The method of correct calculation of regulators based on numerical optimization

V A Zhmud¹, L V Dimitrov²
¹Novosibirsk State Technical University, Karla Marksa ave. h. 20, Novosibirsk, 630073, Russia
²Technical University of Sofia, Sofia, Kliment Ohridski Boulevard, h. 8, 1756 Bulgaria

Abstract. Automation as a branch of engineering science forms and studies methods of control of objects, the main essence of which is to form the structure of the regulator and calculate its coefficients. The number of regulator structures is limited, the most common of them is a serial PID, i.e. a regulator having a proportional, integral and derivative channels connected in parallel, and the regulator itself is switched in series with the object. Thus, most of the publications on automation are connected with the calculation of numerical values of the regulator's coefficient and analysis of the results of its use. Along with the publication of the results of such a calculation for increasingly complex objects, there are still rather big stream of papers, that solve the problem of calculating the regulator for objects, the control of which is no longer of great complexity due to the level of the achieved technology. Therefore, the question of dividing the classes of tasks into those that have already been solved, and those whose solution has yet to be found is relevant. It is required to establish a minimum level of complexity of such problems, and it is also useful to compile a list of obsolete methods, the use of which without cardinal modification is fruitless in comparison with the potential of other methods, which became possible due to the latest achievements of theory and rapid development of computing facilities. Some scientific schools traditionally use obsolete approaches. One of the reasons for this may be their lack of awareness of the latest achievements; another reason may be inept use of these methods, which generates errors and disappointments in them. This paper reveals similar errors and suggests methods of correct calculation. When used correctly, the numerical optimization is the most effective tool for the said problems.

1. Introduction
Automation deals with methods of control of objects. In the analytical form, the control problem can be solved only for the simplest single-channel objects without delay and partly for multichannel objects. For more complex cases, approximate methods or approximations of the object model by simpler model are used. In the both cases, the result solution often cannot be used, because it not correspond to the correct calculation on the base of the correct model.
There are also objects described by such complex mathematical models that even some simplifications do not allow calculating of acceptable regulators for them by analytical methods, for example, multi-channel objects with delay and non-linear links in every channel. The control of such objects is very difficult. There are a number of publications with the solution of such problems by numerical optimization methods, for example [1, 2]. Analytical methods for solving this problem are not known. This situation has changed due to the emergence of powerful computers and effective software tools for modeling and optimization. This allowed applying of numerical optimization. Some authors perform calculation not very correctly, for example, in paper [3] the use of PI(D)² regulator is...
redundant for the object of the second order. In the book [4] test problems for numerical calculations contain a set of variants of objects with a transfer function in the operation form in the form of a fraction, the order of the numerator being equal to or greater than the order of the denominator, which is mistake.

In this paper, we analyze some problems and propose ways of correct solution of the problems of regulator design by numerical optimization methods.

2. Statement of the Task
The task is developing of complete list of recommendations that make it possible to obtain not redundant results that can be used without additional adjustment operation on the real system. This should be done based on the analysis of experience in solving the problems of the regulator design. The table methods of optimization include the Ziegler-Nicholson [5], the Cohen-Kuhn method [6] and some others [7]. The Ziegler-Nicholson method is already not such relevant as in time when it was the first officially published method for calculating PID regulators in 1942 [5].

Without going into the details of each of the tabular methods, we will consider them in aggregate.

1. It is assumed that the object model is a first- or second-order low-pass filter connected in series with a link of pure delay.
2. The response of the real object to a step jump or the response of the object in the system with proportional regulator is analyzed. In the second case, such a value of the proportional regulator coefficient is sought by means of probes, in which there will be undamped automatic oscillations in the system. In other cases, a coefficient is selected for which the overshoot has a predetermined value or the oscillations are damped with a predetermined attenuation coefficient.
3. According to the obtained process, it is proposed to remove its characteristic parameters, and also to fix the coefficient at which such a process takes place.
4. Based on the obtained characteristics, according to empirically found relationships, which are summarized in the table, it is proposed to calculate all three coefficients of the PID regulator or two for PI regulator.

