Effect of adaptation rate value on convergence of gradient descent-based identification methods

A.I. Glushchenko, V.A. Petrov and K.A. Lastochkin
Stary Oskol Technological Institute n.a. A.A. Ugarov (branch of National University of Science and Technology “MISIS”), Stary Oskol
E-mail: a.glushchenko@sf-misis.ru

Abstract. In this study, one of the main problems of practical implementation of adaptive control and identification systems is considered. This is the problem of choice of an adaptation rate value to provide the required convergence rate of linear regression parameters. The most widely applied method to identify such regression parameter values is the gradient descent (GD) one. But the adaptation rate is usually chosen manually for it and kept constant in the course of the plant functioning. The recursive least-squares (RLS) method can be used instead of GD as it provides the developer with the equation to adjust the adaptation rate automatically, taking into account the current value of the regressor. Therefore, the aim of this research is to compare the known properties of GD and RLS methods and demonstrate the capability of the RLS to provide a constant convergence rate under the condition of the time-varying regressor.

1. Introduction
Somehow or other, most of the known methods of the direct and indirect adaptive control can come down to the task of estimating parameters of a linear regression $y = \theta^T \omega$ [1]. Here $\theta$ is a vector of parameters to be estimated, $\omega$ is a regressor of the dependence under consideration. In this case, the estimation (adaptive) laws are usually derived using the gradient descent, speed-gradient or the second Lyapunov methods [1, 2].

A well-known problem of such estimation laws is that the exponential convergence of the regression parameters to their ideal values $\theta$ is guaranteed only if the condition of the persistent excitation (PE) of the regressor is met [1].

The regressor $\omega(t) \in \mathbb{R}^n$ is persistently exciting ($\omega \in \text{PE}$) if $T > 0$ and $\alpha > 0$ exist, such that

$$\int_t^{t+T} \omega(\tau)\omega^T(\tau) \geq \alpha I, \forall t \geq 0.$$ (1)

In practice, such a condition can be met if the setpoint value (or the control action one) is permanently changed, for instance, because it includes some kind of periodic components such as sine or cosine functions. However, their application may cause some negative effects like excessive wear of actuators of the real industrial plants (production units).

This being mentioned, nowadays results have been obtained [3, 4], which ease the persistent excitation condition in some ways. They are based on: 1) application of the filtration of the above-mentioned signals $y$ and $\omega$, 2) the dynamic and memory regressor extension methods (DRE and MRE) including integration operation. All of them allow one to ease the PE strict condition to the regressor initial excitation (IE) one, which is met by almost all types of the setpoint signals.
Another problem of the considered identification loops, which is less covered in the books and research papers, is what value of the adaptation rate $\Gamma$ to choose. Its essence is that, as it is proved in [5], there is an optimal value of the parameter $\Gamma$ for the current value of the regressor $\omega$ of the function $y$, even when the PE condition is fulfilled. This optimal value provides the maximum possible convergence rate of the parameter error of the identification process. This means that such rate of convergence of the regression parameters to their ideal values cannot be made arbitrarily high. In addition to this, each new value of the regressor $\omega_{new}$ implies that we need to find a new optimal value of the adaptation rate $\Gamma_{new}$.

As it is shown in [8], the methods of filtration and extension of the regressor, which were proposed more than 40 years ago [6, 7], solve the first of the above-mentioned problems and allow one to make the convergence rate arbitrarily high as a result of the adaptation rate value improvement. However, the $\Gamma$ value is usually chosen once only and remains fixed. This does not allow keeping the required convergence rate of the parameter error when the regressor is time-varying.

So, methods of filtration and extension of the regressor do not solve the second problem - to keep the pre-defined rate of regression parameters convergence to their ideal values under the condition of the time-varying regressor. From the practical point of view, the one would like to have an estimation law to identify the ideal regression parameters, which meets the stated requirement. To the best of the authors' knowledge, such estimation laws have not been proposed yet.

It should be noted that, in most cases, the loop of regression parameters adjustment is based on the classic gradient descent method and its known modifications. In contrast to this, in this paper, we propose to pay attention to the recursive least-squares (RLS) method, which can be used instead of the gradient one. It differs from the classic gradient loop as it includes the law of adjustment of the adaptation rate $\Gamma$. As it will be shown in this study, this allows the recursive least-squares method to solve partially the above-mentioned problem of choosing the regression parameters adaptation rate.

Focusing on the problem of choice of the adaptation rate, the aim of this research is to compare the convergence rate of the parameter error of the estimation loops based on the following methods: 1) the classic gradient descent, 2) the gradient one with the regressor extension using the MRE procedure, 3) the recursive least-squares method with the exponential forgetting factor.

