Average relaxations of extremal problems

Anatoliy M. Tsirlin
Program System Institute, Russian Academy of Science, set."Botic",Perejaslavl-Zalesky, Russia 152020. e-mail: tsirlin@sarc.botik.ru
and Vladimir Kazakov
School of Finance and Economics, University of Technology, PO Box 123, Broadway NSW 2007 Australia, e-mail:Vladimir.Kazakov@uts.edu.au

May 1, 2003

Abstract

In this paper extremal problems that include averaging operation in constraints and objective are considered. The relaxation caused by a replacement of a problem without averaging with a problem that includes averaging is formally defined and investigated. Canonical form for nonlinear programming problem with averaging is constructed and its conditions of optimality are derived. It is shown how optimality conditions for optimal control problems with various types of objectives and constraints can be derived using its averaged relaxation.
Keywords: Averaging, constraint relaxation, nonlinear programming, optimal control problem, optimality conditions
AMS subject classification: 49K19

1 Introduction

One of the main approaches to solution of an extremal problem is by replacing it with some other (auxiliary) extremal problem with a larger set of feasible solutions. There are a number of cases when this approach is used.

The first case occurs when a solution of the auxiliary problem is simpler than a solution of the original problem, the conditions in the auxiliary problem depend on some parameters and, for some values of these parameters, the optimal solution or optimal value of the auxiliary and original problems are the same [1]. The best known examples of this approach are the use of penalty functions to reduce constrained to unconstrained optimization and the sufficient conditions of optimality for optimal control problems that are based on Krotov's lemma [2].

The second common case when the replacement of the original problem with its averaged relaxation is used is when there is no certainty that a solution of the original problem exists. If the optimial solution of an auxiliary problem with a larger feasible set is found and it turns out to be non-feasible in the original problem then it often can be approximated by some feasible sequence with an arbitrary accuracy. Sliding regimes in optimal control are the best known examples of this case. If the solution of an auxiliary problem turns out to be feasible with respect to the conditions of the original problem then it is also a solution of the original problem.

And finally it is possible that a problem with an enlarged set of feasible solutions could be interesting by itself, the objective then being to investigate whether its solution belongs to some subset of its feasible solutions.

First we consider the following definition of the extension of the extremal problem, which was given in [1]. Extremal problem B is called an extension of the original problem A (or an extended problem A) [1] is the following two conditions hold:
1. The sets of the feasible solutions of the problems A and B relay to each other as

$$D_B \supset D_A. \tag{1.1}$$

where D_A and D_B are the feasible sets of the problems A and B correspondingly.
2. The optimality criterion I_B of the problem B coincides with optimality criterion of the problem A on the set D_A

$$I_B(x) = I_A(x), \quad x \in D_A. \tag{1.2}$$

¿From (1.1) and (1.2) it follows that the values of the original and the extended problems $I_A(x_A^*)$ and $I_B(x_B^*)$ obey the following inequality

$$I_A^* = I_A(x_A^*) \leq I_B^* = I_B^*(x_B^*), \tag{1.3}$$

where x_A^* and x_B^* are the solutions of the problems A and B, and I_A^* and I_B^* are the values of these problems.

A number of extremal problems' formulations, in addition to vector and functional variables, includes their average values or average values of the functions that depend on these variables [3], [4]. Later we will demonstrate that a problem that includes averaging can be viewed as an extention (relaxation) of an extremal problem without averaging. We will compare this way of constructing extention of an extremal problem with the other ways of doing this for the nonlinear programming problem and for the variational problem with the scalar argument. In order to do this we we have to generalize the definition of extension (1.1)-(1.2).

All the theorems and lemmas in the main body of this paper are presented without proofs. The proofs are then given in the Appendix.

2 Definition of an extension

Let us call the problem

$$f(x) \to \max, \quad x \in D \tag{2.1}$$

the original problem, whose solution is the vector $x^* \in D \supset V \supset R^n$.

Assume that $x \in V$ is a random variable, its probability density distribution is $P(x)$ and this distribution is to be chosen. The evaluation of the quality of $P(x)$ can be carried out using some criterion. In particular, $P(x)$ can be chosen from the condition of function $f(\overline{x})$ maximum, where $\overline{x}$ is the ensemble average value of x. In this case the problem of how to choose $P(x)$ takes the form

$$f(\overline{x}) = f\Big[\int_V xP(x)dx\Big] \to \max_{P(x)} \tag{2.2}$$

subject to constraints

$$\int_V P(x)dx = 1, \quad P(x) \geq 0. \tag{2.3}$$

The other way of evaluating the quality of a probability distribution is via the average value of the function f over the ensemble of solutions

$$\overline{f(x)} = \int_V f(x)P(x)dx \to \max_{P(x)} \tag{2.4}$$

subject to the same conditions (2.3). Here it is not required that each element of the sample belongs to D. It is sufficient that, for example, the average value of x belongs to D

$$\int_V xP(x)dx \in D. \tag{2.5}$$

If the problem (2.1) is convex and its optimal solution is x^* then the optimal solutions $P^*(x)$ in problems (2.2) and (2.4) are the same. In this case

$$P^*(x) = \delta(x - x^*), \tag{2.6}$$

where δ is Dirac function. The values of all problems (2.1), (2.2) and (2.4) are also the same. In the general case, the values of the problems (2.2) or (2.5) and the problem (2.1) obey the inequality similar to (1.3).

