ISSN 1995-0802, Lobachevskii Journal of Mathematics, 2018, Vol. 39, No. 7, pp. 916–928. © Pleiades Publishing, Ltd., 2018.

Iterative Method for Solving Parabolic Linear-Quadratic Optimal Control Problem with Constraints on the Time Derivative of the State

A. V. Lapin[1,2*] and A. D. Romanenko[1]**

(Submitted by A. M. Elizarov)

[1] *Institute of Computational Mathematics and Information Technologies, Kazan (Volga Region) Federal University, Kremlevskaya ul. 18, Kazan, Tatarstan, 420008 Russia*

[2] *Coordinated Innovation Center for Computable Modeling in Management Science Tianjin University of Finance and Economics, Tianjin, 300222 P. R. China*

Received October 31, 2017

Abstract—We consider a linear-quadratic optimal control problem of a system governed by parabolic equation with distributed in right-hand side control and control in Neumann boundary condition. Pointwise constraints for control functions and for time derivative of the state function are imposed. We construct a mesh approximation of this problem using two different approximations of the objective functional. Iterative solution methods are investigated for the constructed approximations of the optimal control problems. Numerical results confirm the effectiveness of the proposed methods.

DOI: 10.1134/S199508021807017X

Keywords and phrases: *Parabolic optimal control, state constraints, finite difference method, constrained saddle point problem, iterative method.*

1. INTRODUCTION

Optimal control of time-dependent production processes plays an important role in many industrial applications such as continuous casting problem, crystal growth, cooling of glass melts etc. (see [1–5]). These processes are described by parabolic partial differential equations involving the temperature as a state variable. A need to avoid the defects of the product leads to pointwise constraints on the temperature variable. As a result, mathematical models of the processes appear as state constrained parabolic optimal control problems. Theoretical aspects of parabolic optimal control problems with pointwise constraints for state function are studied in [6–11] (see also the bibliography therein). Numerical solution methods are analysed in [12–20]. In particular, in the articles [12–15] error estimates are derived for discretizations of several classes of these problem, in [16, 17] convergence of regularization methods are proved. Uzawa-type iterative solution methods for finite dimensional problems approximating state constrained parabolic optimal control problems with pointwise constraints for state function are developed in the articles [18–20]. The parabolic optimal control problems with pointwise constraints for time derivative of the state function are considered in [21–23]. In these papers, convergence of Uzawa-type iterative solution methods for finite dimensional approximations of the mentioned problems are proved.

In this paper we consider a parabolic optimal control problem with distributed and boundary control and with observation in the domain. Constraints on the control and on time derivative of state are imposed. We approximate this problem by backward Euler finite difference scheme, prove the existence of a solution and develop iterative solution methods. We construct preconditioned Uzawa-type iterative

* E-mail: avlapine@mail.ru

** E-mail: romart92@mail.ru

solution method with block diagonal preconditioner for the corresponding saddle point problem. The preconditioner is energy equivalent to the "main" matrix of the problem with the constants of the equivalence which don't depend on mesh parameters.

2. FORMULATION OF THE PROBLEM AND ITS APPROXIMATION

2.1. Formulation of Optimal Control Problem

Let $\Omega \subset \mathbb{R}^2$ be a a rectangular domain with the boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$, where Γ_N is a side of $\partial\Omega$, $Q_T = \Omega \times (0,T]$, $\Sigma_D = \Gamma_D \times (0,T]$ and $\Sigma_N = \Gamma_N \times (0,T]$. We denote by $V = \{u \in H^1(\Omega) : u(x) = 0 \text{ on } \Gamma_D\}$ Sobolev space with inner product $(u,v)_V = \int\limits_\Omega \nabla u \cdot \nabla v dx$ and norm $||u||_V = (u,u)^{1/2}$. We consider a parabolic initial-boundary value problem that is used as a state problem:

$$\frac{\partial y}{\partial t} - \Delta y = u \text{ in } Q_T; \quad y = 0 \text{ on } \Sigma_D, \quad \frac{\partial y}{\partial n} = q \text{ on } \Sigma_N; \quad y = 0 \text{ for } t = 0, \quad x \in \Omega. \tag{1}$$

The functions $u = u(x,t)$ and $q = q(x,t)$ are variable control functions, and the solution $y(x,t)$ of (1) is a state function.

If $q \in W = L_2(0,T; H^{1/2}(\Gamma_N)) \cap H^{1/4}(0,T; L_2(\Gamma_N))$, then there exists a unique solution y of problem (1), such that $y \in L_\infty(0,T;V) \cap H^1(0,T;L_2(\Omega))$ and the following stability inequality takes place ([25]):

$$\sup_{0 \leqslant t \leqslant T} ||y(t)||_V + \left|\left|\frac{\partial y(t)}{\partial t}\right|\right|_{L_2(Q_T)} \leqslant C\left(||u(t)||_{L_2(Q_T)} + ||q(t)||_W\right), \quad C = \text{const}. \tag{2}$$

The mentioned regularity properties of state function y allow us to define pointwise constraints for y and for $\partial y/\partial t$ as well. We will consider the following sets of the constraints:

$$U_{ad} = \{u \in L_2(Q_T) : |u(x,t)| \leqslant u_{\max} \text{ a.e. } (x,t) \in Q_T\}, \quad Q_{ad} = \{q \in W : |q| \leqslant \bar{q} \text{ a.e. } \Sigma_N\},$$

$$Y_{ad} = \{y \in L_2(0,T; H_0^1(\Omega)) : \frac{\partial y}{\partial t} \in L_2(Q_T), \quad dy_{\min} \leqslant \frac{\partial y}{\partial t} \leqslant dy_{\max} \text{ a.e. } Q_T\}.$$

Above constants $\bar{u} > 0$, $\bar{q} > 0$ and $-\infty \leqslant dy_{\min} < 0 < dy_{\max} \leqslant \infty$.

Let the objective function be defined as

$$J(y,u,q) = \frac{1}{2}\int\limits_{Q_T} (y(x,t) - y_d(x,t))^2 dxdt + \frac{\varepsilon}{2}\int\limits_{Q_T}\left(\frac{\partial y}{\partial t} - z_d(x,t)\right)^2 dxdt + \frac{1}{2}\int\limits_{Q_T} u^2 dxdt + \frac{1}{2}\int\limits_{\Sigma_N} q^2 d\Gamma dt, \tag{3}$$

where $y_d(x,t), z_d(x,t) \in L_2(Q_T)$ are given observation functions and $\varepsilon \geqslant 0$.

