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

March 19, 2004

Abstract

In this paper extremal problems that include averaging operation are considered. Canonical forms for nonlinear programming problem and for general-type variational problems with averaging are constructed. Their optimality optimality are derived. Examples are given of these conditions applications for particular problems.

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 three 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 case occurs when there is not certainty that a solution of the original problem exists. If the optimial solution of an auxiliary

*This work is supported by Russian Foundation for Basic Reseach (grants No. 01-01-00020, 02-06-80445)

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.

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]. As a rule, an introduction of such averaging into the problem's conditions changes its feasible set Dinto a larger $\bar{D}$: $D \in \bar{D}$.

There are many forms of extremal problems that contain not only vector and functional variables but also their averaged values or averaged values of a function, which depends on these variables [1]. As a rule, incorporation of the averaging operation into formulation of the problem extends the set of feasible solutions D to the set $\bar{D}$.

Extremal problem B is called an extension of the original problem A (or an extended problem A) [1], [2] is the following two conditions hold:
1. The sets of the feasible solutions of the problems A and B relay to each other as

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

2. The optimality criterion I_B of the problem B coincides with optimality criterion of the problem A on the set D_A

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

It turns out that in a problem with averaging the unknown variables are distributions of the variables of the original problem. For example, if the unknown variable in the original problem A is vector $x \in V_x$ then the unknown variable in the averaged problem B is such a distribution $P(x)$ that

$$\int_{V_x} P(x)dx = 1; \quad P(x) \geq 0; \tag{1.3}$$

Thus, the nature of the sets D_B is different from the nature of the set D_A and (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.
1.1 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

$$I_A(y) \geq I_A(z), \quad (y, z) \in D_A \tag{1.4}$$

follows that

$$I_{A_1}(y_1) \geq I_{A_1}(z_1), \quad (y_1, z_1) \in D_{A_1}, \tag{1.5}$$

where y_1 and z_1 correspond to y and z. The equivalent class $\bar{A}$ is defined as a set of all the problems which are isomorphic to the problem A_1.

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

It turns out that for the nonlinear programming problem A is isomorphic to the averaged problem A_1 where distribution $P(x)$ of the form

$$P(x) = \delta(x - x_0). \tag{1.6}$$

corresponds to each vector $x_0 \in V_x$.
1.2 We will call any problem B an extension of the problem A if the conditions (1.1) and (1.2) hold for some problem from the class $\bar{A}$.

Since an averaged problem in which unknown variables are distributions $P(x)$ is an extension of the problem with condition of the type (1.6), it is also an extension of the nonlinear programming problem.

2 Averaged extensions of nonlinear programming problem. Structure of the optimal solution.

Let us consider the following nonlinear programming problem (NP)

$$f_0(x) \to \max \Big/ f(x) = 0, \quad x \in V_x, \tag{2.1}$$

where $x \in R^n$, f_0 is a scalar function and f is an m-dimensional vector function, $m < n$. This is our original problem A.

The averaging is frequently introduced by replacing functions f_0 and f with their averaged values $\overline{f_\nu}$. Here function $\overline{f_\nu(x)}$ is defined as

$$\overline{f_j(x)} = \frac{1}{T}\int_0^T f_j(x(t))dt = \int_{V_x} f_j(x)P(x)dx, \quad j = \overline{0,m}. \tag{2.2}$$

In the latter case the distribution $P(x)$ obeys the condition (1.3). The will call the problem

$$\overline{f_0(x)} \to \max_{P(x)} \Big/ \overline{f(x)} = 0, \tag{2.3}$$

as $\overline{NP}$ problem.

