EURASIAN JOURNAL OF MATHEMATICAL
AND COMPUTER APPLICATIONS
ISSN 2306–6172
Volume 13, Issue 1 (2025) 4 – 16

NEWTON'S METHOD FOR SOLVING A CONTROL PROBLEM UNDER PIECEWISE-IN-TIME DEFINED ACTIONS

Aida-zade K. , Handzel A. [1]

Abstract In this paper we propose an approach to the numerical solution of one class of optimal control problems based on the use of Newton's method. The problem is described by a system of ordinary differential equations. The parameters of control actions from the class of piecewise defined functions as well as the boundaries of intervals of constancy of control actions are to be optimized. Formulas for gradient components and Hesse matrix of the objective functional in the space of optimized parameters are obtained. Comparative results of computer experiments are given.

Key words: second-order method, functional gradient, Hesse matrix, Newton's method, Runge-Kutta method.

AMS Mathematics Subject Classification: 34H05.

DOI: 10.32523/2306-6172-2025-13-1-4-16

1 Introduction

The paper studies the numerical solution of the problem of optimal control of an object (process) described by the Cauchy problem with respect to a system of nonlinear ordinary differential equations.

The control actions, which are piecewise continuous functions, are determined from a parametrically specified class. The unknown parameters of these functions may differ at different time intervals. The boundaries of time intervals and their number as well as parameters of functions at each interval which are optimizable, are determined from the condition of minimization of the given objective functional of the optimal control problem.

Applied nature of the considered class of control actions is explained by the fact that when controlling many objects (processes) it is undesirable to change frequently the modes of functioning of both the object itself and the unit controlling it. Frequent change of modes can lead to rapid wear of technological equipment, on the one hand, and on the other hand, frequent change of control parameters can be simply technically unimplementable. Technological equipment of an oil or gas field is an example of such an object. As it is known, for the equipment for production and transportation of hydrocarbons, frequent change of operation modes, in particular, wells, on the one hand, is undesirable, on the other hand, the implementation of this change is associated with technical difficulties.

[1] Corresponding Author.

Regional power supply systems represent another object of application of the considered class of controls. The operation of this object is described by systems of a large number of ordinary differential equations. When optimally controlling the operation modes of power supply equipment, it is very important to ensure maximum timewise stability of these modes, taking into account the instable, in particular, the periodical power consumption by end users.

As a numerical method for solving the considered optimal control problem it is proposed to use the second-order methods based on Newton's method. It should be noted that the numerical methods of high order of accuracy for solving various problems of computational mathematics have become the object of both theoretical research and their practical application [1]–[9]. An undeniable positive quality that these methods possess is a large convergence rate, especially in cases where the initial point is close enough to the desired solution. For some classes of computational problems these methods are exact, i.e. they do not require iterative procedures to obtain a solution with a given accuracy, but there are still certain problems of application of high-order accuracy methods for solving different classes of computational problems, for example, for finite-dimensional optimization and optimal control problems. There are two main reasons which hold back their widespread use for solving real practical problems. The first reason is related to the difficulty of obtaining formulas for the gradient components and Hesse matrix of the objective function. Their difference approximations, firstly, require a large amount of computational work of the processor, and secondly, the error in calculating the first and second derivatives of the objective function significantly reduce the efficiency of the high-order methods. The second problem is related to the large amount of required computer memory, especially for numerical solution of optimal control problems.

The class of optimal control problems considered in this paper practically does not face these problems. This is due to the specific feature of admissible control actions, which is as follows. The entire time interval is divided into a relatively small number of separate subintervals, at each of which the control actions are searched in the form of parametrically given functions whose parameters need to be optimized with respect to the objective functional of the considered problem. The boundaries of constancy intervals of the parameters involved in the above control functions are also optimizable.

Thus, the initial problem is reduced to finding the optimal values of the parameters involved in the formation of control actions and the parameters defining the boundaries of the control actions constancy intervals.

The considered problem, on the one hand, is a parametric optimal control problem, on the other hand, it can be referred to finite-dimensional optimization problems. The dimension of the being optimized vector of parameters in practical cases is small, rarely reaching several tens. The number of intervals of constancy of parameters of control functions can be both specified and determined on the basis of the necessity of their reduction. Reducing the intervals of constancy of parameters improves the quality of control of the object, its robustness and reliability of functioning of the object as a whole.

For numerical solution of the control problem it is supposed to apply its finite-dimensional approximation using methods and schemes with high approximation accuracy. For numerical solution of the obtained finite-difference control problem by second-order optimization methods, fast differentiation technique is used to derive exact formulas for the gradient and Hesse matrix [9]–[11]. The formulas for the derivatives of the approximated functional obtained by this approach are exact because the approximation schemes we use for the direct and conjugate discrete Cauchy problems are inter-consistent.

Special cases of the proposed approach to solving optimal control problems are relay control problems on the class of piecewise constant functions with optimizable switching times [12], the problems in the class of piecewise linear functions et al. Also, note that a large class of inverse problems can be reduced to the considered problems [13]–[15].

The results of numerical experiments presented in the paper illustrate the scheme of the proposed approach and confirm the efficiency of its application.

2 Problem formulation

We study the numerical solving the problem of control of a dynamic process described in general by the Cauchy problem with respect to a system of nonlinear ordinary differential equations [16]:

$$\dot{y}(t) = f(t, y(t), u(t)), \quad t \in (t_0, t_f], \tag{1}$$

$$y(t_0) = y_0, \tag{2}$$

$$J(u) = \int_{t_0}^{t_f} f_0(t, y(t), u(t))dt + \Phi(x(t_f)) \ \rightarrow \min_{u(t)\in U} . \tag{3}$$

Here $y(t) = (y_1(t), \dots, y_n(t))^\top$ is an n-dimensional vector-function which describes the phase state of the process at time instance t, $t \in [t_0, t_f]$; scalar function $f_0(t, y, u)$ and n-dimensional vector-function $f(t, y, u) = (f_1(t, y, u), \dots, f_n(t, y, u))^\top$ are given and piecewise continuous with respect to first argument and twice continuously differentiable with respect to second and third argument; $u(t) = (u_1(t), \dots, u_r(t))^\top$ is r-dimensional piecewise continuous vector-function which describes control actions of the process at time instance t, $t \in [t_0, t_f]$; $\Phi(\cdot)$ is a given, twice continuously differentiable function; t_0, t_f are given, $^\top$ is transposition sign.