Remark 1. When considering an optimal control problem which objective functional doesn't contain summand with $\partial y_d/\partial t$, we approximate it by a mesh objective function with such kind of summand with a small $\varepsilon > 0$ depending on mesh parameters. This allows us to construct a new iterative solution algorithm for the corresponding mesh optimal control problem.

We will solve the following optimal control problem:

$$\min_{(y,u,q) \in K} J(y,u,q), \quad K = \{(y,u,q) \in Y_{ad} \times U_{ad} \times Q_{ad} : \text{equation (1) holds}\}. \tag{4}$$

Lemma 1. *Problem (4) has a unique solution* (y,u,q).

Proof. The sets of constraints U_{ad}, Q_{ad} and Y_{ad} are convex, closed and contain zero elements, moreover U_{ad} and Q_{ad} are bounded. These properties together with linearity of state equation and stability inequality (2) ensure that the set K is convex, closed, bounded and nonempty. Functional $J = J(y,u,q)$ is convex and continuous. The established properties of J and K ensure the existence of a solution to problem (4). Its uniqueness follows from the strict convexity of the functional J on the set K. To prove this property of J we observe that it is convex in y and strictly convex in u and q, and the equalities $u_1 = u_2$ and $q_1 = q_2$ imply $y_1 = y_2$ for the solutions of problem (1). □

2.2. Finite Difference Approximation

We approximate problem (4) by using Q_1 finite elements and simplest quadrature formulas (cf., e.g., [26]).

Let T_h be the set of mesh cells—nonoverlapping closed rectangales e (finite elements) with maximal diameter h,—which composes a conforming and regular triangulation $\overline{\Omega} = \bigcup\limits_{e\in T_h} e$ of $\overline{\Omega}$. We consider for the simplicity that the function y_d is continuous (otherwise we can deal with mean values of y_d on the cells). We suppose that T_h generates the triangulation ∂T_h on $\overline{\Gamma}_N$, i.e. $\overline{\Gamma}_N$ consists of integer number of sides ∂e of elements $e \in T_h$. We define the finite element space $V_h \subset V$ of the continuous and piecewise bilinear functions (bilinear on each e) which vanish on the boundary Γ_D. By Q_h we denote the space of continuous piecewise linear functions on Γ_N (linear on each $\partial e \in \Gamma_N$), which are the traces on Γ_N of the functions from V_h.

To approximate the integrals of a continuous function $g(x)$ over a finite element $e \in T_h$ or its side ∂e we use the quadrature formulas

$$\int\limits_e g(x)dx \approx S_e(g) = \frac{1}{4}\mathrm{meas}(e)\sum_{\alpha=1}^{4} g(x_\alpha), \quad x_\alpha \text{ are the vertices of } e,$$

$$\int\limits_{\partial e} g(x)d\Gamma \approx S_{\partial e}(g) = \frac{1}{2}\mathrm{meas}(\partial e)\sum_{\alpha=1}^{2} g(x_\alpha), \quad x_\alpha \text{ are the vertices of } \partial e.$$

The corresponding composite quadrature formulas are $S_\Omega(g) = \sum_{e\in T_h} S_e(g)$, $S_\Gamma(g) = \sum_{\partial e\in \partial T_h} S_{\partial e}(g)$. Let further $\omega_t = \{t_j = j\tau, j = 0, 1, \dots N_t; N_t\tau = T\}$ be a uniform mesh on the segment $[0, T]$. We denote by y_h with subscript h a mesh function from the space V_h or Q_h and by y_h^j a time depending mesh function at a time level $t_j \in \omega_t$. Let also y_{dh}^j and z_{dh}^j be the continuous and piecewise linear in x functions which coincide, respectively, with $y_d(x, t_j)$ and $z_d(x, t_j)$ at the nodes of the triangulation T_h.

We approximate state problem (1) by the implicit (backward Euler) scheme:

$$S_\Omega\left(\frac{y_h^j - y_h^{j-1}}{\tau} z_h\right) + S_\Omega\left(\nabla y_h^j \cdot \nabla z_h\right) = S_\Omega(u_h^j z_h) + S_\Gamma(q_h^j z_h) \quad \forall z_h \in V_h \tag{5}$$

for $j = 1, 2, \dots, N_t$ and with initial value $y_h^0 = 0$. It is well-known that problem (5) is unconditionally stable and the following stability inequality holds:

$$\sum_{j=1}^{N_t} S_\Omega\left(|y_h^j|^2\right) \leqslant C_T\left(\sum_{j=1}^{N_t} S_\Omega\left(|u_h^j|^2\right) + \sum_{j=1}^{N_t} S_\Gamma\left(|q_h^j|^2\right)\right), \quad C_T = \mathrm{const}. \tag{6}$$

Remark 2. We can use another approximations of the state equation, such as explicit in time, ADI or fractional steps methods, which satisfy stability inequality like (6) (possibly, under some conditions for mesh steps h and τ, as for the explicit in time approximations). All forthcoming results on the existence of the solutions to saddle point problems and convergence of their iterative solution methods remain valid. Below we consider the specific case $\varepsilon = 0$. The investigations in the case of $\varepsilon > 0$ doesn't differ from this one. Let us introduce the auxiliary mesh functions and constants:

$$p_h^j = y_h^j - y_h^{j-1}, \quad z_{dh}^j = y_{dh}^j - y_{dh}^{j-1}, \quad p_1 = \tau dy_{\min}, \quad p_2 = \tau dy_{\max}.$$

The initial mesh objective function in the case $\varepsilon = 0$ reads as follows:

$$J_{0h}(y_h, u_h, q_h) = \frac{\tau}{2}\sum_{j=1}^{N_t}\left(S_\Omega((y_h^j - y_{dh}^j)^2) + S_\Omega(u_h^j)^2 + S_\Gamma(q_h^j)^2\right). \tag{7}$$

The function J_{0h} doesn't depend explicitly on p_h. Constructing Lagrange function for the corresponding mesh optimal problem we obtain a kind of degenerate saddle point problem. Equivalent transformation of this saddle point problem and iterative solution method was developed in [23].

