BANACH CENTER PUBLICATIONS
VOLUME 1

PROJECTION ON A CONE AND GENERALIZED PENALTY FUNCTIONALS FOR PROBLEMS WITH CONSTRAINTS IN A HILBERT SPACE*

ANDRZEJ P. WIERZBICKI and STANISŁAW KURCYUSZ

Institute of Automatic Control, Technical University of Warsaw, Warsaw

1. Introduction

The penalty functional technique is one of the most general tools in the numerical solution of constrained minimization problems. A rather full treatment of its theory and applications is given in [2], for the case of finitely many constraints.

This method has a known numerical disadvantage: computational effort per iteration usually increases. To avoid this, the shifted penalty functional technique was introduced in 1969 by Hestenes and Powell, and has been broadly investigated since then; see, for instance, [4], [6], [7]. In a series of recent papers Rockafellar presented a theory of augmented Lagrangians, closely related to this topic. All the preceding considerations were carried out for the case of constraints in $\boldsymbol{R}^n$.

The numerical practice in applying the above-mentioned techniques to infinitely-dimensional problems, and the development of the discretization theory for extremal problems, make clear the need for a more general treatment of penalty functionals including the case of infinitely many constraints. Although such penalty functionals have been studied for particular cases [1], [3], the first general formulation and results were obtained in [8]. The present paper contains some new results in that direction.

2. Projection on a cone

While the notion of the projection on a convex closed set in a Hilbert space was known long ago, its properties were not investigated until relatively recently (Moreau, Zarantonello).

The basic facts needed in the sequel were given in [8].

* The work was supported by the National Science Foundation Grant No. GF-37298.

Assume throughout that H is a *Hilbert space*, $D \subset H$ *a nonempty closed convex cone*. By D^* denote the *dual cone* of D, i.e.,

$$D^* = \{d^* \in H\colon \langle d^*, d\rangle \geqslant 0 \ \forall d \in D\}.$$

The unique element $p^D \in D$ minimizing $\|p-d\|$, $d \in D$, is called the *projection of* p *onto* D. Similarly, we have p^{-D^*}.

The basic properties of these projections can be summarized as follows:

(2.1) (Moreau) *Any element* $p \in H$ *can be represented in the form*

$$p = p^D + p^{-D^*}$$

with

$$\langle p^D, p^{-D^*}\rangle = 0$$

and therefore

$$\|p\|^2 = \|p^D\|^2 + \|p^{-D^*}\|^2.$$

This decomposition is unique in the sense that, if $p = p_1 + p_2$, $p_1 \in D$, $p_2 \in -D^*$, $\langle p_1, p_2\rangle = 0$, *then* $p_1 = p^D$, $p_2 = p^{-D^*}$,

(2.2) $\|p^D\| \leqslant \|p\| \ \forall p \in H$; $\|p_1^D - p_2^D\| \leqslant \|p_1 - p_2\| \ \forall p_1, p_2 \in H$.

(2.3) *Functional* $q\colon H \to \boldsymbol{R}$ *defined by* $q(p) = \frac{1}{2}\|p^{-D^*}\|^2$ *is convex and Fréchet differentiable. Its gradient is equal to* p^{-D^*}.

3. General optimization problem

In the sequel E will denote a *Banach space*, $Q\colon E \to \boldsymbol{R}$ a functional and $P\colon E \to H$ an operator. Consider the following problem:

$$(3.1)\qquad \begin{cases} \textit{minimize} \quad Q(y), \\ \textit{subject to} \ \ y \in Y_p = \{y \in E\colon p - P(y) \in D\}. \end{cases}$$

Here, p is fixed in H.

Define the functional $K\colon E \to \boldsymbol{R}$ by $K(y) = \frac{1}{2}\|(p - P(y))^{-D^*}\|^2$. Then, by (2.1), problem (3.1) is equivalent to

$$(3.2)\qquad \begin{cases} \textit{minimize} \quad Q(y), \\ \textit{subject to} \ \ y \in A_0 = \{y \in E\colon K(y) \leqslant 0\}. \end{cases}$$

Thus, a fairly general operator-constrained problem (3.1) can be converted into a problem with a single functional constraint. Moreover, K "behaves well" when the operator P does.