It is assumed that in view of technical and technological consideration function $u(t)$ must satisfy some certain constraints, i.e. $u(t) \in U$. In particular, we will assume that the admissible set for $u(t)$ is as follows

$$U = \{u(t) \in R^r : \underline{u}_i \le u_i(t) \le \overline{u}_i, \quad i = 1, \dots, r, \}, \tag{4}$$

where $\underline{u}_i, \overline{u}_i$, $i = 1, \dots, r$ are given.

The problem of controlling of the considered process consists in determining admissible control $u^*(t) \in U$ and corresponding vector-function $y^*(t)$ such that the pair $(u^*(t), y^*(t))$ minimizes (3), i.e. $J(u^*) = \min_{u(t)\in U} J(u)$.

Let us assume that the being optimized control $u(t)$ is searched in the following parametrically given class of functions:

$$u(t) = \lambda^k \psi(t - \tau_{k-1}), \quad t \in [\tau_{k-1}, \tau_k). \tag{5}$$

Here $\tau_k \in [t_0, \ t_f]$, $k = 0, \dots, \mu$, $\tau_0 = t_0$, $\tau_\mu = t_f$, r-dimensional piecewise continuous vector-function $\psi(t)$ is given and its components are linearly independent; $r \times m$ matrix

λ^k consists of constant parameters which determine the value of control $u(t)$ at time interval $[\tau_{k-1}, \tau_k)$: $\lambda^k = \left\| \lambda^k_{ij} \right\|$, $i = 1, \dots, r$, $j = 1, \dots, m$.

Thus, the original problem (1)-(4) on determining optimal control $u(t)$ is reduced to finding optimal values of the elements of the matrix of parameters λ^k , $k = 1, \dots, \mu$ and switching points τ_k, $k = 1, \dots, \mu - 1$. The total number of the optimized parameters is $(rm\mu + \mu - 1)$.

It is evident that switching points must satisfy the obvious requirements:

$$t_0 \le \tau_{k-1} \le \tau_k \le t_f, \quad k = 1, \dots, \mu. \tag{6}$$

A special case of the control actions of kind (5) are piecewise constant control actions which are frequently encountered with in real life applications. In this case $m = 1$, $\psi_1(t) \equiv 1$ and parameters λ^k are scalar ones with the fixed values at half-open interval $[\tau_{k-1}, \tau_k)$. Another commonly occurring in practice cases are related to piecewise linear in time controls when $m = 2$, $\psi_1(t) \equiv 1$, $\psi_2(t) = t$.

Let us note that functions $\psi(t)$ themselves can be piecewise continuous ones at half-open intervals $[\tau_{k-1}, \tau_k)$, $k = 1, \dots, \mu$, hence in general case controls $u(t)$ can have discontinuity both at $t = \tau_k$, $k = 1, \dots, \mu - 1$ and at a finite number of points at half-open intervals $[\tau_{k-1}, \tau_k)$, $k = 1, \dots, \mu$.

The obtained problem (1)-(6) on finding a finite-dimensional vector $\nu = (\lambda, \tau) \in R^{(mr+1)\mu - 1}$ belongs to parametric problems of optimal control of systems with lumped parameters. Also, this problem can be viewed as a special optimization problem of finite dimension.

It is easy to prove that the functional of the problem (1)-(4) is differentiable and strictly convex with respect to the parameters to be optimized if the functional of the original problem (1)-(3) is strictly convex with respect to $u(t)$.

A remark. For the problems of optimal control of distributed systems described by partial differential equations, the approach considered in the paper can also be used after pre-applying Rothe's method on spatial variables to approximate the above-mentioned equations [17], [18].

It is significant that the use of the considered class of control actions does not lead to frequent change in the values of the parameters of the control actions resulting in the increase of the stability, robustness, and reliability of the control system operation and goes along with no essential sacrificing in the value of the objective functional attained at an admissible optimal piecewise continuous control action $u(t)$.

3 An approach to numerical solution of the problem

To numerically solve the problem under consideration, we use its finite-difference approximation. Next, to solve the resulting finite-dimensional optimization problem using second-order numerical methods, we obtain formulas for the derivatives of the objective functional of the first and second orders in the space of optimized control action parameters $\nu = (\lambda, \tau)$. To obtain these formulas, we will use fast differentiation technique [9-11].

For the finite-difference approximation of the Cauchy problem (1), (2) and functional (3), we apply any known formulas of a high accuracy.

Let us divide the segment $[t_0, t_f]$ into N segments with regular points $\overline{t}_j = t_0 + jh$, $j = 0, \ldots, N$, $\overline{h} = (t_f - t_0)/N$, $\overline{t}_N = t_f$. Let the j-th switching point τ_j belong to the η_j-th regular half-interval: $\tau_j \in [\overline{t}_{\eta_j - 1}, \overline{t}_{\eta_j})$, $j = 1, \ldots, \mu$.

Let us add switching points τ_j, $j = 1, \ldots, \mu - 1$ to $N + 1$ points $\overline{t}_0, \overline{t}_1, \ldots, \overline{t}_N$ and increase the number of segments to $N_\mu = (N + \mu - 1)$. Let us denote the obtained points by $t_0, t_1, \ldots, t_{N_\mu}$, $h_{j-1} = t_j - t_{j-1}$, $j = 1, \ldots, N_\mu$ are the steps of non-uniform partition of the segment $[t_0, t_f]$. It is clear that $\overline{t}_N = t_{N_\mu}$.

