State Evolution Computation based Minimum Variance Nonlinear Estimation with Application for Robust Control

Endre Nagy[*]

Member, SICE, Japan.

*Corresponding Author: *Endre Nagy, Member, SICE, japan*

Abstract: *State estimation methods for lumped parameter systems are presented in the paper with application for robust control. All the methods are based on a nonlinear predictor, which estimates future states through state evolution computations. The prediction may be made in several successive steps or in one step as limit of multistep computations. The nonlinear estimator is a predictor – corrector type one, correction is made through optimization on a finite horizon, giving possibility for approximate minimum variance estimation for a class of nonlinear plants. At the same time, the presented predictor may replace the predictor of extended Kalman filter, making possible better prediction and overall estimation through modification of the extended Kalman filter algorithm. An example shows how to design robust nonlinear stochastic control with the original nonlinear model and with the nonlinear estimator presented in the paper for a class of nonlinear plants. The solution is achieved in analytical form.*

Keywords: *Optimum prediction, Estimation and filtering, Nonlinear estimation, Minimum variance estimation, Robust estimation and control.*

1. INTRODUCTION

State and parameter estimation are basic themes of control engineering. The solution is worked out for linear systems; however, nonlinear state and parameter estimation raises a lot of questions. To solve the optimum nonlinear filtering problem, propagation of the conditional probability density function, which needs a huge amount of computations, is necessary and thus real time application is not possible or not practical [1], [2]. Instead, suboptimal schemes, as the EKF (extended Kalman filter [3], [4]) etc. are used in practice. It is well known that the Ito equation and the Fokker – Planck or Kolmogorov forward equation play a great role in description of the nonlinear state evolution [5], [6]. The Kolmogorov forward equation describes the evolution of transition probability density of the Markov process generated by the Ito equation. However, the computation is related to special circumstances. Besides, the Kolmogorov forward equation has been solved only in a few simple cases [5]. In the past several attempts have been made to find practical solutions for the nonlinear estimation problem [7], [8] etc. A "dynamic optimization filter" (DOF) and a "dynamic optimization parameter estimator" (DOPE) was also described [9], [10]. This estimator tries to estimate the process and output noise by means of dynamic optimization, for which purpose an optimization method with the name "optimized stochastic trajectory / output sequence tracking" (OSTT, [10]) has been developed. A different approach is the "nonlinear optimization filter" (NOF). This filter predicts the state evolution on a finite horizon and updates the estimates through optimization on the horizon. The filter is of predictor – corrector type. To the prediction the linearized system is used, to the discretization virtual intersample points are defined. If accuracy of discretization is improved, accuracy of the prediction, too, will improve to a limit. The resultant state estimate difference may be given with a sum. This predictor has been called MISLINPRED (multi intersample linearization based nonlinear predictor [11]). MISLINPRED gives possibility for intersample state estimation, too. However, if the number of virtual intersample points reaches infinity, the summation can be replaced with an integral, which may be evaluated. Consequently, for the prediction a closed formula, called **M**-operator, may be obtained. The correction (update) is made through optimization on a finite horizon, in function of assumed states at the bottom of horizon. Predictions with the **M**-operator necessitate rethinking previously developed methods, which use predictions, including the author's own methods for state and parameter estimation. It is shown in the paper that the NOF filter may be an approximation of

minimum variance nonlinear optimization filter (MV-NOF), if expected value of output estimation error is zero. On the other hand, the predictor of EKF may be replaced with the **M**-operator. The resultant filter is the MOD-EKF (modified extended Kalman filter). The MOD-EKF may give better state estimates than EKF. An additional advantage of the NOF or MV-NOF is that optimization on a finite horizon gives possibility to synchronous parameter estimation, even in case of time varying parameters, through extension of the state vector with the unknown parameters and through forgetting factors [11], [12]. This paper includes revision and extension of the earlier developed MISLINPRED algorithm for the case of infinite virtual intersample points and for overall nonlinear state estimation. The paper is extended version of the [13] paper, presented at the 20th IFAC World Congress in Toulouse in 2017. The extension is related to robust nonlinear estimation and robust nonlinear stochastic control design. This is the first time that analytical solution is achieved with the original nonlinear equations, applying the **M**-operator and OSTT through a series of two – stage optimizations; all former solutions by the author were obtained either through optimization with the linearized model or through numerical optimization. The only requirements are differentiability of the model (to compute the **M**-operator) and manageability of the derived equations to get a unique solution.

2. NONLINEAR PREDICTION THROUGH STATE EVOLUTION COMPUTATION

2.1 Prediction with Computing States at Finite Intersample Points

Consider a nonlinear ergodic stochastic process described with the state and output equations

$$\boldsymbol{x}(i+1) = \boldsymbol{f}(\boldsymbol{x}(i), \boldsymbol{u}(i), \boldsymbol{w}(i)), \tag{1}$$

$$\boldsymbol{y}(i) = \boldsymbol{h}(\boldsymbol{x}(i)) + \boldsymbol{n}(i), \quad \boldsymbol{x} \in \boldsymbol{R}^n, \boldsymbol{u} \in \boldsymbol{R}^m, \boldsymbol{y} \in \boldsymbol{R}^p. \tag{2}$$