(3.3) PROPOSITION. (i) *If* P *is* D*-convex, then* K *is convex and weakly lower semicontinuous.*

(ii) *If* P *is weakly continuous, then* K *is weakly lower semicontinuous.*

(iii) *If* P *is Fréchet-differentiable, then so is* K *and*

$$K_y(y) = P_y^*(y)\,(P(y) - p)^{D^*}.$$

Recall that the (normal) *Lagrange functional* for problem (3.1) is defined by

$$\Lambda(y, \eta) = Q(y) + \langle \eta, P(y) - p\rangle$$

for $(y, \eta) \in E \times H$.

Assume that $\hat{y}$ is a local solution of (3.1).

(3.4) DEFINITION. $\eta \in D^*$ is called a (*normal*) *Lagrange multiplier* for problem (3.1) at the point $\hat{y}$ iff

(i) $$\langle \eta, P(\hat{y}) - p\rangle = 0$$

and, if Q, P are Fréchet-differentiable, then

(ii) $$\Lambda_y(\hat{y}, \eta) = Q_y(\hat{y}) + P_y^*(\hat{y})\eta = 0$$

or, if Q is convex and P is D-convex, then

(ii′) $$\Lambda(\hat{y}, \eta) \leqslant \Lambda(y, \eta) \quad \forall\, y \in E.$$

Define the *shifted penalty functional* $\Phi\colon E \times R^+ \times H \to \boldsymbol{R}$ by

$$(3.5)\qquad \Phi(y, \varrho, v) = Q(y) + \tfrac{1}{2}\varrho\|(P(y) - v)^{D^*}\|^2.$$

(3.6) PROPOSITION. *Assume that* $\hat{y}$ *minimizes* $\Phi(\cdot, \varrho, v)$ *over* E. *Write*

$$\hat{p} = (P(\hat{y}) - v)^{D^*} + v, \qquad \hat{\eta} = \varrho(P(\hat{y}) - v)^{D^*}.$$

Then $\hat{y}$ *is a global solution of the problem*:

$$\begin{cases} \textit{minimize} \quad Q(y), \\ \textit{subject to} \ \ y \in Y_{\hat{p}} = \{y \in E\colon \hat{p} - P(y) \in D\} \end{cases}$$

and, if P, Q *are Fréchet-differentiable or* Q *is convex, and* P *is* D*-convex, then* $\hat{\eta}$ *is a normal Lagrange multiplier for the above problem at* $\hat{y}$.

4. Convergence of the nonshifted penalty method

Define the (nonshifted) *penalty functional* $F\colon E \times R^+ \to \boldsymbol{R}$

(4.1) $F(y, \varrho) = \Phi(y, \varrho, p)$.

(4.2) THEOREM. *Suppose that* Q, K *are weakly lower semicontinuous,* $Q(y) \geqslant \beta > -\infty$ $\forall y \in E$, *where* E *is a reflexive Banach space. Suppose that the sets* $A_\delta = \{y \in E\colon Q(y) \leqslant \varepsilon,\ K(y) \leqslant \delta\}$ *are nonempty and bounded for* $0 \leqslant \delta \leqslant \delta_0$, $\delta_0 > 0$ *and some* ε. *Then*

(i) *there is a* $\bar{\varrho} \geqslant 0$ *such that for any* $\varrho \geqslant \bar{\varrho}$ *there exists a* $y_\varrho \in E$ *minimizing* $F(\cdot, \varrho)$ *over* E.

Suppose that $\varrho_n \geqslant \varrho$, $\varrho_n \to \infty$ *and write* $y_n = y_{\varrho_n}$. *Then*:

(ii) $Q(y_n) \underset{n\to\infty}{\to} \inf_{y \in Y_p} Q(y)$, $\varrho_n K(y_n) \underset{n\to\infty}{\to} 0$;

and each of the weak cluster points of $\{y_n\}$ *is a solution of* (3.1).

(iii) *Write* $p_n = (P(y_n) - p)^{D^*} + p$. *Then* $\lim_{n\to\infty} p_n = p$ *and each* y_n *minimizes* Q *over* $Y_{p_n} = \{y \in E\colon p_n - P(y) \in D\}$.

(iv) *Write* $\eta_n = \varrho_n(P(y_n)-p)^{D*}$ *and assume that* Q, P *are differentiable or* Q *is convex,* P *D-convex. Then* η_n *is a normal Lagrange multiplier for the problem of minimizing* Q *over* Y_{p_n}.