Since the feasible set of the problems (2.2) and (2.4) does not belong to R^n, the condition (1.1) is meaningless.

Let us change the definition of an extension of extremal problem. First we introduce the definition of the isomorphism of two extremal problem.

Problems A_0 and A_1 are isomorphic (identical with respect to solutions) if it is possible to find such one-to-one mapping between their sets of feasible solutions that from the inequality

$$I_A(y) \geq I_A(z), \quad (y, z) \in D_A \tag{2.7}$$

follows that

$$I_{A_1}(y_1) \geq I_{A_1}(z_1), \quad (y_1, z_1) \in D_{A_1}, \tag{2.8}$$

where y_1 and z_1 correspond to y and z. Here the feasible sets D_{A_0} and D_{A_1} can belong to the different spaces.

The equivalent class $\bar{A}$ is defined as a set of all the problems which are isomorphic to the problem A.

Assume that the element y_0 can not be improved on some subset $\Delta A_0 \supset D_{A_0}$ (that is, y_0 obeys necessary conditions of optimality of the problem A_0). Then the element y_1^0 which corresponds to y_0 obeys the necessary conditions of optimality of the problem A_1. Therefore, the optimality conditions for all the problems that belong to $\bar{A}$ are obtained if they are obtained for any one problem from this class.

Now we can give the following generalized definition of an extended problem: *the problem B is called an extension of the problem A if the conditions (1.1) and (1.2) hold for any problem from the class $\overline{A}$.*

For example, according to this definition, it is possible to establish a one-to-one correspondence between the elements $\tilde{x}$ of the set D and the set of probablity distributions D_1 of the form

$$P_{\tilde{x}}(x) = \delta(x - \tilde{x}). \tag{2.9}$$

Note, that the values of $f(x)$ in (2.1), $f(\overline{x})$ in (2.2) and $\overline{f(x)}$ in (2.4) coinside here on the corresponding elements. Thus, the inequalities (2.7) and (2.8) hold and the problem (2.1) is equivalent to the problems (2.2) and (2.4), in which the feasible set consists of distributions (2.9) and $x \in D$.

The problems (2.2) and (2.4), where $P(x)$ obeys only the conditions (2.3), (2.5) are extensions of the problem with solution in the form (2.9). Therefore they are also extensions for the problem (2.1).

3 Averaged relaxations of nonlinear programming problem. Structure of the optimal solution.

Consider the case when the problem A is the following nonlinear programming problem (NP) (we assumed for simplicity that all the constraints here are the equality constraints)

$$\begin{gathered} f_0(x) \to \max \\ f(x) = 0, \quad x \in V_x, \end{gathered} \tag{3.1}$$

where $x \in R^n$, f_0 is a scalar function and f is an m-dimensional vector function, $m < n$. The space V_x is closed and bounded.

Let us introduce the following problem

$$\begin{gathered} \overline{f}_0 = \frac{1}{T} \int_0^T f_0(x(t))dt \to \max_{x(t)}, \\ \overline{f} = \frac{1}{T} \int_0^T f(x(t))dt = 0, \\ x(t) \in V_x \subset R^n, \forall t \in [0, T]. \end{gathered} \tag{3.2}$$

The solution of the problem (3.2) is sought in the class of measurable functions.

The problem (3.2) is a relaxation of the problem (3.1) because its feasible set includes the subset of functions, which are constant for almost all $t \in [0, T]$, and for each vector $x_0 \in V_x$ in the problem (3.1) there is fuction $x(t) = x_0$.

For each function $x(t)$ it is possible to construct the probability measure $\mu(y) = \dfrac{1}{T}\mu\{t : x(t) \le y\}$, where μZ is the Lebesgue measure of the set Z. In terms of this probability measure the problem (3.2) can be rewritten in the form

$$\begin{gathered} \overline{f}_0 = \int_{V_x} f_0(x)d\mu(x) \to \max_{\mu(x)}, \\ \overline{f} = \int_{V_x} f(x)d\mu(x) = 0. \end{gathered} \tag{3.3}$$

or, after introducing the density of the measure $P_i(x) = \dfrac{d\mu_i}{dx_i}$, as

$$\begin{gathered} \overline{f}_0 = \int_{V_x} f_0(x)P(x)dx \to \max_{P(x)} \\ \overline{f} = \int_{V_x} f(x)P(x)dx = 0. \end{gathered} \tag{3.4}$$

In the points where the measure $\mu(x)$ has a discontinuity of the first kind its density has a δ function component. From the properties of the function $\mu(x)$ it follows that the solution of the problem (3.4) must obey the conditions (2.3).

