# Serdica

Bulgariacae mathematicae publicationes

# Сердика

Българско математическо списание

# A REGULARIZED CONDITIONAL GRADIENT METHOD APPLIED TO SINGULARLY PERTURBED OPTIMAL CONTROL SYSTEMS

M. G. DMITRIEV*, ASEN L. DONTCHEV, VLADIMIR M. VELIOV

We propose a modification of the conditional gradient method which provides geometric convergence for strongly convex functionals. The influence of the computational errors on the convergence rate is estimated. As an application we consider an optimal control problem for a singularly perturbed linear system for the solving of which we use an iterative procedure separating the slow and the fast subsystems.

**1.** Levitin and Poljak [1] showed that the standard conditional gradient method converges as $1/k$ and that this estimation is exact. Here we propose a modification of this method related to Tichonov's regularization technique, which provides convergence of order $q^k$, $q\in(0,1)$, for strongly convex functionals. This modification is related to the method used by Barnes [2] who considered, however, linear-quadratic optimal control problems only. Furthermore, we analyse the influence of the computational errors, accompanying the method, on the convergence rate. As an applcation we consider an optimal control problem for a linear system with slow and fast subsystems containing small parameter in the derivative. Such systems are stiff for computations, therefore we use an iterative procedure, originally proposed by Dmitriev [3] for approximate solving of the state and the adjoint equations. We show that the global error can be estimated by the sum of two independent terms: the error of the method (geometrically decreasing but may depend on the singular parameter) and error of the iterative procedure. This estimate generalizes the corresponding result of Dmitriev [3] who assumed that the terminal part of the functional depends on the slow states only. Our approach turns out to be quite general and can be used in the analysis of other methods with various computational inaccuracies.

**2.** Let $B_x$ and $B_u$ be Banach spaces with norms $\|.\|$, $B_u$ be reflexive and $U$ be a closed, bounded and convex subset of $B_u$. Let $\mathcal{L}:U\to B_x$ be a linear and bounded operator, $b\in B_x$ be given and $J:U\times B_x\to R^1$ be a functional, which satisfies the following conditions: $J$ is Freshet differentiable and the derivatives $J'_x$ and $J'_u$ are Lipschitz continuous on $(\mathcal{L}U+b)\times U$ with constants $L_x$ and $L_u$. There exists a constant $\mu_0>0$ such that for every $(x_1,u_1)$ and $(x_2,u_2)$ from $(\mathcal{L}U+b)\times U$ the following relation holds

$$J(x_1,u_1)-J(x_2,u_2)\geq\langle J'_x(x_2,u_2),x_1-x_2\rangle+\langle J'_u(x_2,u_2),u_1-u_2\rangle+\mu_0\|u_1-u_2\|^2,$$

where $\langle .,.\rangle$ denotes the duality. The last supposition is equivalent to the convexity of $J$ and the strong convexity with respect to $u$ uniformly in $x$.

For the minimum problem

$$(1)\qquad J(x,u)\to\min,\ x=\mathcal{L}u+b,\ u\in U$$

we consider the following minimization procedure: Let the constants $\delta$ and $\Delta$ be fixed and $\mu\in(0,\mu_0)$. The iteration is defined as

$$u_{k+1}=u_k+\alpha_k(\bar{u}_k-u_k),\quad \|x_{k+1}-\mathscr{L}u_{k+1}-b\|\leq\delta,$$

where $\alpha_k$ is obtained from

$$\alpha_k=\operatorname{argmin}\{J(x_k+\alpha(\bar{x}_k-x_k),\ u_k+\alpha(\bar{u}_k-u_k)),\ \alpha\in[0,1]\},$$

and $(\bar{x}_k,\bar{u}_k)$ satisfies

$$|J_k(\bar{x}_k,\bar{u}_k)-J_k(\hat{x}_k,\hat{u}_k)|\leq\Delta,\quad \|\bar{x}_k-\mathscr{L}\bar{u}_k-b\|\leq\delta,$$

where $J_k(x,u)=\langle J_x'(x_k,u_k),\ x-x_k\rangle+\langle J_u'(x_k,u_k),\ u-u_k\rangle+\mu\|u-u_k\|^2$, and $(\hat{x}_k,\hat{u}_k)$ solves the "small problem"

$$J_k(x,u)\to\min,\quad x=\mathscr{L}u+b,\ u\in U. \tag{2}$$

Note that the method differs from the standard conditional gradient method in the quadratic term included in the functional of the small problem (2).