(v) *Let* Q, P *be continuously differentiable or let* Q, P *be continuous, let* Q *be convex, and let* P *be D-convex. If the sequence* $\{y_n\}$ *converges in norm to a point* $\hat{y}$ *(being hence a solution of* (3.1)), *then each weak cluster point of the sequence* $\{\eta_n\}$ *is a (normal) Lagrange multiplier for problem* (3.1).

It is known that the assumptions made above are not the weakest possible to guarantee (ii). This theorem, however, brings to light two interesting features of the penalty algorithm. First, we see that the penalty algorithm can be treated as a two-level coordination technique: a sequence $\{\varrho_n\}$ is chosen to coordinate $\{p_n\}$ to p.

Secondly, the theorem generalizes Balakrishnan's computational approach to the maximum principle [1]: under appropriate assumptions, $\{\eta_n\}$ approximate the Lagrange multipliers of the original problem.

This theorem has also as a corollary an interesting qualitative result concerning the existence of Lagrange multipliers.

(4.3) COROLLARY. *Under the conditions of Theorem* (4.2 iv), *denote by* $\mathscr{P}$ *the set*

$$\mathscr{P} = \{p \in H: \exists \delta_0 > 0\ \exists \varepsilon\ \forall 0 \leqslant \delta \leqslant \delta_0\ A_\delta^{p,\varepsilon} \neq \emptyset \text{ and bounded}\}$$

where $A_\delta^{p,\varepsilon} = \{y \in E: Q(y) \leqslant \varepsilon,\ ||(P(y)-p)^{D*}|| \leqslant \delta\}$. *Denote by* $\mathscr{P}_0$ *the set of all* $p \in \mathscr{P}$ *such that normal Lagrange multipliers exist for the problem of minimizing* Q *over* Y_p. *Then* $\mathscr{P}_0$ *is dense in* $\mathscr{P}$.

5. Shifted penalty method

The following theorem is the generalization of Rockafellar's result [6] in finite dimensions.

(5.1) DEFINITION. Problem (3.1) is called *stable of degree* 2 iff there is a $\varrho \geqslant 0$ and a $v \in H$ such that

$$\hat{Q}(p+\delta p) \geqslant \hat{Q}(p)+\varrho\langle v, \delta p\rangle-\tfrac{1}{2}\varrho||\delta p||^2 \quad \forall \delta p \in \mathcal{O}$$

where $\mathcal{O}$ is some neighbourhood of zero and $\hat{Q}(p) = \inf_{y \in Y_p} Q(y)$.

(5.2) THEOREM. *Assume that* $\hat{Q}(p+\delta p) \geqslant \beta > -\infty$ *for* δp *in some neighbourhood of zero and suppose that* $\hat{y}$ *is a global solution of* (3.1). *The necessary and sufficient condition for the existence of a pair* $(\varrho, v) \in R^+ \times H$ *such that* $\hat{y}$ *minimizes* $\Phi(\cdot, \varrho, v)$ *over* E, *is that problem* (3.2) *be stable of degree* 2.

Under somewhat stronger assumptions we can find the above-mentioned pair (ϱ, v) in an iterative way. More precisely, the following theorem holds, generalizing the results of [5] [7] to infinite dimensions.

Assume for simplicity that $p = 0$.

(5.3) THEOREM. *Suppose there is a neighbourhood* $\mathcal{O}_p \subset H$ *of zero such that the minimization problems* (3.1) *have unique solutions for each* $p \in \mathcal{O}_p$ *and unique*

Lagrange multipliers $\eta(p)$ *such that the mapping* $p \to \eta(p)$ *is Lipschitz continuous with constant* R_η.

Suppose the shifted penalty functional attains global minimum for each $\varrho > \varrho'$ *and* $v \in \mathcal{O}_v \subset H$, $\mathcal{O}_v$ *being a neighbourhood of zero. Then*

(i) *problem* (3.1) *with* $p = 0$ *is stable of degree* 2.

(ii) *Let a sequence* $\{v_n\}$ *be defined by*:

$$v_1 = 0, \quad v_{n+1} = (P(y_n)+v_n)^{D*},$$

where y_n *minimizes globally* $\Phi(\cdot, \varrho, v_n)$. *Then there is a* ϱ'' *such that for* $\varrho \geqslant \max(\varrho', \varrho'')$, $\{v_n\} \subset \mathcal{O}_v$, $\{p_n\} \subset \mathcal{O}_p$ *where* $p_n = v_{n+1}-v_n$. *Moreover,* $p_n \to 0$ *so that the sequence* $\{\varrho v_n\}$ *is convergent to the unique Lagrange multiplier* η_0 *for problem* (3.1) *with* $p = 0$.