Multiple modeling has convincingly shown that even for an object that, in fact, is described by such a model, which is taken as a basis in the first point, the obtained regulator is far from ideal. The hypothesis about the model of the object (point 1) is erroneous in most problems. If the model of the object is not as expected, then the regulator coefficients found by the transient process are not those that are required. Therefore, table methods are not prospective. Therefore, the Ziegler-Nicholson method and similar tabular methods should be removed from circulation.

Software VisSim is offered for modeling and optimization. The regulator reacts to input signals, from which, according to a given algorithm, it generates its output signals for controlling the object. The regulator cannot generate a signal proportional to the actual derivative or integral of its input signal, but can only calculate the estimate of this derivative or integral by a certain algorithm. VisSim works exactly the same way, so it is most appropriate for modeling and optimization.

3. The Discuss of the Admissible Simplifications
All elements of the system must be physically realizable. For any object, we can always specify frequencies at which the transfer function is not just small, but the response of the object at these frequencies is strictly absent or much less than the noise of its measurement. When investigating a loop with delay, the degree of the denominator of the transfer function must always be greater than the degree of the numerator. When examining a loop without delay, the degree of the denominator of the transfer function must always be at least two higher than the numerator. For the same reason, it is meaningless to optimize the PID regulator for an object in which there is no delay and in which the numerator's degree is only two less than the denominator. Indeed, the derivative path compensates for one order of the denominator; therefore, this system theoretically remains stable for any gain coefficient. Setting an incorrect optimization problem cannot lead to success.

In practice, such objects and such problems do not occur, the incorrectness of the statement of the problem can arise only as a result of an overly simplified model of the object. Thus, it can be
recommended that for any identification of a control object, it is not limited to a second-order model under any circumstances, and to identify the model to at least three major (largest) time constants of the denominator of the transfer function. The model of the object is obtained experimentally, every experiment has limited accuracy, and every response is studied in a limited frequency band. Therefore, when solving optimization problems, the object and its model can be identified only in the specified limited frequency region. Otherwise, there is an apparent contradiction between theory and practice, namely: optimal in theory regulators are not optimal in practice and vice versa.

To ensure that there is no contradiction between theory and practice, if the complete model of the object is not enough, it is necessary to refine the models in the high-frequency region to such a sufficient degree that it would ensure the coincidence of theory and practice. If this is not possible, it is necessary to artificially restrict the frequency band in the model in accordance with how it is naturally limited in the real object [8–11].

In theory, it is impossible to optimally adjust the PI-regulator for a first-order object and a PID-regulator for first-order and second-order objects. Whichever setting we have found, it always remains possible to specify a different setting that will be better.

In practice, for any object without exception, there is always the best setting. This is due to the fact that in practice there are no ideal objects of finite order in an unlimited range of frequencies and without consideration of the limitations of the input effect.

The reasons for the existence of an optimal solution can be called the reasons for the limited possible speed and quality of the transient process. A model that does not take them into account is unsuitable for finding the optimal solution, and, therefore, is generally unsuitable.

An attempt to optimize the regulator for an object of the first or second order (that is, without taking into account higher orders, control constraints, delay, nonlinearity, or even some other additional feature of the object) is an attempt to solve a problem that has no solutions. As a result, a pseudo-optimal solution can be found, which cannot be further improved, but this inability to further improve the result will be due to the appearance of incorrect modeling. For example, the integration step will not be sufficiently small at some speed of the locked system. This will give rise to a noticeable effect of the simulation parameters on the resulting process. The delay born by the time of one integration step and (or) the discreteness of sampling will affect to the result, which is far from the real situation with the object and the resolving task. But if the control of the object in practice is carried out discretely, and this value of the integration step will enter the model of the object or regulator, then such a model will be completely adequate to the object: this model contains the same delay, and for it the optimal settings of the said regulators exist, and their search is completely justified.