Denote by $(\hat{x},\hat{u})$ the solution of (1) (which exists) and let $\hat{J}=J(\hat{x},\hat{u})$. Introduce the constants

$$\alpha_0=\mu/(L_x\|\mathscr{L}\|+L_u)(1+\|\mathscr{L}\|),\quad q=1-\alpha_0,\quad D_u=\sup\|u\|,\ u\in U,$$
$$D_x=\sup\|\mathscr{L}u+b\|,\ u\in U.$$

Observe that $L=(L_x\|\mathscr{L}\|+L_u)(1+\|\mathscr{L}\|)$ estimates the Lipschitz constant of the gradient of $\varphi(u)=J(\mathscr{L}_u+b,u)$, hence $\alpha_0\in(0,1)$ and $q\in(0,1)$.

Remark. In the sequel we assume that $u_{k+1}\neq u_k$. In case $u_{k+1}=u_k$ the algorithm terminates and $u_k$ is the exact solution.

Theorem 1. *The following estimation holds*:

$$J(x_k,u_k)-\hat{J}\leq q^k(J(x_0,u_0)-\hat{J})+\Delta+\alpha_0\delta(L_x(2\delta+2D_x+D_u)+L_uD_u). \tag{3}$$

Proof. We have

$$\hat{J}-J(x_k,u_k)\geq\langle J_x'(x_k,u_k),\ \hat{x}-x_k\rangle+\langle J_u'(x_k,u_k),\ \hat{u}-u_k\rangle+\mu_0\|\hat{u}-u_k\|^2 \tag{4}$$
$$\geq J_k(\hat{x},\hat{u})\geq J_k(\hat{x}_k,\hat{u}_k).$$

Furthermore, for some $\alpha\in(0,1)$

$$J(x_{k+1},u_{k+1})-\hat{J}\leq J(x_k+\alpha(\bar{x}_k-x_k),\ u_k+\alpha(\bar{u}_k-u_k))-\hat{J}$$
$$\leq J(x_k,u_k)-\hat{J}+\alpha\langle J_x'(x_k,u_k),\ \bar{x}_k-x_k\rangle+\alpha\langle J_u'(x_k,u_k),\ \bar{u}_k-u_k\rangle$$
$$+\frac{L_x\alpha^2}{2}(\|\bar{x}_k-x_k\|+\|\bar{u}_k-u_k\|)\|\bar{x}_k-x_k\|+\frac{L_u\alpha^2}{2}(\|\bar{x}_k-x_k\|+\|\bar{u}_k-u_k\|)\|\bar{u}_k-u_k\|$$
$$\leq J(x_k,u_k)-\hat{J}+\alpha J_k(\hat{x}_k,\hat{u}_k)-\alpha\mu\|\bar{u}_k-u_k\|^2+\alpha\Delta$$
$$+\frac{\alpha^2}{2}(L_x+L_u)(\|\bar{x}_k-x_k\|+\|\bar{u}_k-u_k\|)^2.$$

Using (4) and the inequality

$$\| \bar{x}_k - x_k \| \leq \| \bar{x}_k - \mathscr{L} \bar{u}_k - b \| + \| \mathscr{L} (\bar{u}_k - u_k) \| + \| \mathscr{L} u_k - b_k - x_k \|$$

in the above estimate, we come to (3).

The obtained estimation (3) means that the error at each step is not accumulated by the iterations and the global error is a sum of the method's error and the inaccuracy in solving the small problem.

**3.** We are interested in linear differential systems containing a positive parameter in the derivative, namely

$$\dot{y} = A_1(t) y + A_2(t) z + \varphi_1(t, \varepsilon),\ y(0) = y^\circ,\ y \in R^n, \tag{5a}$$

$$\varepsilon \dot{z} = \varepsilon A_3(t) y + A_4(t) z + \varphi_2(t, \varepsilon),\ z(0) = z^\circ,\ z \in R^m,\ t \in [0, 1], \tag{5b}$$

where the eigenvalues of the matrix $A_4(t)$ have strictly negative real parts for all $t \in [0, 1]$ and the parameter $\varepsilon$ is small relative to the other parameters of the system. To be specific, assume that the matrices $A_i(.)$ and the functions $\varphi_i(.)$ are continuous and the norms $\| A_1 \|_c$ and $\| A_2 A_4^{-1} A_3 \|_c$ are near 1. Then $\dot{z}$ may be large compared with $y$ and we refer to $y$ and $z$ as the slow and the fast variables, respectively. Taking $\varepsilon = 0$ the order of the system reduces, therefore the perturbation, represented by $\varepsilon$ is called singular.