The problem (3.4) is a relaxation of the problem (3.1). We shall call it the averaged nonlinear programming problem and denote as $\overline{NP}$. There are infinitely many functions $x^*(t)$ in problem (3.1) which correspond to the single solution of the problem (3.4) $P(x) = P^*(x)$. The only exception is the case when $P^*(x) = \delta(x - x^0)$. Then $x^*(t) = x^0$.

The following statement 3.1 is true:

1. *The optimal solution of* $\overline{NP}$ *problem* $P^*(x)$ *has the following form*

$$P^*(x) = \sum_{\nu=0}^{m} \gamma_\nu \delta(x - x_\nu) \tag{3.5}$$

where

$$\gamma_\nu \geq 0, \quad \sum_{\nu=0}^{m} \gamma_\nu = 1. \tag{3.6}$$

2. *A non-zero vector of Lagrange multipliers* $\lambda = (\lambda_0, ...\lambda_m)$ *can be found such that at points* x_ν *the function*

$$R = \sum_{\nu=0}^{m} \lambda_j f_j(x) \tag{3.7}$$

has its global maximum with respect to $x \in V_x$.

The points x^ν are called the basic values of x. If the optimal solution of $\overline{NP}$ problem as a function of time $x(t)$ exists, then it switches from one basic value to another, being equal to each of them during γ_ν fraction of the total duration of the process T.

The averaging in NP problem can be done not for all variables, but for part of them only. Let us divide the variables of the problem (3.1) into two groups - deterministic x and randomized u. The averaging is done only with respect to u. The $\overline{NP}^u$ problem has the following form

$$\begin{aligned} \overline{f_0(x,u)}^u &\to \max_{x,p(u)}, \\ \overline{f(x,u)}^u &= 0. \end{aligned} \tag{3.8}$$

Here

$$\overline{f_j(x,u)}^u = \frac{1}{T}\int_0^T f_j(x, u(t))dt = \int_{V_u} f_j(x,u)P(u)du. \tag{3.9}$$

$$P(u) \geq 0; \qquad \int_{V_u} P(u)du = 1.$$

It is assumed that functions f_j are continues with respect to u and continuously differentiable on x. The optimality conditions for the problem (3.8) take the following form

Statement 2.2:

1. The optimal distribution of the randomized variable has the following form

$$P^*(u) = \sum_{\nu=0}^{m} \gamma_\nu \delta(u - u^\nu), \tag{3.10}$$

$$\gamma_\nu \geq 0, \qquad \sum_{\nu=0}^{m} \gamma_\nu = 1.$$

2. A non-zero vector $\lambda = (\lambda_0, ...\lambda_m)$ can be found such that the Lagrange function $R = \sum_{j=0}^{m} \lambda_j f_j(x, u)$, which is computed using this λ, can not be improved locally with respect to the deterministic variables and has global maximum on randomized variables on the set V_u at each of the basic points u^ν:

$$\begin{gathered}\frac{\delta}{\delta x}\Big\{\sum_{\nu=0}^{m} \gamma_\nu R(x, u^\nu)\Big\}\delta x \leq 0.\\ u^\nu = \arg\max_{u \in V_u} R(\lambda, x^*, u), \quad \nu = 0, m.\end{gathered} \tag{3.11}$$

here δx is a variation that is allowed by the constraints $x \in V_x$

A large number of different versions of $\overline{NP}$ problem exist, because, for example, not all the constraints may depend on both deterministic and randomized variables; and the problem can include not only the averaging of the functions, but also the functions of the averaged values of the time-dependent variables, etc. Therefore it does not make sense to derive optimality conditions for each one of these versions. It is much more reasonable to write down the canonical form of the average extension of NP problem and to derive its necessary conditions of optimality. The statements 2.1 and 2.2 will follow from these conditions.

The canonical form of the averaged extension of the NP problem has the form

$$F_0\Big[\overline{f(x, u)}, x\Big] \to \max \tag{3.12}$$

subject to constraints

$$F_j\Big[\overline{f(x, u)}, x)\Big] = 0, \quad j = \overline{1, r}; \quad x \in V_x, \tag{3.13}$$

the overline f corresponds to the averaging on u over the closed and bounded set V_u. The dimensionality of the vector-function f is m, function F is continuously differentiable on all of its arguments, and f is continues on u and continuously differentiable on x.