In (1), (2) $w(i)$ and $n(i)$ are stochastic disturbances of known statistics. Expectations of states from $x(i)$ to $x(i+1)$ evolve in a nonlinear manner from $E\{x(i)\}$ to $E\{x(i+1)\}$. In the following it is assumed that expected value of states moves on a fictitious trajectory, approximation of which may be computed with the linearized model. For linearizable plants the linearized model may be given in the form

$$\Delta \boldsymbol{x}(i+1) = \boldsymbol{A}(i)\Delta \boldsymbol{x}(i) + \boldsymbol{B}(i)\Delta \boldsymbol{u}(i) + \boldsymbol{D}(i)\Delta \boldsymbol{w}(i), \tag{3}$$

$$\boldsymbol{y}(i) = \boldsymbol{C}(i)\Delta \boldsymbol{x}(i) + \Delta \boldsymbol{n}(i), \tag{4}$$

$$\Delta \boldsymbol{x}(i+1) = \boldsymbol{x}(i+1) - \boldsymbol{x}(i), \ \Delta \boldsymbol{u}(i) = \boldsymbol{u}(i) - \boldsymbol{u}(i-1),$$
$$\Delta \boldsymbol{y}(i) = \boldsymbol{y}(i) - \boldsymbol{y}(i-1). \tag{5}$$

In (3), (4) **A**, **B**, **C**, **D** are appropriate matrices. If the plant model is not linearizable in the whole domain of states, it is usually linearizable sectionally. On the basis of (3), fictitious intersample states may be defined. If there are NN fictitious intersample points between $x(i)$ and $x(i+1)$, the states and control signals at the end of intersample sections may be denoted as

$$\boldsymbol{x}_1(i), \boldsymbol{x}_2(i), ... \boldsymbol{x}_{NN}(i), \boldsymbol{x}_{NN+1}(i) = \boldsymbol{x}(i+1), \ \boldsymbol{u}_1(i), \ \boldsymbol{u}_2(i)..... \tag{6}$$

The estimated state increment $\Delta \hat{\boldsymbol{x}}(i+1)$ follows from the state increment $\Delta \hat{\boldsymbol{x}}(i)$ if the control signal increment is known. $\Delta' \hat{\boldsymbol{x}}_1(i-1) = \hat{\boldsymbol{x}}_1(i-1) - \hat{\boldsymbol{x}}(i-1)$ results in the estimated state increment $\Delta' \hat{\boldsymbol{x}}_1(i)$, which can be computed with (3), too, and $\hat{\boldsymbol{x}}_1(i) = \hat{\boldsymbol{x}}(i) + \Delta' \hat{\boldsymbol{x}}_1(i)$ represents an estimated intersample state. However, when the new state is computed starting from $\hat{\boldsymbol{x}}_1(i),$ a new linearization has to be made at the $\hat{\boldsymbol{x}}_1(i\text{-}1), \boldsymbol{u}(i\text{-}1)$ operating point (it is assumed that $\hat{\boldsymbol{x}}_1(i-1)$ is in the close neighborhood of the trajectory of nonlinearly evolving states), and zero control signal change has to be taken into consideration (in case of zero order hold). Increasing the number of intersample sections may give a better approximation of the nonlinear trajectory of states, increasing the accuracy of prediction. Assuming zero order hold, the linearized equations between intersample points with estimated states and $\hat{\boldsymbol{x}}(i) = E\{\boldsymbol{x}(i)\}, E(\boldsymbol{w}(i) = 0$ are

$$\Delta' \hat{\boldsymbol{x}}_1(i) = \boldsymbol{A}(\hat{\boldsymbol{x}}(i-1), \boldsymbol{u}(i-1))\Delta' \boldsymbol{x}_1(i-1), \tag{7}$$

$$\Delta' \hat{\boldsymbol{x}}_2(i) = \boldsymbol{A}(\hat{\boldsymbol{x}}_1(i-1), \boldsymbol{u}_1(i-1))\Delta' \hat{\boldsymbol{x}}_2(i-1), \tag{8}$$

$$\Delta' \hat{\boldsymbol{x}}(i+1) = \boldsymbol{A}(\hat{\boldsymbol{x}}_{NN}(i-1), \boldsymbol{u}_{NN}(i-1))\Delta' \hat{\boldsymbol{x}}_{NN+1}(i-1)$$
$$+ \boldsymbol{B}(\hat{\boldsymbol{x}}_{NN}(i-1), \boldsymbol{u}_{NN}(i-1))\Delta' \boldsymbol{u}_{NN+1}(i-1). \tag{9}$$

If **A**, **B** depend on $w(i)$, value of the process noise at the operating point of real sampling instants may be taken into consideration with its expectation. With

$$\hat{\boldsymbol{x}}(i+1) = \hat{\boldsymbol{x}}(i) + \Delta' \hat{\boldsymbol{x}}_1(i) + \Delta' \hat{\boldsymbol{x}}_2(i) + ... \Delta' \hat{\boldsymbol{x}}_{NN+1}(i) \tag{10}$$

the whole trajectory of estimated states may be approximated, resulting in an estimation for the states at the next real sampling time point, through a series of intersample state estimations. At the beginning of computations estimation is made on the expected value of states at the actual and preceding sampling instants. Next an estimation is made on the initial increment to be taken into consideration, e.g. as

$$\Delta' \hat{\boldsymbol{x}}_1(i-1) \approx (\hat{\boldsymbol{x}}(i) - \boldsymbol{x}(i-1))/(NN+1). \tag{11}$$

The prediction is based on the relationship

$$E\{\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})\} = \boldsymbol{f}\{E(\boldsymbol{x}), E(u), E(w)\} \tag{12}$$

for linear systems. The described predictor may be called "multi intersample linearization based nonlinear predictor" (MISLINPRED [11]).