2. Main Results
Let the following identification loops to estimate the unknown regression parameters be described with mathematical rigor: gradient ones with and without the regressor extension using the MRE procedure, and the one based on the recursive least-squares method.

Let the regression (2) be introduced, which is used to approximate the ideal function $y(t) = \theta^T \omega(t)$.

$$\hat{y}(t) = \hat{\theta}^T(t) \omega(t),$$

where $\omega \in \mathbb{R}^n$ is a known regressor vector, $\theta \in \mathbb{R}^n$ is a vector of the unknown constant parameters, $\hat{\theta} \in \mathbb{R}^n$ is a vector of the adjustable parameters. The adaptive (estimation) laws are to be derived for the vector $\hat{\theta}$ with the help of all the above-mentioned methods.

The equation of error between the adjustable (2) and ideal $y(t)$ functions is written as (3).

$$\hat{y}(t) - y(t) = \hat{y}(t) - \hat{\theta}^T(t) \omega(t) - \theta^T \omega(t) = \hat{\theta}^T(t) \omega(t),$$

where $\hat{\theta} \in \mathbb{R}^n$ is a parameter error vector.

First of all, the gradient descent method will be considered. The adaptive laws to adjust the regression parameters will be presented.

2.1. Gradient-based estimation loop
Let the objective criterion (4) be introduced to derive the equations of the gradient descent based estimation loop using the error equation (3) [1].
The gradient of the objective function with respect to the regression parameters is written as (5).

\[ \nabla Q_{\theta^T} = \frac{\partial Q_{\theta^T}}{\partial \theta^T}(t) = \frac{\partial \tilde{y}(t)}{\partial \theta^T}(t) \tilde{y}^T(t) = \omega(t) \omega^T(t) \hat{\theta}(t) \]  

The law of estimation of the ideal parameters of the regression is written as (6) in accordance with the gradient descent method.

\[ \dot{\hat{\theta}}(t) = -\Gamma \nabla Q_{\theta^T} = -\Gamma \omega(t) \omega^T(t) \hat{\theta}(t) \]  

As the values of the ideal parameters \( \theta \) are fixed, then the differential equation of the parameter error \( \hat{\theta} \) is written as (7).

\[ \dot{\hat{\theta}}(t) = -\Gamma \omega(t) \omega^T(t) \hat{\theta}(t) \]  

The solution of the differential equation (7) is presented as (8).

\[ \hat{\theta}(t) = e^{-\int_{0}^{t} \omega(\tau) \omega^T(\tau) d\tau} \hat{\theta}(0) \]  

If the estimation law (7) is used, then the behavior of the parameter error has the following properties:

1) if the PE condition is met, then the parameter error (8) converges to zero exponentially [1];
2) there is an optimal value of the adaptation rate for the current value of the regressor \( \omega \), which provides the maximum possible rate of parameter error (8) convergence to zero [5, 8].

However, the value of the adaptation rate itself is chosen manually and just once. In most cases, it remains fixed in the course of the estimation loop functioning.

Next, the gradient estimation loop with the extension of the regressor according to the MRE procedure is considered.

2.2. Gradient-based estimation loop with MRE

Following the MRE procedure, the ideal function \( y(t) = \theta^T \omega(t) \) is extended with the help of its multiplication by the regressor \( \omega^T(t) \):

\[ y(t) \omega^T(t) = \theta^T \omega(t) \omega^T(t). \]  

Let the following aperiodic filter be introduced:

\[ \dot{u}_f = -\beta u_f + u, \quad u_f(0) = 0, \]  

where \( \beta > 0 \) is the memory parameter, \( u_f \) and \( u \) are the filter output and input respectively.

The equation (9) is passed through the filter (10) to obtain:

\[ \Omega(t) = e^{-\beta t} \int_{t_0}^{t} e^{\beta \tau} \omega(\tau) \omega^T(\tau) d\tau; \quad Y(t) = e^{-\beta t} \int_{t_0}^{t} e^{\beta \tau} y(\tau) \omega^T(\tau) d\tau. \]  

Using (11), the equation (9) is rewritten as:

\[ Y(t) = \theta^T \Omega(t) \]  

Applying the same technique as (4)-(7), the gradient-based estimation loop for the regression (12) is written as:
\[ \dot{\hat{\theta}}(t) = \hat{\theta}(t) = -\Gamma \Theta(t) \Theta^T(t) \dot{\hat{\theta}}(t) \quad (13) \]

The solution of the differential equation (13) is shown as (14).

\[ \hat{\theta}(t) = e^{-\int_0^t \Gamma(\tau) \Theta^T(\tau) d\tau} \hat{\theta}(0) \quad (14) \]

If the estimation law (13) is used, then the behavior of the parameter error has the following properties:
1) if the PE condition is met, then the parameter error (14) converges to zero exponentially [1];
2) the convergence rate of the parameter error (14) can be made arbitrarily high as a result of the adaptation rate \( \Gamma \) value increase [5, 8].