Let point τ_j be the χ_j-th point of the non-uniform partition of the segment $[t_0, t_f]$: $t_{\chi_j} = \tau_j$, $j = 1, \ldots, \mu - 1$, $t_0 = \tau_0$, $t_N = \tau_\mu$. Then $h_{\chi_j - 1} = \tau_j - t_{\chi_j - 1} = t_{\chi_j} - t_{\chi_j - 1}$ and $h_{\chi_j} = t_{\chi_j + 1} - \tau_j = t_{\chi_j + 1} - t_{\chi_j}$.

To make further formulas less bulky let us introduce the following matrix-function and vector-function:

$$\widetilde{\lambda}(t) = \lambda^k = \text{const}, \quad t \in [\tau_{k-1}, \tau_k), \quad k = 1, \ldots, \mu;$$
$$\widetilde{\psi}(t) = \psi(t - \tau_{k-1}), \quad t \in [\tau_{k-1}, \tau_k), \quad k = 1, \ldots, \mu.$$

We write the finite-dimensional approximation of problem (1), (2) as follows:

$$y^0 = y_0, \quad y^{j+1} = \widehat{\mathbf{F}}^j \left(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)\right) = y^j + h_j F\left(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)\right), \quad (7)$$
$$j = 0, \ldots, N_\mu - 1.$$

In (7), n-dimensional vector-function $\widehat{\mathbf{F}}$ is determined by the approximation scheme for the Cauchy problem (1),(2). For example, for an explicit Euler method

$$F(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)) = f(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)). \quad (8)$$

For the fourth-order Runge-Kutta method, we obtain

$$F(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)) = \frac{1}{6}(k_1^j + 2k_2^j + 2k_3^j + k_4^j), \quad (9)$$

$$k_1^j = f\left(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)\right), \qquad k_2^j = f\left(t_j + \frac{h_j}{2}, y^j + \frac{h_j k_1^j}{2}, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j + \frac{h_j}{2})\right),$$
$$k_3^j = f\left(t_j + \frac{h_j}{2}, y^j + \frac{h_j k_2^j}{2}, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j + \frac{h_j}{2})\right), \quad k_4^j = f\left(t_j + h_j, y^j + h_j k_3^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j + h_j)\right),$$

here k_1^j, k_2^j, k_3^j, k_4^j are n-dimensional vector functions of three arguments.

It is clear that virtually all numerical methods for solving systems of differential equations can be described by (7) and it is evident that function $\widehat{\mathbf{F}}^j$ will be different for each of these methods.

To simplify the given formulas when approximating the functional, we use only regular points $\overline{t}_j$, $j = 0, \ldots, N$:

$$I(v) = \overline{h} \sum_{j=0}^{N} \sigma_j f_0 \left(\overline{t}_j, y^j, \widetilde{\lambda}(\overline{t}_j)\widetilde{\psi}(\overline{t}_j)\right) + \Phi\left(y(\overline{t}_N)\right). \quad (10)$$

In particular, for values σ_j, $j = 0, \ldots, N$ well-known Simpson's method of fourth order accuracy can be used.

Taking into account the specifics of the dependencies between the phase variables y^j, $j = 0, \ldots, N_\mu$, and the optimized variables $v = (\lambda, \tau)$, we use fast differentiation technique to obtain formulas for the first and second order derivatives: $\partial I(v)/\partial\lambda$, $\partial I(v)/\partial\tau$, $\partial^2 I(v)/\partial\lambda^2$, $\partial^2 I(v)/(\partial\lambda\partial\tau)$, $\partial^2 I(v)/\partial\tau^2$.

Let us introduce the following index sets corresponding to dependencies (7)

$$\begin{aligned} &S = \{0,1,\ldots,N\}, \quad \overline{S} = \{0,1,\ldots,N,\ldots,N_\mu\}, \\ &S_{y^j} = \left\{s \in \overline{L} : y^s = \widehat{\mathbf{F}}(\cdot, y^j, \cdot)\right\}, \quad j \in \overline{S}, \\ &S_{\tau_j} = \left\{s \in \overline{L} : y^s = \widehat{\mathbf{F}}(\cdot, \tau_j, \ldots)\right\}, \quad j = 1,\ldots,\mu-1, \\ &S_{\lambda^k} = \left\{s \in \overline{L} : y^s = \widehat{\mathbf{F}}(\cdot, \lambda^k_{ij}, \cdot)\right\}, \quad k = 1,\ldots,\mu. \end{aligned}$$

As it follows from the definition of the sets S_{y^j}, S_{τ_j} and S_{λ^k}, they are determined by the indices of those variables y^s, $s = 0,\ldots,N_\mu$, the values of which are directly affected by y^j, τ_j and λ^k respectively.

For example, in the case of using the Euler (8) and Runge-Kutta methods, we obtain:

$$\begin{aligned} &S_{y^j} = \{j+1\}, \quad S_{\tau_j} = \{\chi_j - 1, \chi_j\}, \quad S_{\lambda^k} = \{\chi_k, \chi_k + 1, \ldots, \chi_{k+1} - 1\}, \\ &h_{\chi_j - 1} = \tau_j - t_{\chi_j - 1}, \quad h_{\chi_j} = t_{\chi_j + 1} - \tau_j, \quad j = 1,\ldots,N_\mu - 1. \end{aligned}$$

Let us introduce the following n-dimensional vector p^i

$$p^i = \frac{dI(v)}{dy^i}, \quad i = 0,\ldots,N_\mu. \tag{11}$$

Here the derivative is understood as total, with regard for the dependence (7), i.e. the influence on the functional of both the values of vectors y^i and vectors y^j, $j \in S_{y^i}$, depending on y^i. This implies

$$p^i = \frac{\partial I(v)}{\partial y^i} + \sum_{j \in S_{y^i}} \frac{\partial y^j}{\partial y^i} p^j. \tag{12}$$

Here and below, expressions for vector and matrix derivatives $\partial I(v)/\partial y^j$, $\partial y^j/\partial y^i$ are understood as partial derivatives, namely the derivatives with respect to the variables explicitly involved in the dependencies. Particularly, in the case of using Euler method (8) we obtain

$$p^j = h_j\sigma_j \frac{\partial f_0}{\partial y}(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)) + \left(E + h_j \frac{\partial f}{\partial y}(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j))\right) p^{j+1}; \; j = N_\mu - 1,\ldots,0, \tag{13}$$

$$p^{N_\mu} = \overline{h}_N \sigma_N \frac{\partial f_0}{\partial y}(\overline{t}_N, y^N, \widetilde{\lambda}(\overline{t}_N)\widetilde{\psi}(\overline{t}_N)) + \frac{\partial \Phi}{\partial y}(y^N). \tag{14}$$

In (14) it is considered that $\overline{t}_N = t_{N_\mu} = t_f$.

