SOME REMARKS ON GENERALIZED LAGRANGIANS

S. Kurcyusz

Technical University of Warsaw
Institute of Automatic Control
ul. Nowowiejska 15/19
00-665 WARSAW, POLAND

Abstract

In the paper a definition is given of a class of generalized Lagrangians, and some simple properties of them are discussed, especially those related to the topology in the set of constraints. A general formulation of the method of multipliers is presented and a theorem characterising convergence of this method in case of linear-quadratic problems in Hilbert space. Numerical examples of computing the optimal control of time lag systems to terminal functions are presented. The results indicate that the effectiveness of the method of multipliers depends on the choice of the norm in the set of constraints.

0. INTRODUCTION

The paper presents a definition and selected simple properties of a class of generalized Lagrangians. Many authors have recently furnished definitions and discussed the properties of generalized Lagrange functionals - see e.g. [1] [6] [7] [8] [22] [26].

However, their research has been primarily limited to nonlinear or convex programming in R^n or to convex problems in more general spaces [14] . A more detailed study of so called shifted penalty functionals in Hilbert space has been presented in [30] [31] . The theoretical investigations are intimately related to practical problems of computational techniques and become especiallny fruitful when applied to the mentioned shifted penalty method (often called the method of multipliers) [10] [16]. A number of papers appeared discussing the convergence, the rate of convergence and the relation to the more abstract duality framework [2] [15] [29] [19] [20] and others. Except [23] [25] [30] [31] all consideration have been carried out in R^n as the space of constraints values.

The goal of this paper is to extend several notions to a general class of optimization problems including those with infinite dimensional or integer constraints. Only the main facts are given in order to rather indicate the possibility of generalization than describe it in detail. The duality theory for these generalized Lagrangians, leads in a natural way to φ-conjugates of functions, [27] [28] which resemble Fenchel conjugates with the scalar product (or duality) substituted by an arbitrary function of two variables. The presentation below has much in common with the thesis of Seidler [26]. However, some simple properties are shown connected with the topology in the set of constraints values: in finite dimensional case essentially one topology is used, while generally it plays non-negligible role. In the last two paragraphs the application of the method of multipliers to a problem with linear operator constraints is discussed and numerical examples are described which show that the proper choice of the penalizing norm, consistent with the topological features of the constraining operator, influences strongly the computational effectiveness of the algorithm. The examples presented are optimal control problems for time-lag systems.

1. PRELIMINARIES

Suppose that two sets are given, Y and P, a family $\{Y_p\}_{p \in P}$ of subsets of Y indexed by $p \in P$ and a functional $Q : Y \to R$. Then a family of optimization problems can be defined:

(1.1) minimize $Q(y)$ over $y \in Y_{p_o}$

where p_o can change over P.

Such a problem statement allows the discussion of linear, nonlinear as well as integer programming problems and extremal problems with infinitely many constraints, for instance those of optimal control theory. Besides, the description is simple and clear.

The family $\{Y_p\}_{p \in P}$ determines a family $\{P_y\}_{y \in Y}$ of "inverse images" in P :

(1.2) $p \in P_y \iff y \in Y_p$.

Problem (1.1) can be equivalently stated as

(1.3) minimize $Q(y)$ over y: $p_o \in P_y$.

The family $\{Y_p\}_{p \in P}$ introduces also in P a natural partial ordering:

$$(1.4) \qquad p \leqslant p' \Longleftrightarrow Y_p \subset Y_{p'} .$$

If we suppose for a while that this partial order is such that for every $y \in Y$ with $P_y \neq \emptyset$ there exists $\inf_{p \in P_y} p \in P_y$ then (1.4) is equivalent to

$$\text{minimize } Q(y) \quad \text{over } y: \ \inf_{p \in P_y} p \leqslant p_0 .$$

This reformulated problem is in no way easier to treat then (1.1). However, it suggests a relaxation of (1.1) into what is called a "surrogate problem" [5] [9] . Given a functional $f : P \to \mathbf{R}$, solve

$$(1.5) \quad \text{minimize } Q(y) \quad \text{over } y: \ \inf_{p \in P_y} f(p) \leq f(p_0).$$

Any solution to (1.1) clearly solves (1.5) and it is always possible to find a function f such that (1.5) is equivalentt to (1.1).

Suppose now that instead of one function f we have a family of such functions, namely, a function $\varphi : P \times W \to R$ where W is another set. Any $w \in W$ defines a function $\varphi(\cdot, w): P \to R$ which can be viewed as the "distribution of prices" on perturbations $p \in P$.

Then the value

$$(1.6) \qquad K(y,w) = \inf_{p \in P_y} \varphi(p,w) - \varphi(p_0,w)$$

can be used to measure the distance from a given y to the set of admissible solutions Y_{p_0} in terms of prices corresponding to the "distribution of prices" $w \in W$. Assume further that

$$(1.7) \qquad \sup_{w \in W} K(y,w) = \begin{cases} +\infty & , \ p_0 \notin P_y \\ 0 & , \ p_0 \in P_y \end{cases} .$$

Then solving (1.1) is clearly equivalent to solving the following problem without constraints:

$$(1.8) \qquad \text{minimize} \quad \sup_{w \in W} L(y,w) \quad \text{over} \quad y \in Y$$

where $L(y,w)$ we define to be the <u>generalized Lagrangian</u> associated with (1.1):

$$(1.9) \qquad L(y,w) = Q(y) + K(y,w) =$$

$$= Q(y) + \inf_{p \in P_y} \varphi(p,w) - \varphi(p_o,w) .$$

For similar definitions see e.g. [6] [22] and especially [26]. All have been given for nonlinear programs with constraints in R^n.

We give now the examples.

