15th Triennial World Congress, Barcelona, Spain

THE GENERALIZED LEAST-SQUARES SYSTEM IDENTIFICATION SUBJECT TO UNKNOWN BOUNDED NOISE

Vyacheslav F. Gubarev and Nataliya V. Panova

Space Research Institute of NAS and NSA of Ukraine
40, Acad. Glushkov Prsp., 03680 Kiev-187, Ukraine
E-mail: gvf@d310.icyb.kiev.ua

Abstract: On the base of the least-squares technique the universal method of regularized identification is developed. For any given input and output in the presence of the worst-case bounded noise it provides the convergent parameter estimate when the order of stable model is increased. This implies that even in the case when available information is not sufficient to identify the system it is possible to evaluate an approximate model using the approach offered in the paper. The recurrent form of regularized least-squares algorithm with adaptive regularization depending on incoming data is constructed. *Copyright © 2002 IFAC*

Keywords: Least-squares identification, Regularization, Bounded noise.

1. INTRODUCTION

The least-squares (LS) algorithm is one of the most widely used algorithms in identification problems. It injoys optimal estimate in the class of linear unbiased estimates and converges under stochastic assumptions about exogenous noise (Hyotyniemi, 1996; Ljung, 1987). The classic question concerning singular matrix inversion also is successfully solved using ridge estimates, truncated estimates, principal component regression, Marquardt estimates etc.

While solving the LS identification problem under the worst-case noise there are new problems. By Akcay and Khargonekar (1993) it is shown that this method robustly converges for estimating of the finite impulse response (FIR) models of systems. The result which demonstrates a divergence of LS H_∞-identification of the infinite impulse response (IIR) models under the bounded noise, was obtained by Akcay and Hjalmarson (1994). In other words it is impossible to use LS identification for IIR-models, as the problem is ill-posed irrespective of its matrix conditionality. By Gubarev and Panova (1999) the regularized LS transfer function estimate of the FIR systems was offered. It eliminates ill-conditionality and H_∞-converges as the order of the stable system is increased, in the presence of the worst-case noise for pseudo-random binary (PRB) input.

In the present paper on the base of the LS technique the universal regularized method of identification is developed. It provides the convergent worst-case error of the transfer function when passaging to approximated IIR-model for any given input sequence including not enough informative. The main idea of proposed regularization method is similar to stable summation of Fourier series where the coefficients are summarized with the specially selected weights. It provides both convergence and matrix invertibility for any input signal. Structure of regularizing matrix and constraints to be satisfied in order to guarantee the convergence of identification algorithm are obtained in section 2.1. In sections 2.2 and 2.3 some specific algorithms addressed to certain class of systems are constructed and convergence condition are obtained.

Section 3 contains a recurrent form of one algorithm under consideration. The worth and distinction of the suggested recurrent algorithm consist in its ability to refine a current estimate by selecting the optimal regularization parameter for new recurring step. The results of computational experiment with proposed recurrent algorithm are given in section 4.

2. IDENTIFICATION PROBLEM

We will consider a SISO BIBO linear time invariant discrete system

$$y_i = G(q,\boldsymbol{\theta})u_i + v_i\,, \qquad (1)$$

where the transfer function $G(q,\boldsymbol{\theta})$ is the function of unit delay operator q^{-1} and parameter vector $\boldsymbol{\theta}$, y_i is the output, u_i is the input, v_i is a bounded disturbance

$$|v_i| \le \varepsilon\,, \quad \forall i\,. \qquad (2)$$

We will consider nonparametric identification of stable plants (1) decomposable in series

$$G(z) = \sum_{k=0}^{\infty} g_k z^{-k} = \lim_{n\to\infty} \sum_{k=0}^{n} g_k z^{-k}\,, \qquad (3)$$

where $\{g_k\}$ are unknown parameters. To estimate them we will approximate a system (1) by FIR model with finite order n and will pass to limit as $n \to \infty$. Then for $\boldsymbol{\theta}_n = [g_0, g_1, \ldots, g_n] \in \Re^{n+1}$ we have

$$y_i = \sum_{k=0}^{n} g_k u_{i-k} + v_i\,,\ i = n+1,\ldots,n+N\,, \qquad (4)$$

where u_j, $j = 1,\ldots,n+N$. For a solvability of (4) there must be at least $N > n$ and informative input sequence. Then the LS estimate is

$$\hat{\boldsymbol{\theta}}_N = \mathbf{R}_N^{-1}\mathbf{S}_N\,. \qquad (5)$$

where $\mathbf{R}_N = \frac{1}{N}\mathbf{F}_N^T\mathbf{F}_N$, $\mathbf{S}_N = \frac{1}{N}\mathbf{F}_N^T\mathbf{Y}_N$. Here $\mathbf{F}_N$ is the $N\times(n+1)$ matrix composed of inputs $\mathbf{F}_N^T = [\mathbf{U}_1,\ldots,\mathbf{U}_N]$, $\mathbf{U}_i^T = [u_{i+n},\ldots,u_i]$, vector $\mathbf{Y}_N \in \Re^N$ is composed of outputs $\mathbf{Y}_N^T = [y_{n+1},\ldots,y_{n+N}]$. If the input is informative the matrix $\mathbf{R}_N$ is invertible and solution (5) exists and is unique.