Observe that the standard form of a singularly perturbed system is

$$\dot{y} = A_1(t) y + A_2(t) w + \varphi_1(t, \varepsilon),\ \varepsilon \dot{w} = A_3(t) y + A_4(t) w + \varphi_2(t, \varepsilon),$$

but, in case of differentiability of $A_4^{-1}(t) A_3(t)$, it can be transformed into (5ab) by the substitution $z = w + A_4^{-1} A_3 y$.

Dmitriev [3] proposed the following iterative procedure for solving (5ab):

$$\varepsilon \dot{z}_k = \varepsilon A_3(t) y_k + A_4(t) z_k + \varphi_2(t, \varepsilon), \quad z_k(0) = z^\circ,$$

$$\dot{y}_{k+1} = A_1(t) y_{k+1} + A_2(t) z_k + \varphi_1(t, \varepsilon),\ y_{k+1}(0) = y^0.$$

It was proved in [3] that if $(y, z)$ is the solution of (5ab) then there exist constants $c, c_1$ such that $\| y_k - y \|_c + \| z_k - z \|_{L_1} \leq c_1 (c\varepsilon)^k$.

Using a direct proof, which turns out to be very simple, we obtain here a refinement of this result. Let $Z_\varepsilon(t, \tau)$ be the fundamental matrix solution of the equation (5b). It is known that there exist constants $\sigma_0, \sigma > 0$ such that

$$\| Z_\varepsilon(t, \tau) \| \leq \sigma_0 \exp\left(-\sigma \frac{t-\tau}{\varepsilon}\right). \tag{6}$$

Analogously, let $Y(t, \tau)$ be the fundamental matrix solution of (5a), and let the constants $\alpha_0, \alpha$ satisfy $\| Y(t, \tau) \| \leq \alpha_0 \exp(\alpha(t - \tau))$.

Lemma 1. *Denote* $c = \frac{\alpha_0}{\alpha} e^\alpha \| A_2 A_4^{-1} A_3 \|_c$ *and* $c_1 = \frac{\sigma_0}{\sigma} \| A_3 \|_c$. *Then*

$$\| y_k - y \|_c \leq (c\varepsilon)^k + O(\varepsilon^{k+1}),\quad \| z_k - z \|_c \leq c_1 (c\varepsilon)^{k+1} + O(\varepsilon^{k+2}).$$

Proof. The first estimate follows from (6) and from the equality

$$\Delta y_{k+1}(t) = \int_0^t Y(t, \tau) A_2(\tau) \int_0^\tau Z_\varepsilon(\tau, s) A_3(s) \Delta y_k(s)\, ds\, d\tau,$$

where $\Delta y_k = y_k - y$. The second estimate uses the Cauchy formula for (5b).

Observe that the convergence rate of this procedure does not depend on the functions $\varphi_i$ and on the initial conditions.

4. Consider the following optimal control problem

$$J(x(.), u(.)) = g(x(1)) + \int_0^1 f(x(t), u(t), t)dt \to \min \tag{7}$$

$$x = (y, z) \in R^n \times R^m, \quad u \in R^r,$$

$$\dot{y} = A_1(t)y + A_2(t)z + B_1(t)u, \quad y(0) = y^\circ, \tag{8a}$$

$$\varepsilon\dot{z} = \varepsilon A_3(t)y + A_4(t)z + B_2(t)u, \quad z(0) = z^\circ \tag{8b}$$

$$u(t) \in U \quad \text{for a. e. } t \in [0, 1], \; u(.) \in L_2^{(r)}[0, 1],$$

where the matrices $A_i(t)$ are the same as in (5ab), the matrices $B_i(t)$ are continuous in [0, 1], the set $U \subset R^r$ is closed, convex and bounded. We assume that the functions $g$ and $f(., t)$ are both convex and differentiable and the derivatives $g'$, $f'_x$, $f'_u$ are Lipschitz continuous (with respect to $x$ and $u$ uniformly in $t$); $f(x, u, t)$ is strongly convex with respect to $u$ with a constant $\mu_0$ uniformly in $x$, $t$; $f$, $f'_x$ and $f'_u$ are continuous.

