Iterative solution methods for parabolic optimal control problem with constraints on time derivative of state function

E. Laitinen, A. Lapin

Abstract—An iterative solution method is proposed and investigated for the finite difference approximation of a parabolic optimal control problem with constraints on time derivative of the state function. Convergence analysis of the iterative methods is made. It is based on the general results on the convergence of iterative methods for constrained saddle point problem ([1], [2], [3]).The main feature of the constructed iterative solution methods is their easy implementation. Computational experiments confirm the theoretical results.

Index Terms—terative methods, saddel point problem, constraints in time derivativeterative methods, saddel point problem, constraints in time derivativei

I. Problem formulation

Let $\Omega = [0,1]^n, n \geqslant 1$, $\partial\Omega$ be its boundary, $Q_T = \Omega \times (0,T]$ and $\Sigma_T = \partial\Omega \times (0,T]$. Define a state problem with distributed control:

$$\frac{\partial y}{\partial t} - \Delta y = f + u \text{ in } Q_T; \quad \begin{array}{l} y = 0 \text{ on } \Sigma_T; \\ y = 0 \text{ for } t = 0,\ x \in \Omega, \end{array} \tag{1}$$

where function $f(x,t) \in L_2(Q_T)$ is given, while $y(x,t)$ and $u(x,t)$ are unknown state and control functions. This problem has a unique weak solution $y \in L_2(0,T; H_0^1(\Omega))$ such that $\frac{\partial y}{\partial t} \in L_2(Q_T)$.

Let objective function be defined by the equality

$$J(y,u) = \frac{1}{2}\int\limits_{Q_T} (y(x,t) - y_d(x,t))^2 dxdt + \frac{\alpha}{2}\int\limits_{Q_T} u^2 dxdt, \quad \alpha > 0, \tag{2}$$

with given observation function $y_d(x,t) \in L_2(Q_T)$.

Finally, define the sets of the constraints:

$$\begin{array}{l} U_{ad} = \{u \in L_2(Q_T) : |u| \leqslant \bar{u} \text{ a.e. } Q_T\}; \\ Y_{ad} = \{y : \frac{\partial y}{\partial t} \in L_2(Q_T) \text{ and } y_{\min} \leqslant \frac{\partial y}{\partial t} \leqslant y_{\max} \text{ a.e. } Q_T\}, \end{array} \tag{3}$$

with given constants $\bar{u} > 0, y_{\min}$ and $y_{\max}$.

We will solve the following optimal control problem:

$$\begin{array}{c} \min\limits_{(y,u)\in K} J(y,u), \\ K = \{(y,u) \in Y_{ad} \times U_{ad} : \text{ equation (1) holds}\}. \end{array} \tag{4}$$

Lemma 1. Problem (4) has a unique solution (y,u) if $K \neq \emptyset$.

E. Laitinen is with the Department of Mathematical Sciences, University of Oulu, Oulu, Finland. E-mail: erkki.laitinen@oulu.fi

A. Lapin is with the Department of Computational Mathematics and Cybernetics, Kazan Federal University, ul. Kremlevskaya, 18, Kazan 420008, Russia. E-mail: avlapine@mail.ru

II. Finite difference approximation of the optimal control problem

We suppose for the simplicity that $f(x,t)$ is a continuous function in $\bar{\Omega} \times [0,T]$. Let ω_x be the uniform mesh of the meshsize h on $\bar{\Omega}$, $\operatorname{card}\omega_x = N_x$. By A we denote the mesh approximation of Laplace operator with homogeneous Dirichlet boundary conditions. Then the spectrum of symmetric and positive definite matrix A belongs to the segment $[\nu_{\min}(A), \nu_{\max}(A)]$, where $\nu_{\max}(A)$ has an order h^{-2}, while $\nu_{\min}(A) > 0$ is limited from below by a constant which doesn't depend on h. For the mesh functions defined on the mesh ω_x and the vectors from $\mathbb{R}^{N_x}$ of their nodal values we will use the same notations. By $(.,.)_x$ and $\|.\|_x$ we denote the inner product and euclidian norm in $\mathbb{R}^{N_x}$. Further, let $\omega_t = \{t_j = j\tau,\ j = 0,1,\dots M;\ M\tau = T\}$ be a uniform mesh on the segment $[0,T]$. Denote by $y_j = y(x,t_j)$ a mesh function on a time level $t_j \in \omega_t$, or equivalently the vector $y_j \in \mathbb{R}^{N_x}$ of its nodal values.