The identification algorithm with the performance measured by the worst-case identification error $e_N = \sup_{\|v\|_\infty \le \varepsilon} \|\hat{G}_N - G\|_\infty$ is convergent if

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\lim_{N\to\infty} e_N = 0\,. \qquad (6)$$

Here $\|G\|_\infty = \max_{-\pi\le\omega\le\pi} |G(e^{j\omega})|$ denotes the uniform norm for the discrete-time transfer function G.

2. GENERALIZED ESTIMATES

The ill-posedness of the problem (4) under the condition (3) is irrespective to properties of matrix $\mathbf{R}_N$ and is caused by worst-case noise $v(t)$. Matrix $\mathbf{R}_N$ may be also ill-conditioned or singular because of noninformative input sequence: due to shortage of data or if the input sequence is not permanently exciting. Therefore it is desirable to construct the estimates which ensure both H_∞-convergence of algorithm as $n \to \infty$ and invertibility of improved matrix $\mathbf{R}_N$.

2.1. General case

It is offered in order to ensure both H_∞-convergence and invertibility to multiply an estimate (5) by matrix $\mathbf{C}$ and like in ridge-estimates instead of $\mathbf{R}_N$ to invert matrix $\tilde{\mathbf{R}}_N = \mathbf{R}_N + \mathbf{K}$ where $\mathbf{K}$ provides well-conditionality. Then a general class of universal linear estimates is

$$\hat{\boldsymbol{\theta}}_N = \mathbf{C}\tilde{\mathbf{R}}_N^{-1}\mathbf{S}_N\,. \qquad (7)$$

For new estimate (7) let's write down a function

$$e_N^n\left(\hat{\boldsymbol{\theta}}_N, \mathbf{C}, \mathbf{K}\right) = \sup_{\|\mathbf{v}\|_\infty \le \varepsilon} \left\|\tilde{G}_N^n\left(\hat{\boldsymbol{\theta}}_N\right)\right\|_\infty\,, \qquad (8)$$

where $\tilde{G}_N^n\left(\hat{\boldsymbol{\theta}}_N\right) = \mathbf{W}_n\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_n\right)$, $\mathbf{W}_n = \left(1\, e^{-j\omega} \ldots e^{-jn\omega}\right)$. As convergence of e_N^n is followed by convergence of e_N for all proposed algorithms bound of error e_N^n instead of bound of error e_N will be calculated.

<u>Theorem 1.</u> Let the true system be an n-th order FIR system, $\mathbf{C}$ is an $(n+1)\times(n+1)$ arbitrary nonsingular matrix, and $\mathbf{K}$ is an $(n+1)\times(n+1)$ arbitrary positive definite matrix guaranteeing invertibility of $\tilde{\mathbf{R}}_N$. Then worst-case identification error (8) of algorithm (7) is bounded as

$$\begin{aligned} e_N^n\left(\hat{\boldsymbol{\theta}}_N, \mathbf{C}, \mathbf{K}\right) &\le M_1 \frac{\varepsilon}{\sqrt{2\lambda_{\min}(\mathbf{K})}} + \\ &+ M_2\|\boldsymbol{\theta}\|_1 + M_1\sqrt{\frac{\lambda_{\max}(\mathbf{K})}{\lambda_{\min}(\mathbf{K})}}\|\boldsymbol{\theta}\|_2\,, \end{aligned} \qquad (9)$$

here $M_1 = \sqrt{\sum_{j=0}^{n}\left(\sum_{i=0}^{n}\left|(\mathbf{C})_{ij}\right|\right)^2}$, $M_2 = \max_{0\le j\le n}\sum_{i=0}^{n}\left|(\mathbf{E}-\mathbf{C})_{ij}\right|$, $\mathbf{E}$ is $(n+1)\times(n+1)$ identity matrix.

Proof. Consider the parameter estimation error

$$\begin{aligned} \hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_n &= \mathbf{C}\tilde{\mathbf{R}}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\left(\mathbf{Y}_N \pm \mathbf{F}_N\boldsymbol{\theta}_n\right) - \boldsymbol{\theta}_n \\ &= \mathbf{C}\tilde{\mathbf{R}}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N - \left(\mathbf{E} - \mathbf{C}\tilde{\mathbf{R}}_N^{-1}\mathbf{R}_N \pm C\right)\boldsymbol{\theta}_n \\ &= \mathbf{C}\tilde{\mathbf{R}}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N - (\mathbf{E}-\mathbf{C})\boldsymbol{\theta}_n - \mathbf{C}\tilde{\mathbf{R}}_N^{-1}\mathbf{K}\boldsymbol{\theta}_n\,. \end{aligned} \qquad (10)$$

Then

$$\tilde{G}_N^n\left(\hat{\boldsymbol\theta}_N\right) = \mathbf{W}_n\left(\hat{\boldsymbol\theta}_N - \boldsymbol\theta_n\right) = \mathbf{W}_n\mathbf{C}\tilde{\mathbf{R}}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N - \\ - \mathbf{W}_n(\mathbf{E}-\mathbf{C})\boldsymbol\theta_n - \mathbf{W}_n\mathbf{C}\tilde{\mathbf{R}}_N^{-1}\mathbf{K}\boldsymbol\theta_n = \overline{G}_1 + \overline{G}_2 + \overline{G}_3 \quad (11)$$

where $\overline{G}_1 = \mathbf{W}_n\mathbf{C}\tilde{\mathbf{R}}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N$, $\overline{G}_2 = -\mathbf{W}_n(\mathbf{E}-\mathbf{C})\boldsymbol\theta_n$, $\overline{G}_3 = -\mathbf{W}_n\mathbf{C}\tilde{\mathbf{R}}_N^{-1}\mathbf{K}\boldsymbol\theta_n$. Let us estimate separately the H_∞-norms of $\overline{G}_1$, $\overline{G}_2$ and $\overline{G}_3$.