In this article we use another approach. Namely, we add to mesh objective function J_{0h} a finite dimensional counterpart of the summand $(\tau^2/2)\int_Q (\partial y/\partial t - z_d(x,t))^2 dxdt$ with $z_d \sim \partial y_d/\partial t$. This results in the modified mesh objective function

$$J_h(y_h, p_h, u_h, q_h) = \frac{\tau}{2}\sum_{j=1}^{N_t}\left(S_\Omega((y_h^j - y_{dh}^j)^2) + S_\Omega((p_h^j - z_{dh}^j)^2) + S_\Omega(u_h^j)^2 + S_\Gamma(q_h^j)^2\right). \tag{8}$$

Let also the sets of constraints for mesh optimal control problem be given by the following equalities:

$$\begin{aligned} U_{ad}^h &= \{u_h : V_h \times \omega_t \to \mathbb{R} : |u_h^j| \leqslant \bar{u} \forall x \in \Omega, j = 1, 2, \ldots, N_t\}, \\ Q_{ad}^h &= \{q_h : Q_h \times \omega_t \to \mathbb{R} : |q_h^j| \leqslant \bar{q} \forall x \in \Gamma_N, j = 1, 2, \ldots, N_t\}, \\ P_{ad}^h &= \{p_h : V_h \times \omega_t \to \mathbb{R} :: p_1 \leqslant p_h^j \leqslant p_2 \forall x \in \ \Omega, j = 1, 2, \ldots, N_t\}. \end{aligned} \tag{9}$$

Approximation procedures result to the following mesh optimal control problem:

$$\begin{aligned} &\text{find} \min_{(y_h, p_h, u_h, q_h) \in K_h} J_h(y_h, p_h, u_h, q_h), \\ &K_h = \{(y_h, p_h, u_h, q_h) : p_h \in P_{ad}^h, u_h \in U_{ad}^h, q_h \in Q_{ad}^h, \text{ equation (5) holds}\}. \end{aligned} \tag{10}$$

Lemma 2. *Mesh optimal control problem (10) has a unique solution* (y_h, p_h, u_h, q_h).

Proof. The result follows from the facts that the set K_h is nonempty, closed, convex and bounded, while the function J_h is continuous and strictly convex on K_h. □

3. SADDLE POINT PROBLEM AND ITERATIVE SOLUTION METHOD

3.1. Algebraic form of Problem (10), Saddle Point Problem

Denote by $y \in \mathbb{R}^{N_y}$ the vector of nodal values of a function $y_h \in V_h$ ($N_y = \dim V_h$). Then we get the "onto" correspondence $y \Leftrightarrow y_h$. Similarly a vector $q \in \mathbb{R}^{N_q}$ corresponds to $q_h \in Q_h$.[1)] By $(.,.)_y$, $(.,.)_q$ and $||.||_y$, $||.||_q$ we denote the inner products and euclidian norms in $\mathbb{R}^{N_y}$ and $\mathbb{R}^{N_q}$, respectively. By $(.,.)$ and $||.||$ we denote the inner product and euclidian norm in $\mathbb{R}^{N_t N_y}$ while $[.,.]$ and $[.]$ mean the inner product and euclidian norm in $\mathbb{R}^{N_t N_q}$.

Let $y \Leftrightarrow y_h \in V_h$, $z \Leftrightarrow z_h \in V_h$ and $q \Leftrightarrow q_h \in Q_h$, $p \Leftrightarrow p_h \in Q_h$. We define stiffness matrix $A \in \mathbb{R}^{N_y \times N_y}$, diagonal matrices $\tilde{M} \in \mathbb{R}^{N_y \times N_y}$ and $\tilde{M}_q \in \mathbb{R}^{N_q \times N_q}$ and rectangular matrix $\tilde{S}_q \in \mathbb{R}^{N_y \times N_q}$ by the following equalities:

$$\begin{gathered} (Ay, z)_y = S_\Omega\left(\nabla y_h \cdot \nabla z_h\right), \quad (\tilde{M}y, z)_y = S_\Omega(y_h z_h), \\ (\tilde{M}_q q, p)_q = S_\Gamma(q_h p_h), \quad (\tilde{S}_q q, z)_y = S_\Gamma(q_h z_h). \end{gathered}$$

With these notations mesh state equation (5) can be written for the vectors of nodal values of mesh functions:

$$\tilde{M}\frac{y^j - y^{j-1}}{\tau} + Ay^j = \tilde{M}u^j + \tilde{S}_q q^j, \quad j = 1, 2, \ldots, N_t, \quad y^0 = 0. \tag{11}$$

Further we use the block diagonal matrices with N_t constant blocks, namely, $M = \mathrm{diag}(\tilde{M}, \tilde{M}, \ldots, \tilde{M})$, $M_q = \mathrm{diag}(\tilde{M}_q, \tilde{M}_q, \ldots, \tilde{M}_q)$ and $S_q = \mathrm{diag}(\tilde{S}_q, \tilde{S}_q, \ldots, \tilde{S}_q)$. We define also matrix $L \in \mathbb{R}^{N_t N_y \times N_t N_y}$ as follows:

$$(Ly)^1 = \tilde{M}\frac{y^1}{\tau} + Ay^1; \quad (Ly)^j = \tilde{M}\frac{y^j - y^{j-1}}{\tau} + Ay^j \quad \text{for} \quad j = 2, \ldots, N_t.$$

Now we can rewrite mesh state equation (11) in the following short form $Ly = Mu + S_q q$. Note, that stability inequality (6) implies the estimate $(My, y) \leqslant C_T\big((Mu, u) + [M_q q, q]\big)$.

[1)] Since hereafter we consider only finite dimensional problems, we use the same notations for the vectors as previously for the functions.

Let us now rewrite mesh objective function (8) (divided by τ) in the algebraic form:

$$I(y,p,u,q)=\frac{1}{2}\sum_{j=1}^{N_t}\big((\tilde{M}(y^j-y_d^j),y^j-y_d^j)_y+(\tilde{M}(p^j-z_d^j),p^j-z_d^j)_y+(\tilde{M}u^j,u^j)_y$$
$$+(\tilde{M}_q q^j,q^j)_q\big)\equiv\frac{1}{2}(M(y-y_d),y-y_d)+\frac{1}{2}(M(p-z_d),p-z_d)+\frac{1}{2}(Mu,u)+\frac{1}{2}[M_q q,q].$$

Pointwise constraints (9) can be obviously rewritten for the vectors of nodal values of mesh functions, and we denote by θ, φ_u and φ_q the indicator functions of the sets P_{ad}^h, U_{ad}^h and Q_{ad}^h, respectively. Recall that a function $\phi(w):S\to\mathbb{R}\cup\{+\infty\}$ is the indicator function of a closed and convex set S if $\phi(w)=\{0\ \text{ for }\ w\in S;+\infty\ \text{otherwise}\}$.