Let us approximate state equation (1) by weighted finite difference:

$$\frac{1}{\tau}(y_j - y_{j-1}) + A(\delta y_j + (1-\delta)y_{j-1}) = f_j + u_j,\ j = 1,\dots,M,\ y_0 = 0 \tag{5}$$

with $\delta \in [0,1]$. We suppose that the stability condition $\tau < 2(\nu_{\max}(A)(1-2\delta))^{-1}$ in the case $\delta < 1/2$ is satisfied. In the case $\delta \geqslant 1/2$ this finite difference problem is unconditionally stable.

Matrix $L \in \mathbb{R}^{MN_x \times MN_x}$:

$$\begin{array}{c} (Ly)_j = \{\frac{1}{\tau}(y_j - y_{j-1}) + A(\delta y_j + (1-\delta)y_{j-1}) \text{ for} \\ j = 2,\dots,M; \frac{1}{\tau}y_1 + \delta A y_1 \text{ for } j = 1\} \end{array}$$

is positive definite (the stability condition condition is supposed to be satisfied in the case $\delta < 1/2$).

The objective function (2) is approximated by the mesh objective function

$$I(y,u) = \frac{1}{2}\sum_{j=1}^{M} \|y_j - y_{dj}\|_x^2 + \frac{\alpha}{2}\sum_{j=1}^{M} \|u_j\|_x^2, \tag{6}$$

while the mesh approximations of the constraints sets (3) are

$$\begin{array}{l} U_{ad}^h = \{u : |u(x,t)| \leqslant \bar{u}\ \forall x \in \omega_x, \forall t \in \omega_t\}, \\ Y_{ad}^h = \{y : \tau y_{\min} \leqslant y_j - y_{j-1} \leqslant \tau y_{\max}\ (y_0 = 0)\ \forall x \in \omega_x, \forall t \in \omega_t\}. \end{array}$$

Now mesh optimal control problem reads as follows:

$$\begin{array}{c} \min\limits_{(y,u)\in K_h} I(y,u), \\ K_h = \{(y,u) \in Y_{ad}^h \times U_{ad}^h : \text{ equation (5) holds}\}. \end{array} \tag{7}$$

Lemma 2. Problem (7) has a unique solution if $K_h \neq \emptyset$.

ISBN: 978-1-61804-323-8

III. Saddle point problem

Let us define matrix $R \in \mathbb{R}^{MN_x \times MN_x}$, $(Ry)_j = \{y_j - y_{j-1}$ for $j = 2, \dots, M$; y_1 for $j = 1\}$, and vector $p = Ry$. Then we can replace the constraint $y \in Y^h_{ad}$ in the optimal control problem by the following constraint: $p \in P^h_{ad} = \{\tau y_{\min} \leqslant p_j \leqslant \tau y_{\max},\ j = 1, 2, \dots M\}$. Let further θ and φ be indicator functions of the sets P^h_{ad} and U^h_{ad}: $\theta(p) = \{0$ if $p \in P^h_{ad}$; $+\infty$ otherwise$\}$, $\varphi(u) = \{0$ if $u \in U^h_{ad}$; $+\infty$ otherwise$\}$. Then mesh optimal control problem (6) can be written as

$$\min_{Ly = f+u,\ p = Ry} \{I(y,u) + \theta(p) + \varphi(u)\}.$$

Define Lagrange function

$$\mathcal{L}(y,u,\lambda) = I(y,u) + \theta(p) + \varphi(u) + (\lambda, Ly - u - f) + (\mu, Ry - p),$$

where $(.,.)$ is the inner product in $\mathbb{R}^{MN_x}$. Its saddle point satisfies the following system (cf. e.g. [4]):

$$\begin{pmatrix} E & 0 & 0 & L^T & R^T \\ 0 & \alpha E & 0 & -E & 0 \\ 0 & 0 & 0 & 0 & -E \\ L & -E & 0 & 0 & 0 \\ R & 0 & -E & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ u \\ p \\ \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} 0 \\ \partial\varphi(u) \\ \partial\theta(p) \\ 0 \\ 0 \end{pmatrix} \ni \begin{pmatrix} y_d \\ 0 \\ 0 \\ f \\ 0 \end{pmatrix}, \tag{8}$$

where $E \in \mathbb{R}^{MN_x \times MN_x}$ is unit matrix, $\partial\varphi(u)$ and $\partial\theta(p)$ are the subdifferentials of φ and θ respectively.