4. The Discuss of the System Quality Criteria
Book [12] proposes as the best value function is the integral over the time interval $T$ from the error square $e(t)$:

$$\Psi_0(T) = \int_0^T e^2(t) dt.$$  \hspace{1cm} (1)

Our research [8–11] has shown that in the case of applying this cost function in the system there is an excessive number of oscillations, overshoots reach 30% and higher [8]. In addition, the objective function (1) does not provide a sufficiently fast aspiration of the error to zero, that is, it is not effective for calculating of the integral channel coefficient of the PID-regulator. In [8], a cost function is recommended in the general form:

$$\Psi_1(M, N, T) = \int_0^T e(t) |M t^N| dt.$$  \hspace{1cm} (2)

with the remark that the most effective variants are $\{N = M = 1\}$, $\{N = 4, M = 2\}$ and several others. Variant $\{N = 0, M = 2\}$, corresponding situation of (1) is not effective.
Paper [13] proposes minimization of (2) with $N = 0, M = 1$, without absolute value, only integral from $e(t)$, which is not correct, because $e(t)$ can be negative too, absolute value should be used. Additional terms can be introduced to achieve special requirements, for example, to reduce overshoot, to eliminate areas where the error increases in time, and so on [8].

The value of $T$, that is, the time of simulation and integration, is also not fixed. It should be ensured that such a time $T$ is chosen so that the transient process terminates almost in an interval equal to 70–90% of this value. Composite quality criteria are adequately described in [8, 9]. The most effective composite criteria work if each term is responsible for a separate feature of the transient process, for example, one term increases with increasing static error, the other with an increase in overshoot, the third with an increase in the transient process time, and so on.

It is mistakenly considered that the optimal system is unique for a given object. Repeatedly there are statements that the optimal system is unique, since "optimal" means "the best", and there cannot be several best settings.

Those who think so, forget that in technology the term "optimal" means "providing the desired extremum of a given cost function". And since there can be several cost functions, then there can be several optimal settings. But even with a single form of the cost function, any of the weight coefficients can be changed, which will give another optimization result. The integration time $T$ can also be changed, and this will also give rise to a new optimal result.

The designer of the system should, in addition to solving the problem of optimizing the regulator coefficients, also to solve the problem of "optimizing the optimization criterion". And this task is often solved by choosing the best result from the sets of optimal settings [8–11].

5. Automatic Optimization of Locked-Loop Regulators

To run optimization in VisSim, the developer needs the following.

1. Define the structure of the project in the form of a regulator model and an object model, connected in series to the loop. Also, control and (or) perturbing influences with the help of blocks generating the required functions. As a rule, step jump generators are used.
2. Set the values of the optimized parameters using the "parameter unknown" blocks.
3. Set the start values to each of these parameters, giving a constant to the input of these blocks.
4. Ensure the use of these parameters as regulator coefficients, which should be optimized. For this purpose it is useful to give them the names of functional variables (for example, $P$, $I$, $D$ are the coefficients of the proportional, integral and differentiating paths). Assigned names should be used to call these functions and use them as appropriate gains. Developer can use multiplier blocks or directly blocks of gain factors.
5. Provide calculation of the cost function and send the result to the "cost" block, which should be the only one in the project. To this end, a calculator is formed, the output of which is connected to the input of this block.
6. Provide an indication of the optimization result, for which it is advisable to use the value indicator block (at the output of each of the "parameter unknown" blocks).

In addition, it is possible, but not necessary, to reflect the value of the resulting transient process on the oscilloscope block and the value of the cost function on such a block and (or) on the block of the indicator of a numerical value.

To optimize the cost function, it is necessary to solve the optimization problem, obtain transient processes and decide whether they are satisfactory or not. If they are unsatisfactory, such corrections should be made to the cost function, which would presumably eliminate the shortcomings of transient processes that make them unsatisfactory [9].