$$\left\|\overline{G}_1\right\|_\infty = \max_{-\pi\le\omega\le\pi}\left|\mathbf{W}_n\mathbf{C}\tilde{\mathbf{R}}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N\right| \\ \le \max_{-\pi\le\omega\le\pi}\left\|\mathbf{W}_n\mathbf{C}\right\|_2\left\|\tilde{\mathbf{R}}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N\right\|_2 \quad (12) \\ \le M_1\left\|\left(\tfrac{1}{N}\mathbf{F}_N^T\mathbf{F}_N + \mathbf{K}\right)^{-1}\tfrac{1}{\sqrt{N}}\mathbf{F}_N^T\right\|_2\left\|\tfrac{1}{\sqrt{N}}\mathbf{V}_N\right\|_2.$$

Denote $\overline{\mathbf{F}}_N = \frac{1}{\sqrt{N}}\mathbf{K}^{-1/2}\mathbf{F}_N$. $\mathbf{K}^{-1/2}$ exists since matrix $\mathbf{K}$ is square, symmetric and positive definite. So

$$\left\|\overline{G}_1\right\|_\infty \le M_1\left\|\mathbf{K}^{-1/2}\left(\overline{\mathbf{F}}_N^T\overline{\mathbf{F}}_N + \mathbf{E}\right)^{-1}\overline{\mathbf{F}}_N^T\right\|_2\left\|\tfrac{1}{\sqrt{N}}\mathbf{V}_N\right\|_2 \\ \le M_1\left\|\mathbf{K}^{-1/2}\right\|_2\left\|\left(\overline{\mathbf{F}}_N^T\overline{\mathbf{F}}_N + \mathbf{E}\right)^{-1}\overline{\mathbf{F}}_N^T\right\|_2\left\|\tfrac{1}{\sqrt{N}}\mathbf{V}_N\right\|_2 \quad (13) \\ \le M_1\frac{\left\|\mathbf{V}_N\right\|_2}{\sqrt{2\lambda_{\min}(\mathbf{K})}\sqrt{N}},$$

where it was taken in account that $\left\|\left(\overline{\mathbf{F}}_N^T\overline{\mathbf{F}}_N + \mathbf{E}\right)^{-1}\overline{\mathbf{F}}_N^T\right\|_2 \le \frac{1}{\sqrt{2}}$. Then

$$\left\|\overline{G}_2\right\|_\infty = \max_{-\pi\le\omega\le\pi}\left|\mathbf{W}_n(\mathbf{E}-\mathbf{C})\boldsymbol\theta_n\right| \\ \le \max_{-\pi\le\omega\le\pi}\max_{0\le j\le n}\left|\left\{\mathbf{W}_n(\mathbf{E}-\mathbf{C})\right\}_j\right|\left\|\boldsymbol\theta_n\right\|_1 \quad (14) \\ \le \max_{0\le j\le n}\sum_{i=0}^{n}\left|(\mathbf{E}-\mathbf{C})_{ij}\right|\left\|\boldsymbol\theta_n\right\|_1 \le M_2\left\|\boldsymbol\theta\right\|_1,$$

$$\left\|\overline{G}_3\right\|_\infty = \max_{-\pi\le\omega\le\pi}\left|\mathbf{W}_n\mathbf{C}\tilde{\mathbf{R}}_N^{-1}\mathbf{K}\boldsymbol\theta_n\right| \le \max_{-\pi\le\omega\le\pi}\left\|\mathbf{W}_n\mathbf{C}\right\|_2 \\ \times\left\|\tilde{\mathbf{R}}_N^{-1}\mathbf{K}\boldsymbol\theta_n\right\|_2 \le M_1\left\|\mathbf{K}^{-1/2}\right\|_2\left\|\left(\overline{\mathbf{F}}_N^T\overline{\mathbf{F}}_N + \mathbf{E}\right)^{-1}\right\|_2 \quad (15) \\ \times\left\|\mathbf{K}^{1/2}\boldsymbol\theta_n\right\|_2 \le M_1\sqrt{\frac{\lambda_{\max}(\mathbf{K})}{\lambda_{\min}(\mathbf{K})}}\left\|\boldsymbol\theta\right\|_2.$$

Here the inequality $\left\|\left(\overline{\mathbf{F}}_N^T\overline{\mathbf{F}}_N + \mathbf{E}\right)^{-1}\right\|_2 \le 1$ was used. As $\sup_{\|v\|_\infty\le\varepsilon}\left\|\mathbf{V}_N\right\|_2 \le \varepsilon$ we come to the final result. ▲

The estimate (9) does not depend on number of data N and holds true as $N\to\infty$. Each of its terms must be restricted as $n\to\infty$. For the first term it is enough

$$\frac{M_1}{\sqrt{\lambda_{\min}(\mathbf{K})}} = O(1),\ n\to\infty. \quad (16)$$

Let's consider an asymptotic behavior of M_2

$$M_2 = \max_{0\le j\le n}\sum_{i=0}^{n}\left|(\mathbf{E}-\mathbf{C})_{ij}\right| \le \max_{0\le j\le n}\left(\sum_{i=0}^{n}\left|(\mathbf{C})_{ij}\right| + 1\right) = \\ = \max_{0\le j\le n}\sum_{i=0}^{n}\left|(\mathbf{C})_{ij}\right| + 1 = M_1 + 1. \quad (17)$$