Theorem 3.3 (The optimality conditions of the canonical form of the averaged nonlinear programming problem $\overline{NP}$):
1. The optimal distributions of the randomized variables have the following form

$$P^*(u) = \sum_{\nu=0}^{m} \gamma_\nu \delta(u - u^\nu) \tag{3.14}$$

where γ_ν obey the conditions (3.6).
2. A non-zero vector $\lambda = (\lambda_0, \lambda_{j\nu})(j = \overline{1,r}; \quad \nu = \overline{0,m})$ can be found such that for each basic value u^ν of vector u the function

$$L_1 = \lambda_0 \frac{\delta F_0}{\delta \bar{f}} f(x,u) + \sum_{j=1}^{r} \lambda_j \frac{\delta F_j}{\delta \bar{f}} f(x,u),$$

attains its maximum on V_u. Here

$$\bar{f} = \sum_{\nu=0}^{m} \gamma_\nu f(x, u^\nu).$$

Hence

$$u^\nu = \arg\max_{u \in V_u} L(x^*, \lambda, u). \tag{3.15}$$

3. The function

$$R = \sum_{j=0}^{r} \lambda_j F_j \tag{3.16}$$

can not be improved locally with respect to its deterministic arguments

$$\frac{\delta R}{\delta x} \delta x \le 0. \tag{3.17}$$

The proof of this theorem is given in the Appendix.

It is easy to show that after reduction of the problems $\overline{NP}$ and $\overline{NP}^u$ to the form (3.12), (3.13) their optimality conditions follow from the theorem 3.3. Note that the number of the basic solutions is determined by the dimension of the fector-fuction f_0.

4 Averaging in variational problems

Introduction of averaging in variational problems, where unknown variables depend on the scalar argument t, allows to obtain a solution in a form of maximizing sequences and to formulate optimality conditions in a from of maximum principle for any arbitrary form of optimality criterion and constraints. We shall start by giving some auxiliary statements and definitions.

Assume that in the problem A: $I_A(y) \to \max, y \in D_A$ there are a finite number J of constraints, which determines the feasible set D_A. Assume that for j-th condition ($j = \overline{1, J}$) the norm Δ_j can be introduced for the deviation of this constraint from the nominal value.

Definition 4.1. *The problem A is correct with respect to its value if for any $\epsilon > 0$ such δ can be fount that from the inequality $\max\limits_j(\Delta_j) \leq \delta$ follows that the absolute value of the deviation of I_A^* is less or equal to ϵ.*

Nonlinear programming problem is correct in terms of the definition 4.1. if the Slater's complementary slackness conditions are satisfied.

Definition 4.2. *The relaxation of the problem A B: $I_B(y) \to \max, y \in D_B$ is equivalent if*

$$I_{\bar{A}}^* = \sup_{y \in D_{\bar{A}}} I_{\bar{A}}(y) = I_B^* = \sup_{y \in D_B} I_B(y) \tag{4.1}$$

Lemma 4.3. *The sufficient condition for the relaxation to be equivalent is the possibility for any solution of the extended problem $y^0 \in D_B$ to find such a sequence of the solutions of the original problem $\{y_i\} \subset D_A$ that*

$$\lim_{i \to \infty} I_{\bar{A}}(y_i) = I_B(y^0) \tag{4.2}$$

If the problem is correct with the respect to its value then it is possible that the sequence $\{y_i\}$ does not belong to $D_{\bar{A}}$. The only requirement is that any constraint of the original problem is satisfied with arbitrary accuracy in the limit $i \to \infty$ (in accordance with the definition 4.1).

Lemma 4.4. *If y_A^* is an optimal solution of the problem A, the relaxation B is equivalent to A and $D_B \supset D_A$, then y_A^* obeys necessary conditions of optimality of the relaxation problem.*

We will call the following problem the canonical form of variational problem

$$I = \int_0^T \Big[f_{01}(t, x(t), u(t), a) + \tag{4.3}$$

$$+ \sum_l f_{02}(t, x(t), a)\delta(t - t_l) \Big] dt \to \max$$

subject to constraints

$$J_j(\tau) = \int_0^T \Big[f_{j1}(t, x(t), u(t), a, \tau) +$$

$$+f_{j2}(t,x(t),a,\tau)\delta(t-\tau)\Big]dt=0; \tag{4.4}$$

$$\forall\tau\in[0,T], \quad j=\overline{1,m}, \quad u\in V_u, \quad a\in V_a.$$

Here a is the vector of parameters, which are constant on $[0,T]$, functions f_{j1} and f_{j2} are continuously differentiable on x, a and t and continuous on u.