Lemma 3. Let the strengthened variant of the assumption $K_h \neq \emptyset$ be satisfied:

There exists a pair $(y^*, u^*) \in \operatorname{int} Y^h_{ad} \times \operatorname{int} U^h_{ad}$ such that $Ly^* = f + u^*$.

Then saddle point problem (8) has a nonempty solution set $X = \{(w, \eta)\}$ and w is unique.

IV. Iterative methods

Using the notations $w = (y,u,p)^T$, $\eta = (\lambda,\mu)^T$, $g_1 = (y_d, 0, 0)^T, g_2 = (f, 0)^T$, $\partial\psi(w) = (0, \partial\varphi(u), \partial\theta(p))^T$ and

$$\mathcal{A} = \begin{pmatrix} E & 0 & 0 \\ 0 & \alpha E & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} L & -E & 0 \\ R & 0 & -E \end{pmatrix}$$

problem (8) can be written in the following compact form:

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^T \\ \mathcal{B} & 0 \end{pmatrix} \begin{pmatrix} w \\ \eta \end{pmatrix} + \begin{pmatrix} \partial\psi(w) \\ 0 \end{pmatrix} \ni \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}. \tag{9}$$

The degenerate matrix $\mathcal{A}$ is an obstacle to the application of Uzawa-type iterative methods for solving (9). To overcome this deficiency we use two equivalent transformations of (8) and obtain the saddle point problems with positive definite matrices instead of $\mathcal{A}$. In both transformations we use the last equation of system (8), and obtain the variants of saddle point problem (9) with the matrices

$$\mathcal{A}_{1r} = \begin{pmatrix} E & 0 & 0 \\ 0 & \alpha E & 0 \\ -rR & 0 & rE \end{pmatrix} \text{ or}$$

$$\mathcal{A}_{2r} = \begin{pmatrix} E + rR^TR & 0 & -rR^T \\ 0 & \alpha E & 0 \\ -rR & 0 & rE \end{pmatrix},\ r > 0,$$

instead of $\mathcal{A}$ and with the same matrix $\mathcal{B}$, function ψ and vectors g_1, g_2.

Lemma 4. Matrix $\mathcal{A}_{1r}$ is positive definite for $0 < r < 1$ and matrix $\mathcal{A}_{2r}$ is positive definite for any $r > 0$. Moreover, for these parameters r they are energy equivalent to block-diagonal matrix $\mathcal{A}_0 = \operatorname{diag}\begin{pmatrix} E & \alpha E & rE \end{pmatrix}$ with constants of the equivalence, which depend only on r:

$$(1 - \sqrt{r})(\mathcal{A}_0 z, z) \leqslant (\mathcal{A}_{1r} z, z) \leqslant (1 + \sqrt{r})(\mathcal{A}_0 z, z),$$

$$\sigma_0(r)(\mathcal{A}_0 z, z) \leqslant (\mathcal{A}_{2r} z, z) \leqslant \sigma_2(r)(\mathcal{A}_0 z, z),\ \forall z = (y,u,p),$$

where $\sigma_0(r) = (1 + 2r + 2\sqrt{r + r^2})^{-1}$, $\sigma_1(r) = 2r(1 + 5r + \sqrt{1 + 6r + 25r^2})^{-1}$.

A preconditioned Uzawa-type iterative method for solving saddle point problem (9) reads as

$$\begin{aligned} &\mathcal{A}w^{k+1} + \partial\psi(w^{k+1}) \ni \mathcal{B}^T\eta^k + g_1, \\ &\frac{1}{\rho} D(\eta^{k+1} - \eta^k) + \mathcal{B}w^{k+1} = g_2, \end{aligned} \tag{10}$$

where D is a symmetric and positive definite matrix (preconditioner), $\rho > 0$ is an iterative parameter.