The above problem can be considered in the framework of the abstract setting (1), where one may take $B_x = L_2^{(n+m)}[0, 1] \times R^{n+m}$, $B_u = L_2^{(r)}[0, 1]$; the operator $\mathscr{L}$ will be determined by the differential equation (8ab) and will depend on the singular parameter, and all the conditions for (1) will hold. It turns out that the conditional gradient method descibed in the first part of this work is very convenient to be applied to (7) since the small problem will be an optimal control problem which is linear with respect to the state, hence, it can be solved "at once" by Pontryagin's maximum principle. One has to solve the adjoint equation

$$\dot{\zeta} = -A_1^T(t)\zeta - \varepsilon A_3^T(t)\eta + f'_y(x_k(t), u_k(t), t), \tag{9a}$$

$$\varepsilon\dot{\eta} = -A_2^T(t)\zeta - A_4^T(t)\eta + f'_z(x_k(t), u_k(t), t), \tag{9b}$$

$$\zeta(1) = -g'_y(x_k(1)), \quad \eta(1) = -g'_z(x_k(1))/\varepsilon.$$

Then $\widehat{u}_k(t)$ will be determined by

$$(f'_u(x_k(t), u_k(t), t)^T + \psi(t)^T B(t))u + \mu |u - u_k(t)|^2 \to \min, \; u \in U, \tag{10}$$

where $t \in [0, 1]$, $\psi = (\zeta, \eta)$, $B = (B_1, B_2)$ and $\mu \in (0, \mu_0)$. The control in the next step will be

$$u_{k+1}(t) = u_k(t) + \alpha_k(\widehat{u}_k(t) - u_k(t)),$$

where $\alpha_k = \arg\min\{J(x_k(.) + \alpha(\widehat{x}_k(.) - x_k(.)), u_k(.) + \alpha(\widehat{u}_k(.) - u_k(.))), \alpha \in [0, 1]\}$.

This is the case when the computations are exact, i. e. $\delta = \Delta = 0$, in the general setting. Then, according to Theorem 1, the method converges as geometric progression. The ratio of this progression, however, may depend on the singular perturbation parameter. In order to find this dependence one should estimate the norm of the operator $\mathscr{L}_\varepsilon$.

In the sequel we omit the somewhat lengthy proofs which are entirely based on the mathematical technique presented in [4, Chapter 3].

Lemma 2. *Consider the operator*

$$\mathcal{L}_\varepsilon : u(.) \in L_2^{(r)}[0,1] \to (y(1), z(1), y(.), z(.)) \in R^n \times R^m \times L_2^{(n)}[0,1] \times L_2^{(m)}[0,1],$$

*where* $(y(.), z(.))$ *are determined by* (8ab) *with zero initial conditions, and let* $\|\mathcal{L}_\varepsilon\|$ *be the usual operator norm. There exist constants* $c_2$ *and* $c_3$ *such that*

$$\|\mathcal{L}_\varepsilon\| \le c_2 \varepsilon^{-0,5} + c_3. \tag{11}$$

The following example shows that $c_2 \neq 0$, in general

$$\varepsilon \dot{z} = -z + u_\varepsilon(t), \quad z(0) = 0,$$

$$u_\varepsilon(t) = \begin{cases} 0, & t \in [0, 1-\varepsilon], \\ 1, & t \in (1-\varepsilon, 1], \end{cases}$$

Then $\|\mathcal{L}_\varepsilon\| \ge |z_\varepsilon(1)| / \|u_\varepsilon(.)\|_{L_2} \sim \varepsilon^{-0.5}$.

The estimate (11) applied to (3) gives us the following convergence of the method

$$J(x_k(.), u_k(.)) - \hat{J} \le (1 - c_4 \varepsilon)^k (J(x_0(.), u_0(.)) - \hat{J})$$

for sufficiently small ε. If the function $g$ is independent of $z$ then one can improve this estimate by changing the spaces and redefining the operator $\mathcal{L}_\varepsilon$.

Lemma 3. *Consider the operator*

$$\mathcal{L}_\varepsilon : u(.) \in L_2^{(r)}[0,1] \to (y(.), z(.)) \in C^{(n)}[0,1] \times L_1^{(m)}[0,1],$$

*where* $(y(.), z(.))$ *are determined by* (8ab) *with zero initial conditions. There exists a constant* $c_5$ *such that* $\|\mathcal{L}_\varepsilon\| \le c_5$.

If $g$ depends on $z$ linearly, then the Lipschitz constant of $g'$ is zero, and the number $\alpha_0$ in (3) can be estimated from below by const. $\varepsilon^{0.5}$.