Lemma 4.5. *Assume that the problem (4.3), (4.4) is correct with respect to its value (according to the definition 4.1, where a closeness of each initial and variated condition (4.4) is defined as* $\Delta_j = max_\tau |J_j(\tau)|$*) then the averaged relaxation of this problem*

$$\bar{I}=\int_0^T\Big[\overline{f_{01}(t,x,u,a)}^u+\sum_l f_{02}(t,x,a)\delta(t-t_l)\Big]dt\to\max \tag{4.5}$$

subject to constrains

$$\bar{J}_j(\tau)=\int_0^T\Big[\overline{f_{j1}(t,x,u,a,\tau)}^u+f_{j2}(t,x,a,\tau)\delta(t-\tau)\Big]dt=0, \tag{4.6}$$

$$\forall\tau\in[0,T], j=\overline{1,m}, u\in V_\lambda, a\in V_a$$

is equivalent to the problem (4.3), (4.4).

Here

$$\overline{f_{j1}}^u=\int_{V_u} f_{j1}(t,x,A,u,a,\tau)P(u,t)du. \tag{4.7}$$

Distribution $P(u,t)$ obeys the conditions

$$p(u,t)\geq 0; \quad \int_{V_u} P(u,t)du=1 \quad \forall t\in[0,T]. \tag{4.8}$$

The proof of this statement is based on Lemma 4.3, because for any function $P(u,t)$ in the problem (4.5), (4.6) such a sequence of solutions of the problem (4.3), (4.4) can be constructed that the functionals I and J_j tend to $\overline{I}$ and $\overline{J}_j(\tau)$ correspondingly.

Sliding regimes are examples of such sequences inoptimal control problems.

The solution of the problem (4.5) – (4.6) is the distribution $P^*(u,t)$, the function $x(t)$ and the vector a. It obeys the following conditions (**Theorem 4.5.**):

1. Assume that the optimal solution of the problem (4.5)–(4.6) is

$$P^*(u,t)=\sum_{\nu=0}^m \gamma_\nu(t)\delta(u-u^\nu(t)), \tag{4.9}$$

where the piece-wise continues functions $\gamma_\nu(t) \geq 0 \ \forall t \in [0,T]$ *and* $\sum_{\nu=0}^{m} \gamma_\nu(t) = 1.$

2. A scalar $\lambda_0 \geq 0$ and a *vector function* $\lambda(\tau) = (\lambda_1(\tau), ...\lambda_m(\tau))$, *which is piece-wise continuous for almost everywhere on* $[0,T]$ *and is defined and non-zero simultaneously with* λ_0 *on the interval* $[0,T]$ *and equal zero outside of this interval can be found such that the functional*

$$S = \lambda_0 \bar{I} + \sum_{j=1}^{m} \int_0^T \lambda_j(\tau) \bar{J}_j(\tau) d\tau = \int_0^T R dt \tag{4.10}$$

and its integrand

$$R = \lambda_0 R_0 + \sum_{j=1}^{m} R_j^{cn}, \tag{4.11}$$

$$R_0 = \sum_{\nu=0}^{m} \gamma_\nu(t) f_{01}(t, x(t), u^\nu(t), a) + \sum_l f_{02}(t, x(t), a) \delta(t - t_l),$$

$$R_j^{cn} = \int_0^T \lambda_j(\tau) \Big[\sum_{\nu=0}^{m} \gamma_\nu(t) f_{j1}(t, x(t), u^\nu, a, \tau) + $$
$$+ f_{j2}(t, x(t), a, \tau) \delta(\tau - t) \Big] d\tau \tag{4.12}$$

obey the following conditions

$$\frac{\delta S}{\delta a} \delta a \leq 0, \tag{4.13}$$

$$\frac{\delta R}{\delta x} = 0, \tag{4.14}$$

$$u^\nu(t) = \arg \max_{u \in V_u} R(x, \lambda, a^*, u). \tag{4.15}$$

Since the relaxation (4.5), (4.6) is equivalent to the problem (4.3), (4.4), from the Lemma 4.4 it follows that if the optimal solution of the latter one $(u^*(t), x^*(t), a)$ exists, then it obeys the optimality conditions (4.13) –(4.15).

In the conditions of existence for the optimal solution of the problem (4.3), (4.4) it is assumed that $\gamma_0(t) = 1$, and the other multipliers $\gamma_j(t)$ in (4.9) are equal zero.

Conditions (4.13) – (4.15) allow to derive necessary conditions of optimality in a form of maximum principle for a problem with arbitrarily combination of criterion type and constraints. This can be done simply by writing down items R_0 and R_j^{cn} for each type of criterion and constraints, denoting $u(t)$ these variables, which after reducing the problem to the canonical form, are present in function f_{01} only (variables of the first group),and then writing down function R according to (4.11) and substituting it into (4.13), (4.15).