As a result we obtain the following algebraic form of mesh optimal control problem (10):

$$\min_{Ly=Mu+S_q q,p=Ry}\{I(y,p,u,q)+\theta(p)+\varphi_u(u)+\varphi_q(q)\}. \tag{12}$$

We take following Lagrange function for problem (12):

$$\mathcal{L}(y,u,q,p,\lambda,\mu)=I(y,p,u,q)+\theta(p)+\varphi_u(u)+\varphi_q(q)+(\lambda,Ly-Mu-S_q q)+(\mu,Ry-p).$$

Its saddle point satisfies [27] the system (saddle point problem)

$$\begin{pmatrix}M&0&0&0&L^T&R^T\\0&M&0&0&-M&0\\0&0&M_q&0&-S_q^T&0\\0&0&0&M&0&-E\\L&-M&-S_q&0&0&0\\R&0&0&-E&0&0\end{pmatrix}\begin{pmatrix}y\\u\\q\\p\\\lambda\\\mu\end{pmatrix}+\begin{pmatrix}0\\\partial\varphi_u(u)\\\partial\varphi_q(q)\\\partial\theta(p)\\0\\0\end{pmatrix}\ni\begin{pmatrix}My_d\\0\\0\\Mz_d\\0\\0\end{pmatrix}, \tag{13}$$

where $\partial\varphi_u$, $\partial\varphi_q$ and $\partial\theta$ are the subdifferentials of the corresponding functions and E is identity matrix. With the notations $z=(y,u,q,p)^T$, $\eta=(\lambda,\mu)^T$, $f=(My_d,0,0,Mz_d)^T$, $\Psi(z)=\theta(p)+\varphi_u(u)+\varphi_q(q)$, and

$$\mathcal{A}=\operatorname{diag}(M,M,M_q,M),\quad B=\begin{pmatrix}L&-M&-S_q&0\\R&0&0&-E\end{pmatrix},$$

problem (13) can be rewritten in a compact form:

$$\begin{pmatrix}\mathcal{A}&B^T\\B&0\end{pmatrix}\begin{pmatrix}z\\\eta\end{pmatrix}+\begin{pmatrix}\partial\Psi(z)\\0\end{pmatrix}\ni\begin{pmatrix}f\\0\end{pmatrix}. \tag{14}$$

Theorem 1. *Problem (13) has a solution (y,u,q,p,λ,μ) with unique y, u, q, p, which coincide with the solution of problem (12).*

Proof. Matrix $\mathcal{A}$ is positive definite, matrix B has a full column rank and function Ψ is convex, proper and lower semicontinuous. Moreover, zero vector satisfies the equation $Bz=0$ and belongs to int domΨ. All these properties ensure the validity of the assumptions of proposition 1 from [24], whence the result. □

3.2. Iterative Solution Method

We consider a preconditioned Uzawa-type iterative method for solving saddle point problem (13):

$$\mathcal{A}z^{k+1}+\partial\Psi(z^{k+1})\ni B^T\eta^k+f,\quad \frac{1}{\rho}D(\eta^{k+1}-\eta^k)+Bz^{k+1}=0, \tag{15}$$

where D is a symmetric and positive definite matrix (preconditioner), $\rho > 0$ is an iterative parameter. Iterative method (15) converges for any initial guess η^0 if the pair (D, ρ) of the preconditioner D and the iterative parameter ρ satisfies the following assumption (24): $D > (\rho/2)B\mathcal{A}^{-1}B^T$. It is easy to see that

$$B\mathcal{A}^{-1}B^T = \begin{pmatrix} LM^{-1}L^T + M + S_q M_q^{-1} S_q^T & LM^{-1}R^T \\ RM^{-1}L^T & RM^{-1}R^T + M^{-1} \end{pmatrix}.$$

An easy invertible block diagonal preconditioner D which is spectrally equivalent to $B\mathcal{A}^{-1}B^T$ with the constants, which don't depend on meshsizes h and τ, is constructed in [23]. More precisely, the following result is true:

Lemma 3. *Matrix* $D = \begin{pmatrix} LM^{-1}L^T & 0 \\ 0 & M^{-1} \end{pmatrix}$ *is spectrally equivalent to* $B\mathcal{A}^{-1}B^T$ *with constants, which don't depend on meshsizes* h *and* τ. *In particular,*

$$(B\mathcal{A}^{-1}B^T\eta, \eta) \leqslant \left(1 + C_T + \sqrt{C_T^2 + 4}\right)(D\eta, \eta) \forall \eta = (\lambda, \mu).$$

As a consequence of this lemma and Theorem 1 from [24] the following statement holds:

Theorem 2. *Method (15) for problem (13) converges if* $0 < \rho < 2/\left(1 + C_T + \sqrt{C_T^2 + 4}\right)$.

Expanded form of method (15) for problem (13) with preconditioner $D = \begin{pmatrix} LM^{-1}L^T & 0 \\ 0 & M^{-1} \end{pmatrix}$ reads as follows:

$$\begin{cases} My^{k+1} = My_d - L^T\lambda^k - R^T\mu^k, \quad Mu^{k+1} + \partial\varphi_u(u^{k+1}) \ni M\lambda^k, \\ M_q q^{k+1} + \partial\varphi_q(q^{k+1}) \ni S_q^T\lambda^k, \quad Mp^{k+1} + \partial\theta(p^{k+1}) \ni Mz_d + \mu^k, \end{cases} \tag{16}$$

$$LM^{-1}L^T\frac{\lambda^{k+1} - \lambda^k}{\rho} = Ly^{k+1} - Mu^{k+1} - S_q q^{k+1}, \quad \frac{\mu^{k+1} - \mu^k}{\rho} = MRy^{k+1} - Mp^{k+1}. \tag{17}$$

On every step of method we have to solve three inclusions in system (16) with diagonal matrices and diagonal operators. The solution of the inclusions is reduced to simple pointwise projections (for all coordinates of nodal vectors on every time level) on the corresponding sets of the constraints.