Suppose P is a topological vector space, $S : Y \to P$ - an operator and $D \subset P$ a (closed) convex cone with vertex at zero. Consider the following nonlinear program:

$$(1.10) \qquad \text{minimize} \quad Q(y) \quad \text{subject to} \qquad p_o - S(y) \in D.$$

Then for $p \in P$, $Y_p = \{ y \in Y: \ p - S(y) \in D \}$ and the partial order (1.4) coincides with the partial ordering introduced by the cone D.

(i) Let $W = D^*$, the dual cone, and $\varphi(p,w) = \langle w, p \rangle$. Then

$$K(y, w) = \langle w, S(y) - p_o \rangle$$

and L is the classical Lagrange functional. (1.7) clearly holds.

(ii) Suppose P is a lattice with respect to the order introduced by D, i.e. for each $p \in P$ $\max(0,p) = p^+$ exists (all function spaces are lattices with respect to the cone of nonnegative functions). Suppose $\Psi : P \to R$ is a functional satisfying the following conditions ($p \leq p'$ means $p' - p \in D$):

(a) $0 \leq p \leq p' \Rightarrow \Psi(p) \leq \Psi(p')$

(b) $0 \leqslant p < p' \Rightarrow \Psi(p) < \Psi(p')$ where $p < p' \Leftrightarrow p \leqslant p'$, $p \neq p'$.

(c) $\Psi(p) = \Psi(p^+ + (-p)^+)$.

Let $W = R_+ \times P$, so that $w \in W$ is a pair $w = (\varrho, v)$, $\varrho \geqslant 0$, $v \in P$. Define

(1.11) $\qquad \varphi(p,w) = \varrho \Psi(p - v).$

According to the definition of Y_p,

(1.12) $\quad P_y = \{p \in P:\ p - S(y) \in D\} = \{p \in P:\ p \geqslant S(y)\}$.

Hence

$$\inf_{p \in P_y} \varphi(p,w) = \varrho \inf_{p \geqslant S(y)} \Psi(p-v) = \varrho \inf_{p' \geqslant S(y) - v} \Psi(p') = \varrho \Psi((S(y) - v)^+) ,$$

(1.13) $\quad K(y,w) = \varrho\Psi((S(y) - v)^+) - \varrho\Psi(p_o - v).$

It remains to verify (1.7). If $p_o \notin P_y$ then

$$(S(y) - p_o)^+ > 0 \quad \text{and by (b)}$$

$$K(y,(\varrho, p_o)) = \varrho[\Psi((S(y) - p_o)^+) - \Psi(0)] \xrightarrow[\varrho \to \infty]{} +\infty$$

If $p_o \in P_y$ then $S(y) \leq p_o$, $(S(y) - v)^+ \leqslant (p_o - v)^+$ and in virtue of (a) and (c)

$$K(y,w) = \varrho[\Psi((S(y) - v)^+) - \Psi(p_o - v)] \leqslant$$

$$\leqslant \varrho[\Psi((S(y) - v)^+) - \Psi((p_o - v)^+)] \leqslant 0 ,$$

while

$$K(y, (\varrho, p_o)) = \varrho[\Psi(0) - \Psi(0)] = 0.$$

(iii) Suppose P is a Hilbert space. Set $W = R_+ \times P$, as before and take $\alpha > 0$. Define

$$\varphi(p,w) = \varrho \| p - v \|^{\alpha} .$$

We have from (1.12):

$$\inf_{p \in P_y} \varphi(p,w) = \varrho \inf_{p' \in S(y) - v + D} \| p' \|^{\alpha} = \varrho \| (S(y) - v)^{D^*} \|^{\alpha}$$

where p^{D^*} denotes the projection of p onto D^* - see [31].

Then

(1.14) $$K(y,w) = \varrho \| (S(y) - v)^{D^*} \|^{\alpha} - \varrho \| p_0 - v \|^{\alpha} .$$

Property (1.7) is verified similarly as above. For $\alpha = 2$, $L(y,w)$ with this K is the augmented Lagrangian of Wierzbicki [30] [31].

(iv) We now specialize the two preceding examples to the case of $P = R^n$, $D = \{p = (p^1,\dots,p^n): p^i \geqslant 0,\ i = 1,\dots,m,\ p^i = 0,\ i = m+1,\dots,n\}$.

Suppose $\zeta : R \to R$ is a monotone (strictly increasing) in R_+, nonnegative function with $\zeta(0) = 0$.

Define ψ under (ii) by

$$\psi(p) = \sum_{i=1}^{n} \zeta(p^i) .$$

Then the generalized Lagrangian (1.9) with K as in (1.13) is practically the generalized Lagrangian employed by Mangasarian [15].

Assume now for simplicity that $m = n = 1$ and $p_0 = 0$ so that we have only one inequality constraint and that $\zeta(a) = a^2$, $a \in R$. Then the above $K(y,w)$ becomes a special case of (1.14) with $P = R$, $D = D^* = R_+$, $\alpha = 2$:

$$K(y,w) = \varrho((S(y) - v)^+)^2 - \varrho v^2 .$$

Substitute $\lambda = 2\varrho v$ and note that

$$K(y,w) = \varrho\, \gamma(S(y), \lambda/\varrho)$$

where

$$\gamma(a,b) = ((a - \tfrac{1}{2} b)^+)^2 - \tfrac{1}{4} b^2 = \begin{cases} a^2 - ab, & a \geqslant \tfrac{1}{2} b \\ -\tfrac{1}{4} b^2, & a \leqslant \tfrac{1}{2} b . \end{cases}$$

One can therefore obtain also the augmented Lagrangian of Rockafellar [19].