So $M_1 = O(1)$ is followed by $M_2 = O(1)$. For stable plants $\left\|\boldsymbol\theta\right\|_1 = \sum_{k=0}^{\infty}\left|g_k\right| < \infty$ and therefore the second term is restricted as $n\to\infty$. $\left\|\boldsymbol\theta\right\|_1 = O(1)$ as $n\to\infty$ is followed by $\left\|\boldsymbol\theta\right\|_2 = O(1)$. As to the third term one must require

$$\sqrt{\frac{\lambda_{\max}(\mathbf{K})}{\lambda_{\min}(\mathbf{K})}} = O(1),\ n\to\infty. \quad (18)$$

Finally we have H_∞-convergence conditions for stable plants

$$M_1 = O(1),\ \frac{1}{\sqrt{\lambda_{\min}(\mathbf{K})}} = O(1),\ \sqrt{\lambda_{\max}(\mathbf{K})} = O(1) \quad (19)$$

Sequence $\bar{c}_j = \sum_{i=0}^{n}\left|(\mathbf{C})_{ij}\right|$ must be square summable, for example $\bar{c}_j = j^{-(0,5+\alpha)}$, α>0; $\bar{c}_j = \rho^j$, 0<ρ<1. The diagonal matrix $\mathbf{C} = \mathrm{diag}\left\{1,1,2^{-1},\ldots,n^{-1}\right\}$ meets the given requirements. The second and third terms of (19) can be fulfilled with matrix $\mathbf{K} = k\mathbf{E}$ by means of fitting scalar coefficient $k > 0$.

2.2. Nonsingular case

Let us consider a typical case when the finite problem (4) is well-posed. Then for existent matrix $\mathbf{R}_N^{-1}$ we'll define the class of regularized estimates

$$\hat{\boldsymbol\theta}_N = \mathbf{C}\mathbf{R}_N^{-1}\mathbf{S}_N. \quad (20)$$

Theorem 2. Let the true system be an n-th order FIR system, $\mathbf{C}$ is an $(n+1)\times(n+1)$ arbitrary nonsingular matrix. Then worst-case error of the identification algorithm (20) is bounded as

$$e_N^n\left(\hat{\boldsymbol\theta}_N, \mathbf{C}\right) \le M_1\frac{\varepsilon}{\sqrt{\lambda_{\min}\left(\mathbf{R}_N\right)}} + M_2\left\|\boldsymbol\theta\right\|_1. \quad (21)$$

Proof. The error of the transfer function estimate is

$$\tilde{G}_N^n\left(\hat{\boldsymbol\theta}_N\right) = \mathbf{W}_n\left(\hat{\boldsymbol\theta}_N - \boldsymbol\theta_n \pm \mathbf{C}\boldsymbol\theta_n\right) = \mathbf{W}_n \\ \times\left(\mathbf{C}_N\mathbf{R}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N - (\mathbf{E}-\mathbf{C}_N)\boldsymbol\theta_n\right) = G_1 + G_2, \quad (22)$$

where $G_1 = \mathbf{W}_n\mathbf{C}\mathbf{R}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N$, $G_2 = -\mathbf{W}_n(\mathbf{E}-\mathbf{C})\boldsymbol\theta_n$. For G_1 we have

$$\left\|G_1\right\|_\infty = \max_{-\pi\le\omega\le\pi}\left|\mathbf{W}_n\mathbf{C}\mathbf{R}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N\right| \le \max_{-\pi\le\omega\le\pi}\left\|\mathbf{W}_n\mathbf{C}\right\|_2 \times\left\|\mathbf{R}_N^{-1}\tfrac{1}{N}\mathbf{F}_N^T\mathbf{V}_N\right\|_2 \le M_1\frac{\left\|\mathbf{V}_N\right\|_2}{\sqrt{\lambda_{\min}(\mathbf{R}_N)}\sqrt{N}}. \quad (23)$$

The bound of $\left\|G_2\right\|_\infty$ coincides with (14). As far as $\sup_{\|v\|_\infty\le\varepsilon}\left\|\mathbf{V}_N\right\|_2 \le \varepsilon$ we come to the result (21). ▲

Each term of (21) must be restricted as $n\to\infty$. For any fixed n and bounded input the convergence is ensured if

$$\lim_{N\to\infty}\lambda_{\min}(\mathbf{R}_N) = a, \quad (24)$$

where $a>0$ depends on n in general. Then the finiteness of the first term is achieved when

$$\frac{M_1}{\sqrt{a}} = O(1) \text{ as } n\to\infty. \quad (25)$$

The second term converges if $M_2 = O(1)$ and matrix $\mathbf{C}$ is selected like in algorithm (7). So for H_∞-convergence of the algorithm (20) it is required

$$\sqrt{a} = O(1),\ M_1 = O(1),\ M_2 = O(1) \text{ as } n\to\infty. \quad (26)$$

2.3. H_∞ - convergent ridge estimates

Let's assume that in the estimate (7) $\mathbf{C}$ is identity matrix and define the conditions which guarantee the H_∞-convergence and correctness when data is not completely informative. In this case (7) is reduced to

$$\hat{\boldsymbol{\theta}}_N = \tilde{\mathbf{R}}_N^{-1}\mathbf{S}_N. \quad (27)$$

For stable plants the H_∞-convergence condition for (27) can be obtained as a corollary of the theorem 1.