2.2 Prediction with Assumption of Infinite Intersample Points

With (7)~(10), state increment prediction on a real sampling interval through assumption of finite virtual sampling intervals may be given in the form

$$\Delta\hat{\boldsymbol{x}}(i+1) = \sum_{k=0}^{NN} \boldsymbol{f}'_{\hat{\boldsymbol{x}}}(\hat{\boldsymbol{x}}_k(i-1), \boldsymbol{u}_k(i-1))\Delta' \hat{\boldsymbol{x}}_{k+1}(i-1)$$
$$+ \sum_{k=0}^{NN} \boldsymbol{f}'_{\boldsymbol{u}}(\hat{\boldsymbol{x}}_k(i-1), \boldsymbol{u}_k(i-1))\Delta' \boldsymbol{u}_{k+1}(i-1), \tag{13}$$
$$\hat{\boldsymbol{x}}_0(i-1) = \hat{\boldsymbol{x}}(i-1), \quad \boldsymbol{u}_0(i-1) = \boldsymbol{u}(i-1).$$

In (13) $0 \le k \le NN$ is serial number of virtual sampling instants. Since $E\{w(i)\}=0$, a disturbance term is lacking in (13). Introducing the

$$\hat{\boldsymbol{x}}^* = [\hat{x}_1, \hat{x}_2, ... \hat{x}_n, u_1, u_2, ... u_m]^{\mathrm{T}} \tag{14}$$

extended state vector, for $k=\infty$ the above sum may be replaced with

$$\Delta\hat{\boldsymbol{x}}(i+1) = \int_{\hat{\boldsymbol{x}}^*(i-1)}^{\hat{\boldsymbol{x}}^*(i)} \boldsymbol{f}'_{\hat{\boldsymbol{x}}^*}(\hat{\boldsymbol{x}}^*) d\,\hat{\boldsymbol{x}}^* = \boldsymbol{M}(\varDelta\hat{\boldsymbol{x}}(i), \varDelta\boldsymbol{u}(i)), \tag{15}$$

symbolized with the $\boldsymbol{M}$ - operator. Elements of (14) are sums of integrals of multivariable scalar functions. The variables are independent, but related to each other through the sampling instants. It can be seen that the above algorithm may be extended for the case of continuous – discrete Kalman filter [3]. However, the real sampling time points have to be replaced with virtual ones in this case. The result may be verified for linear systems as follows. Consider a SISO system with the state equation

$$\boldsymbol{x}(i+1) = \boldsymbol{A}\boldsymbol{x}(i) + bu(i). \tag{16}$$

With (13) for the real state increment estimation

$$\Delta\hat{\boldsymbol{x}}(i+1) = \sum_{k=0}^{NN} \boldsymbol{A}^T \Delta' \hat{\boldsymbol{x}}_{k+1}(i-1) + \sum_{k=0}^{NN} b\Delta' u_{k+1}(i-1). \tag{17}$$

For ∞ virtual intersample points

$$\Delta\hat{\boldsymbol{x}}(i+1) = \int_{\hat{\boldsymbol{x}}(i-1)}^{\hat{\boldsymbol{x}}(i)} \boldsymbol{A}^T d\,\boldsymbol{x}(i) + \int_{u(i-1)}^{u(i)} bdu(i)$$
$$= [\boldsymbol{A}\boldsymbol{x}(i)]_{\hat{x}(i-1)}^{\hat{x}(i)} + [bu(i)]_{u(i-1)}^{u(i)}$$
$$= \boldsymbol{A}(\hat{\boldsymbol{x}}(i) - \hat{\boldsymbol{x}}(i-1)) + b(u(i) - u(i-1))$$
$$= \boldsymbol{A}\Delta\hat{\boldsymbol{x}}(i) + b\,\Delta u(i). \tag{18}$$

However, subtracting the state equation

$$\boldsymbol{x}(i) = \boldsymbol{A}\boldsymbol{x}(i-1) + bu(i-1) \tag{19}$$

from (16), we get back (18) for $\hat{x} = x$. The prediction based on state evolution computation gives back what follows from the linear model. However, the nonlinear estimation is more complex and the *M*-operator based estimation may be better – especially for large disturbances – than increment computation from the nonlinear model.

3. MINIMUM VARIANCE ESTIMATION

3.1. Relationship between State and Output Variances

The mean of a random variable [14]

$$y=h(x) \tag{20}$$

can be expressed as

$$E\{h(x)\} = \int_{-\infty}^{\infty} h(x)\, p(x)\, dx, \tag{21}$$

where $p(x)$ is the density function of x. If x is concentrated around its mean, $E\{h(\text{x})\}$ may be expressed with the moments of x. Suppose that $p(x)$ is negligible outside the interval $[\alpha-\varepsilon, \alpha+\varepsilon]$ and in this interval $h(x))=h(\alpha)$. In this case

$$E\{h(x)\}=h(\alpha). \tag{22}$$

To get a better approximation, $h(x)$ may be approximated with the polynomial

$$h(x)=h(\alpha)+h'(\alpha)(x-\alpha)+h''(\alpha)(x-\alpha)^2/2+... \tag{23}$$

If $h(x)$ is substituted with a parabola, from (21) and (23)

$$\mu_y = E\{h(x)\} = h(\alpha) + h''(\alpha)\sigma^2 / 2, \tag{24}$$

where σ^2 is variance of x. On the basis of (24), the mean of $h^2(x)$ can be expressed as

$$\sigma_y^2 + \mu_y^2 = E\{h^2(x)\}$$
$$= h^2(\alpha) + [h'(\alpha)^2 + h(\alpha)h''(\alpha)]\sigma^2. \tag{25}$$

Substituting (24) for μ_y and discarding the term of $\sigma^{4,}$, we get

$$\sigma_y^2 = |h'(\alpha)|^2 \sigma,^2 \tag{26}$$

i.e. first order estimate of the output variance is proportional to the input variance. If there are two state variables, the output can be given in the form

$$y=h(x_1,x_2) \tag{27}$$

Expectation of the output can be computed [13] as

$$E\{h(x_1, x_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x_1, x_2)\, p(x_1, x_2)\, dx_1 dx_2. \tag{28}$$

In (28) $p(x_1,x_2)$ is the joint density of x_1 and x_2. (28) may be given in the form of a single integral

$$E\{h(x_1)\} = \int_{-\infty}^{\infty} h(x_1)\, p_x(x_1)\, dx_1, \tag{29}$$

$$p_x(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2)\, dx_2. \tag{30}$$

The similarity between (29) and (21) can be noticed. If the same assumptions can be made for x_1 and $p_x(x_1)$, as well as x_2 and $p_x(x_2)$ as in the single variable case, proportionality between the input and output variances may be seen. If there are more than two state variables, with the same reasoning we get the same conclusion.