The next step is to consider the procedure of the estimation loop development using the recursive least-squares method.

2.3. Estimation loop based on recursive least-squares method

The objective criterion is chosen as (15) to develop the estimation loop on the basis of the recursive least-squares method with the exponential forgetting factor [1, 9].

\[ J = 0.5 \int_0^t e^{-\lambda(t-\tau)} \tilde{y}(\tau) \tilde{y}^T(\tau) d\tau, \quad (15) \]

where \( \lambda \) is the exponential forgetting factor.

Then the least-squares estimates of the ideal values of the regression parameters \( \theta \) are to be derived from the equation (16) using (15).

\[ \nabla \hat{\theta}^T(\tau) J^T = \frac{\partial J^T}{\partial \hat{\theta}^T(\tau)} = \int_0^t e^{-\lambda(t-\tau)} \omega(\tau) [\omega^T(\tau) \dot{\hat{\theta}}(t) - \tilde{y}^T(\tau)] d\tau = 0 \quad (16) \]

The summands, which include \( y \), are moved to the right part of the equation to obtain (17).

\[ \int_0^t e^{-\lambda(t-\tau)} \omega(\tau) \omega^T(\tau) \dot{\hat{\theta}}(t) d\tau = \int_0^t e^{-\lambda(t-\tau)} \omega(\tau) y^T(\tau) d\tau \quad (17) \]

Then the following equation is obtained for the regression adjustable parameters:

\[ \dot{\hat{\theta}}(t) = \left[ \int_0^t e^{-\lambda(t-\tau)} \omega(\tau) \omega^T(\tau) d\tau \right]^{-1} \int_0^t e^{-\lambda(t-\tau)} \omega(\tau) y^T(\tau) d\tau. \quad (18) \]

The loop of ideal parameter estimation is obtained from the equation (18) according to the procedure, which is described in [9]. In this case, the estimation loop is a system of differential equations (19).

\[ \begin{align*}
    \dot{\hat{\theta}}(t) &= -\Gamma(t) \omega(t) \omega^T(t) \dot{\hat{\theta}}(t) \\
    \dot{\Gamma}(t) &= \lambda \Gamma(t) - \Gamma(t) \omega(t) \omega^T(t) \Gamma(t)
\end{align*} \quad (19) \]

As it follows from the system of equations (19), the estimation loop of unknown regression parameters, which is developed according to the recursive least-squares method, differs from the gradient ones (7) and (13). The loop (19) includes the law of adjustment of the adaptation rate value.
The solutions of the differential equations (19) are written as (20).

\[
\tilde{\theta}(t) = e^{-\frac{1}{2} \int^t_0 \Gamma' \omega(\tau) \omega^T(\tau) d\tau} \tilde{\theta}(0);
\]

\[
\Gamma'(t) = \frac{e^{\lambda t} \Gamma(0)}{1 + \int^t_0 e^{\lambda \tau} \omega(\tau) \omega^T(\tau) d\tau}.
\]

If the estimation law (19) is used, then the behavior of the parameter error has the following properties:

1) if the PE condition is met, then the parameter error (20) converges to zero exponentially [1];

2) the convergence rate of the parameter error (20) can be made arbitrarily high as a result of the forgetting factor value increase [9].

3. Experimental Results
The Matlab/Simulink has been used to conduct experiments to illustrate the properties of the gradient identification algorithms without (equation (7)) and with (equation (13)) the regressor extension, and the one developed on the basis of the recursive least-squares method (equation (19)). The parameter error convergence rate was to be assessed. The regressor equation, filter constant $\beta$, initial values of the parameter error and the adaptation rate matrix were chosen according to the values shown in (21).

\[
\Gamma(0) = I; \quad \lambda = 10; \quad \beta = 2; \quad \omega = \begin{bmatrix} A \sin(t) & A \cos \left( \frac{\pi}{4} t \right) \end{bmatrix}^T; \quad \tilde{\theta}(0) = \begin{bmatrix} 10 & 10 \end{bmatrix}
\]

Since there were two parameters in the considered regression, the regressor was chosen as a vector containing two spectral components to fulfill the condition of the persistent excitation.

Considering the experiments with the gradient loops (7) and (13), the constant adaptation rate, which coincided with the initial one for the recursive least-squares loop, was used.