Corollary 1. Let the true system be an n-th order FIR system, $\mathbf{K}$ is an $(n+1)\times(n+1)$ arbitrary positive semidefinite matrix that provides the invertibility of $\tilde{\mathbf{R}}_N$. Then for any $\varepsilon>0$ the worst-case identification error of algorithm (27) is bounded as

$$e_N^n\left(\hat{\boldsymbol{\theta}}_N,\mathbf{K}\right) \le \overline{M}_1\left(\varepsilon + \left\|\mathbf{K}^{1/2}\boldsymbol{\theta}_n\right\|_2\right), \quad (28)$$

$$\overline{M}_1 = \sqrt{\sum_{j=0}^{n}\left(\sum_{i=0}^{n}\left|\left(\mathbf{K}^{-1/2}\right)_{ij}\right|\right)^2}, \quad (29)$$

and H_∞-converges as $n\to\infty$ if it is fulfilled

$$\overline{M}_1 = O(1),\ \left\|\mathbf{K}^{1/2}\boldsymbol{\theta}_n\right\|_2 = O(1) \text{ as } n\to\infty. \quad (30)$$

The first term in (30) means that matrix $\mathbf{K}^{-1/2}$ is similar to the matrix $\mathbf{C}$ from (7) and (20) and meets the same requirement. For diagonal matrix $\mathbf{K} = \mathrm{diag}\{k_0,k_1,\ldots,k_n\}$, $k_i>0$ it is enough to put $k_j = o(j)$, $j\to\infty$, for example $k_j = j^{1+\alpha}$, $\alpha>0$; or $k_j = \rho^{-j}$, $0<\rho<1$. The second term in (30) restricts the class of identifiable systems up to the systems with bounded weighted norm

$$\lim_{n\to\infty}\left\|\mathbf{K}^{1/2}\boldsymbol{\theta}_n\right\|_2 = \lim_{n\to\infty}\sqrt{\sum_{j=0}^{n}k_j g_j^2} < \infty. \quad (31)$$

So if $k_j = j^{1+\alpha}$, $0<\alpha<1$, $\forall j$ then method (27) is admissible for strictly stable plants: $\lim_{n\to\infty}\sum_{j=0}^{n} j\left|g_j\right| < \infty$.

3. RECURRENT LS ALGORITHM (RLSA) WITH ADAPTIVE REGULARIZATION

There are recurrent forms of regularized LS algorithms (see e.g. Arnold, 1972) but they don't adapt regularization parameters to entry of new data. In this section a generalized RLSA with adaptive to incoming data regularization is described. Let us consider the following estimate

$$\hat{\boldsymbol{\theta}}_N^\delta = \left(\mathbf{F}_N^T\mathbf{F}_N + \delta_N\boldsymbol{\Xi}\right)^{-1}\mathbf{F}_N^T\mathbf{Y}_N,\ \delta_N>0, \quad (32)$$

$$\boldsymbol{\Xi} = \mathrm{diag}[\xi_0,\xi_1,\ldots,\xi_n],\ \xi_i>0,\ i=0,1,\ldots,n \quad (33)$$

based on (7) obtained as $\hat{\boldsymbol{\theta}}_N^\delta = \arg\min_\theta\{(\mathbf{Y}_N - \mathbf{F}_N\boldsymbol{\theta})^2 + \delta_N\boldsymbol{\theta}^T\boldsymbol{\Xi}\boldsymbol{\theta}\}$. A regularization parameter δ_N will be recurrently adjusted not only with respect to ε but with respect to incoming data also. We define

$$\mathbf{P}_N^{-1} = \mathbf{F}_N^T\mathbf{F}_N + \delta_N\boldsymbol{\Xi}. \quad (34)$$

Lemma 1. The vector $\hat{\boldsymbol{\theta}}_N^\delta = \arg\min_\theta\{(\mathbf{Y}_N - \mathbf{F}_N\boldsymbol{\theta})^2 + \delta_N\boldsymbol{\theta}^T\boldsymbol{\Xi}\boldsymbol{\theta}\}$ and matrix $\mathbf{P}_N^{-1}$ obey the recurrences

$$\hat{\boldsymbol{\theta}}_N^\delta = \left(\mathbf{E} - \Delta\delta_N\mathbf{P}_N\boldsymbol{\Xi}\right)\hat{\boldsymbol{\theta}}_{N-1}^\delta + \mathbf{P}_N x_N\left(y_N - \mathbf{x}_N^T\hat{\boldsymbol{\theta}}_{N-1}^\delta\right), \quad (35)$$

$$\mathbf{P}_N^{-1} = \mathbf{P}_{N-1}^{-1} + \mathbf{x}_N\mathbf{x}_N^T + \Delta\delta_N\boldsymbol{\Xi}, \quad (36)$$

where $\Delta\delta_N = \delta_N - \delta_{N-1}$.