To obtain a robust controller, parallel modeling of several systems should be used, in which objects differ in coefficients taken from the limits of the permissible values of these values in accordance with the task, and the regulators are identical. The cost function in this case is the sum of the cost functions calculated from the errors in all these systems. The minimum number of such systems should be two, the maximum – as much as the software and hardware allow.
6. Ensuring the Correctness of Modeling

The most common cause of incorrect simulation is the insufficiently small step of integration [3–6]. As the regulator coefficients increase, the speed of the system increases, so after some achieved speed, the integration step becomes already insufficiently small. This will cause a violation of the stability of the system with a further increase in the coefficients. As a result, the delay for the time of one integration step and (or) the discreteness of sampling, which does not have a place in the real object, will be affected. In order to correctly model in this case, it is necessary to reduce the integration step, and as a result, it will be possible to further increase the regulator coefficients.

The adequacy of the model to the object is preliminary checked not only by sufficient proximity of transient processes in the open loop, but also by the presence of at least two factors (inertial, transcendental and nonlinear links) in each optimized loop. The final check is carried out by the coincidence of the results of analytical or model optimization with the results of application of the received systems in practice.

The validity of the criterion requires the presence of a technical sense in the problem of minimizing the chosen cost function and the existence of a minimum of this function.

The correctness of the simulation is provided by the right choice of the software product, the knowledge of the fundamentals of the theory of automatic control and the application of a number of critical checks of the correctness of the selected software product.

The suppression of the effect of noise, static error and other undesirable factors of a real object on the result of real optimization is ensured by introducing in the cost function of some insensitivity to errors and noise below the allowable value for them. A stabilization error within acceptable levels should not increase the cost function. Its value under these conditions also should not affect the result of optimization of the regulator.

Preservation of correctness in the modeling of the obtained nonlinear systems is achieved by taking into account and applying the real values of all signals.

The validation of the model and the correctness of the simulation should be finally implemented according to the following main principle: Transient processes in the model and in the practical system with regulators obtained by the optimization procedure must coincide; this should also be the case if any of its coefficients are changed in the calculated optimal regulator (by less than 0.1%).

Insensitivity of the result to small deviations of the regulator parameters is called "robustness of the solution". It is a positive property of this solution, in contrast to the "non-robust solution", sensitive to fine-tuning the parameters. Verification of the robustness of the solution can also be carried out by introducing test small deviations of the parameters, which should not significantly change the result of the real system operation.

A test of the optimality of the resulting system can be carried out by introducing trial large deviations of the parameters, which is more significant than for testing the robustness of the solution. Any change of parameters in relation to the optimal setting should only worsen the set of positive properties of the system in comparison with the properties of the optimal system. Formally, any change in any factor of the regulator should lead to an increase in the cost function. Such test is very revealing, but, unfortunately, there are practically no references to the fact that it was conducted in the literature. This check should not be applied to the model (or just to the model), but to the real object, which is much more important.

The validity of the cost function is checked by an estimate of the resulting transient processes and by comparing them with the variants obtained by other methods or with other cost functions.

7. The Numerical Examples

Paper [5] resolves the problem of controlling of an object described as two successive connected filters of the first order and a link of pure delay. For example, the object model is adopted in the following form:

\[ W_1(s) = \frac{\exp(-\tau, s)}{(1+T_1 s)^2}. \] (3)
Here $T_s = 4.7s$, $\tau = -1$.

To demonstrate the existence of a sufficient arsenal of means for solving the problem considered in [5], we solve the following two problems.

**Example 1.** Let's solve the problem of optimizing the PID-regulator for the object (3), provided that its parameters change: $2.35 < T < 7.05$ and $-1 > \tau > -9$.

In accordance with the proposed method of designing robust regulators [8–11], we will optimize the regulator for the worst combinations of object parameters, after which we will check how the system behaves with the best combination of parameters.

Figure 1 shows the structure for optimizing the PID-regulator. The object corresponds to the model (3); the regulator contains a proportional, integrating and derivative links with its coefficients, which should be found. The structure of these blocks is not shown, since it is trivial. The cost function is used the same as in [8–11]. The results of optimization are the found PID-regulator coefficients: $k_p = 1.05$; $k_i = 0.053$; $k_d = 5.38$. The transient processes are shown in Fig. 2.