Solving a system of linear equations with the matrix $LM^{-1}L^T$ consists of sequential solution of the systems with the matrices L and L^T (solution of forward and backward in time linear finite difference equations).

3.3. One More Saddle Point Problem and Iterative Method

Along with method (16), (17) we use in the numerical experiments the iterative method constructed in [23]. For the reader's convenience we give this method and the conditions of its convergence (slightly improved for the problem considered here).

A saddle point problem with the degenerate matrix $\mathcal{A}_0 = \mathrm{diag}(M, M, M_q, 0)$ that appears when using objective function J_{0h} is equivalently transformed to a saddle point problem of the form (14) with the matrix

$$\mathcal{A}_r = \begin{pmatrix} (1 + r_1)M & -r_1 ML^{-1}M & -r_1 ML^{-1}S_q & 0 \\ 0 & M & 0 & 0 \\ 0 & 0 & M_q & 0 \\ -r_2 MR & 0 & 0 & r_2 M \end{pmatrix}, \quad r_1 > 0, \quad r_2 > 0,$$

and B defined above.

For $(r_1, r_2) \in \omega = \{0 < r_1 < (1 + \sqrt{1 + 2C_T})/C_T, 0 < r_2 < 1 + r_1 - r_1^2 C_T/2\}$ matrix $\mathcal{A}_r$ is positive definite and energy equivalent to the matrix $\mathcal{A} = \mathrm{diag}(M, M, M_q, M)$ with constants of the equivalence, which depend only on r_1, r_2:

$$c_0(\mathcal{A}z, z) \leqslant (\mathcal{A}_r z, z) \leqslant c_1(\mathcal{A}z, z), \quad z = (y, u, q, p)^T. \tag{18}$$

Above $0 < c_0 < c_1$ are minimal and maximal eigenvalues of the quadratical form $(1 + r_1)y^2 + u^2 + q^2 + r_2 p^2 - r_1 C_T^{1/2} uy - r_1 C_T^{1/2} qy - 2r_2 yp$. Uzawa-type method for the corresponding saddle point problem has the form:

$$\begin{cases} Mu^{k+1} + \partial\varphi_u(u^{k+1}) \ni M\lambda^k, \quad M_q q^{k+1} + \partial\varphi_q(q^{k+1}) \ni S_q^T \lambda^k, \\ (1 + r_1)My^{k+1} = My_d + r_1 ML^{-1} Mu^{k+1} + r_1 ML^{-1} S_q q^{k+1} - L^T \lambda^k - R^T \mu^k, \\ r_2 Mp^{k+1} + \partial\theta(p^{k+1}) \ni r_2 MRy^{k+1} + \mu^k, \end{cases} \tag{19}$$

$$LM^{-1}L^T \frac{\lambda^{k+1} - \lambda^k}{\rho} = Ly^{k+1} - Mu^{k+1} - S_q q^{k+1}, \quad \frac{\mu^{k+1} - \mu^k}{\rho} = MRy^{k+1} - Mp^{k+1}. \tag{20}$$

Method (19), (20) converges if $(r_1, r_2) \in \omega$ and $0 < \rho < 2c_0 \left[1 + C_T + \sqrt{C_T^2 + 4}\right]^{-1}$ with c_0 is defined in (18).

The implementation of (19), (20) is similar to the implementation of (16), (17).

Remarks on the accuracy control. Let $z = (y, u, q, p)$ and $z^k = (y^k, u^k, q^k, p^k)$ be exact solution and k-th iteration of the corresponding saddle point problem and iterative method. Let also $\mathcal{A} = \mathrm{diag}(M, M, M_q, M)$ and residual vector r^k of the k-th iteration equals $r^k = (r_\lambda^k, r_\mu^k)$ with $r_\lambda^k = Ly^k - Mu^k - S_q q^k$, $r_\mu^k = MRy^k - Mp^k$. According to general estimate from [28] the following estimate holds:

$$||z - z^k||_{\mathcal{A}} = o(||r^k||_{D^{-1}}^{1/2}) = o\left(\left(||r_\lambda^k||_{L^{-T}ML^{-1}}^2 + ||r_\mu^k||_M^2\right)^{1/2}\right). \tag{21}$$

When iterative parameter ρ is close to 1 we can use the vector $(\lambda^k - \lambda^{k-1}, \lambda^k - \lambda^{k-1})$ instead of r^k in this estimate. Note also that energy norm of matrix $\mathcal{A}$ in the left side of (21) corresponds to mesh analog of $L^2(0, T; L^2(\Omega))$-norm for the components $y^k - y, u^k - u$ and $p^k - p$ of the error and to mesh analog of $L^2(0, T; L^2(\Gamma_N))$-norm for the component $q^k - q$.

4. NUMERICAL RESULTS

4.1. Problem with Distributed Control Function

First series of numerical tests were carried out for the problem with distributed control, when the state problem was Dirichlet initial-boundary value problem

$$\frac{\partial y}{\partial t} - \Delta y = u \ \text{ in } \ Q_1, \quad y = 0 \ \text{ on } \ \Sigma_D, \quad y = 0 \ \text{ for } \ t = 0, \quad x \in \Omega. \tag{22}$$

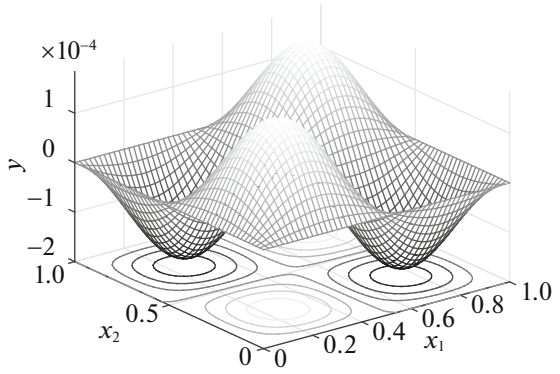

Fig. 1. Plot of state y at $t = T$.