Proof. Let's divide matrix $\mathbf{F}_N$ and vector $\mathbf{Y}_N$ as follows

$$\mathbf{F}_N^T = \left[\mathbf{F}_{N-1}^T \vdots\, \mathbf{x}_N\right],\ \mathbf{Y}_N^T = \left[\mathbf{Y}_{N-1}^T \vdots\, y_N\right], \quad (37)$$

where $\mathbf{F}_{N-1}$ is $(N-1)\times(n+1)$ matrix, $\mathbf{x}_N\in\Re^{n+1}$, $\mathbf{Y}_{N-1}\in\Re^{N-1}$. Then the matrix $\mathbf{P}_N^{-1}$ may be written as

$$\mathbf{P}_N^{-1} = \mathbf{P}_{N-1}^{-1} + \mathbf{x}_N\mathbf{x}_N^T + (\delta_N - \delta_{N-1})\boldsymbol{\Xi} = \mathbf{P}_{N-1}^{-1} + \mathbf{x}_N\mathbf{x}_N^T + \Delta\delta_N\boldsymbol{\Xi}.$$

Hence, we have

$$\hat{\boldsymbol{\theta}}_N^\delta = \mathbf{P}_N \mathbf{F}_N^T \mathbf{Y}_N = \mathbf{P}_N \left(\mathbf{P}_{N-1}^{-1}\hat{\boldsymbol{\theta}}_{N-1}^\delta + \mathbf{x}_N y_N\right) =$$
$$= \mathbf{P}_N \left(\left(\mathbf{P}_N^{-1} - \mathbf{x}_N \mathbf{x}_N^T - \Delta\delta_N \boldsymbol{\Xi}\right)\hat{\boldsymbol{\theta}}_{N-1}^\delta + \mathbf{x}_N y_N\right) = \quad (38)$$
$$= \left(\mathbf{E} - \Delta\delta_N \mathbf{P}_N \boldsymbol{\Xi}\right)\hat{\boldsymbol{\theta}}_{N-1}^\delta + \mathbf{P}_N \mathbf{x}_N \left(y_N - \mathbf{x}_N^T \hat{\boldsymbol{\theta}}_{N-1}^\delta\right).$$

▲

Let $\mathbf{P}(*,i)$ and $\mathbf{P}(i,*)$ are the i-th column and the i-th row of matrix $\mathbf{P}$ respectively; $\mathbf{I}_m = \lfloor 0 \ \ldots \ 1_{(m)} \ \ldots \ 0 \rfloor$, $m = 0,\ldots,n$; $\mathbf{P}_{N-1,k}^{-1} = \mathbf{P}_{N-1}^{-1} + \Delta\delta_N \sum_{m=0}^{k} \xi_m \mathbf{I}_m^T \mathbf{I}_m$, $k = 0,\ldots,n$;

$$\Delta\mathbf{P}_{N-1,m} = \frac{\xi_m \mathbf{P}_{N-1,m-1}(*,m)\mathbf{P}_{N-1,n-1}(m,*)}{1 + \Delta\delta_N \xi_m \mathbf{P}_{N-1,m-1}(m,m)}, \ m = 1,\ldots,n;$$

$$\Delta\mathbf{P}_{N-1,0} = \frac{\xi_0 \mathbf{P}_{N-1}(*,0)\mathbf{P}_{N-1}(0,*)}{1 + \Delta\delta_N \xi_0 \mathbf{P}_{N-1}(0,0)}.$$

<u>Theorem 3.</u> The vector $\hat{\boldsymbol{\theta}}_N^\delta = \arg\min_\theta \left\{(\mathbf{Y}_N - \mathbf{F}_N \boldsymbol{\theta})^2 + \delta_N \boldsymbol{\theta}^T \boldsymbol{\Xi} \boldsymbol{\theta}\right\}$ and matrix $\mathbf{P}_N$ obey the recurrences

$$\hat{\boldsymbol{\theta}}_N^\delta = \hat{\boldsymbol{\theta}}_{N-1}^\delta + \frac{\mathbf{P}_{N-1,n}\mathbf{x}_N}{1 + \mathbf{x}_N^T \mathbf{P}_{N-1,n}\mathbf{x}_N}\left(y_N - \mathbf{x}_N^T \hat{\boldsymbol{\theta}}_{N-1}^\delta\right) - \Delta\delta_N \left(\mathbf{P}_{N-1,n} - \frac{\mathbf{P}_{N-1,n}\mathbf{x}_N \mathbf{x}_N^T \mathbf{P}_{N-1,n}}{1 + \mathbf{x}_N^T \mathbf{P}_{N-1,n}\mathbf{x}_N}\right)\boldsymbol{\Xi}\hat{\boldsymbol{\theta}}_{N-1}^\delta, \quad (39)$$

$$\mathbf{P}_{N-1,i} = \mathbf{P}_{N-1,i-1} - \Delta\delta_N \Delta\mathbf{P}_{N-1,i}, \ i = n,\ldots,1$$
$$\mathbf{P}_{N-1,0} = \mathbf{P}_{N-1} - \Delta\delta_N \Delta\mathbf{P}_{N-1,0}. \quad (40)$$

Proof. Diagonal matrix $\boldsymbol{\Xi}$ may be represented as a sum of vector products

$$\boldsymbol{\Xi} = \sum_{m=0}^{n} \xi_m \mathbf{I}_m^T \mathbf{I}_m \ . \quad (41)$$

For $\mathbf{P}_N$, $\mathbf{P}_{N-1,i}$, $i = 0,\ldots,n$ we use the matrix inversion lemma

$$\mathbf{P}_N = \left(\mathbf{P}_{N-1,n}^{-1} + \mathbf{x}_N \mathbf{x}_N^T\right)^{-1} = \mathbf{P}_{N-1,n} - \mathbf{P}_{N-1,n}\frac{\mathbf{x}_N \mathbf{x}_N^T}{1 + \mathbf{x}_N^T \mathbf{P}_{N-1,n}\mathbf{x}_N}\mathbf{P}_{N-1,n} \quad (42)$$

$$\mathbf{P}_{N-1,n} = \left(\mathbf{P}_{N-1,n-1}^{-1} + \Delta\delta_N \xi_n \mathbf{I}_n^T \mathbf{I}_n\right)^{-1} = \mathbf{P}_{N-1,n-1} - \Delta\delta_N \Delta\mathbf{P}_{N-1,n}, \quad (43)$$

$$\mathbf{P}_{N-1,i} = \mathbf{P}_{N-1,i-1} - \Delta\delta_N \Delta\mathbf{P}_{N-1,i}, \ i = n,\ldots,1. \quad (44)$$

$$\mathbf{P}_{N-1,0} = \left(\mathbf{P}_{N-1}^{-1} + \Delta\delta_N \xi_0 \mathbf{I}_0^T \mathbf{I}_0\right)^{-1} = \mathbf{P}_{N-1} - \Delta\delta_N \Delta\mathbf{P}_{N-1,0}. \quad (45)$$