In the case of using the Runge-Kutta method (9) to approximate (1), (2), it is easy to see that any $j = N_\mu - 1,\ldots,0$

$$p^j = h_j\sigma_j \frac{\partial f_0}{\partial y}(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)) + \left(E_n + \frac{h_j}{6}\frac{\partial k_1^j}{\partial y^j} + \frac{h_j}{3}\frac{\partial k_2^j}{\partial y^j} + \frac{h_j}{3}\frac{\partial k_3^j}{\partial y^j} + \frac{h_j}{2}\frac{\partial k_4^j}{\partial y^j}\right) p^{j+1}; \tag{15}$$

where

$$\begin{aligned}
\frac{\partial k_1^j}{\partial y^j} &= \frac{\partial f}{\partial y}(t_j, y^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)),\\
\frac{\partial k_2^j}{\partial y^j} &= \frac{\partial f}{\partial y}(t^j + \frac{h_j}{2}, y^j + \frac{h_j k_1^j}{2}, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)) \left[E_n + \frac{1}{2}\frac{\partial k_1^j}{\partial y^j}\right],\\
\frac{\partial k_3^j}{\partial y^j} &= \frac{\partial f}{\partial y}(t^j + \frac{h_j}{2}, y^j + \frac{h_j k_2^j}{2}, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)) \left[E_n + \frac{1}{2}\frac{\partial k_2^j}{\partial y^j}\right],\\
\frac{\partial k_4^j}{\partial y^j} &= \frac{\partial f}{\partial y}(t^j + h_{j,}, y^j + h_j k_3^j, \widetilde{\lambda}(t_j)\widetilde{\psi}(t_j)) \left[E_n + h_j\frac{\partial k_1^j}{\partial y^j}\right],
\end{aligned}$$

here $\partial k_s^j/\partial y$, $s = 1, \dots, 4$ are n-dimensional square matrix functions, E_n is an n-dimensional unit square matrix.

It is clear that if some other different method is applied to solve the Cauchy problem (1)–(2), other specific formulas will be obtained from (12) that do not coincide with (13), (15).

Then from (7), (9), (15) we obtain:

$$p^{N_\mu} = \overline{h}_N \sigma_N \frac{\partial f_0}{\partial y}(\overline{t}_N, y^N, \widetilde{\lambda}(\overline{t}_N)\widetilde{\psi}(\overline{t}_N)) + \frac{\partial \Phi}{\partial y}(y^N). \tag{16}$$

It is evident that condition (16) for the adjoint system at the right end of time interval for the case of using the Runge-Kutta method coincides with condition (14) obtained when applying Euler's method.

Then the following formulas hold for the components of the gradient of the functional $I(\nu)$:

$$\frac{dI(v)}{d\lambda^k} = \frac{\partial I(v)}{\partial \lambda^k} + \sum_{s \in S_{\beta_k}} \frac{\partial y^s}{\partial \lambda^k}\frac{dI(v)}{dy^s} = \frac{\partial I(v)}{\partial \lambda^k} + \sum_{s=\chi_k}^{\chi_{k+1}-1} \frac{\partial y^s}{\partial \lambda^k} p^s, \quad k = 1, \dots, \mu, \tag{17}$$

$$\begin{aligned}
\frac{dI(v)}{d\tau_k} &= \frac{\partial I(v)}{\partial \tau_k} + \sum_{s \in S_{\tau_k}} \sum_{\nu=1}^{n} \frac{\partial y^s}{\partial \tau_k}\frac{dI(v)}{dy^s}\\
&= \sum_{\nu=1}^{n} \left[(\tau_k - t_{\chi_{k-1}}) p_\nu^{\chi_{k-1}} - (t_{\chi_k} - \tau_k) p_\nu^{\chi_k}\right], \quad k = 1, \dots, \mu - 1.
\end{aligned} \tag{18}$$

In (17), the derivatives $\partial y_s^k/\partial \lambda^k$, $\partial I(\nu)/\partial \lambda^k$ are determined by the approximation schemes of problem (1), (2) and functional (3). Since only regular points t_i, $i = 1, \dots, N_\mu$, are used when approximating the functional (3) then $\partial I(v)/\partial \tau_k = 0$ in (18).

Let us consider obtaining calculation formulas for the elements of the matrix $d^2 I(v)/d\lambda^2$:

$$\begin{aligned}
\frac{d}{d\lambda^{k_2}}\left(\frac{dI(v)}{d\lambda^{k_1}}\right) &= \frac{\partial}{\partial \lambda^{k_2}}\left(\frac{dI(v)}{d\lambda^{k_1}}\right) + \sum_{s_2 \in S_{\lambda^{k_2}}} \frac{\partial y^{s_2}}{\partial \lambda^{k_2}}\frac{d^2 I(v)}{d\lambda^{k_1} dy^{s_2}}, \quad k_1, k_2 = 1, \dots, \mu - 1,\\
\frac{\partial}{\partial \lambda^{k_2}}\left(\frac{dI(v)}{d\lambda^{k_1}}\right) &= \frac{\partial^2 I(\nu)}{\partial \lambda^{k_2}\partial \lambda^{k_1}} + \sum_{s_1 \in S_{\lambda^{k_1}} \cap S_{\lambda^{k_2}}} \frac{\partial^2 y^{s_1}}{\partial \lambda^{k_1}\partial \lambda^{k_2}}\frac{dI(v)}{dy^{s_1}} + \sum_{s_1 \in S_{\lambda^{k_1}}} \frac{\partial y^{s_1}}{\partial \lambda^{k_1}}\frac{\partial}{\partial \lambda^{k_2}}\frac{dI(v)}{dy^{s_1}},\\
\frac{d^2 I(v)}{d\lambda^{k_1} dy^{s_2}} &= \frac{\partial}{\partial \lambda^{k_1}}\frac{dI(v)}{dy^{s_2}} + \sum_{s_3 \in S_{\lambda^{k_2}}} \frac{\partial y^{s_3}}{\partial \lambda^{k_1}}\frac{d^2 I(v)}{dy^{s_2} dy^{s_3}}.
\end{aligned}$$