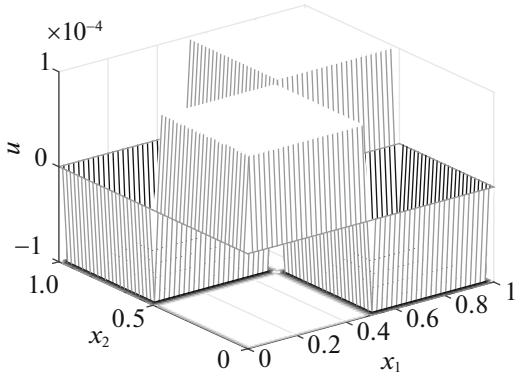

Fig. 2. Plot of control u at $t = T$.

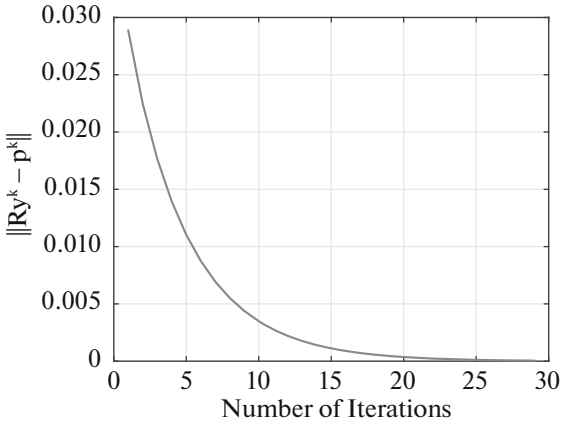

Fig. 3. Plot of residual $||r_\mu^k||$ for $\varepsilon = 1$.

Above $Q_1 = \Omega \times (0,1)$ with square domain $\Omega = (0,1) \times (0,1) \subset \mathbb{R}^2$. We consider objective functional

$$J(y,u) = \frac{1}{2}\int\limits_{Q_1} (y(x,t) - y_d(x,t))^2 dxdt + \frac{\varepsilon}{2}\int\limits_{Q_1} \left(\frac{\partial y}{\partial t} - z_d(x,t)\right)^2 dxdt + \frac{1}{2}\int\limits_{Q_1} u^2 dxdt \tag{23}$$

Table 1. Norm of residual vectors for distributed control problem

k	$\varepsilon = 1$		$\varepsilon = 0$	
	$\|\|r_\lambda^k\|\|$	$\|\|r_\mu^k\|\|$	$\|\|r_\lambda^k\|\|$	$\|\|r_\mu^k\|\|$
1	0.2347	0.0288	0.2347	0.0288
2	0.1882	0.0224	0.1888	0.0231
3	0.1509	0.0176	0.1519	0.0185
4	0.1211	0.0139	0.1222	0.0149
5	0.0971	0.011	0.0984	0.0119
10	0.0323	0.0034	0.0332	0.0039
15	0.0107	0.0011	0.0112	0.0013
20	0.0035	3.5936×10^{-4}	0.0037	4.3532×10^{-4}
25	0.0012	1.145×10^{-4}	0.0013	1.4310×10^{-4}
30	4.0101×10^{-4}	4.4644×10^{-5}	4.2895×10^{-4}	4.5566×10^{-5}
35	1.3371×10^{-4}	–	1.4424×10^{-4}	–
41	4.5393×10^{-5}	–	4.8967×10^{-5}	–
$J(y^k, u^k)$	0.02626		0.02812	

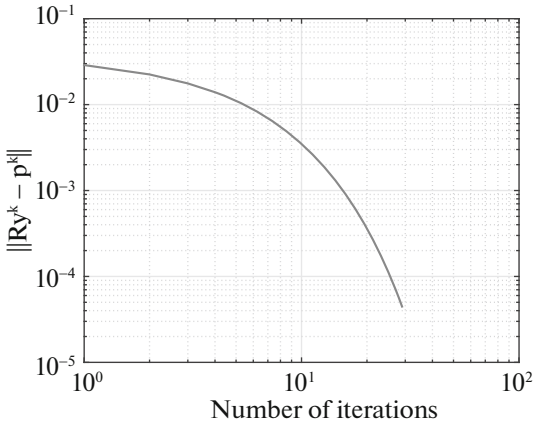

Fig. 4. Plot of logarithm residual $||r_\mu^k||$ for $\varepsilon = 1$.

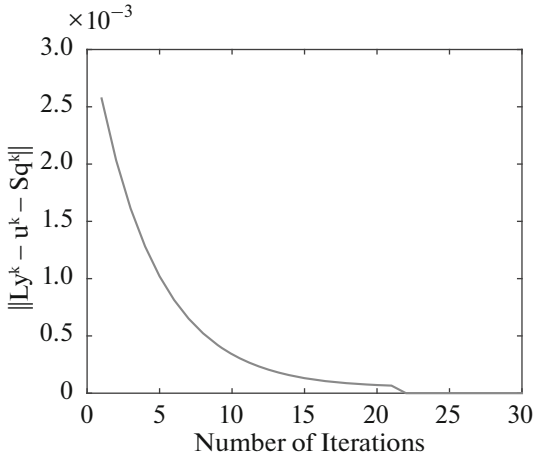

Fig. 5. Plot of residual $||Ly^k - u^k - S_q q^k||_{D^{-1}}$.

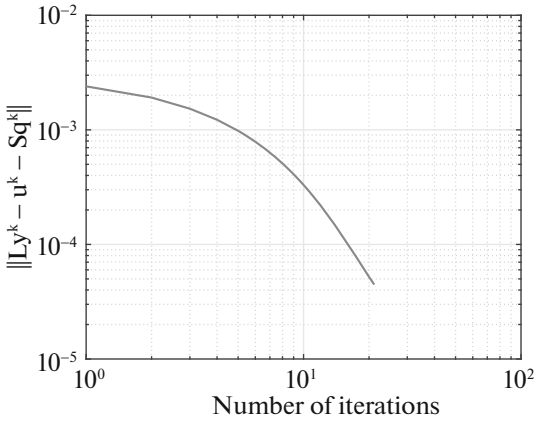

Fig. 6. Plot of logarithm residual $||Ly^k - u^k - S_q q^k||_{D^{-1}}$.

with $\varepsilon = 0$ and $\varepsilon = 1$. The observation function is $y_d(x,t) = t^2 \sin(2\pi x_1)\sin(2\pi x_2)$ and its time derivative $z_d(x,t) = 2t\sin(2\pi x_1)\sin(2\pi x_2)$. Pointwise constraints on control and time derivative of state are taken as follows: $|u| \leqslant 10^{-4}$, $\left|\partial y/\partial t\right| \leqslant 10^{-4}$. These very hard constraints ensure the appearance of a significant number of active control and state constraints.