3.2. Conditions for Minimum Variance Estimation

To make the estimation a minimum variance one, optimum prediction and optimum correction in a minimum variance sense is necessary. The objective is minimization of the reconstruction error variance

$$var\tilde{x}(i) = E\{(\tilde{x}(i) - E(\tilde{x}(i))(\tilde{x}(i) - E(\tilde{x}(i))^{\mathrm{T}}\}, \tag{31}$$

$$\tilde{x}(i) = x(i) - \hat{x}(i). \tag{32}$$

Substituting (32) into (31),

$$var\tilde{x}(i) = E\{(x(i) - \hat{x}(i) - E(x(i) - \hat{x}(i)) \tag{33}$$
$$(x(i) - \hat{x}(i) - E(x(i) - \hat{x}(i))^{\mathrm{T}}\}.$$

Consider a SISO system. According to (26), the estimated output variance – with some assumptions – is proportional to the estimated state variance (provided, the output disturbance is not taken into consideration). If there is a unique relationship between the estimated states and estimated output, minimization of the estimated output variance results in minimization of the estimated state variance, provided, there is a strictly monotone increasing relationship (sufficient condition) between the variances. This relationship can be examined through simulations. Minimization of the estimated output variance may be approximated from the performance index (PI)

$$\frac{1}{N-1}\sum_{i=1}^{N}(y(i) - \hat{y}(i))^2 \tag{34}$$

through optimization on a finite horizon in function of the estimated states at the bottom of horizon. For the state estimation error variance

$$var\{x(i) - \hat{x}(i)\} = var\,x(i) + var\,\hat{x}(i) - 2cov(x(i), \hat{x}(i)), \tag{35}$$

and for the output estimation error variance

$$var\{y(i) - \hat{y}(i)\} = var\,y(i) + var\,\hat{y}(i) - 2cov(y(i), \hat{y}(i)). \tag{36}$$

If there is only one state variable and the output estimation is unbiased, the (34) PI can be used for minimization of (35). Otherwise, minimization of the output estimation error variance

$$var\{y(i) - \hat{y}(i)\} \approx \frac{1}{N-1}\sum_{i=1}^{N}\{y(i) - \hat{y}(i) - E(y(i) - \hat{y}(i))\}^2, \tag{37}$$

$$E\{y(i) - \hat{y}(i)\} \approx \frac{1}{N}\sum_{i=1}^{N}\{y(i) - \hat{y}(i)\} \tag{38}$$

(approximation of (36)) gives possibility for better state estimation, if the relationship between (35) and (36) is strictly monotone increasing for some suitable norm of (35). This relationship can also be scrutinized with simulations. Alternatively, if there are several state variables, expectation of the output can be expressed with each state variable separately, as shown in (28) ~ (30), and investigation of relationship between (35) and (36) may be led back to the single state variable case. However, if there is additional disturbance on the output, it has to be corrected to the estimation. The disturbance free output can be given as

$$y^*(i) = y(i) - n(i). \tag{39}$$

Since value of $n(i)$ is not known, we may replace the disturbance with its expectation:

$$y^*(i) \approx y(i) - E\{n(i)\}. \tag{40}$$

If $E\{n(i)\}=0$, $y^*(i) \approx y(i)$ may be applied.

3.3. Correction through Optimization on a Finite Horizon

The proposed estimator may be viewed as a predictor – corrector [3] one. Prediction is made with MISLINPRED or the **M** - operator on a finite horizon in function of selected starting states at the bottom, and correction is made through minimization of a suitable PI, which may have the form

$$F(N) = \frac{1}{Nm-1}\sum_{j=1}^{N}\sum_{k=1}^{m}\{y_k(i-j) - \hat{y}_k(i-j)\}^2. \tag{41}$$

To the computations we assume that the state and output equations of the model are analytic and the stochastic process is ergodic. (41) is the output variance, in practice the $1/Nm$-1 factor may be omitted. If

$$E\{y_k(i-j) - \hat{y}_k(i-j)\} = 0; \quad k = 1,...m; \quad j = 0,... N\text{-}1, \tag{42}$$

(41) is approximation of the output estimation error variance, too. The filter computed through minimization of (41) may be called "nonlinear optimization filter" (NOF). If (42) is subtracted from

the output estimation error in (41) (like in (37) for a SISO plant), the resulting filter may be called " minimum variance nonlinear optimization filter" (MV-NOF). However, NOF is equivalent with MV-NOF, if expected value of the output estimation error is zero. Due to complexity of the PI, analytical solutions may be problematical. The optimization problem, however, may be solved through iterations [10] or through discretizing the states at the bottom of horizon and evaluating the PI for each discretized state combination. The global minimum is determined from comparison of the PI values belonging to different independent variable combinations at the bottom. To the iterative solution assume that the PI has the form of (41). Define a vector as

$$\boldsymbol{Q}(N) = [F(N), F(N-1),... F(N-n+1)]^{\mathrm{T}}.$$ (43)