Finally, for the elements of the matrix $d^2I(v)/d\lambda^2$ we obtain:

$$\begin{aligned}\frac{d^2I(v)}{d\lambda^{k_1}d\lambda^{k_2}} &= \frac{\partial^2 I(v)}{\partial\lambda^{k_2}\partial\lambda^{k_1}} + \sum_{s_1\in S_{\lambda^{k_1}}} \frac{\partial^2 y^{s_1}}{\partial\lambda^{k_1}\partial\lambda^{k_2}}\frac{dI(v)}{dy^{s_1}} + \sum_{s_1\in S_{\lambda^{k_1}}} \frac{\partial y^{s_1}}{\partial\lambda^{k_1}}\frac{\partial}{\partial\lambda^{k_2}}\frac{dI(v)}{dy^{s_1}} \\ &+ \sum_{s_2\in S_{\lambda^{k_2}}} \frac{\partial y^{s_2}}{\partial\lambda^{k_2}}\frac{\partial}{\partial\lambda^{k_1}}\frac{dI(v)}{dy^{s_2}} + \sum_{s_2\in S_{\lambda^{k_2}}}\sum_{s_3\in S_{\lambda^{k_1}}} \frac{\partial y^{s_2}}{\partial\lambda^{k_2}}\frac{\partial y^{s_3}}{\partial\lambda^{k_1}}\frac{d^2I(v)}{dy^{s_2}dy^{s_3}}. \qquad (19)\end{aligned}$$

As can be seen from (19), when calculating matrix $d^2I(v)/d\lambda^2$, it is required to calculate matrices $\partial(dI(v)/dy)/\partial\lambda$ and $d^2I(v)/dy^2$. Let us derive recurrent formulas for these matrices. Let us introduce the matrices:

$$A_{gq} = \frac{d^2I(v)}{dy^g dy^q}, \quad B_{kg} = \frac{\partial}{\partial\lambda^k}\frac{dI(v)}{dy^g}, \quad g,q = 1,\ldots,N_\mu, \quad k = 1,\ldots,\mu.$$

It is clear that

$$\frac{d^2I(v)}{dy^g dy^q} = \frac{\partial}{\partial y^g}\frac{dI(v)}{dy^q} + \sum_{i\in S_{y^g}} \frac{\partial y^i}{\partial y^g}\frac{d^2I(v)}{dy^i dy^q}.$$

For the elements of matrix A we obtain recurrent relations:

$$A_{gq} = \frac{\partial}{\partial y^g}\frac{dI(v)}{dy^q} + \sum_{i\in S_{y^\nu}} \frac{\partial y^i}{\partial y^g}A_{iq}.$$

Let us introduce similar relations for matrix $\partial(dI(v)/dy)/\partial y$. Let us denote:

$$\begin{aligned}C_{gq} &= \frac{\partial}{\partial y^g}\frac{dI(v)}{dy^q} = \frac{\partial}{\partial y^g}\left(\frac{\partial I(v)}{\partial y^q} + \sum_{i\in S_{x^q}} \frac{\partial y^i}{\partial y^q}\frac{dI(v)}{dy^i}\right) \\ &= \frac{\partial^2 I(v)}{\partial y^g\partial y^q} + \sum_{i\in S_{y^q}\cap S_{y^g}} \frac{\partial^2 y^i}{\partial y^g\partial y^q}\frac{dI(v)}{dy^i} + \sum_{i\in S_{y^q}} \frac{\partial y^i}{\partial y^q}\frac{\partial}{\partial y^g}\frac{dI(v)}{dy^i} \\ &= \frac{\partial^2 I(v)}{\partial y^g\partial y^q} + \sum_{i\in S_{y^q}\cap S_{y^g}} \frac{\partial^2 y^i}{\partial y^g\partial y^q}\frac{dI(v)}{dy^i} + \sum_{i\in S_{y^q}} \frac{\partial y^i}{\partial y^q}C_{gi}.\end{aligned}$$

Finally, we introduce recurrent relations for the matrix B:

$$\begin{aligned}B_{kg} &= \frac{\partial}{\partial\lambda^k}\frac{dI(v)}{dy^g} = \frac{\partial}{\partial\lambda^k}\left(\frac{\partial I(v)}{\partial y^g} + \sum_{s\in S_{y^g}}^{n} \frac{\partial y^s}{\partial y^g}\frac{dI(v)}{dy^s}\right) \\ &= \frac{\partial^2 I(v)}{\partial\lambda^k\partial y^g} + \sum_{s\in S_{y^g}\cap S_{\lambda^k}} \frac{\partial^2 y^s}{\partial\lambda^k\partial y^g}\frac{dI(v)}{dy^s} + \sum_{s\in S_{y^g}} \frac{\partial y^s}{\partial y^g}\frac{\partial}{\partial\lambda^k}\frac{dI(v)}{dy^s} \\ &= \frac{\partial^2 I(v)}{\partial\lambda^k\partial y^g} + \sum_{s\in S_{y^g}\cap S_{\lambda^k}} \frac{\partial^2 y^s}{\partial\lambda^k\partial y^g}\frac{dI(v)}{dy^s} + \sum_{s\in S_{y^g}} \frac{\partial y^s}{\partial y^g}C_{ks}.\end{aligned}$$

Using the obtained recurrent relations, we calculate the matrices A, C, B, and then matrix $d^2I(v)/d\lambda^2$.