As we can see, the system remains stable under all the tests carried out, the tendency of changes in the properties of the transient process is quite evident. This allows us to state that for intermediate values of the model parameters of an object; transient processes will have an intermediate form.

The parameters of the model of the object have changed together, while the combinations of, for example, the largest value of the filter time constant with the lowest value of the time constant of the delay link, or on the contrary, have been tested, but in this case it completely coincides with the research conditions described in the paper [5]. A detailed investigation requires that the system be tested with the specified combinations of parameters.

In two cases, there is no overshooting; in the worst case it does not exceed 2%. The duration of the process does not exceed 90 s. This result is much better than the results obtained in [5], despite the fact that the structure of the regulator is much simpler, since a simple PID-regulator is used, which is simpler than the regulator in [5]. Indeed, in all the results in this paper the duration of the transient process is more than 200 s, except for the result "with the model predictive regulator", which, according to the authors, does not possess robustness and is characterized by a large static error. Our solution has not such significant disadvantages.
Example 2. Let's solve the problem of optimizing the PID controller for the next object from [5] under the same conditions.

\[ W_2(s) = \frac{\exp(-\tau_2 s)}{1 + T_2 s}. \]  
(4)

Here \( \tau_2 = -2.61 \), \( T_2 = 6.386 \).
In accordance with the proposed method of designing robust regulators [8–11], we will optimize the regulator for the worst combinations of object parameters, after which we will check how the system behaves with the best combination of parameters.

The structure for optimizing the PID regulator is the same as in the previous example. The structure of the regulator is the same. The structure of the object is not shown, since it is sufficiently clear from relation (4). The cost function is the same as in the previous example. The results of optimization are the found PID-regulator coefficients: $k_p = 1.03; k_i = 0.069; k_d = 2.89$. Transients are shown in Fig. 3. These processes are similar to the processes of Example 1 only in general, but in details are significantly different. The result of calculating the regulator also differs, in particular the coefficient of the derivative path, which differs almost twice. At low values of the object's time constants, high-frequency oscillations are observed in the transient process, although the system remains stable and, according to formal criteria, the transient process is quite acceptable: the duration of the process does not exceed 80 s; there is no overshooting in two cases, in the worst case there is an insignificant overshooting about 1%.

The oscillations can be reduced. For this we can optimize the object not by the worst parameters, but with a slightly improved one, namely: the time constant of the delay link is taken to be not 9 s, but 7 s. The found PID-regulator coefficients: $k_p = 1; k_i = 0.08; k_d = 2.2$. The corresponding transient processes are shown in Fig. 4. In processes with smaller values of the parameters of the object, the oscillations are practically eliminated. But at large values, the overshoot has increased to about 10%. Visually, this result is better, but formally one can prefer the previous result, since there is less overshooting in it. The choice depends on the technical requirements and operating conditions of the system.

This result is also better than the one given in [5].

![Figure 4. Transients in the system of Example 2 with the corrected values of the controller parameters at the maximum values of time constants (line 1), minimum values (line 3) and average values (line 2).](image-url)
Figure 5. Two transient processes in the system according to Example 2: line 1 – when disturbance is applied to the output of the object; line 2 – when disturbance is applied to the input of the object.

In the paper [5], among other things, the point of application of the disturbance is chosen at the input of the object. In the general case, for the analysis of system properties, such a "test impact" is not correct. Usually the point of application of the disturbance is at the output of the object, that is, through the adder, located between the output of the object and the output of the system. This is more correct, because in the first case the object acting as a filter suppresses this disturbance by itself, so the load on the feedback loop is substantially reduced. For illustration, in the system of Example 2, we simulate the transient process with the two indicated types of disturbance. Figure 5 shows two transient processes: a) when the disturbance is applied to the input of the object (line 2); b) when disturbance is applied to the output of the object (line 1).

In the first case, the error abruptly becomes equal to unit, which is 100% of the interference. This error occurs over a time equal to the time constant of the delay link. After this time, the error starts to decrease slowly and then it reaches zero, changes sign and then drops to zero.