Ideal value of $Q(N)$ is zero. Therefore, for the deviations

$$\alpha \boldsymbol{Q}(N) \approx -\boldsymbol{J}(\hat{\boldsymbol{x}}(i-N)^{n+1} - \hat{\boldsymbol{x}}(i-N)^{n}), \qquad 0 < \alpha \le 1,$$ (44)

where

$x(i-N)^{n}$ - estimate of the state vector after the nth iteration,

$$\boldsymbol{J} = \frac{\partial \boldsymbol{Q}(N)}{\partial \hat{\boldsymbol{x}}(i-N)}.$$ (45)

The derivatives in the (45) Jacobian matrix can be evaluated numerically. From (44)

$$\hat{\boldsymbol{x}}(i-N)^{n+1} \approx \hat{\boldsymbol{x}}(i-N)^{n} - \alpha \boldsymbol{J}^{-1}\boldsymbol{Q}(N).$$ (46)

In practice J^{-1} is not computed, but (44) is solved for the new state estimate. (46) gives possibility for recursive computations, when the procedure is repeated at the new sampling instant, and starting estimates of states at the bottom of new horizon are the states computed on the preceding horizon. A relatively simple computation may be used for state estimation through fixed point iteration [15]. For the ideal case from (43)

$$F(N) = 0,$$
$$\vdots$$
$$F(N-n+1) = 0.$$ (47)

The formula for iteration for the (47) system of equations is

$$\hat{\boldsymbol{x}}^{n+1} = \boldsymbol{Q}(N) + \alpha \hat{\boldsymbol{x}}^{n},$$ (48)

where $0 < \alpha \le 1$ is the learning rate. If the iteration is not convergent, for another initial point and / or smaller α it may converge. To complete the computation for several initial state values, the trap of finding the neighbourhood of a local minimum may be avoided. However, in case of recursive computations, this may be important only at the start of computations.

3.4. Some Conditions for Robustness

To have a robust estimator, either the design has to be made in accordance with robustness principles, or a non-robust estimator may be made robust with some transformation. Robustness means that small disturbances or changes in the design hypothesis result in only small estimation errors. The first step for robust estimation is to get a process model as accurate as possible. Estimation with MISLINPRED or the $\boldsymbol{M}$ - operator gives inherently some robustness. In addition there are several possibilities to make a non – robust estimator to a robust one: the method of weighted least squares (WLS [16]), least trimmed squares regression (LTS [17]), maximum trimmed likelihood estimation (MTLE [18]), M-estimation [19], S-estimation [20], MM-estimation [21]. In case of WLS (41) is modified to

$$F(N) = \sum_{j=1}^{N} \sum_{k=1}^{m} w_{i-j,k} \{y_k(i-j) - \hat{y}_k(i-j)\}^2,$$ (49)

where $w_{i-j,k}$ is scalar weight. Most accurate results are obtained when the weights are inversely proportional to the output variances at each combination of predictor values, i.e. for a SISO plant

$$w_{i-j} = \frac{1}{\sigma_{i-j}^2}.$$ (50)

However, the variances are not known, instead estimated weights are used based on a set of replicate measurements:

$$w_{i-j} = \frac{1}{\dfrac{\sum\limits_{s=1}^{r}(y_{i-j,s} - \bar{y}_{i-j})^2}{r-1}}. \tag{51}$$

In (51) r is number of replicate measurements at the j-th stage, $y_{i-j,s}$ is output indexed by stage number and replicate measurement serial number, and $\bar{y}_{i-j}$ is mean of the output. To get relatively adequate values for w_{i-j}, a large number of replicate measurements are needed. In case of no replicate measurements for a set of predictor variable values, repeated measurements may be applied. A better approximation is to relate the variance of response to the predictor variables:

$$\sigma^2_{i-j} = f(\hat{x}(i-j)), \tag{52}$$

$$w_{i-j} = \frac{1}{f(\hat{x}(i-j))}. \tag{53}$$

A power function is often a good choice:

$$\sigma^2_{i-j} = c_1 f(x(i-j))^{c_2}, \tag{54}$$

where c_1, c_2 are suitable constants. Another approach is to extend the state vector with the unknown weights and compute their estimates from optimization. A disadvantage of WLS is that the applied weights are only approximations. LTS is a good alternative and best choice of some researchers. The LTS estimates are defined as

$$\hat{x}_N = \underset{x}{argmin} \sum_{i=1}^{h} \varepsilon_i^2, \quad N/2 < h \le N, \tag{55}$$

where ε_i^2 represents the i-th order squared error. The h trimming constant determines the robustness of the LTS estimator, since N-h observations with the largest errors don't influence the sum. If $h=N$, the resultant estimator is the traditional LS and the robustness is the smallest. MTLE may be derived from the maximum likelihood estimator similarly to derivation of LTS estimation [22].