Then we calculate

$$\mathbf{P}_N \mathbf{x}_N = \mathbf{P}_{N-1,n}\mathbf{x}_N - \frac{\mathbf{P}_{N-1,n}\mathbf{x}_N \mathbf{x}_N^T \mathbf{P}_{N-1,n}}{1 + \mathbf{x}_N^T \mathbf{P}_{N-1,n}\mathbf{x}_N}\mathbf{x}_N = \frac{\mathbf{P}_{N-1,n}\mathbf{x}_N}{1 + \mathbf{x}_N^T \mathbf{P}_{N-1,n}\mathbf{x}_N} \quad (46)$$

and substitute (46) and (42) in (35). ▲

Complete recurrence of matrix $\mathbf{P}_N$ can be written as

$$\mathbf{P}_N = \mathbf{P}_{N-1} - \Delta\delta_N \sum_{m=0}^{n} \Delta\mathbf{P}_{N-1,m} - \frac{\mathbf{P}_{N-1,n}\mathbf{x}_N \mathbf{x}_N^T \mathbf{P}_{N-1,n}}{1 + \mathbf{x}_N^T \mathbf{P}_{N-1,n}\mathbf{x}_N}. \quad (47)$$

The expression (47) shows that the revising of the regularization parameter δ_N when N increases has resulted in complication of recurrent algorithm. If $\delta_N = \delta_{N-1}$ and $\mathbf{P}_{N-1,n} = \mathbf{P}_{N-1}$ the standard RLSA with regularization (Arnold, 1972) is obtained. The offered recurrent procedure on step N contains $n+2$ identical embedded procedures. For diagonal matrix $\boldsymbol{\Xi}$ owing to expansion (41) it is possible to reduce computational burden up to $(n+1)^2$ operations. For comparison $(n+1)^3$ operations are required in nonrecurrent algorithm.

The parameter δ_N is sensitive to data entry only when we have a shortage of them. When available data approach to informative ones the regularization parameter δ_N varies slowly and asymptotically verges towards steady-state value. Then standard RLSA can be used. It is the way to spare computational resource at the large N and to obtain good enough estimates at small N.

When first $n+1$ inputs are accumulated and $Y_1 = y_{n+1}$ is obtained it is possible to evaluate parameter vector $\hat{\boldsymbol{\theta}}_1^\delta$ having initial values $\hat{\boldsymbol{\theta}}_0$ and $\mathbf{P}_0$. An explicit expressions of vectors $\hat{\boldsymbol{\theta}}_N^\delta$ and matrices $\mathbf{P}_N$ subject to initial data are established in the following theorem.

<u>Theorem 4.</u> For arbitrary $\hat{\boldsymbol{\theta}}_N^\delta$ and $\mathbf{P}_N$ of system (39) (or (35)) in view of (4) it is satisfied

$$\mathbf{P}_N^{-1}\left(\hat{\boldsymbol{\theta}}_N^\delta - \boldsymbol{\theta}_n\right) = \mathbf{P}_{N-1}^{-1}\left(\hat{\boldsymbol{\theta}}_{N-1}^\delta - \boldsymbol{\theta}_n\right) + \mathbf{x}_N v_N - \Delta\delta_N \boldsymbol{\Xi}\boldsymbol{\theta}_n, \quad (48)$$

$$\mathbf{P}_N^{-1} = \mathbf{P}_0^{-1} + \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T + (\delta_N - \delta_0)\boldsymbol{\Xi}, \quad (49)$$

$$\hat{\boldsymbol{\theta}}_N^\delta = \left(\mathbf{P}_0^{-1} + \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T + (\delta_N - \delta_0)\boldsymbol{\Xi}\right)^{-1}\left(\mathbf{P}_0^{-1}\hat{\boldsymbol{\theta}}_0 + \sum_{i=1}^{N} \mathbf{x}_i y_i\right). \quad (50)$$

The matrix $\mathbf{P}_N$ for all N is symmetric and positive definite if $\mathbf{P}_0^T = \mathbf{P}_0 > 0$. Let's select $\mathbf{P}_0$ as follows

$$\mathbf{P}_0^{-1} = \delta_0 \boldsymbol{\Xi}, \ \boldsymbol{\Xi} > 0, \ \delta_0 > 0. \quad (51)$$