Table 2. Norm of residual vectors for distributed control problem

k	$\varepsilon = 1$		$\varepsilon = 0$	
	$\|\|r_\lambda^k\|\|$	$\|\|r_\mu^k\|\|$	$\|\|r_\lambda^k\|\|$	$\|\|r_\mu^k\|\|$
1	0.0023	4.784×10^{-4}	0.1715	3.731×10^{-4}
2	0.0019	3.285×10^{-4}	0.0014	3.273×10^{-4}
3	0.0015	2.399×10^{-4}	0.0012	2.885×10^{-4}
4	0.0012	1.96×10^{-4}	0.0011	2.55×10^{-4}
5	9.814×10^{-4}	1.85×10^{-4}	0.001	2.259×10^{-4}
10	3.307×10^{-4}	2.449×10^{-4}	6.12×10^{-4}	1.253×10^{-4}
15	1.203×10^{-4}	2.731×10^{-4}	3.274×10^{-4}	7.025×10^{-5}
20	5.209×10^{-5}	2.788×10^{-4}	2.162×10^{-4}	4.963×10^{-5}
100	–	1.695×10^{-4}	–	–
250	–	1.232×10^{-4}	–	–
500	–	5.032×10^{-5}	–	–
$J(y^k, u^k, q^k)$	9.814×10^{-5}		1.0886×10^{-4}	

We approximate the problem by a mesh scheme using uniform in space mesh with mesh step $h = 10^{-2}$ and mesh in time with $\tau = h$. Let

$$Ay(x) = h^{-2}(4y(x_1, x_2) - y(x_1 + h, x_2) - y(x_1, x_2 + h) - y(x_1 - h, x_2) - y(x_1, x_2 - h))$$

for an internal mesh point x be mesh Laplacian with homogeneous Dirichlet boundary conditions. Then mesh state problem is backward Euler scheme

$$\frac{y^k - y^{k-1}}{\tau} + Ay^k = u^k \; \forall k \geqslant 1.$$

We take iterative parameter $\rho = 0.2$ and use estimate (21) for stopping criterion. Namely, we stop iterations when $||r^k||_{D^{-1}} < \epsilon = 5 \cdot 10^{-5}$. Table 1 shows the results of tests for (22), (23) and consists of two residual columns for each ε. For the brevity we use in the table the notations $||r_\lambda^k||$ and $||r_\mu^k||$ instead of $||r_\lambda^k||_{L^{-T}ML^{-1}}$ and $||r_\mu^k||_M$, respectively.

We emphasize the following features of the numerical results that can be seen from the table:

1. balanced speed reduction of norms for two components of the residual vector;
2. the difference between calculated optimal values for two objective functions is of order τ^2.

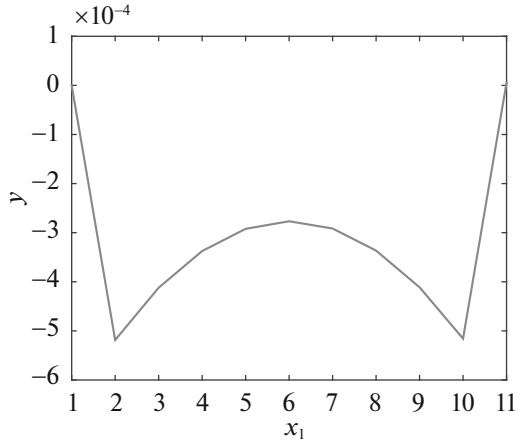

Fig. 7. Section y at $t = \tau$.

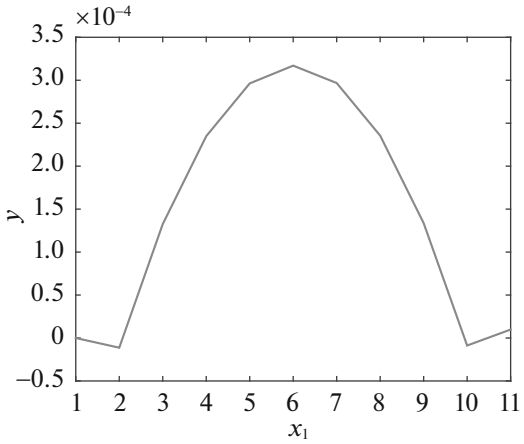

Fig. 8. Section y at $t = 2\tau$.

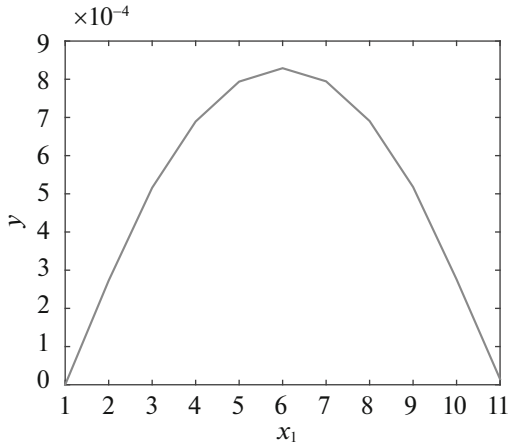

Fig. 9. Section y at $t = 0.5T$.

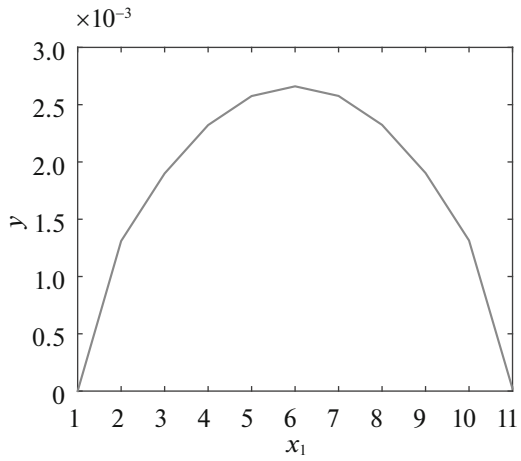

Fig. 10. Section y at $t = T$.