4. MODIFIED EXTENDED KALMAN FILTER

An approximate solution to the nonlinear filtering problem is the EKF [3], [4]. The EKF algorithm calculates the gain and variance matrices and updates from the linearized model, but the original nonlinear equations are are used to state propagation. However, the standard state propagation of EKF may be replaced with prediction by the **M** - operator. The resulting filter may be called "modified extended Kalman filter"(MOD-EKF). The state prediction equation for discrete EKF has the form[3]

$$\hat{x}(i/i-1) = f(\hat{x}(i-1/i-1), u(i-1)). \tag{56}$$

However, the **M** - operator uses infinite virtual sampling points to estimate the state evolution as shown in section 2.2, and the result may be better prediction. The other steps of the computation (innovation, innovation covariance and Kalman gain computation, state and covariance correction [3]) are the same for MOD-EKF as for EKF. The MOD-EKF algorithm may be summarized in the following manner:

State propagation:

$$\hat{x}(i/i-1) = \hat{x}(i-1/i-1) + M(\Delta x(i), \Delta u(i))$$

Covariance prediction:

$$\begin{aligned} P(i/i-1) &= A(\hat{x}(i/i-1))P(i-1/i-1)A^{\mathrm{T}}(x(i/i-1)) \\ &\quad + R_w(i-1) \end{aligned} \tag{57}$$

Innovation:

$$e(i) = y(i) - \hat{y}(i/i-1) \tag{58}$$

Innovation covariance:

$$R_e(i) = C(\hat{x}(i/i-1))P(i/i-1)C^T(\hat{x}(i/i-1))$$
$$+ R_n(i) \tag{59}$$

Kalman gain:

$$K(i) = P(i/i-1)C^T(\hat{x}(i/i-1))R_e^{-1} \tag{60}$$

State correction:

$$\hat{x}(i/i) = \hat{x}(i/i-1) + K(i)e(i) \tag{61}$$

Covariance correction:

$$P(i/i) = [I - K(i)C(\hat{x}(i/i-1))]P(i/i-1) \tag{62}$$

Initial conditions: $\hat{x}(0,0)$; $P(0,0)$

Assumed disturbances: $N(0, R_w(i))$, $N(0, R_n(i))$, white and Gaussian.

5. SIMULATION RESULTS

Example 1:

Consider a plant given with the equations

$$x(i+1) = a\,exp(-x(i))\,u(i) + bu(i) + w(i), \tag{63}$$

$$y(i+1) = x(i+1) + n(i+1). \tag{64}$$

In (63) a, b are parameters. It is assumed that the disturbances are of normal distribution and the expectations, variances and parameter values are

$$E\{w(i)\} = 0;\ E\{n(i)\} = 0;\ a = 0.3;$$
$$b = 0.6;\ \sigma_{w(i)}^2 = 0.15;\ \sigma_{n(i)}^2 = 0.075. \tag{65}$$

The linearized state equation is

$$\Delta\hat{x}(i+1) = -a\,exp(-\hat{x}(i))\,u(i)\Delta\hat{x}(i)$$
$$+ (a\,exp(-\hat{x}(i)) + b)\Delta u(i) = A(i)\Delta\hat{x}(i) + B(i)\Delta u(i). \tag{66}$$

Assuming 2 finite intersample points, the state equations on the first, second and third virtual intersample sections are

$$\Delta'\hat{x}_1(i) = -a\,exp(-\hat{x}(i-1))\,u(i-1)\Delta'\hat{x}_1(i-1), \tag{67}$$

$$\Delta'\hat{x}_2(i) = -a\,exp(-\hat{x}_1(i-1))\,u_1(i-1)\Delta'\hat{x}_2(i-1), \tag{68}$$

$$\Delta'\hat{x}_3(i) = -a\,exp(-\hat{x}_2(i-1))\,u_2(i-1)\Delta'\hat{x}_3(i-1)$$
$$+ (a\,exp(-\hat{x}_2(i-1)) + b)\Delta u(i). \tag{69}$$

With (67) ~ (69)

$$\Delta\hat{x}(i+1) = \Delta'x_1(i) + \Delta'x_2(i) + \Delta'x_3(i). \tag{70}$$

Measurement data was generated in accordance with (65). If ∞ intersample points are assumed, through use of (15)

$$\Delta\hat{x}(i+1) = 2\,a\,exp(-\hat{x}(i))\,u(i)$$
$$- 2\,a\,exp(-\hat{x}(i-1))\,u(i-1)).. + b(u(i) - u(i-1)). \tag{71}$$

To optimization on a finite horizon, the (37) PI and one, two and three stages horizons were taken into consideration. The optimum was approximated through discretizing the states at the bottom of horizon with the step size $\Delta x = 0.1$, evaluating the PI for each discretized state and picking up the optimum, assuming that the optimum state at the bottom is in the range

$$x(i-N) - 0.1 \le \hat{x}(i-N) \le x(i-N) + 0.1,$$

where N is length of horizon in steps. For computations on finite intersample sections, (11) was assumed. Fig. 1 shows simulation results with MOD-EKF with ∞ intersample points and EKF:

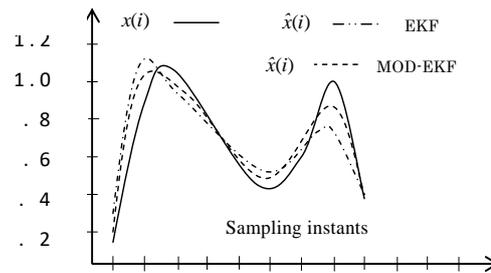**Fig1.** *Real and estimated states with MOD-EKF and EKF*

Improvement of estimation with MOD-EKF based on comparison of sums of squared errors on a 9 step horizon is about 40%. The result with MV-NOF may be comparable with the one obtained with MOD-EKF, however, tuning the filter is necessary. The estimation is more accurate if the length of finite horizon is 2 stages compared with the case of one stage, and approximately the same for 3 stages. For the case of finite virtual intersample points, the result is better for 2 assumed intersample points than for only one, and better for one intersample point than without intersample point. At the same time, the simulation gives better result for ∞virtual intersample points than only for two ones.