For further use we shall need the following definition. The primal functional $\hat{Q} : P \to R$ is defined by

$$\hat{Q}(p) = \inf_{y \in Y_p} Q(y). \tag{1.15}$$

In virtue of (1.7) the optimal value for (1.1) is

$$\hat{Q}(p_o) = \inf_{y \in Y} \sup_{w \in W} L(y,w). \tag{1.16}$$

2. DUALITY

The theory of Lagrange multipliers for convex problems is strongly related to the theory of Fenchel conjugate functions in convex analysis [11] [14] [18] . During last several years some attempts have been made to extend the tools of convex analysis to nonlinear problems (e.g. [11]). In particular, the notion of φ -conjugate functions has been introduced by Weiss [27] and Vogel [28] and applied to the study of augmented Lagrangians similar to (1.9) by Seidler [26]. Given the primal problem (1.16) its dual may be formulated:

$$\text{Find} \quad \sup_{w \in W}(D) = \sup_{w \in W} \inf_{y \in Y} L(y,w) \tag{2.1}$$

Always

$$\sup(D) \leqslant \hat{Q}(p_o). \tag{2.2}$$

Define

$$\hat{L}(w) = \inf_{y \in Y} L(y,w)$$

and note that

$$\hat{L}(w) = \inf_{y \in Y} \inf_{p \in P_y} \{ Q(y) + \varphi(p,w) - \varphi(p_o,w) \} =$$

$$= \inf_{p \in P} \inf_{y \in Y_p} \{ Q(y) + \varphi(p,w) - \varphi(p_o,w) \} =$$

$$= - \varphi(p_o,w) + \inf_{p \in P} \{ \varphi(p,w) + \hat{Q}(p) \} .$$

For a given function $F : P \to R$ Weiss [27] defines its $-\varphi$ - conjugate by

(2.3) $$F^{*}(w) = \sup_{p\in P}\left\{-\varphi(p,w) - F(p)\right\} = -\inf_{p\in P}\left\{\varphi(p,w) + F(p)\right\}$$

and the second $-\varphi$-conjugate of F by

(2.4) $$F^{**}(p) = \sup_{w\in W}\left\{-\varphi(p,w) - F^{*}(w)\right\}.$$

Therefore we may write:

(2.5) $$\hat{L}(w) = -\varphi(p_o,w) - \hat{Q}^{*}(w)$$

(2.6) $$\sup (D) = \hat{Q}^{**}(p_o).$$

Therefore the problem when inequality (2.2) becomes an equality, so that we have the weak duality

(2.7) $$\inf_{y}\sup_{w} L(y,w) = \sup_{w}\inf_{y} L(y,w) ;$$

is equivalent to asking when

(2.8) $$\hat{Q}(p_o) = \hat{Q}^{**}(p_o).$$

The well known theorem of Moreau-Fenchel [11] states that (2.8) holds for a convex function $\hat{Q}$ defined on a topological vector space P at any p_o if and only if $\hat{Q}$ is l.s.c. Somehow similar requirements are needed in our general case.

<u>Theorem 2.1</u>

(compare [7])

(3.7) or, equivalently, (2.8) holds iff there is a sequence $\{w_n\}_{n=1}^{\infty} \subset W$ satisfying

(2.9) $$\hat{Q}(p) - \hat{Q}(p_o) \geqslant \varphi(p_o,w_n) - \varphi(p,w_n) - \frac{1}{n} \qquad \forall p\in P.$$

<u>Proof.</u>

$\hat{Q}(p_o) = \hat{Q}^{**}(p_o) = \sup\limits_{w}\inf\limits_{p}\left\{\varphi(p,w)-\varphi(p_o,w) + \hat{Q}(p)\right\}$ iff for each integer $n \geqslant 0$ an $w_n \in W$ exists such that

$$\hat{Q}(p_o) \leqslant \inf_{p}\left\{\varphi(p,w) - \varphi(p_o,w) + \hat{Q}(p)\right\} + \frac{1}{n}$$

or, which is the same

$$\hat{Q}(p_o) \leq \varphi(p,w) - \varphi(p_o,w) + \hat{Q}(p) + \frac{1}{n} \qquad \forall p \in P$$

as stated in (2.9).

In the theorem of Moreau-Fenchel the two properties are distinguished: convexity and lower semicontinuity. Lying aside the question of what may be here called " φ-convexity" (some definitions exist [28] but they are not very constructive) we observe some simple facts concerning the lower semicontinuity of Q and its conjugates.

In the sequel let τ_w (resp. τ_p) denote any topology in W (resp. P) such that all functions $\varphi(p, \cdot)$, $p \in P$ (resp. $\varphi(\cdot, w)$, $w \in W$) are u.s.c. in τ_w (τ_p). Observe that functions $\varphi(\cdot,w)$, $w \in W$ generate at least one such topology in P, for instance the weakest possessing this property. It is interesting to note that the convergence in P in this weakest topology is characterized as follows:

$$p_n \xrightarrow[n]{} p_o \Longleftrightarrow \overline{\lim_n}\, \varphi(p_n,w) \leqslant \varphi(p_o,w) \qquad w \in W .$$

so that $\{p_n\}$ converges to p_o iff it converges in terms of all "distributions of prices" $\varphi(\cdot,w)$, $w \in W$. In the examples described in the preceding section this topology is equivalent to weak topology in P (example (i)) and to norm topology (example (iii)).

<u>Proposition 2.1</u>

For every function $F: P \rightarrow R$, F^* is l.s.c. in τ_w and F^{**} in τ_p.