In the second case, the entire period of time equal to the time constant of the delay, the error is zero. Only after that, it gradually increases to 60% of the disturbance and then gradually falls to zero value. The transient process ends approximately in the same time, but since the error occurs later, the time of its existence is shorter. That is, the second option is much softer, this is a gentle mode of operation of the system, and it gives less information about the quality of the system. If a step input is applied to the control input, the error will change in the same way as in the first case (with the opposite sign, but with the same value and with the same shape of the graph). Therefore, only this method of testing the system is sufficiently correct and obvious.

8. Conclusions
This paper has discussed some characteristic errors in solving the problem of numerical optimization of regulators for a single-loop automatic control system. It gives the criticism of table methods and shows on numerical examples that out-of-date methods of solving problems yield worse results with a higher price. Once again, the efficiency of the method for design of the robust regulators has been confirmed by using the hypothesis of the worst combination of the parameters of the object model.

9. References
[1] Abdullin A A, Mamatov A G, Konstantin S. Gorshkov K S 2018 Discrete-Time State Feedback Control Design for Linear Objects with Delay IEEE Conference of Russian Young Researchers in
[2] Li Junhui, Lu Jieying and Su Weizhou 2017 Stabilizing control of a class of discrete-time systems with random delay 36th Chinese Control Conference (CCC) https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8028588

[3] Frantsuzova G A, Kotova E P 2017 Calculation and Research of Possibilities of Automatic Control Systems with Standard PID- and Modified PI2Dcontroller. Automatics & Software Engineering 1(19) P 10–15 http://jurnal.nips.ru/en/getenglish

[4] Nikulin E A 2004 Fundamentals of the theory of automatic control. Frequency methods of analysis and synthesis of systems https://www.twirpx.com/file/352592/

[5] Stopakevich A A and Stopakevich A A 2016 Vostochno-Yevropeyskiy Zhurnal Peredovoykh Tekhnologiy 1/2 (79) 48–56 (in Russian). URL: http://dspace.opu.ua:8080/jspui/bitstream/123456789/3839/1/Stopakevych.pdf

[6] Mishanov M S, Gladyshev I S and Baydali S A 2011 Metody Rascheta Regulyatorov v Odnokonturnykh SAU pp 396–397 URL: www.lib.tpu.ru/fulltext/c/2011/C01/V02/193.pdf

[7] VanDoren V 2018 Understanding PID control and loop tuning fundamentals https://www.controleng.com/single-article/understanding-pid-control-and-loop-tuning-fundamentals/61530560e8cf5044c7291bda7e44b655.html

[8] Zhmud V A and Liapidevskiy A V 2010 The Design of the Feedback Systems by Means of the Modeling and Optimization in the Program VisSim 5.6/6.0 Proc. Of The 30th IASTED Conference on Modelling, Identification, and Control – AsiaMIC 2010 (Phuket, Thailand) pp 27–32 URL: http://www.actapress.com/Abstract.aspx?paperId=41995

[9] Zhmud V, Yadrishnikov O, Poloshchuk A and Zavorin A. 2012 Modern Key Technologies in Automatics: Structures and Numerical Optimization of Regulators 7th Int. Forum on Strategic Technology (IFOST). URL: https://ieeexplore.ieee.org/document/6357804/

[10] Zhmud V A, Dimitrov L V, Yadrishnikov O D 2014 Calculation of regulators for the problem of mechatronics by means of numerical optimization method. 12th Int. Conf. on Actual Problems on Electr. Intrum. Eng. APEIE URL: https://ieeexplore.ieee.org/document/7040784/

[11] Zhmud V and Zavorin A 2015 The design of the control system for object with delay and interval-given parameters SIBCON–2015: proc. (Omsk, 2015) IEEE. URL: https://ieeexplore.ieee.org/document/7147060/

[12] Dyakonov V P 2004 VisSim+Mathcad+MATLAB. Visual mathematic modeling (in Russian).

[13] Romero J A and Balaguer P 2010 Tuning of PID Controllers based on simplified single parameter optimization Roberto Sanchis International Journal of Control March 3, 2010 https://core.ac.uk/download/pdf/61398931.pdf