Example 2:

Consider a discrete nonlinear process given by the state and output equations

$$x(i+1) = ax(i) + bx^2(i) + cu(i) + w(i). \tag{72}$$

$$y(i+1) = x(i+1) + n(i+1). \tag{73}$$

The parameter values are a=0.3, b=0.5 and c=0.6. The assumed disturbances are as in Example 1, with $R_w(i)$=0.15, $R_n(i)$=0.075 and $R_w(i)$=0.006, $R_n(i)$=0.003, repectively. The state estimation algorithm with MOD-EKF can be given with the following equations:

$$\hat{x}(i+1/i) = \hat{x}(i/i) + a(\hat{x}(i/i) - \hat{x}(i-1/i-1))$$
$$+ b(\hat{x}^2(i/i) - \hat{x}^2(i-1/i-1)) + c(u(i) - u(i-1))$$
$$P(i+1/i) = (a + 2b\hat{x}(i/i))^2 \, P(i/i) + R_w(i)$$
$$e(i+1) = y(i+1) - \hat{y}(i+1/i)$$
$$R_e(i+1/i) = P(i+1/i) + R_n(i)$$
$$K(i+1) = P(i+1/i) / R_e(i+1)$$
$$\hat{x}(i+1/i+1) = \hat{x}(i+1/i) + K(i+1)\,e(i+1)$$
$$P(i+1/i+1) = (1 - K(i+1))\, P(i+1/i).$$

This algorithm was tested with various inputs and compared with the EKF algorithm. Improvement of estimation with comparison of sum of squared errors is about 31% in average. Change of disturbance variances did not influence essentially the result. The Unscented Kalman Filter (UKF), too, can provide good accuracies, comparing with EKF [23]. However, UKF is based on a set of trial points, and thus cannot be considered as a global approximation. Moreover, it does not work well with nearly deterministic systems and needs more computation than the EKF. In comparison, computational cost of MOD-EKF is comparable with that of EKF. Particle filter depends on an initial guess of the prior distribution and it is sensitive to the initial values [7]. Consequently, recovery from poor initial guess needs much more measurements. Besides, in consequence of limited number of particles, the obtained result is often less accurate than that of obtained with MOD-EKF or even EKF. However, the accuracy can be increased as necessary by increasing the number of particles, but the computation may become expensive.

Example 3:

This example concerns robust stochastic nonlinear tracking control design with the original nonlinear equations. Assume an unknown plant where the true plant model is given with (63), (64), (65). Let the model structure to the design be

$$x(i+1) = a1\, x^2(i) + b1\, u(i) + w(i), \tag{74}$$

$$y(i+1) = x(i+1) + n(i+1). \tag{75}$$

where $a1$, $a2$ are parameters. The motivation for selection of the (74), (75) model may be the easy manageability and the fact that arbitrary accuracy can be achieved with polynomials, although high degree polynomials may be numerically unstable [15]. To the solution first the model parameters are identified; this can be done e.g. as shown in [11]. Various former investigations ([11], etc.) prove that the optimum solution is two – stage optimization based one. Consequently, for optimization the PI

$$F_{i+1}^{i+2} = \sum_{j=i+1}^{i+2} [\{y(j) - r(j)\}^2 + \lambda\, u^2(j-1)] \tag{76}$$

is used. In (76) $r(j)$ is the reference and λ is control weight. First optimum tracking is computed with the model. However, output of the plant is different. The error is fed back through the negative inverse of the model to the input of the plant, forcing the plant output nearer to the model output. Simulations show that an additional K feedback gain, determined through trials, may improve the solution. The model output estimation was achieved with the algorithm shown in Example 2, although MV – NOF may also be used. To computation of the right inverse of the model, the model input is expressed with model variable estimates, and the estimated output is forwarded with one step. With OSTT from

$$\frac{\partial F_{i+1}^{i+2}}{\partial u(i)} = 0 \tag{77}$$

the next control signal can be computed in knowledge of $u(i)$. The initial control signal may be estimated from some considerations, e.g. from one – stage optimization. Completing the computations, we get

$$u(i+1) = -(2(\hat{x}(i/i) + a1(\hat{x}(i/i)^2 - \hat{x}(i-1/i-1)^2) + b1(u(i) - u(i-1)) - r(i+1))\, b1 + 2\lambda\, u(i)$$
$$+ 2(\text{DER}(i+1) + 2\,a1\,\hat{x}(i+1/i+1)\text{DER}(i+1) - b1)(\hat{x}(i+1/i+1) + a1(\hat{x}(i+1/i+1)^2 - \hat{x}(i/i)^2)$$
$$- b1\,u(i) - r(i+2)))/(2(\text{DER}(i+1) + 2\,a1\,\hat{x}(i+1/i+1)\text{DER}(i+1) - b1)\,b1), \tag{78}$$

$$\text{DER}(i+1) = \frac{\partial \hat{x}(i+1/i+1)}{\partial u(i-1)}.$$

$\text{DER}(i+1)$ was numerically approximated to the simulation. The correction on the input of the plant can be computed as

$$ucorr(i+1) = -K(y(i+1) - \hat{x}(i+1/i+1) - a1\,\hat{x}(i+1/i+1)^2)/b1. \tag{79}$$

To the simulation (78) and (79) was used. Various investigations have been done through simulations. If process parameters are $a=0.2$ and $b=0.5$, the parameter estimator finds minimum at $a1=0.463$ and $b1=0.563$. If $\lambda=0.5$, the control is unstable without feedback. The control is unstable for $a=0.3$ and $b=0.6$, too, without feedback, but it becomes stable with feedback in both cases, making possible robust stability. However, deterioration of the process output is about 45% , based on comparison of sum of error squares on a finite horizon. Accuracy of tracking may be improved for processes of time varying parameters if the model parameters are identified in real time, resulting in robust performance. Optimum value of feedback gain is about $K\approx1.5$. The simulations show that better tracking accuracy can be achieved with two – stage control then with one – stage one, and with MOD –EKF instead of EKF.