<u>Proof.</u>

It is sufficient to show [11] that for any $\alpha \in R$ the level set $\{ w \in W : F^*(w) \leqslant \alpha \}$ is closed. It is, as a product of closed sets:

$$\{w : F^*(w) \leq \alpha\} = \{w: -\varphi(p,w) - F(p) \leq \alpha \quad \forall p \in P\} = \\ = \bigcap_{p \in P} \{w: \varphi(p,w) \geq -F(p) - \alpha\} .$$

The proof for F^{**} is analogous.

Proposition 2.2

If $\hat{Q}(p_o) = \hat{Q}^{**}(p_o)$ then $\hat{Q}$ is l.s.c. in τ_p at p_o.
Proof.
For any integer $n \geq 1$ take w_{2n} as in (2.9). Since $\varphi(\cdot, w_{2n})$ is u.s.c. in τ_p, for some neighborhooed U of p_o

$$\varphi(p,w_{2n}) \leq \varphi(p_o,w_{2n}) + \frac{1}{2n} \qquad \forall p \in U.$$

Combining this with (2.9) we have

$$\hat{Q}(p) \geq \hat{Q}(p_o) - \frac{1}{n} \qquad \forall p \in U .$$

Lower semicontinuity of the primal functional $\hat{Q}(p)$ over Y is characterized in the following way (compare Dolecki [3] for the case of linear operator equality constraints).

Theorem 2.2

For $\alpha \in R$ denote $A_\alpha = \{p: \hat{Q}(p) \leq \alpha\}$, $PB_\alpha = \{p: p \in P_y, Q(y) \leq \alpha\}$.
Then $\hat{Q}$ is l.s.c. on P (in any topology τ) iff

$$A_\alpha = \bigcap_{\varepsilon > 0} \overline{PB}_{\alpha+\varepsilon} \qquad \forall \alpha \in R. \tag{2.10}$$

Proof.
Sufficiency of (2.10) is clear, since all level sets of $\hat{Q}$ are closed.

Necessity. Observe that for $\alpha \in R$

$$\hat{Q}(p) \leq \alpha \iff \forall \varepsilon > 0 \quad \exists y \in Y_p \quad Q(y) \leq \alpha + \varepsilon$$

so that

$$A_\alpha = \bigcap_{\varepsilon > 0} PB_{\alpha+\varepsilon} \tag{2.11}$$

and

$$A_\alpha \subset \bigcap_{\varepsilon > 0} \overline{P\,B}_{\alpha+\varepsilon}$$

To prove the converse inclusion, take $p \in \bigcap_{\varepsilon > 0} \overline{P\,B}_{\alpha+\varepsilon}$. Since $\hat{Q}$ is l.s.c., for any $\delta > 0$ there is a neighborhood U of p such that

$$(2.12) \qquad \hat{Q}(p) \leq \hat{Q}(p') + \delta \qquad \forall p \in U$$

Now let $\varepsilon > 0$; there is $p' \in U \cap P\,B_{\alpha+\varepsilon}$, i.e.

$$Q(y) \leq \alpha + \varepsilon \qquad \text{for some} \quad y \in Y_{p'}.$$

Together with (2.12) this yields

$$\hat{Q}(p) \leq \hat{Q}(p') + \delta \leq Q(y) + \delta \leq \alpha + \delta + \varepsilon .$$

Since $\delta, \varepsilon > 0$ were arbitrary, $\hat{Q}(p) \leq \alpha$ and $p \in A_\alpha$. Therefore (2.10) holds.

Example

Suppose Y is a topological, P a topological Hausdorff vector space and we consider the nonlinear program (1.10) described previously

$$\text{minimize} \quad Q(y) \qquad \text{subject to} \quad p - S(y) \in D$$

where $S : Y \longrightarrow P$ is continuous and $D \subset P$ is a closed set. If the level sets B_α, $\alpha \in R$ of Q are compact then the sets

$$P\,B_\alpha = \{p : p = S(y) + d, \ d \in D, \ y \in B_\alpha\} = S(B_\alpha) + D$$

are closed for any α, being a sum of a closed set and a compact. In virtue of (2.11) this means that (2.10) holds and $\hat{Q}$ is l.s.c.

In the context of generalized Lagrange functionals (1.9) also other questions can be discussed, for instance the problem of strong duality instead of weak duality (2.7):

$$(2.13) \qquad \inf_y \sup_w L(y,w) = \max_w \inf_y L(y,w).$$

Supposing a maximizing $\hat{w} \in W$ exists (and is known!), it would be possible to compute the optimal value $\hat{Q}(p_o)$ in one unconstrained minimization of $L(\cdot, w)$, which is a very desirable property. It turns out that (2.13) is equivalent to supporting $\hat{Q}$ at p_o by $\varphi(p_o, \hat{w}) - \varphi(\cdot, \hat{w})$ for some (maximizing) $\hat{w} \in W$:

$$\hat{Q}(p) - \hat{Q}(p_o) \geqslant \varphi(p_o, \hat{w}) - \varphi(p, \hat{w}) \qquad \forall p \in P.$$

(Compare [19] [31]).

One may further proceed with deriving different problems from (1.1) with help of functions $\varphi(\cdot, w)$ (constrained Lagrangian minimization, surrogate problems etc.) and discussing their properties. See [1] [6] [7] [8] [9].