It is also important that this allows to trace easily how changes or addition of some condition effect optimality conditions - the changes it causes in one of the terms in function R and in participation of some variables in the first group.

Example: Let us consider the following optimal control problem

$$I = \int_0^T f_{01}(x,u,t)dt \to \max \tag{4.16}$$

$$\dot{x}_j = f_{j1}(x,u,t), \quad u \in V_u, \quad j = \overline{1,m}, \quad x(0) = x_0.$$

with the usual assumptions about the functions f_0 and f_j. ¿From the comparison of the problems (4.16) and (4.3), (4.4) it is clear that $R_0 = f_{01}(x,u,t)$. Differential equations can be rewritten in (4.4) form as

$$J_j(\tau) = \int_0^T \Big[f_{j1}(x(t),u(t),t)h(\tau - t) - x(t)\delta(\tau - t) \Big] dt = 0.$$

Here $h(t)$ is Heaviside function and $\delta(t)$ is Dirac function. The term

$$\begin{aligned} R_j^{cn} &= \int_0^T \lambda_j(\tau) \Big[f_{j1}(x,u,t)h(\tau - t) - x_j(t)\delta(\tau - t) \Big] dt = \\ &= f_{j1}(x,u,t) \int_t^T \lambda_j(\tau)d\tau - \lambda_j(t)x(t) = \\ &= f_{j1}(x,u,t)\psi_j(t) + \dot{\psi}_j(t)x_j(t), \end{aligned} \tag{4.17}$$

where $\psi_j(t) = \int_t^T \lambda_j(\tau)d\tau$. The function R is

$$R = \lambda_0 f_{01}(x,u,t) + \sum_j \psi_j(t) f_{j1}(x,u,t) + \sum_j \dot{\psi}_j(t)x_j(t).$$

For the bounded function $\lambda_j(\tau)$ $\psi_j(T) = 0$.

Conditions (4.13), (4.14) yield equations of the Pontrygin's maximum principle. Note that inclusion of various constraints at the final instance of time yields transversality conditions directly, without any special derivations.

Indeed, assume that the constraint

$$F(x(T)) = 0$$

is added to the problem (4.16). This constraint adds the term

$$R_F^{cn} = \lambda_F F(x(t))\delta(t-T)$$

to the function R. Then from the condition (4.14) it follows that

$$\psi_j(T) = \lambda_F \frac{\delta F}{\delta x_j}.$$

Other applications of this approach can be found in [6]–[7].

ACKNOWLEDGEMENT

This work is supported by RFFI (grant 01-01-00020 and 02-06-80445), the School of Finance and Economics, UTS, AC3 and the Capital Markets CRC.

REFERENCES

1. V. I. Gurman, Extension principle in extremal problems, *Moscow-Fizmatlit*, 1997.

2. V. F. Krotov, V. I. Gurman, Methods and problems of optimal control, *Moscow-Nauka*, 1973.

3. S. Fromovitz, Non – Linear programming with randomization, Man. Sci., **11**, 9A (1965).

4. D. A. Ovsianikov, Mathematical methods of beam control, Leningrad–Univ.of Len., 1980.

5. A. M.Tsirlin, Optimization in average and sliding regimes in optimal control problems, Izv. AN SSSR, Tech. Ciber., 2 (1974), 27 – 33.

6. A.M. Tsirlin, Optimality conditions of the averaged mathematical programming problems, Doc. AN SSSR, **323** 1 (1992), 43 – 47.

7. A. M. Tsirlin, Metods of averaged optimization and their application. *Moscow–Fizmatlit*, 1997.

Appendix.

The proof of the Theorem 3.3. For the fixed $x \in V_x$ the values of the vector-function f in the problem (3.12), (3.13) belong to the set Q, which represents the mapping of the set V_u onto an m-dimensional space f. The values of the vector $\bar{f}$ belong to the convex hull of Q. According to the Carateodory theorem each element of the convex hull can be represented as a linear combination of not more than $(m+1)$ elements of Q. Therefore, $\overline{f(x,u)}$ can be represented as

$$\overline{f(x,u)} = \sum_{\nu=0}^{m} \gamma_\nu f(x, u^\nu), \qquad (A.1)$$

where γ_ν obey the conditions (3.5). The equality (A.1) allows to re-represent the problem (3.12), (3.13) as a standard nonlinear programming problem with the unknown variables x, u^ν, γ_ν ($\nu = \overline{0,m}$) and the constraints (3.13) and (3.5). We will use the Kuhn-Tucker theorem to derive its contitions of optimality.