Due to [1] iterative method (10) converges for any initial guess η^0 (convergence means $(w^k, \eta^k) \to (w^*, \eta^*) \in X$ for $k \to \infty$) if the pair "preconditioner D - parameter ρ" satisfies one of the following (equivalent) assumptions:

$$\mathcal{A}_s \geqslant \frac{(1+\varepsilon)\rho}{2}\mathcal{B}^T D^{-1}\mathcal{B} \text{ or } D \geqslant \frac{(1+\varepsilon)\rho}{2}\mathcal{B}\mathcal{A}_s^{-1}\mathcal{B}^T,\ \ \varepsilon > 0,$$

where $\mathcal{A}_s = 0.5(\mathcal{A} + \mathcal{A}^T)$ is the symmetric part of $\mathcal{A}$. The optimal preconditioner D is a matrix which is spectrally equivalent to $\mathcal{B}\mathcal{A}_s^{-1}\mathcal{B}^T$: $c_0\mathcal{B}\mathcal{A}_s^{-1}\mathcal{B}^T \leqslant D \leqslant c_1\mathcal{B}\mathcal{A}_s^{-1}\mathcal{B}^T$, with smallest ratio $\dfrac{c_1}{c_0}$.

Our goal is to construct a preconditioner D such that the constants c_0, c_1 don't depend on meshsizes h and τ and on the parameter α, while D is "easily invertible".

Due to Lemma 4 the matrix $B\mathcal{A}_s^{-1}B^T$ is spectrally equivalent to $\mathcal{B}\mathcal{A}_0^{-1}\mathcal{B}^T = \begin{pmatrix} LL^T + \alpha^{-1}E & LR^T \\ RL^T & RR^T + r^{-1}E \end{pmatrix}$ for any choice $\mathcal{A} = \mathcal{A}_{1r}$ or $\mathcal{A} = \mathcal{A}_{2r}$. In turn, this matrix is spectrally equivalent to a block-diagonal one. More precisely, the following statement takes place:

Lemma 5. Matrix

$$D = \begin{pmatrix} (L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E) & 0 \\ 0 & r^{-1}E \end{pmatrix}$$

is spectrally equivalent to $\mathcal{B}\mathcal{A}_0^{-1}\mathcal{B}^T$ with constants, which depend only on r.

Method (10) for problem (8) with $\mathcal{A} = \mathcal{A}_{1r}$ and with preconditioner D reads as follows:

$$\begin{aligned} &y^{k+1} = y_d - L^T\lambda^k - R^T\mu^k, \\ &\alpha u^{k+1} + \partial\varphi(u^{k+1}) \ni \lambda^k, \\ &rp^{k+1} + \partial\theta(p^{k+1}) \ni rRy^{k+1} + \mu^k, \\ &(L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E)\frac{\lambda^{k+1} - \lambda^k}{\rho} = Ly^{k+1} - u^{k+1} - f, \\ &\frac{\mu^{k+1} - \mu^k}{r\rho} = Ry^{k+1} - p^{k+1}. \end{aligned} \tag{11}$$

Theorem 1. Method (11) converges if $r \in (0,1)$ and $0 < \rho < 2(1 - \sqrt{r})(\sqrt{1+r} - r)^2$.

Implementation. On every step of method (11) we have to solve two inclusions, for u^{k+1} and for p^{k+1}, and the system of equations with the matrix $(L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E)$. Solving the inclusions reduces to pointwise projections on the corresponding sets of the constraints. On the other hand, solving a system of linear equations with the matrix $(L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E)$ consists of sequential solving the systems with the matrices $L + \alpha^{-1/2}E$ and $L^T + \alpha^{-1/2}E$. In the case of explicit finite difference scheme ($\sigma = 0$) these matrices

ISBN: 978-1-61804-323-8

are triangle ones and the solutions can be found by explicit calculations.