3. EVERETT THEOREM AND THE METHOD OF MULTIPLIERS

<u>Theorem</u>

Suppose $\bar{y}$ satisfies

$$(3.1) \qquad L(\bar{y}, w) \leqslant L(y, w) + \varepsilon \qquad \forall y \in Y$$

and $\bar{p} \in P_{\bar{y}}$ is such that

$$(3.2) \qquad K(\bar{y}, w) = \varphi(\bar{p}, w) - \varphi(p_o, w)$$

Then $\bar{y}$ is an ε-solution to the problem (1.1) with p_o changed to $\bar{p}$, i.e.

$$Q(\bar{y}) \leq \hat{Q}(\bar{p}) + \varepsilon \ , \qquad \bar{y} \in Y_{\bar{p}}.$$

<u>Proof.</u> is classical [4] [1]

From the algorithmic point of view it would be most desirable to establish the existence of a $w \in W$ such that $\bar{y}$ determined by (3.1) would satisfy the constraints, i.e. $p_o \in P_{\bar{y}}$. Then the whole constrained problem (1.1) would reduce to the single minimization of $L(\cdot, w)$ without constraints. Since such a $w = \hat{w}$ would be generally unknown, one should apply an iterative scheme to find it.

Assume from now on that for any $w \in W$ and $y \in Y$ with $P_y \neq \emptyset$ the set $P(y, w)$ is nonempty, where

$$P(y, w) = \{ p \in P_y : K(y, w) = \varphi(p, w) - \varphi(p_0, w) \} .$$

Given $\varepsilon \geq 0$ and $w \in W$ denote by $Y(\varepsilon, w)$ the set of all $\bar{y}$'s satisfying (3.1), i.e. being ε-minimal points for $L(\cdot, w)$ over Y. Finally, denote

$$P_{\varepsilon}(w) = \bigcup_{\bar{y} \in Y(\varepsilon, w)} P(\bar{y}, w) .$$

Then one can find an ε-solution to (1.1) minimizing ε-approximately $L(\cdot, w)$ over Y iff a $\hat{w} \in W$ exists solving the inclusion

(3.3) $$p_0 \in P_{\varepsilon}(w) .$$

Suppose now that:

(i) P is a Banach space, $W = V \subset P$ and for $w = v \in V$

$$\varphi(p, v) = \varrho \Psi(p - v)$$

where $\varrho > 0$ is treated as a parameter.

(ii) For some $\varrho = \varrho_0 > 0$ and each $v \in V$ an element y_v exists minimizing $L(\cdot, v)$ over Y.

(iii) (for simplicity) For this $\varrho = \varrho_0$ and all $v \in V$ the set $P_0(v)$ contains precisely one element, say p_v.

Under these assumptions (3.3) becomes:

$$p_0 = p_v$$

or

(3.4) $$v = v + p_0 - p_v = T(v).$$

This is a fixed point problem and one may try to solve it iteratively using the method of successive approximations: given initial $v_1 \in V$, one takes

(3.5) $$v_n = T(v_{n-1}), \qquad n = 2,3,\ldots .$$

The algorithm therefore works as follows:

1. Select $\varrho = \varrho_o$ and initial v_1. Set $n := 1$.
2. Minimize $L(\cdot, v_n)$ over Y obtaining $y_n = y_{vn}$ and $p_n = p_{v_n}$.
3. Update v_n by setting

(3.6) $$v_{n+1} := v_n + \eta(p_o - p_n), \qquad \eta > 0 \text{ fixed.}$$

4. Set $n := n+1$ and go to 2.

It should be noted that the formula (3.6) is independent of the function Ψ and ϱ_o chosen, although y_n and p_n clearly depend on them.

This is the algorithm (shifted penalty method, method of multipliers) of Hestenes [10] and Powell [16] for equality constraints in R^n, generalized to inequality constraints in R^n and in Hilbert space by Wierzbicki [29] [30]. The rate of convergence of this algorithm for $P = R^n$ has been investigated in 1971 by Wierzbicki [29] and later by several authors, most completely by Bertsekas [2]. All of them used quadratic functionals. For the theoretical discussion of this algorithm see also [10] [19] [20] for $P = R^n$ [30] [31] for $P =$ a Hilbert space, [23] [24] [25] for variational and optimal control problems and others.

It is clear that one may attempt to solve (3.4) by other methods than successive approximations, e.g. Newton or variable metric method, provided the operation $v \longmapsto v - T(v)$ possesses desirable properties.

4. THE SHIFTED PENALTY METHOD FOR LINEAR OPERATOR CONSTRAINTS

In this section we shall investigate briefly the application of the shifted penalty method to an important special case of the optimization problem (1.1) and present a theorem precising the conditions under which this method is convergent. The next section contains numerical examples.

Suppose Y, P are Hilbert spaces and $S \in \mathcal{L}(Y, P)$. The problem is

(4.1) $$\text{minimize } \|y\|^2 \text{ subject to } Sy = p_o .$$

This is a special case of (1.10) with the cone D reduced to $\{0\}$. For every $p_o \in \text{im}S$ (4.1) has a unique solution $\hat{y}(p_o)$. We set $W = V = \text{im}S$ and $\varphi(p,v) = \varrho\|p - v\|^2$, as in the preceding section, treating here ϱ as a parameter. The shifted penalty method has been formulated above and the Lagrangian is here

(4.2) $$L(y, v) = \|y\|^2 + \varrho\|Sy - v\|^2 - \varrho\|p_o - v\|^2 .$$

<u>Theorem 4.1</u>

The shifted penalty method with $\eta = 1$ converges for each $p_o \in \text{im}S$, initial $v_1 \in \text{im}S$ and $\varrho > 0$ if, and only if $\text{im}S$ is a closed subspace of P. In this case the rate of convergence is estimated by:

(4.3) $$\frac{\|v_{n+1} - v_n\|}{\|v_n - v_{n-1}\|} \leqslant \frac{1}{1 + \varrho\varkappa} \qquad \text{where } \varkappa > 0.$$