Combining the above results we obtain

Theorem 1. *The regularized conditional gradient method converges in the following way*

$$J(x_k(.), u_k(.)) - \hat{J} \le \begin{cases} (1-c_7)^k (J(x_0(.), u_0(.)) - \hat{J}) & \text{when } g'_z = 0; \\ (1-c_8 \varepsilon^{0.5})^k (J(x_0(.), u_0(.)) - \hat{J} & \text{when } g \text{ is linear in } z; \\ (1-c_9 \varepsilon)^k (J(x_0(.), \hat{u}_0(.)) - \hat{J}) & \text{otherwise.} \end{cases}$$

Let us consider now a more general situation when the subprocedures in the method are not exact. To be concrete let us assume that the state and the adjoint equations are solved approximately by the procedure presented in part 3 of this paper.

Suppose we are on the $k$-th step having the control $u_k$ and the state $x_k$. Then we have to solve the adjoint equation (9ab) for $x_k$ and $u_k$. Since the convergence of the iterative procedure is independent of the free terms and of the initial conditions after $i$ steps of the procedure we get $\bar{\psi}_k$ which, according to Lemma 1, differs from the exact solution $\psi_k$ as $\|\bar{\psi}_k - \psi_k\|_c \le c_{10}\varepsilon^i + O(\varepsilon^{i+1})$

where the constant $c_{10}$ is independent of $k$. The error of the adjoint variable represents a perturbation of the small problem (10), which is strongly convex. Applying Proposition 1.3 from [4] to (10) we get that the approximate solution $\bar{u}_k$ of the small problem fulfils $\|\bar{u}_k-\hat{u}_k\|_c\leq c_{11}\varepsilon^i+O(\varepsilon^{i+1})$ (note that $\bar{u}_k(.)$ and $\hat{u}_k(.)$ are continuous functions). Furthermore, if $\bar{x}_k$ approximates the optimal trajectory by the iterative procedure, then $\|\bar{x}_k-\hat{x}_k\|_c\leq c_{12}\varepsilon^i+O(\varepsilon^{i+1})$. The last two inequalities imply

$$J_k(\bar{x}_k(.),\bar{u}_k(.))-J_k(\hat{x}_k(.),\hat{u}_k(.))\leq c_{13}\varepsilon^i+O(\varepsilon^{i+1})=\Delta.$$

It remains to take $\delta=c_{14}\varepsilon^i$ and to apply the general estimate (3). We have already estimated $\alpha_0$ and $q$ in lemmas 1,2 and Theorem 1. The constant $D_x$ does not depend on $\varepsilon$ since the solution of (8ab) is bounded uniformly in $t$ and $\varepsilon$ for uniformly bounded controls, see the proof of Lemma 3.1 in [4]. Finally, as an extension of Theorem 1 we get

Theorem 2. *Suppose that the regularized conditional gradient method together with the iterative procedure for solving differential equations (with $i$ steps) are applied to the problem (7). The following estimation holds for small $\varepsilon$:*

$$J(x_k(.),u_k(.))-\hat{J}\leq(1-c_7-c_8\varepsilon^{0.5}-c_9\varepsilon)^k(J(x_0(.),u_0(.))-\hat{J})+c_{15}\varepsilon^i.$$

*If $g$ does not depend on the fast states $z$, then $c_7>0$; if $g$ depends linearly on $z$, then $c_8>0$.*

**5.** As a general conclusion we obtain that the proposed regularized gradient method converges geometrically, but, when applied to optimal control problems, the presence of singular perturbations (ill-posedness of the state matrix) may lead to slowing down the convergence rate. However, if the fast states are not included in the terminal part of the functional, the convergence is uniform in singular perturbations. The iterative procedure we use to solve the state and the adjoint equations does not affect the convergence, moreover, the computational error is not accumulated along the iterations. The resulting error can be estimated by the sum of the error of the method and the error of the iterative procedure. This means practically that if we are not near the minimum we do not need to solve precisely the differential equations.

## REFERENCES

1. E. Levitin, B. T. Poljak. Minimization methods in presence of constraints. *Journ. Vych. Math. i Math. Phys.*, **6**, No 5, 1966.
2. E. R. Barnes. A geometrically convergent algorithm for solving optimal control systems. *SIAM J. Control*, **10**, No 3, 1972.
3. M. G. Dmitriev. Iterative solving of optimal control problems with slow and fast motions. *Doklady AN SSSR*, **272**, No 2, 1983.
4. A. L. Dontchev. Perturbations, approximations and sensitivity analysis of optimal control systems. *Lecture Notes in Control and Inf. Sc.*, **52**, Springer, New York.

*Received 24. 1. 1984*

**Siberian Division of the Academy of Sciences of USSR Computing Centre Krasnoyarsk, USSR*

*Centre for Mathematics and Mechanics Sofia 1090 P. O. Box 373*