The following statement is true 2.1: *The optimal solution of $\overline{NP}$ problem $P^*(x)$ has the following form*

$$P^*(x) = \sum_{\nu=0}^{m} \gamma_\nu \delta(x - x_\nu) \tag{2.4}$$

where

$$\gamma_\nu \geq 0, \quad \sum_{\nu=0}^{m} \gamma_\nu = 1. \tag{2.5}$$

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

$$R = \sum_{\nu=0}^{m} \lambda_j f_j(x) \tag{2.6}$$

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

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 (2.1) 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

$$\overline{f_0(x,u)}^u \to \max_{x,p(u)} \Big/ \overline{f(x,u)}^u = 0. \tag{2.7}$$

Here

$$\overline{f_j(x,u)}^u = \frac{1}{T}\int_0^T f_j(x,u(t))dt = \int_{V_u} f_j(x,u)P(u)du. \tag{2.8}$$

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

It is assumed that functions f_j are continues with respect to u and continuously differentiable on x. (Statement 2.2) Optimality conditions of the problem (2.7) have the form:

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

$$P^*(u) = \sum_{\nu=0}^{m} \gamma_\nu \delta(u - u^\nu), \tag{2.9}$$

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

1. 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:

$$u^\nu = \arg\max_{u \in V_u} R(\lambda, x^*, u), \quad \nu = 0, m. \tag{2.10}$$

$$\frac{\delta}{\delta x}\Big\{\sum_{\nu=0}^{m} \gamma_\nu R(x, u^\nu)\Big\}\delta x \leq 0. \tag{2.11}$$

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

Since it is possible that not all the constraints in NP problem depend on both deterministic and randomized variables, and it can include not only averaging of the functions but also functions of the averaged values of the time-dependent variables etc., many different averaged extensions of the NP problem can be found. 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

$$F_0\left[\overline{f(x,u)}, x\right] \to \max \tag{2.12}$$

subject to constraints

$$F_j\left[\overline{f(x,u)}, x)\right] = 0, \quad j = \overline{1,r}; \quad x \in V_x, \tag{2.13}$$

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 2.3 (The optimality conditions of the averaged extension of the NP problem):
1. The optimal distributions of the randomized variables have the following form

$$P^*(u) = \sum_{\nu=0}^{m} \gamma_\nu \delta(u - u^\nu) \tag{2.14}$$

where γ_ν obey the conditions (2.5).
2. 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 function

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

attains its maximum on V_u. Here

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

Hence

$$u^\nu = \arg\max_{u \in V_u} L(x^*, \lambda, u). \tag{2.15}$$

3. *Function*

$$R = \sum_{j=0}^{r} \lambda_j F_j \tag{2.16}$$

can not be improved locally with respect to its deterministic arguments

$$\frac{\delta R}{\delta x} \delta x \le 0. \tag{2.17}$$

It is easy to show that after reduction of the problems $\overline{NP}$ and $\overline{NP}^u$ to the form (2.12), (2.13) the optimality conditions for them follow from the theorem 2.3.

3 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.

Definition 3.1. *Problem A:* $I_A(y) \to \max \big/ y \in D_A$ *is called correct with respect to its value if infinitesimal change of any of the constraints* D_A *leads to infinitesimal change of the value of the problem* I_A^*.

Naturally, this definition requires a definition of how to measure the closeness to each other of two conditions which define D_A. If this is done then it can be formulated in terms of $E \sim \delta$.

Nonlinear programming problem is correct in terms of the definition 3.1. if the Slater's complementary slackness conditions are satisfied.

Definition 3.2. *Extension B:* $I_B(y) \to \max \big/ y \in D_B$ *of the problem A is equivalent if*

$$I_{\bar{A}}^* = \sup_{y \in D_{\bar{A}}} I_{\bar{A}}(y) = I_B^* = \sup_{y \in D_B} I_B(y) \tag{3.1}$$

Lemma 3.3. *The sufficient condition for the extension 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*

$$\lim_{i \to \infty} I_{\bar{A}}(y_i) = I_B(y^0) \tag{3.2}$$

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 3.1).

Lemma 3.4. *If* y_A^* *is the optimal solution of the problem A, extension B is equivalent to A and* $D_B \supset D_A$ *then* y_A^* *obeys necessary conditions of optimality of the extension problem.*

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

$$I = \int_0^T \Big[f_{01}(t, x(t), u(t), a) + \tag{3.3}$$

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

subject to constraints

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

$$+ f_{j2}(t, x(t), a, \tau)\delta(t - \tau) \Big] dt = 0; \tag{3.4}$$

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

Here a is 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 3.5. *Assume that the problem (3.3), (3.4) is correct with respect to its value (according to the definition 3.1, where a closeness of each initial and variated condition (3.4) should be understood in uniform metrics) then the averaged extension of this problem is*

$$\bar{I} = \int_0^T \left[\overline{f_{01}(t,x,u,a)}^u + \sum_l f_{02}(t,x,a)\delta(t-t_l)\right] dt \to \max \qquad (3.5)$$

subject to constrains

$$\bar{J}_j(\tau) = \int_0^T \left[\overline{f_{j1}(t,x,u,a,\tau)}^u + f_{j2}(t,x,a,\tau)\delta(t-\tau)\right] dt = 0, \qquad (3.6)$$

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

is equivalent to (3.3), (3.4).