The matrices $d^2I(v)/d\tau d\lambda$ and $d^2I(v)/d\tau^2$ are calculated in a similar way.

Formulas (11)-(18) make it possible to use various types of efficient methods of first order finite-dimensional optimization to solve the approximated problem (5), (7), (10).

Furthermore, if the formulas for Hesse matrix are obtained then second-order methods can be applied.

It is essential that given formulas are accurate and consistent with the used schemes of methods for finite-dimensional approximation of Cauchy problems with respect to direct and adjoint systems of differential equations obtained in the continuous case. For example, conditions (14), (16) are more accurate in comparison with the frequently used formula of approximation of the condition for the adjoint variable by the formula $p^N = \partial\Phi(y(t_{N_\mu}))/\partial y$ [11, 17, 18].

4 Optimization of the number of subintervals

As it is known, frequent switching of operation modes is undesirable for real control systems of technical objects and technological processes. Frequent switching of control modes reduces reliability, robustness of control, worsens the characteristics of the controlled object. Therefore, it is important to determine such a wise number of switching of control modes, at which the increase of this number does not lead to a significant decrease in the main control criterion.

Let $(v^*(t), y^*(t))$ be the optimal pair, which is a solution of the problem (1)-(4) under piecewise continuous control $u(t)$, at which objective functional (3) attains minimal value $J^* = \min_u J(u)$.

It is obvious that the optimal value of the objective functional on the class of piecewise defined functions (5) depends on μ, the number of subintervals of constancy of parameters of control actions, and it is clear that the following relation holds $J^* \leq J^*_{\mu_1}(\nu) \leq J^*_{\mu_2}(\nu)$ if $\mu_2 < \mu_1$, and the following condition is satisfied (see Fig. 1).

$$J^* = \lim_{\mu\to\infty} J^*_\mu(\nu). \tag{20}$$

Due to the differentiability of objective functional (10), it is easy to show that for an arbitrary positive value ε there is $\overline{\mu}^*$, such that the following condition holds:

$$J^*_\mu(\nu) - J^* \leq \varepsilon \quad \text{for} \quad \mu \geq \overline{\mu}^*. \tag{21}$$

The value of ε is determined from practical considerations. To find the appropriate minimum number $\overline{\mu}_\varepsilon$, at which the condition (21) is satisfied, one can use efficient methods of one-dimensional search (the method of the golden section, dichotomy, etc.).

After solving the problem (1)-(4) with a given value of the number of subintervals μ, the number of switching of control parameters can be reduced by decreasing the number of subintervals. For this purpose, the following two methods are used.

If for any subinterval its length is less than given value ε_τ

$$|\tau_s - \tau_{s-1}| \leq \varepsilon_\tau, \tag{22}$$

then subinterval $[\tau_{s-1}, \tau_s)$ is ignored and merged with the previous subinterval $[\tau_{s-2}, \tau_{s-1})$.

If all control parameters on any s-th subinterval differ from the parameters of the $(s-1)$-th subinterval by an amount smaller than the specified value ε_λ, i.e.

$$|\lambda^s_{ij} - \lambda^{s-1}_{ij}| \leq \varepsilon_\lambda, \quad i = 1, \dots, r, \quad j = 1, \dots, m, \tag{23}$$

then, as suggested above, subinterval $[\tau_{s-1}, \tau_s)$ is merged with subinterval $[\tau_{s-2}, \tau_{s-1})$. The application of the above procedure for reducing the number of subintervals will be illustrated in the next section by an example of solving a test problem.

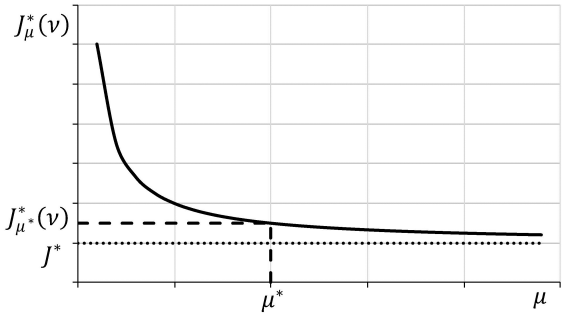

Figure 1: Diagram of $J^*_\mu(\nu)$ with respect to μ.

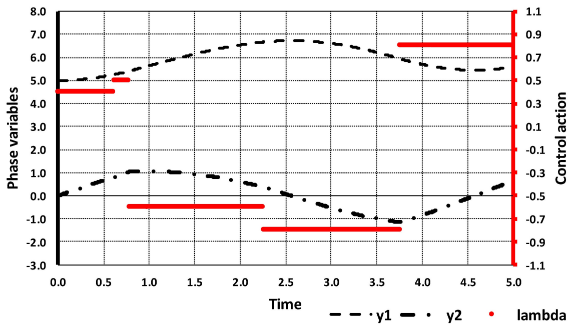

Figure 2: Initial phase trajectory and control action (case 5)

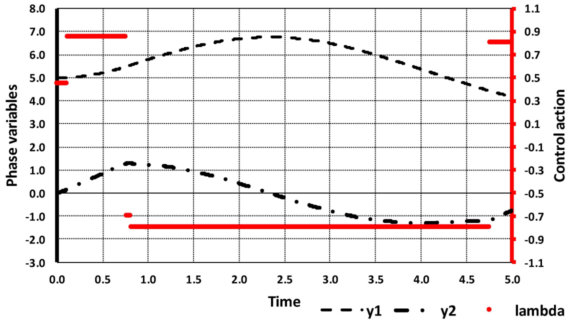

Figure 3: Initial phase trajectory and control action (case 6)