<u>Proof.</u>

Sufficiency. The mapping T defined by (3.4) can be here explicity expressed. For any $v \in P$, unique y_v exists minimizing $L(\cdot\,, v)$ over P. By (4.2), this y_v satisfies

$$L'(y_v, v) = y_v + \varrho S^*(S\,y_v - v) = 0.$$

The corresponding p_v is equal to $p_v = S\, y_v$ and hence from the above equation

$$y_v = -\varrho S^*(p_v - v),$$
$$p_v = S\, y_v = -\varrho\, SS^*(p_v - v)$$

and

$$(4.4) \qquad (I + \varrho S S^*)\, p_v = \varrho SS^* v\,.$$

Since imS is closed, by Banach closed range theorem [32] $\mathrm{im}S^*$ also is. Operator S^* is an injection from imS onto $\mathrm{im}S^*$; both these subspaces being complete, S possesses a continuous inverse $S^{*-1} : \mathrm{im}S^* \longrightarrow \mathrm{im}S$ in virtue of the open mapping theorem [32]. Denote $\varkappa = \| S^{*-1}\|^{-2}$ (norm computed with respect to $\mathrm{im}S^*$) and $A = I + \varrho SS^*$. Clearly, $A^* = A$ and $A : \mathrm{im}S \longrightarrow \mathrm{im}S$. Moreover, for each $p \in \mathrm{im}S$

$$\langle p, A\, p\rangle = \langle p, (I + \varrho SS^*)p\rangle = \|p\|^2 + \varrho\|S^*p\|^2 \geqslant$$
$$\geqslant \|p\|^2 + \varrho\varkappa\|p\|^2 = (1 + \varrho\varkappa)\,\|p\|^2.$$

In virtue of Lax-Milgram theorem [32], $A \in \mathcal{L}$ (imS, imS) admits the continuous inverse on imS and

$$(4.5) \qquad \|A^{-1}\| \leqslant \frac{1}{1 + \varrho\varkappa}\,.$$

Now, for $v \in \mathrm{im}S$, p_v may be calculated from (4.4):

$$p_v = A^{-1} \varrho\, SS^* v = A^{-1}(Av - v) = v - A^{-1} v.$$

The mapping $T: \mathrm{im}S \longrightarrow P$ is defined by (3.4):

$$T(v) = p_o - p_v + v = p_o + A^{-1}v$$

Hence for every $p_o, v \in \mathrm{im}S$, $T(v) \in \mathrm{im}S$ and for $v', v'' \in \mathrm{im}S$

$$\| T(v') - T(v'')\| \leq \|A^{-1}\|\, \|v' - v''\| \leq \frac{1}{1 + \varrho\varkappa}\|v' - v''\|\,.$$

Since $\varkappa > 0$, T is a contraction mapping in a complete metric space imS. Hence the algorithm converges to a point $v = T(v)$

with the rate of convergence (4.3).

Necessity. If the algorithm is convergent, then the equation

$$v = T(v)$$

admits a solution for each $p_o \in \text{im}S$. According to the definition of T (see the preceding section), for each $p_o \in \text{im}S$ an element $v \in P$ exists such that the unique optimal solution $\hat{y}(p_o)$ to (4.1) minimizes $L(\cdot, v)$ over Y. By (4.2), $\hat{y}(p_o)$ must satisfy

$$L'(\hat{y}(p_o), v) = \hat{y}(p_o) + \varrho S^*(S\hat{y}(p_o) - v) = 0$$

so that

$$(4.5) \qquad \hat{y}(p_o) \in \text{Im}S^*. \qquad \forall p_o \in \text{im}S .$$

Suppose imS is not closed, so that by closed range theorem [32] $(\ker S)^{\perp} \setminus \text{im}S^*$ contains an element y_o. Select $p_o = Sy_o$. For any y such that $Sy = p_o$ obviously $\langle y_o, y - y_o \rangle = 0$, so that

$$\|y\|^2 - \|y_o\|^2 = \langle y + y_o, y - y_o\rangle = \langle y - y_o, y - y_o\rangle = \|y - y_o\|^2 .$$

Hence $y_o = \hat{y}(p_o)$, the unique solution to (4.1). Then (4.5) yields a contradiction, since $y_o \notin \text{im}S^*$. Therefore imS must be closed.

The "if" part of theorem 4.1 can be also obtained from a more general theorem [31]. The condition that imS be closed in P is intimately related to the existence of Lagrange multipliers for the problem (4.1.) with any $p_o \in \text{im}S$ [13] [21]. In the convex case the existence of Lagrange multipliers is sufficient for the convergence of the method of multipliers, as has been shown by Rockafellar for $P = R^n$ [20].

In the course of sufficiency proof the constant in (4.3) was taken to be $\|S^{*-1}\|^{-2}$. Thus, the smaller is the norm of

$\| S^{*-1} \|$, the quicker the convergence. Therefore the rate of convergence depends on the norm chosen; this is confirmed also by the numerical results.

It is also possible to apply the shifted penalty method with $\eta \neq 1$; but then η and ϱ must satisfy

$$0 < \eta < 1 + \frac{\varkappa \varrho}{2 + \varkappa\varrho} ,$$

the rate of convergence is expressed by

$$\frac{\| v_{n+1} - v_n \|}{\| v_n - v_{n-1} \|} \leq | 1 - \eta | + \frac{\varrho}{1 + \varkappa\varrho}$$

and is the best at $\eta = 1$.

5. NUMERICAL EXAMPLES

Two optimal control problems for linear time-lag systems with fixed final function were solved numerically.