Here

$$\overline{f_{j1}}^u = \int_{V_u} f_{j1}(t,x,A,u,a,\tau)P(u,t)du. \qquad (3.7)$$

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

$$p(u,t) \geq 0; \quad \int_{V_u} P(u,t)du = 1 \quad \forall t \in [0,T]. \qquad (3.8)$$

The proof of this statement is based on Lemma 3.3.

The solution of the problem (3.5) – (3.6) consists of distribution $P^*(u,t)$, function $x(t)$ and vector a. It obeys the following conditions (**Theorem 3.5.**):

1. Optimal distribution has the form

$$P^*(u,t) = \sum_{\nu=0}^{m} \gamma_\nu(t)\delta(u-u^\nu(t)), \qquad (3.9)$$

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

2. Scalar $\lambda_0 \geq 0$ *vector function* $\lambda(\tau) = (\lambda_1(\tau), ...\lambda_m(\tau))$, *piece-wise continuous for almost everywhere on* $[0,T]$ *that 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*

$$S = \lambda_0 \bar{I} + \sum_{j=1}^{m} \int_0^T \lambda_j(\tau) \bar{J}_j(\tau) d\tau = \int_0^T R dt \tag{3.10}$$

and its integrand

$$R = \lambda_0 R_0 + \sum_{j=1}^{m} R_j^{cn}, \tag{3.11}$$

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

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

$$+ f_{j2}(t, x(t), a, \tau)\delta(\tau - t) \Big] d\tau \tag{3.12}$$

obey the following conditions

$$\frac{\delta S}{\delta a} \delta a \le 0, \tag{3.13}$$

$$\frac{\delta R}{\delta x} = 0, \tag{3.14}$$

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

Since the extension (3.5), (3.6) is equivalent to the problem (3.3), (3.4), from the Lemma 3.4 it follows that if the optimal solution of the latter one $(u^*(t), x^*(t), a)$ exists then it obeys the optimality conditions (3.13) –(3.15).

Conditions for existence of the optimal solution of the problem (3.3), (3.4) assume that $\gamma_0(t) = 1$, and the other multipliers $\gamma_j(t)$ in (3.9) are equal zero.

Conditions (3.13) – (3.15) 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), writing down function R according to (3.11) and substituting it into (3.13), (3.15).

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 items in function R and in participation of some variables in the first group.
Example: Let us consider the following optimal control problem

$$I = \int_0^T f_{01}(x,u,t)dt \to \max \tag{3.16}$$

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

with the usual assumptions about the functions f_0 and f_j. ¿From the comparison of the problems (3.16) and (3.3), (3.4) it is clear that $R_0 = f_{01}(x,u,t)$. Differential equations can be rewritten in (3.4) 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

$$R_j^{cn} = \int_0^T \lambda_j(\tau)\Big[f_{j1}(x,u,t)h(\tau - t) - x_j(t)\delta(\tau - t)\Big]dt =$$

$$= f_{j1}(x,u,t)\int_t^T \lambda_j(\tau)d\tau - \lambda_j(t)x(t) =$$

$$= f_{j1}(x,u,t)\psi_j(t) + \dot{\psi}_j(t)x_j(t), \tag{3.17}$$

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

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

Conditions (3.13), (3.14) 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.

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

4 Acknowledgments

This work is directly supported by grants from RFFI (grant 01-01-00020 and 02-06-80445) and by the School of Finance and Economics of University of Technology, Sydney.

References

[1] V. I. Gurna, 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 averaging optimization and their application. *Moscow–Fizmatlit*, 1997.