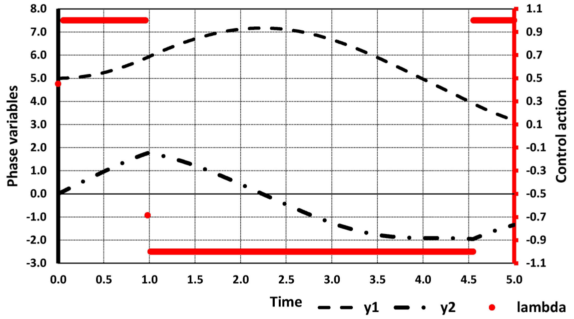

Figure 4: Optimal phase trajectory and control action (case 6)

5 Results of computer experiments

Computer experiments were conducted to numerically analyze the approach described in this paper for applying second-order methods to solve optimal control problems. The following non-linear problem was used as a test problem [19]–[21]:

$$\begin{aligned}
&\dot{y}_1(t) = y_2(t), \quad \dot{y}_2(t) = u(t) - \sin y_1(t), \qquad t \in (0, 5],\\
&y_1(0) = 5, \quad y_2(0) = 0,\\
&J(u) = y_1^2(5) + y_1^2(5) \to \min, \quad |u(t)| \le 1.
\end{aligned}$$

The control is searched among piecewise constant functions with optimizable switching points. In [19]-[21], the optimal solutions obtained by means of various numerical methods are presented. In spite of the fact that the exact solution of this non-linear problem is unknown, the solutions obtained in these works are close enough to each other and they can be considered close to the optimal one.

When solving the direct and adjoint Cauchy problems, the Runge-Kutta method of the fourth order was used, the integration step was chosen equal to $h = 0.01$.

In the numerical solution, different initial values of optimization parameters were used to start the iterative process. The results of the problem solution obtained by the authors are highlighted bold in Tabs. 1 and 2 (non-bold values are taken from [20]).

In the computer experiments, the number of subintervals μ was taken equal to two and four. The tables show the comparative results obtained by the approach proposed in the paper and the results obtained in [20]. In order to reasonably compare the values of the objective functional, the Cauchy problems were additionally solved using

Table 1: Numerical results of solving test problem for initial number of switchings set to 2.

Case	(τ^0, λ^0)	(τ^*, λ^*)	I^*
1	0.780, 3.460, 0.70, -0.60, 0.50	**0.982, 4.550, 1.0, -1.0, 1.0**	**11.908**
		0.950, 4.511, 1.0, -1.0, 1.0	11.962
2	0.780, 3.460, 2.00, -2.00, 0.50	**0.982, 4.550, 1.0, -1.0, 1.0**	**11.908**
		0.950, 4.501, 1.0, -1.0, 1.0	11.967
3	0.520, 2.730, 0.80, -0.80, 0.40	**0.982, 4.550, 1.0, -1.0, 1.0**	**11.908**
		0.950, 4.528, 1.0, -1.0, 1.0	11.944
4	0.280, 3.260, 0.26, -0.40, 0.32	**0.982, 4.550, 1.0, -1.0, 1.0**	**11.908**
		0.950, 4.512, 1.0, -1.0, 1.0	11.961

Table 2: Numerical results of solving test problem for initial number of switchings set to 4.

Case	(τ^0, λ^0)	(τ^*, λ^*)	I^*
5	0.62, 0.78, 2.26, 3.76, 0.4, 0.5, -0.6, -0.8, 0.8	**0.561, 0.982, 2.156, 4.550, 1.00, 1.00, -1.00, -1.00, 1.00**	**11.908**
		0.949, 0.950, 1.858, 4.512, 1.00, 0.584, -1.00, -1.00, 1.00	11.961
6	0.11, 0.76, 0.81, 4.76, 0.45, 0.85, -0.7, -0.8, 0.8	**10^{-17}, 0.982, 0.982, 4.550, 0.45, 1.00, -0.687, -1.00, 1.00**	**11.908**
		0.950, 0.950, 0.950, 4.509, 1.00, 0.835, -0.657, -1.00, 1.00	11.961

Table 3: Numerical results of solving test problem when the number of switching points is less than 2. $-(1)$ – no switching points.

Case	Number of switching points	(τ^0, λ^0)	(τ^*, λ^*)	I^*
7	0	-(1), 0.00	-(1), -1.00	21.80
8	0	-(1), +1.00	-(1), -1.00	21.80
9	1	2.50, +1.00, -1.00	0.00; +1.00; -1.00	21.80
10	1	2.50, -1.00, +1.00	5.00; -1.00; +1.00	21.80
11	1	2.50, 00.00, 00.00	0.00; -0.61; -1.00	21.80

optimal parameters given in [20] (the Runge-Kutta method with step $h = 0.01$ was used for this purpose).

Figs. 2-4 show the plots of initial and obtained controls, as well as the plots of phase trajectories at initial and obtained control.

When analyzing the numerical results, we see at once that the values of the obtained control actions and objective functional are the same for all the four cases 1-4 when the number of switchings is two (see Tab. 1).

For the case of four initial switchings, it is easy to check that the solutions for both cases 5 and 6 can be converted to the results shown in Tab. 1 in view of the reasonings considered in Sec. 3. In case 5, we can apply condition (23) and merge subintervals $[0, 0.561]$ with $[0.561, 0.982]$ and $[0.982, 2.156]$ with $[2.156, 4.550]$. In case 6, condition

(22) can be used and we can delete two "microscopic" subintervals where control action is equal to 0.450 and -0.687 (see Tab. 2 and Fig. 4). After doing that we obtain the solutions which are the same as given for cases 1-4.

To analyze the influence of the number of switching points of control action, additional numerical experiments were carried out: with no switching points (cases 7, 8) and with one switching point (cases 9, 10, 11). The results are shown in Tab. 3.

