related Stochastic Processes. Sample Function Properties and Their Applications," John Wiley and Sons, Inc., Hoboken, 1967.