Hence the recurrent matrix $\mathbf{P}_N^{-1}$ does not depend on initial conditions and looks like $\mathbf{P}_N^{-1} = \mathbf{F}_N^T \mathbf{F}_N + \delta_N \boldsymbol{\Xi}$. Here the item $\delta_N \boldsymbol{\Xi}$ can be considered as initial condition of $\mathbf{P}_N^{-1}$ at the moment N. To select $\hat{\boldsymbol{\theta}}_0$ we substitute $\mathbf{P}_0^{-1} = \delta_0 \boldsymbol{\Xi}$ in (50) and obtain

$$\hat{\boldsymbol{\theta}}_N^{\delta} = \left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T + \delta_N \, \boldsymbol{\Xi} \right)^{-1} \left(\delta_0 \boldsymbol{\Xi} \hat{\boldsymbol{\theta}}_0 + \sum_{i=1}^{N} \mathbf{x}_i y_i \right). \quad (52)$$

The additional term $\left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T + \delta_N \, \boldsymbol{\Xi} \right)^{-1} \delta_0 \boldsymbol{\Xi} \hat{\boldsymbol{\theta}}_0$ distinguishes the obtained recurrent estimate $\hat{\boldsymbol{\theta}}_N^{\delta}$ from a nonrecurrent one. It is recommended to select small δ_0 and arbitrary $\hat{\boldsymbol{\theta}}_0$ with small norm or to select $\hat{\boldsymbol{\theta}}_0$ =0.

4. EXAMPLE

The method of section 3 was computationally tested. The estimation errors of regularized RLSA with constant and varying regularization parameter were compared. The table 1 holds the outcome of such experiment. The model of second order $G(z) = \dfrac{0.8z + 0.4737}{(z-0.9)(z-0.6)}$ is selected. The first $n+1$ =81 coefficients of expansion in series (3) were evaluated by standard regularized RLSA and by proposed regularized RLSA for optimal and quasioptimal regularization parameters (Tichonov and Arsenin, 1977) and for quasioptimal stabilizer $\beta = 0.8$ (Gubarev and Panova, 1999). The noise $v_i = \varepsilon \cdot \operatorname{sign} y_i$, $i = n+1, n+2, \ldots, n+N$, $\varepsilon = 0.5$, pseudorandom binary inputs of degree $p = 8$ and the measuring data are used. For sequence N = 50, 60, 70, 80 the Euclidean norms of parameter estimation errors $\hat{\boldsymbol{\theta}}_N^{\delta_N}$ at varying regularization parameter and $\hat{\boldsymbol{\theta}}_N^{\delta}$ at constant one are obtained

$$\sigma_0(N) = \max \left\| \hat{\boldsymbol{\theta}}_N^{\delta_N} - \boldsymbol{\theta}_n \right\|_2, \; \sigma_1(N) = \max \left\| \hat{\boldsymbol{\theta}}_N^{\delta} - \boldsymbol{\theta}_n \right\|_2 . (53)$$

The tables show that optimal parameter δ_o diminishes when data renew. The optimal parameter δ_o is properly approximated by quasioptimal parameter $\delta_{q.o}$. Hence having selected a fixed parameter δ (even quasioptimal) for initial $N = N_0$ the subsequent comes out worth than for algorithm (32). The proposed adaptive algorithm yields in experiments tenfold decrease of the estimation error in comparison with standard regularized RLSA.

Table 1. $\sigma_0(N)$ for $\delta_{q.o}$, δ_o and $\sigma_1(N)$ for $\delta = 10^{-3}, 10^{-5}, 10^{-7}$

	N=50	N=60	N=70	N=80
$\delta = 10^{-3}$	0.743	0.645	0.616	0.577
$\delta = 10^{-5}$	0.828	0.608	0.536	0.532
$\delta = 10^{-7}$	1.088	0.795	0.608	0.564
$\delta_{q.o}$	0.828	0.608	0.550	0.516
δ_o	0.743	0.602	0.536	0.516

Table 2. $\delta_{q.o}$ and δ_o for N=50,60,70,80

	N=50	N=60	N=70	N=80
$\lg \delta_{q.o}$	-5	-5	-6	-6
$\lg \delta_o$	-3	-4	-5	-6

5. CONCLUSION

In the present paper using regularization methods the generalized LS algorithms are obtained and their properties are investigated. It was shown how the LS technique should be generally modified in order that the approximate IIR-models may be identified even in the case of data lack and in addition H_∞-convergence under bounded unknown noise should be provided. Regularized recurrent identification with adaptive regularization parameter is apparently general and effective method which is indifferent to whether the problem under consideration is well-posed or not. Automatically we obtain trivial regularization parameter when available information provides correctness of the problem setting.

ACKNOWLEDGEMENT

The authors are grateful to Professor Petros A. Ioannou for discussions and fruitful cooperation in this work.

REFERENCES

Akcay H., Hjalmarson H. (1994). The least-squares identification of FIR systems subject to worst-case noise. *Systems and Control Letters*, **23**, 329–338.

Akcay H., Khargonekar P.P. (1993). The least squares algorithm, parametric system identification, and bounded noise. *Automatica*, **29**, 1535–1540.

Arnold A. (1972). *Regression and the Moor-Penrose Pseudoinverse*. Academic Press, New York and London.

Gubarev V.F., Panova N.V. (1999). Properties of the regularized least-squares method in transfer function identification. In: *Proc of the 14-th IFAC World Congress*, **H**, pp. 433–438. Bejing, China.

Hyotyniemi, H. (1996). Regularization of parameter estimation. In: *Proc of the 13-th Trennial IFAC World Congress*, pp. 227–232. San Francisco, USA.

Ljung L. (1987). *System identification: Theory for the user*. Prentice-Hall, Englewood Cliffs, New Jersey.

Tichonov A.N., Arsenin V.Ya. (1977). *Solutions methods of ill-posed problems*, Winston & Sons: Washington DC.