If the solution is searched in the class of constant functions then optimal value of control action is -1. If the number of switching instances is one, then actually the same result is obtained. In case 9, we observe a needle-shaped value $\lambda = 1$ at the starting time instance $t = 0$, and then the value of λ is set to -1 till the end of the time interval. In case 10, a needle-shaped value $\lambda = 1$ is observed at the end time instance $t = 5$, while the value of λ is set to -1 at the whole interval with the exception for the last point $t = 5$. As for case 11, its solution is in effect identical to that for case 9. The only difference is in the value of the needle-shaped control action, in this case $\lambda = -0.61$

Summing up, the solutions obtained by the proposed method gave a value of the objective functional lesser than in [20], which is the result of more accurate finding of the values of the switching times.

Conclusion

The paper proposes to use second-order optimization methods for numerical solution of one class of optimal control problems. The specific feature of the considered class of problems is as follows. The time interval considered in the problem is divided into subintervals, at each of which the control actions are represented as a linear combination of the given basis functions. The parameters to be optimized are 1) the coefficients at the basis functions at each subinterval and 2) the subinterva1s' boundaries (i.e., the switching times of the control parameters).

The formulas of the first and second order derivatives of the objective functional with respect to the parameters to be optimized are obtained in the paper. This allows us to use efficient numerical methods of Newtonian type optimization. Due to the specific features of the used class of control actions, the dimension of the optimized vector is small even for practical problems.

The paper presents an algorithm for determining the number of subintervals at which the control of the process is robust and efficient. The results of computer experiments for a known test problem are given.

References

[1] Mordukhovich B.S., *Second-Order Variational Analysis in Optimization, Variational Stability, and Control: Theory, Algorithms, Applications. Cham.* Switzerland: Springer, Series in Operations Research and Financial Engineering, 2024.

[2] Tyatyushkin A.I., Zarodnyuk T.S., Gornov A.Y., *Algorithms for nonlinear optimal control problems based on the first and second order necessary conditions.* Journal of Mathematical Sciences, 239. 2 (2019), 185-196.

[3] Abbasov M.E., *Calculus of second order coexhausters.* Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes, 14.4 (2018), 276-285.

[4] Karmitsa N., Bagirov A., Mäkelä M.M., *Comparing different nonsmooth minimization methods and software.* Optimization Methods and Software, 27.1 (2012), 131-153.

[5] Abbasov M.E., *Second-order minimization method for nonsmooth functions allowing convex quadratic approximations of the augment.* Journal of Optimization Theory and Applications, 171.2 (2015), 666-674.

[6] Anikin A., Gornov A., *An implementation of Newton's method for Keating's potential optimization problems.* Studia Informatica Universalis, 9.3 (2011), 11-20.

[7] Skorik G.G., Vasin V.V., *Regularized Newton type method for retrieval of heavy water in atmosphere by Ir-spectra of the solar light transmission.* Eurasian Journal of Mathematical and Computer Applications, 7.2 (2019), 79-88.

[8] Handzel A.V., *Second-order method in network problems*, Bulletin of the National Academy of Sciences of Azerbaijan. Series: Physics, Engineering and Mathematics, 1 (1989), 89-91. (in Russian)

[9] Aida-zade K.R., Evtushenko Yu.G., *Fast automatic differentiation on a computer.* Mathematical Models and Computer Simulations, 1.1 (1989), 120-131. (in Russian)

[10] Evtushenko Yu.G., *Optimization and fast automatic differentiation.* Dorodnicyn Computing Centre of Russian Academy of Sciences, Moscow, 2013. (in Russian)

[11] Aida-zade K.R., Talybov S.G., *On consistency of schemes for finite difference approximation of boundary value problems in optimal control.* Bulletin of the National Academy of Sciences of Azerbaijan. Series: Physics, Engineering and Mathematics, 6 (1998), 21-25.(in Russian)

[12] Li R., Teo K.L., Wong K.H., Duan G.R., *Control parameterization enhancing transform for optimal control of switched systems*, Mathematical and Computer Modelling, 43.11-12 (2006), 1393-1403.

[13] Kabanikhin S.I., *Inverse and Ill-posed Problems*, Theory and Applications, Berlin, Boston: De Gruyter, 2011.

[14] Karchevsky A.L., *Reformulation of an inverse problem statement that reduces computational costs.* Eurasian Journal of Mathematical and Computer Applications, 1.2 (2013), 4-20.

[15] Romanov V.G., Karchevsky A.L., *Determination of permittivity and conductivity of medium in a vicinity of a well having complex profile.* Eurasian Journal of Mathematical and Computer Applications, 6.4 (2018), 62-.72.

[16] Vasil'ev F.P., *Optimization methods.* MCCME, Moscow, 2011. (in Russian)

[17] Aida-zade K.R., Evtushenko Yu.G, Talybov S.G., *Numerical schemes for solving problems of optimal control of objects with distributed parameters.* Bulletin of the National Academy of Sciences of Azerbaijan. Series: Physics, Engineering and Mathematics, 5. (1985), 34-40. (in Russian)

[18] Aida-zade K.R., *Study and numerical solution of finite difference approximations of distributed-system control problems.* Computational Mathematics and Mathematical Physics, 29.2 (1989), 15-21.

[19] Rahimov A.B., *On an approach to solution to optimal control problems on the classes of piecewise constant, piecewise linear, and piecewise given functions.* Tomsk State University Journal of Control and Computer Science, 19.2 (2012), 20-30. (in Russian)

[20] Aida-zade K.R., Rahimov A.B., *Solution of Optimal Control Problem in Class of Piecewise-Constant Functions.* Automatic Control and Computer Science, 41.1 (2007), 18-24.

[21] Vasil'ev O.V., Tyatyushkin A.I., *A method for solving optimal control problems based on the maximum principle.* Computational Mathematics and Mathematical Physics, 21.6 (1981), 14-22. (in Russian)

Kamil Aida-zade,
Institute of Control Systems, Baku, Azerbaijan,
Azerbaijan University of Architecture and Construction, Baku, Azerbaijan,
Email: kamil_aydazade@rambler.ru,

Alexander Handzel,
North-Caucasus Federal University,
Stavropol, Russia,
Email: hndalxvld@yahoo.com

Received 02.09.2024 , Accepted 15.01.2025, Available online 31.03.2025.