Example 1

minimize $Q(y) = \int_0^2 y^2(t)dt$

for the control $y \in L^2(0,t)$

subject to constraints

$$(5.1) \qquad \begin{cases} \dot{x}(t) = - x(t-1) + y(t) & \text{a.e. in } [0,2] \\ x(t) = 0 & \forall t \in [-1,0] \end{cases}$$

$$(5.2) \qquad x(t) = - \frac{1}{2} (t-1)^2 \qquad \forall t \in [1,2]$$

This problem was taken from [12]. This is the special case of (4.1). The operator S is defined as follows: given control $y(\cdot)$, solve (5.1) and put $Sy = x|_{[1,2]}$. The element p_0 is defined by (5.2). For each $y \in L^2(0,2)$, $Sy \in W_1^2(1,2)$, the Sobolev space of absolutely continuous functions with square integrable desirative. Since $W_1^2(1,2) \subset L^2(1,2)$, one may use here at least two spaces of constraints: $P_1 = L^2(1,2)$ and $P_2 = W_1^2(1,2)$. imS is closed in P_2 and not closed in P_1. Consequently, two Lagrangians can be used (we neglect the term $\varrho \| p_o - v \|^2$ since it does not influence the computations):

$$L_1(y,v) = \int_0^2 y^2(t)dt + \varrho \int_1^2 (x(t) - v(t))^2 \, dt$$

$$L_2(y,v) = \int_0^2 y^2(t)dt + \varrho (x(2) - v(2))^2 + \varrho \int_1^2 (\dot{x}(t) - \dot{v}(t))^2 dt$$

The method used was (after the problem was discretized) the original algorithm of Powell [16], which increases ϱ if the improvement in the constraints violation was too small. Computations have been carried out for several values of initial ϱ_o. Both algorithms were the same and the computational effort per one evaluation of L_1 or L_2 was practically the same. The results are:

ϱ_o	Number of evaluation of L		Final constraint violation	
	$P_1 = L^2(1,2)$	$P_2 = W_1^2(1,2)$	$P_1 = L^2(1,2)$	$P_2 = W_1^2(1,2)$
0.01	742	152	$0.921 \cdot 10^{-3}$	$0.44 \cdot 10^{-3}$
1.0	816	157	$0.983 \cdot 10^{-3}$	$0.92 \cdot 10^{-3}$
10	1153	161	$0.452 \cdot 10^{-3}$	$0.64 \cdot 10^{-3}$

Computations have been performed on an IBM-360 in Fortran. The constraint violation was in both cases measured as

$$\max_{t \in [1,2]} \left| x(t) + \frac{1}{2} (t-1)^2 \right| .$$

<u>Example 2</u>. (K.M. Przyłuski [17]).

$$\text{minimize} \quad Q(y) = \int_0^3 (y(t) - z(t))^2 \, dt \quad {}^{1)}$$

over $y \in L^2(0,3)$ subject to the constraints:

$$(5.3) \quad \begin{cases} \dot{x}_1(t) = u(t) & \text{a.e. in } [0,3] \\ \dot{x}_2(t) = x_1(t-1) & \\ x_1(t) = x_2(t) = 0 & \forall\, t \in [-1,0] \end{cases}$$

$$(5.4) \quad \begin{cases} x_1(t) = t - 1 & \\ x_2(t) = \frac{1}{2}(t-2)^2 & \forall\, t \in [2,3] \end{cases}$$

where
$$z(t) = \begin{cases} 0 & , \ t \in [0, 3/2] \\ 1 & , \ t \in (3/2, 3] . \end{cases}$$

For given control $y \in L^2(0,3)$ the operator Sy is defined by $S\,y = (x_1, x_2)\big|_{[2,3]}$ where (x_1, x_2) solve (5.3). The element p_0 is determined by (5.4). Since $x_2(\cdot)$ is absolutely continuous, one can put either $P_1 = W_1^2(2,3\,;\,R^2)$ or $P_2 = W_1^2(2,3) \times W_2^2(2,3)$. imS is not closed in P_1 and closed in P_2 [13]. The corresponding Lagrangians are (also without the term $\varrho \| p_0 - v \|^2$):

$$L_1(y,w) = \int_0^3 (y(t) - z(t))^2 dt + \varrho |x(3) - v(3)|^2 + \int_2^3 |\dot{x}(3) - \dot{v}(3)|^2 dt$$

$$L_2(y,w) = \int_0^3 (y(t)-z(t))^2 dt + \varrho |x(3)-v(3)|^2 + \varrho |\dot{x}_2(3) - \dot{v}_2(3)|^2 +$$

$$+ \varrho \int_2^3 \left(|\dot{x}_1(t) - \dot{v}_1(t)|^2 + |\ddot{x}_2(t) - \ddot{v}_2(t)|^2 \right) dt.$$

The computational effort per one evaluation of L_1 or L_2 was more less the same. The results are:

---

1) This is not exactly the problem of (4.1) type but can be transformed to (4.1) by a simple shift of zero in Y.

Discretization of [0,3]	ϱ_0	Number of evaluation of L		Final constraint violation	
		$P_1 = W_1^2$	$P_2=W_1^2 x W_2^2$	$P_1 = W_1^2$	$P_2= W_1^2 x W_2^2$
12	1	200	88	$0.649\cdot 10^{-3}$	$0.960\cdot 10^{-3}$
	10	184	60	$0.624\cdot 10^{-3}$	$0.912\cdot 10^{-3}$
	100	146	50	$0.682\cdot 10^{-3}$	$0.466\cdot 10^{-3}$
90	1	1090	44	> 0.012	$0.246\cdot 10^{-2}$
	10		40		$0.177\cdot 10^{-2}$
	100		30		$0.177\cdot 10^{-2}$

The constraint violation was here measured by

$$c = \max(|x_1(3)-2|, |x_2(3)- \tfrac{1}{2}|, \max_{t\in[2,3]} (|\dot{x}_1(t)-1|, |\dot{x}_2(t)- t + 2|))$$

Summarising, in these two cases the computational effort for solving the problem with similar accuracy was 2-3 times smaller for the Lagrangian L_2, employing the proper norm. For more detailed description and discussion of these results see [13] [17].