Indeed, after the substitution of the expression (A.1), the problem (3.12), (3.13) takes the following form

$$F_0 \left[\sum_{\nu=0}^{m} \gamma_\nu f(x, u^\nu), x \right] \to \max \qquad (A.2)$$

subject to constraints

$$F_j \left[\sum_{\nu=0}^{m} \gamma_\nu f(x, u^\nu), x \right] = 0, \quad j = \overline{1,r}, \quad x \in V_x \qquad (A.3)$$

$$\sum_{\nu=0}^{m} \gamma_\nu = 1, \quad \gamma_\nu \geq 0. \qquad (A.4)$$

Its Lagrange function is

$$\bar{R} = \sum_{j=0}^{r} \lambda_j F_j \left[\sum_{\nu=0}^{m} \gamma_\nu f(x, u^\nu), x \right] - \Lambda \left(\sum_{\nu=0}^{m} \gamma_\nu - 1 \right).$$

γ_ν are non-negative and the condition of $\bar{R}$ optimality with respect to γ_ν is

$$\frac{\partial \bar{R}}{\partial \gamma_\nu} \delta\gamma_\nu \leq 0, \quad \delta\gamma_\nu \geq 0.$$

Therefore, for the basic values $u = u^\nu$ the function

$$L = \sum_{j=0}^{r} \lambda_j \frac{\partial F_j}{\partial \bar{f}} f(x,u) = \Lambda \quad \text{for} \quad \gamma_\nu^* > 0 \qquad (A.5)$$

and

$$L \leq \Lambda \quad \text{for} \quad \gamma_\nu = 0.$$

Thus, for all the values u^ν, that have positive weights in the solution of the problem (A.2)–(A.4), the function L attains its maximum on u over the set V_u. That is, the condition (3.15) holds. If $f(x, u^*) \in Q$ on the optimal solution then L attains its maximum at u^*.

The condition of optimality of $\bar{R}$ on x has the following form

$$\frac{\partial \bar{R}}{\partial x}\delta x \leq 0 \Rightarrow \sum_{j=0}^{r} \lambda_j \left[\frac{\partial F_j}{\partial \bar{f}} \sum_{\nu=0}^{m} \gamma_\nu \frac{\partial f(x, u^\nu)}{\partial x} + \frac{\partial F_j}{\partial x} \right] \delta x \leq 0. \qquad (A.6)$$

In (A.5), (A.6) we denote $\overline{f(x,u)} = \sum_{\nu=0}^{m} \gamma_\nu f(x, u^\nu)$. Thus $\sum_{\nu=0}^{m} \gamma_\nu \frac{\partial f(x, u^\nu)}{\partial x} = \frac{\partial \bar{f}}{\partial x}$.

After using the denotion (3.16) the condition (A.6) yields the inequalities (3.17).

Let us show how *the statements 3.1 and 3.2 follow from this theorem.* Indeed, in the problem (3.3) the averaging is done over all its variables, the dimension of the function f is m and the function L, which is used in formulas (3.15) and (A.5), coinsices with the Lagrange function R of the problem (3.1).

In problem (3.7) the averaging is done on u, the derivatives $\frac{\partial F}{\partial \bar{f}}$ in (A.5) are equal one, because F_j and f_j are identical. Therefore the functions L and R coinsise with each other and with the Lagrange function of the problem (3.8) without averaging. Here the conditions (3.15) and (3.17) coinsise with (3.11).

The proof of the Lemma 4.3 follows from the definition (4.2). If the optimal solution y^* of the extended problem exists then y^0 can be replased with y^*.

The *Lemma 4.4 follows from from the condition that y^*_A can not be improved on D_A*, and D_A is a subset of the fesible set of the extended problem.

The *proof of the Lemma 4.5 follows from the Lemma 4.3* and from the fact that for any solution $P^0(u,t)$, $x^0(t)$, a^0 it is possible to find a sequence of solutions

$$\{z_i\} = \{u_i(t), x^0(t), a^0\},$$

for which

$$I(z_i) \to \overline{I^*}, \quad J_j(\tau, z_i) \to \overline{J_j}(\tau) = 0.$$

Indeed, one can prove the first statement of the Theorem 4.5 by completely repeating the derivations, which were employed above to prove the Theorem 3.3: that for any fixed values of t, x, a, τ the value of the vector $\bar{f} = (\bar{f}_0, \bar{f}_1, ..., \bar{f}_m)$ belongs to the convex hull of the set Q that is produced by the mapping of V_u onto $(m+1)$-dimensional space f. Since the solution maximizes f_0 with respect to U, it belongs to the upper bound of Q and can be represented as a linear combination of not more than $(m+1)$ (and not $(m+2)$) elements of Q. This proves the (4.9).