Let now

$$\mathcal{A}_{2r} = \mathcal{A}_1 + \mathcal{A}_2, \quad \mathcal{A}_1 = \begin{pmatrix} E + rR^TR & 0 & 0 \\ 0 & \alpha E & 0 \\ -rR & 0 & rE \end{pmatrix}, \quad \mathcal{A}_2 = \begin{pmatrix} 0 & 0 & -rR^T \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{12}$$

Block relaxation-Uzawa iterative method for solving saddle point problem (9) reads as follows:

$$\begin{aligned} &\mathcal{A}_1 w^{k+1} + \mathcal{A}_2 w^k - \mathcal{B}^T \eta^k + \partial\psi(w^{k+1}) \ni g_1, \\ &\frac{1}{\rho} D(\eta^{k+1} - \eta^k) + \mathcal{B} w^{k+1} = g_2. \end{aligned} \tag{13}$$

Due to [2] this method converges for any initial guess (w^0, η^0) if there exist constants $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ and a continuous and non-negative function ρ, $\rho(0) = 0$, such that

$$\begin{aligned} (\mathcal{A}_1 w, w) + (\mathcal{A}_2 v, w) \geqslant \quad & \varepsilon_1 \|w\|^2 + \frac{(1+\varepsilon_2)\rho}{2} (D^{-1}\mathcal{B}w, \mathcal{B}w) \\ & + \rho(w) - \rho(v) \ \forall w, v. \end{aligned} \tag{14}$$

Method (13) for problem (9) with the matrix $\mathcal{A} = \mathcal{A}_{2r}$ splitted into the sum as mentioned in (12) with the same preconditioner D as above takes the form:

$$\begin{aligned} &y^{k+1} + rR^TRy^{k+1} = y_d - L^T\lambda^k - R^T\mu^k + rR^Tp^k, \\ &\alpha u^{k+1} + \partial\varphi(u^{k+1}) \ni \lambda^k, \\ &rp^{k+1} + \partial\theta(p^{k+1}) \ni rRy^{k+1} + \mu^k, \\ &(L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E)\frac{\lambda^{k+1} - \lambda^k}{\rho} = Ly^{k+1} - u^{k+1} - f, \\ &\frac{\mu^{k+1} - \mu^k}{r\rho} = Ry^{k+1} - p^{k+1}. \end{aligned} \tag{15}$$

Theorem 2. Method (15) converges if $r > 0$, $0 < \rho < 1$.

Implementation. The implementation of method (15) differs from the implementation of method (11) only in the equation for y^{k+1}. Namely, now we have to solve a system of linear equations with the matrix $E + rR^TR$ for finding y^{k+1}. The corresponding calculations reduce to solving for every fixed node of ω_x a system with tridiagonal matrix (with respect to time variable), so, can be implemented by Thomas algorithm.

Acknowledgment

In this work, the first and second authors were supported by grants n:o 278488 and n:o 278029 from Academy of Finland.

References

[1] A. Lapin: Preconditioned Uzawa type methods for finite-dimensional constrained saddle point problems, Lobachevskii J. Math. - V.31, 4 - P.309-322 (2010).

[2] N.S. Kashtanov, A.V. Lapin: Efficiently implementable iterative methods for linear elliptic variational inequalities with constraints on the gradient of solution, Matematica. - N.7 -P.10-24 (2015) (in Russian).

[3] E. Laitinen, A. Lapin and S. Lapin: Easily implementable iterative methods for variational inequalities with nonlinear diffusion-convection operator and constraints to gradient of solution, Russian J. Numer. Analysis Math. Modeling - V.30,1 - P.43 - 54 (2015).

[4] I. Ekeland and R. Temam: Convex analysis and variational problems –Amsterdam: North- Holland. – 1976.

Ph.D Erkki Laitinen is a university lecturer at the Department of Mathematical Sciences of University of Oulu, and an adjunct professor of Computer Science at the University of Jyväskylä, Finland. His research interests include numerical analysis, optimization and optimal control. He is active in promoting these techniques in practical problem solving in engineering, manufacturing, and industrial process optimization. He has published more than hundred peer reviewed scientific papers in international journals and conferences. He has participated in several applied projects dealing with optimization and control of production processes or wireless telecommunication systems.

D. Sc. Alexander Lapin is a professor of Kazan Federal University (Russia) at the Institute of Computational Mathematics and Information Technology. His research interests include numerical analysis of free boundary problems and optimal control. He participated in the implementation of numerus applications including dam problem, nonlinear filtration problem, Stefan problem etc. He has published more than hundred peer reviewed scientific papers in journals and conferences proceedings.