Figure 1 shows a comparison of the norms of parameter errors obtained using the equations (7) and (19) with different values of the regressor amplitude $A$ (21).

![Figure 1. Comparison of the norms of the parameter errors obtained using the gradient-based loop without the regressor extension (A) and the recursive least-squares based one (B).](image)

As it follows from Fig.1A, the gradient loop (7) with the constant adaptation rate $\Gamma = I$ provided different rates of parameter error norm convergence to zero for different amplitude $A$ of the regressor. In this case, the existence of the optimal value of the adaptation rate, which was noted in the
introduction and section 2.1, was clearly demonstrated. The regressor of lower amplitude $A = 1$ provided a faster rate of convergence than the regressor $A = 10$. This means that the adaptation rate value $\Gamma^*=I$ was not optimal for the regressor of amplitude $A = 10$.

As it follows from Fig.1B, the estimation loop (19) also provided different convergence rates for the different amplitudes $A$ of the regressor. However, comparing Fig.1A and Fig.1B, it was concluded that this effect was much weaker. The adjustment process in Fig.1B was completed faster than in 1.5 seconds for all three regressors, whereas in Fig.1A none of the processes converged during this period of time.

Also, the adjustment of the adaptation rate allowed to provide a directly proportional dependence of the convergence rate on the amplitude of the regressor: the higher the amplitude of the regressor, the higher the convergence rate.

Figure 2 shows a comparison of norms of the parameter error obtained by application of the equation (13) to estimate the regression parameters under the condition of the different values of the regressor amplitude $A$ (21).

![Figure 2. Comparison of norms of the parameter errors obtained as a result of the experiments with the gradient estimation loop with the regressor extension.](image)

Figure 2 demonstrates that the higher the amplitude of the regressor, the higher the convergence rate, i.e., there is no optimal value of the adaptation rate $\Gamma$ in this case. This was stated in the introduction and section 2.2. It is an advantage of the MRE based loop as compared to the classic one (7). However, as it can be seen from the comparison of Fig.2 and Fig.1B, in contrast to the estimation loop based on the recursive least-squares method, in this one, as well as in the gradient one, there were different convergence rate values for the different amplitude $A$ of the regressor. So these results also demonstrated the advantages of the recursive least-squares based estimation loop.

In general, the results of the experiments showed that, in practice, in terms of the convergence rate for the different amplitude values of the regressor, it was preferable to use the recursive least-squares based estimation loop (19) with the adjustable adaptation rate comparing to the gradient loops without (7) and with (13) the regressor extension.

4. Conclusion
The automatic choice of the adaptation rate (using RLS) in the course of the regression parameters adjustment allowed taking into consideration the current value of the regressor. In its turn, it meant that the adaptation rate was improved at those moments when the regressor value became lower. This resulted in faster convergence of the parameter error of the identification loop to zero as a whole.

Further research will be to solve the problem stated in the introduction section. Namely, it is planned to use the recursive least-squares method to derive the estimation loop equations, which will provide the pre-determined and constant (despite time-varying values of the regressor) rate of the
parameter error convergence to zero. In addition, the recursive least-squares method will be applied under the conditions when the PE requirement is not met. In this case, the main aim will be to ensure the finiteness of the adaptation rate value.

Acknowledgments
This research was financially supported by the Russian Foundation for Basic Research (project № 18-47-310003 r_a).

References
[1] Ioannou P and Sun J 2013 Robust Adaptive Control (New York: Dover)
[2] Fradkov A L, Miroshnik I V and Nikiforov V O 1999 Nonlinear and adaptive control of complex systems (Springer Science & Business Media)
[3] Aranovskiy S, Bobtsov A, Ortega R and Pyrkin A 2016 Performance enhancement of parameter estimators via dynamic regressor extension and mixing. IEEE Transactions on Automatic Control 62 3546-3550
[4] Chowdhary G, Yucelen T, Mühlegg M and Johnson E 2013 Concurrent learning adaptive control of linear systems with exponentially convergent bounds. International Journal of Adaptive Control and Signal Processing 27 280-301
[5] Narendra K S and Annaswamy A M 1987 Persistent excitation in adaptive systems. International Journal of Control 45 127-160
[6] Lion P M 1967 Rapid identification of linear and nonlinear systems. AIAA Journal 5 1835-1842
[7] Kreisselmeier G 1977 Adaptive observers with exponential rate of convergence. IEEE transactions on automatic control 22 2-8
[8] Ortega R, Nikiforov V and Gerasimov D 2020 On modified parameter estimators for identification and adaptive control. A unified framework and some new schemes. Annual Reviews in Control. In press
[9] Sastry S and Bodson M 2011 Adaptive control: stability, convergence and robustness (Courier Corporation)