For any solution $P^0(u,t)$ that has the form (4.9) it is possible to construct the sequence $\{u_i(t)\}$ of solutions of the problem (4.3), (4.4) using the following algorithm: let us divide the interval $[0,\tau]$ into i sub-intervals $\Delta_1, \Delta_2, ..., \Delta_i$ and assume that on each of these intervals the functions $\gamma_\nu(t)$ and $u^\nu(t)$ are constant. Assume that for the r-th interval their values are $\gamma_{\nu r}$ and u_r^ν $(\nu = \overline{0,m})$. We shall call the correspondent problem a discretization of the problem (4.5), (4.6).

Let us construct a similar division of the interval $[0,\tau]$ in the problem (4.3), (4.4). The only difference here is that each sub-interval of the first division is futher subdivided into $(m+1)$ smaller pieses. So, Δ_r is divided into Δ_{r0}, Δ_{r1}, ..., Δ_{rm}, such that the following equality holds

$$\frac{\Delta_{r\nu}}{\Delta_r} = \gamma_{\nu r}.$$

We then assume that the variables $u(t)$ in the problem (4.3), (4.4) are piece-wise constant functions of time and that are equal to U_r^ν on the interval $\Delta_{r\nu}$. The values of the functionals I and $J(\tau)$ in the problem (4.3), (4.4) on the solution, which is constructed by this algorithm, are equal to the values of the corresponding functionals in the discretization problem (4.5), (4.6). If $i \to \infty$ and Δ_r approaching zero uniformally on r then the values of I_D and $J_D(\tau)$ in discretization of the averaged problem are arbitrary close to $\overline{I}$ and $\overline{J}(\tau)$, because the problem (4.3), (4.4) is correct with respect to its value. Lemma 3.5 is proven.

In order to prove the Statement 2 of the Theorem 4.5, we will use the following theorem [5]:
Let $y^*(t)$ be a solution of the following maximization problem

$$I = \int_0^T f_0(y,t)dt \qquad (A.7)$$

subject to constraints

$$J_j(\tau) = \int_0^T f_j(y,t,\tau) = 0, \quad j = \overline{1,m}, \quad \tau \in [0,T], \qquad (A.8)$$

where f is continuous and continuosly differentiable function on all its arguments. Then such non-zero vector

$$\lambda = (\lambda_0, \lambda_1(\tau), ..., \lambda_m(\tau)), \quad \lambda_0 \geq 0,$$

can be found that for $y = y^*$ the following inequality holds

$$\left(\frac{\partial R}{\partial y}\right) \delta y \leq 0, \qquad (A.9)$$

where

$$R = R_0 + \sum_{j=1}^{m} R_j = \lambda_0 f_0 + \sum_{j=1}^{m} \int_0^T \lambda_j(\tau) f_j(y, t, \tau) d\tau,$$

and δy is a variation of $y(t)$ that does not violate the condition $y \in V_y(t)$.

For a distribution $P(u, t)$ that has the form (4.9) the problem (4.5), (4.6) takes the form (A.7), (A.8) with

$$\bar{R} = R + R_{m+1} = \lambda_0 R_0 + \sum_j R_j^{\text{cn}} - R_{m+1},$$

where R_0 and R_j^{cn} have the form (4.12), and the term R_{m+1} corresponds to the condition

$$\sum_{\nu=0}^{m} \gamma_\nu(t) - 1 = 0, \quad \forall t \in [0, T], \qquad (A.10)$$

which can be rewritten in the form (A.8) as

$$J_{m+1}(\tau) = \int_0^T \left(\sum_{\nu=0}^{m} \gamma_\nu(t) - 1\right) \delta(t - \tau) d\tau = 0 \quad \forall \tau \in [0, T].$$

Thus,

$$R_{m+1} = \int_0^T \lambda_{m+1}(\tau) \left(\sum_{\nu=0}^{m} \gamma_\nu(t) - 1\right) \delta(t - \tau) d\tau = \lambda_{m+1}(t) \left(\sum_{\nu=0}^{m} \gamma_\nu(t) - 1\right).$$

The constraint (A.9) on the the variables γ_ν

$$\frac{\partial \bar{R}}{\partial \gamma_\nu} \delta \gamma_\nu \leq 0, \quad \gamma_\nu \geq 0$$

yields the following condition for the basic values $u^{\nu}(t)$ (only for these values $\gamma_{\nu}(t) > 0$),

$$R(x, \lambda, a^*, u^{\nu}) = \lambda_{m+1}(t), \quad \nu = \overline{0, m},$$

and for $u \neq u^{\nu}(t) \quad \gamma_{\nu}(t) = 0$ and $\delta\gamma_{\nu} > 0$, and therefore

$$R(x, \lambda, a^*, u) \leq \lambda_{m+1}(t).$$

from which the condition of (4.15) maximum follows.

The condition (4.14) follows from (A.9) after taking into account the absence of constraints on x, and the conditions (4.13) are the consequence of the fact that the problem (3.5), (3.6) is a nonlinear programming problem with respect to its parameters a and S is its Lagrange function.