4.2. Problem with Boundary Control Function

In this example we solved optimal control problem with distributed and boundary control. We took the objective functional (3) with parameter $\varepsilon = 1$. Numerical tests were were conducted with the following input data: the observation functions $y_d(x,t) = t^2 x_1(1-x_1)x_2(1-x_2)$, $T = 0.5$, mesh steps $h = 10^{-2}$ and $\tau = h$. Control constraints were the same as in previous example, while the pointwise constraints on time derivative of state function were $\left|\partial y/\partial t\right| \leqslant 10^{-2}$. We took iterative parameter $\rho = 0.2$ and the same stopping criterion as before: $\epsilon = 5 \cdot 10^{-5}$. The calculated results are presented in Table 2 and Figs. 5–10.

ACKNOWLEDGMENTS

This work was supported by Russian Foundation for Basic Research, grant 16-01-00408. First author's research was partly supported by Major Research Plan of the National Natural Science Foundation of China (91430108).

REFERENCES

1. R. Dautov, R. Kadyrov, E. Laitinen, A. Lapin, J. Pieskä, and V. Toivonen, "On 3D dynamic control of secondary cooling in continuous casting process," Lobachevskii J. Math. **13**, 3–13 (2003).
2. M. Gunzburger, E. Ozugurlu, J. Turner, and H. Zhang, "Controlling transport phenomena in the Czochralski crystal growth process," J. Cryst. Growth **234**, 47–62 (2002).
3. D. Clever and J. Lang, "Optimal control of radiative heat transfer in glass cooling with restrictions on the temperature gradient," Preprint SPP1253-20-01 (2008).
4. R. Pinnau, "Analysis of optimal boundary control for radiative heat transfer model led by the SP1-System," Commun. Math. Soc. **5**, 951–969 (2007).
5. M. Hinze and S. Ziegenbalg, "Solidification of a GaAs melt—optimal control of the phase interface," J. Cryst. Growth (2009). doi 10.1016/j.jcrysgro.2009.02.031
6. J. F. Bonnans and E. Casas, "On the choice of the function spaces for some state-constrained control problems," Numer. Funct. Anal. Optim. **7**, 333–348 (1984).
7. E. Casas, "Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations," SIAM J. Control Optim. **35**, 1297–1327 (1997).
8. J. P. Raymond and H. Zidani, "Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls," SIAM J. Control Optim. **36**, 1853–1879 (1998).
9. N. Arada and J. P. Raymond, "Dirichlet boundary control of semilinear parabolic equations, II. Problems with pointwise state constraints," Appl. Math. Optim. **45**, 145–167 (2002).
10. E. Casas, J. C. Reyes and F. Troltzsch, "Sufficient second-order optimality conditions for semilinear control problems with pointwise state constraints," SIAM J. Optim. **19**, 616–643 (2008).
11. J. F. Bonnans and P. Jaisson, "Optimal control of a parabolic equation with time-dependent state constraints," SIAM J. Control Optim. **48**, 4550–4571 (2010).
12. K. Deckelnick and M. Hinze, "Variational discretization of parabolic control problems in the presence of pointwise state constraints," J. Comp. Math. **29**, 1–15 (2011).
13. W. Gong and M. Hinze, "Error estimates for parabolic optimal control problems with control and state constraints," Comput. Optim. Appl. **56**, 131–151 (2013)
14. D. Meidner and B. Vexler, "A priori error estimates for space-time finite element approximation of parabolic optimal control problems. Part II: Problems with control constraints," SIAM J. Control Optim. **47**, 1301–1329 (2008).
15. D. Meidner, R. Rannacher and B. Vexler, "A priori error estimates for finite element discretizations of parabolic optimization problems with pointwise state constraints in time," SIAM J. Control Optim. **49**, 1961–1997 (2011).
16. I. Neitzel and F. Troltzsch, "On convergence of regularization methods for nonlinear parabolic optimal control problems with control and state constraints," Control Cybernet. **37**, 1013–1043 (2008).
17. I. Neitzel and F. Troltzsch, "On regularization methods for the numerical solution of parabolic control problems with pointwise state constraint," ESAIM: Control, Optimiz Calculus Variati. **15**, 426-452 (2009). doi 10.1051/cocv:2008038
18. A. Lapin and A. Romanenko, "Uzawa-type iterative method with parareal preconditioner for a parabolic optimal control problem," IOP Conf. Ser.: Mater. Sci. Eng. **58** (2016).
19. A. Romanenko, "Explicit scheme with variable in time steps for a parabolic optimal control problem," Uch. Zap. Kazan. Univ., Ser. Fiz.-Mat. Nauki **158**, 376–387 (2016).

20. A. Lapin and E. Laitinen, "Iterative solution of mesh constrained optimal control problems with two-level mesh approximations of parabolic state equation," J. Appl. Math. Phys. **6** 58–68 (2018).
21. A. Lapin, E. Laitinen, and S. Lapin, "Explicit algorithms to solve a class of state constrained parabolic optimal control problems," Russ. J. Numer. Anal. Math. Model. **30**, 351–362 (2015).
22. A. Lapin and E. Laitinen, "Iterative solution methods for parabolic optimal control problem with constraints on time derivative of state function," WSEAS Recent Adv. Math.: Math. Comput. Sci. Eng. Ser. **48**, 72–74 (2015).
23. A. Lapin and E. Laitinen, "Preconditioned Uzawa-type method for a state constrained parabolic optimal control problem with boundary control," Lobachevskii J. Math. **37**, 561–569 (2016).
24. A. Lapin, "Preconditioned Uzawa type methods for finite-dimensional constrained saddle point problems," Lobachevskii J. Math. **31**, 309–322 (2010).
25. J.-L. Lions and E. Magenes, *Non-Homogeneous Boundary Value Problems and Applications* (Springer, Berlin, 1972).
26. Ph. G. Ciarlet, *The Finite Element Method for Elliptic Problems* (North-Holland, Amsterdam, 1978).
27. I. Ekeland and R. Temam, *Convex Analysis and Variational Problems* (North-Holland, Amsterdam, 1976).
28. E. Laitinen, A. Lapin, and S. Lapin, "Iterative solution methods for variational inequalities with nonlinear main operator and constraints to gradient of solution," Lobachevskii J. Math. **33**, 364–371 (2012).