6. CONCLUSION

The paper presents methods for nonlinear state estimation by way of nonlinear state evolution prediction and correction. A predictor with the name MISLINPRED has been developed for state evolution computation. This predictor can estimate future states in several steps assuming finite virtual intersample points, while the M-operator makes estimates in one step assuming ∞ intersample points. The finite or infinite intersample point assumption gives possibility for intersample state estimation, too. Simulation shows that in the case of assumed ∞ intersample points, accuracy of

prediction is better than in case of assumed finite intersample points. Correction is achieved through optimization on a finite horizon. With suitable selection of the performance index, minimum variance nonlinear filtering may be approximated for a class of nonlinear plants. Moreover, MISLINPRED or the *M*-operator can be used in the EKF algorithm for prediction of future states, instead of the EKF predictor. Simulation shows that accuracy of estimation improves in this case. Similar or better accuracy can be achieved with optimization on a finite horizon, with the advantage that sphere of applicability is larger. The paper shows that analytical solution of robust nonlinear stochastic tracking problems is possible with the original nonlinear model and with the presented estimator for a class of nonlinear plants. The solution is optimum for the model with the given estimation method and is working efficiently for the plant.

REFERENCES

[1] T. Söderström, Discrete – time stochastic systems, estimation and control, Cambridge, UK: University Press, 1984, ch. 9, p. 233.

[2] T. Glad and L. Ljung, Control theory, London and New York: Taylor & Francis, 2000, ch. 15, p.379.

[3] J.V. Candy, Signal processing, New York: McGraw – Hill Book Company, 1987, ch. 6.6, pp. 154-159.

[4] R.F. Stengel, Optimal control and estimation, New York: Dover Publications, 1994, ch. 4.6, pp. 386-388.

[5] A.H. Jazwinsky, Stochastic processes and filtering theory, Mineola, New York: Dover Publications, 2007, ch.4, pp. 105-110.

[6] K.J. Åström, Stochastic control theory, Mineola, New York: Dover, 2006, ch. 7, pp. 71-72.

[7] Rawlings, J.B. and Bakshi, B.R., Particle filtering and moving horizon estimation, Computers & Chemical Engineering, 30(10-12) 1529 (2006).

[8] Daum, F., Nonlinear filters: beyond the Kalman filter, IEEE A & E Systems Magazine, Part 2: Tutorials, 20(8) 57 (2005).

[9] Nagy E., Estimation on lumped parameter systems: dynamic optimization approach, Conference Proceeding, International Conference on Control and Automation, Pp. 101-106, (2004).

[10] Nagy E., Estimation and control: numerical solutions for hard nonlinear plants, Conference Proceeding, International Conference on Cybernetics and Information Technology, Systems and Applications and International Conference on Computing, Communications and Control Technologies, II, Pp. 268-273, (2007).

[11] Nagy E., Nonlinear stochastic adaptive control: a new approach, Symposium Preprint, 9th International Symposium on Robot Control (SYROCO' 09), Pp. 201-207, (2009).

[12] K.J. Åström and B. Wittenmark, Adaptive control, New York: Addison - Wesley, 1995, ch.2.2, pp. 52-53.

[13] Nagy E., Minimum Variance Nonlinear Estimation: State Evolution Approach, IFAC PapersOnLine 50-1 (2017) 3786-3792.

[14] A. Papoulis, Probability, random variables, and stochastic processes, Singapore: McGraw-Hill Series in Electrical Engineering, 1984, ch. 5-3, pp. 106-107.

[15] J.R. Rice, Numerical methods, software, and analysis, Singapore: McGraw-Hill, 1983, ch. 8.7, pp. 251-253.

[16] J.L. Crassidis and J.L. Junkins, Optimum estimation of dynamic systems, Boca Raton: Chapman & Hall / CRC, 2004, ch. 1, pp. 14-15.

[17] Hubert, M. and Rousseeuw, P.J., Robust regression with both continuous and binary regressors, Journal of Statistical Planning and Inference. 57 153 (1997).

[18] Hadi, A.S. and Lucano, A., Maximum trimmed likelihood estimators: a unified approach, examples and algorithms, Computational Statistics & Data Analysis. 25(3) 251 (1997).

[19] Perrachi, F., Robust M-estimators, Economic Reviews, 9(1) 1 (1990).

[20] P.J. Rousseeuw,and V.J. Yohai, Robust regression by means of S-estimators, Lecture Notes in Statistics No. 26, Springer, Berlin / New York, 256 (1984).

[21] V.J. Yohai, High breakdown - point and high efficiency robust estimates for regression, The Annals of Statistics, 15(2) 642 (1987).

[22] Nagy, E., Robust nonlinear estimation in case of noises of unknown statistics, Conference Proceeding, Eights International Conference on Mathematical Problems in Engineering Aerospace and Sciences, Pp. 529 – 533, (2011).

[23] Romanenko, A. and Castro, J.A., The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study, Computers & Chemical Engineering. 28, 347 (2004).

AUTHOR'S BIOGRAPHY

Endre Nagy, He graduated in 1966 at the Budapest Technical University, Faculty of Electrical Engineering. After having passed a PhD course at the Computing and Automation Research Institute of Hungarian Academy of Sciences in 1973, he worked at the Research Institute for Electric Energy Industry as scientific associate, and later as an associated professor in Budapest. Since 1984 he lives in Japan. He was a guest researcher at the Chuo University, Tokyo, between 1984 and 1986, and worked on computer control related fields as a researcher. He is requested reviewer of scientific papers and received several awards, including H. Prof. Engineering. His main interest is development of control theory.