(iii) *Given any* $\alpha > 0$, *there exists a* $\varrho_\alpha \geqslant \varrho''$, $\varrho_\alpha \geqslant \dfrac{1+\alpha}{\alpha} R_\eta$ *such that* $\varrho \geqslant \varrho_\alpha$ *implies*

$$||p_{n+1}|| \leqslant \alpha||p_n||,$$

$$\left\|v_{n+1}-\frac{1}{\varrho}\eta_0\right\| \leqslant \alpha\left\|v_n-\frac{1}{\varrho}\eta_0\right\|.$$

Combining this result with (3.6) one can obtain a strong convergence theorem for the shifted penalty technique.

6. Conclusions

The penalty functional for a general infinite-dimensional operator constrained extremal problem has been defined and its properties have been investigated. Two penalty techniques: increased nonshifted and nonincreased shifted, have been considered. A theorem has been stated generalizing Balakrishnan's ε-technique approach to the necessary optimality conditions. Further two theorems generalize known results (obtained formerly for the case of a finite number of constraints) concerning the shifted penalty technique. Thus, the projection on a cone is a useful tool for a general theory of penalty methods in a Hilbert space.

References

[1] A. V. Balakrishnan, *A computational approach to the maximum principle*, J. Comp. Syst. Sci. 5 (1971), pp. 163–191.

[2] A. Fiacco, G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, J. Wiley, 1968.

[3] E. S. Levitin, B. T. Poljak, *Convergence of minimizing sequences in conditional extremum problems*, Soviet Math. Dokl. 7, pp. 764–76.

[4] O. L. Mangasarian, *Unconstrained Lagrangians in nonlinear programming*, Computer Sci. Tech. Rep. No. 174, University of Wisconsin, Madison 1973.

[5] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In: *Optimization*, ed. by R. Fletcher, Academic Press 1969, pp. 283–298.
[6] R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, to appear in SIAM J. Control.
[7] A. P. Wierzbicki, *A penalty function shifting method in constrained static optimization and its convergence properties*, Archiwum Automatyki i Telemechaniki 16 (1971), pp. 395–416.
[8] —, A. Hatko, *Computational methods in Hilbert space for optimal control problems with delays*, Proc. of 5-th IFIP Conference on Optimization, Rome 1973.

BANACH CENTER PUBLICATIONS
VOLUME 1

ON THE MAXIMUM PRINCIPLE FOR OPTIMAL CONTROL PROCESSES DESCRIBED BY HAMMERSTEIN INTEGRAL EQUATIONS WITH WEAKLY SINGULAR KERNELS

L. v. WOLFERSDORF
Sektion Mathematik, Bergakademie Freiberg, Freiberg, DDR

1

Let X be a real Banach space with norm $\|\cdot\|$, X_0 an open subset of X, and W a certain set. Let $T\colon X_0\times W\to X$ be a transformation and $f\colon X_0\times W\to R$ a real-valued functional. We consider the following *optimal control problem*:

(1) $$\operatorname{Min} f(x, w), \qquad (x, w)\in X_0\times W$$

subject to the process equation

(2) $$x = T(x, w).$$

For this problem L. Bittner [1] has developed a simple exact method of deriving maximum principles. He has applied his method especially to optimal control processes governed by Urysohn integral equations with variable integration domains and continuous kernels. I want to give here at first a modification of Bittner's basic theorem for this model of an optimal control process. Then I will apply this modification to a derivation of a maximum principle in the case where the process equation is a Hammerstein integral equation with a completely continuous kernel in a space of continuous functions.

2

Let Y be another real Banach space with norm $\|\cdot\|_y$, which has the following properties:

(a) *X is a dense subset of Y,*

(b) *the embedding operator $X\to Y$ is continuous*, i.e., *there exists a constant $C>0$ such that*

$$\|x\|_y \leqslant C\|x\| \qquad \forall x\in X.$$

THEOREM. *Let the following assumptions hold*:

(i) *$T(x, w)$ has a partial Fréchet derivative $T_x(x, w)$ at every point (x, w) of the set $B(x_0, r_0)\times W_0$, $r_0>0$, where $B(x_0, r_0)$ denotes the ball with centre x_0 and*