6. CONCLUSIONS

A class of generalized Lagrangians has been defined associated with a family of extremal problems with general constraining set. The relation of these Lagrangians to others found in literature was indicated; it occurs that many facts proved for generalized Lagrangians of nonlinear programs with R^n constraints remain true in much more general setting. Some simple properties of these Lagrangians have been shown, especially those related to the topology in the set of constraints values; also the Everett theorem and an abstract formulation of the method of multipliers (shifted penalty technique) were given. A theorem has been provided characterizing the convergence of this method in the case of linear-quadratic problems in Hilbert space. The last section contains numerical examples of the application of the algorithm to time-delay optimal control problems.

The results show that the behavior of the finitedimensional algorithm applied to the discretized version of inifinite-dimensional problem may depend on topological properties of the original problem.

REFERENCES

[1] M. Bellmore, H.J. Greenberg, J.J. Jarvis - Generalized penalty - function concepts in mathematical optimization - Opns. Res., 18 (1970), No 2.

[2] D.P. Bertsekas - Combined primal-dual and penalty method for constrained minimization - SIAM J. Control 13 (1975), No 3.

[3] S. Dolecki - Bounded controlling sequences, inf - stability and certain penalty procedures, to appear.

[4] H. Everett III - Generalized Lagrange multipliers method for solving problems of optimum allocation of resources - Opns. Res. 11 (1963), No 3.

[5] F. Glover - Surrogate constraints - Opns. Res. 16 (1968), 741-769

[6] F.J. Gould - Extensions of Lagrange multipliers in nonlinear programming - SIAM J. Appl. Math. 17 (1969) No 6.

[7] F.J. Gould - Nonlinear pricing: applications to concave programming - Opns. Res. 19 (1971), No 4.

[8] H.J. Greenberg - The generalized penalty function/surrogate model - Opns. Res. 21 (1973), No 1.

[9] H.J. Greenberg, W.P. Pierskalla - Surrogate mathematical programming - Opns. Res. 18 (1970), 924-939.

[10] M.R. Hestenes - Multiplier and gradient methods. In: Computing methods in optimization problems - 2, ed. by: L.A. Zadeh, L.W.: Neustadt, A.V. Balakrishnan, Academic Press 1969, 143-164.

[11] A.D. Ioffe, W.M. Tikhomirov - Theory of extremal problems (in Russian) - Nauka, Moscow 1974.

[12] M.Q. Jacobs, T.J. Kao - An optimum setting problem for time-lag systems - J. Math. Anal. Appl, 40 (1973), 687-707.

[13] S. Kurcyusz - Necessary optimality conditions for problems with function space constraints (in Polish) - Ph. D. Thesis, Instytut Automatyki, Politechnika Warszawska, Warsaw 1974.

[14] P.J. Laurent - Approximation et optimization - Herrman, Paris 1972

[15] O.L. Mangasarian - Unconstrained Lagrangians in nonlinear programming - SIAM J. Control 13 (1975), No 4.

[16] M.J.D. Powell - A method for nonlinear constraints in minimization problems, in: Optimization, ed. by R. Fletcher, Academic Press 1969, 283-298.

[17] K.M. Przyłuski - Application of the shifted penalty method to dynamic optimization of delay processes (in Polish) - M. Sc. Thesis, Instytut Automatyki, Politechnika Warszawska, Warszawa 1974.

[18] R.T. Rockafellar - Convex analysis - Princeton University Press, Princeton 1970.

[19] R.T. Rockafellar - Augmented Lagrange multiplier functions and duality in nonconvex programming - SIAM J. Control, 12 (1974), No 2.

[20] R.T. Rockafellar - The multiplier method of Hestenes and Powell applied to convex programming - J. Opt. Theory Appl. 10 (1973).

[21] S. Rolewicz - Functional analysis and control theory (in Polish) - PWN, Warsaw 1974.

[22] J.D. Roode - Generalized Lagrangian functions and mathematical programming - in: Optimization, ed. by R. Fletcher, Academic Press 1969.

[23] R.D. Rupp - A method for solving a quadratic optimal control problem - J. Opt. Theory Appl. 9 (1972), No 4.

[24] R.D. Rupp - Approximation of the Classical Isoperimetric Problem - J.Opt.Theory Appl. 9 /1972/,pp.251-264

[25] R.D. Rupp - A nonlinear optimal control minimization technique - Trans. AMS, 178 (1973), 357-381.

[26] K.H. Seidler - Zur Dualisierung in der nichtlinearen optimierung - Ph. D. Thesis, Technische Hochschule Ilmenau, Ilmenau 1972.

[27] W. Vogel - Duale Optimierungsaufgaben und Sattelpunktsatze - Unternehmensforschung 13 (1969), 1-28.

[28] E.A. Weiss - Konjugierte Funktionen - Arch. Math. 20 (1969), 538-545.

[29] A.P. Wierzbicki - A penalty function shiffting method in constrained static optimization and its convergence properties - Archiwum Automatyki i Telemechaniki 16 (1971), 395-416.

[30] A.P. Wierzbicki, A. Hatko - Computational methods in Hilbert space for optimal control problems with delays - Proc. of 5-th IFIP Conference on Optimization Techniques, Rome 1973.

[31] A.P. Wierzbicki, S. Kurcyusz - Projection on a cone, generalized penalty functionals and duality theory - Institute of Automatic Control, Technical University of Warsaw Technical Report No 1/1974.

[32] K. Yosida - Functional analysis - Springer Verlag, Berlin 1966.